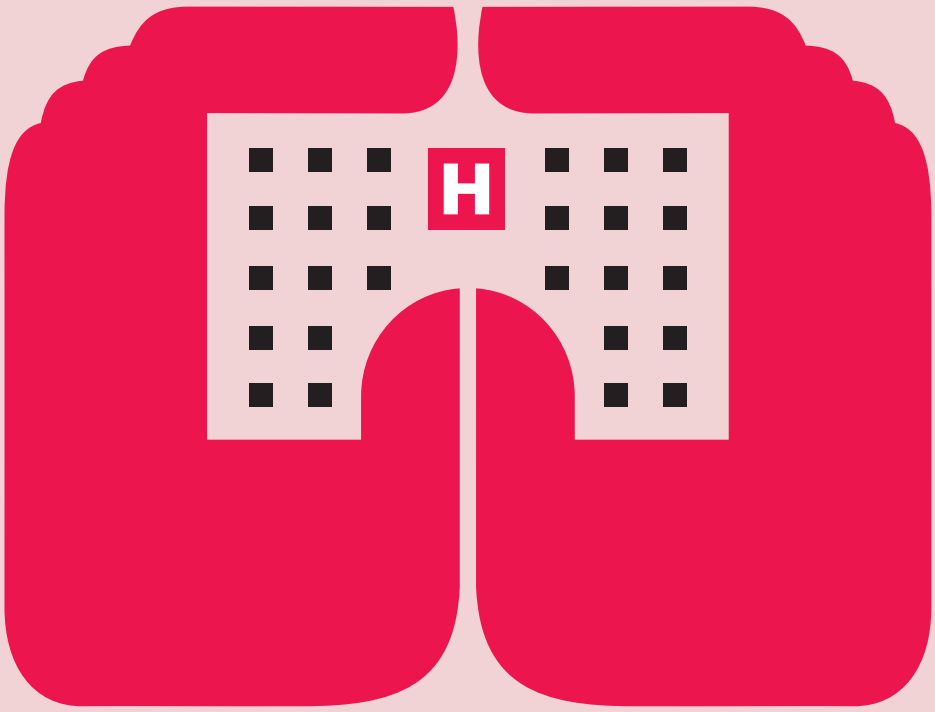
Evidence:

Safer Patients Initiative phase two

A controlled evaluation of the second phase of a complex patient safety intervention implemented in English hospitals

February 2011



The Health Foundation
Inspiring Improvement

Identify Innovate Demonstrate Encourage

Evidence: Safer Patients Initiative phase two is published by the Health Foundation
90 Long Acre, London WC2E 9RA

ISBN 978-1-906461-22-5

Contributors

Amirta Benning, programme manager, School of Health and Population Science, University of Birmingham.

Nick Barber, professor of the practice of pharmacy, Department of Practice and Policy, School of Pharmacy, University of London.

Martin Carmalt, consultant physician, Royal Orthopaedic Hospital, Birmingham. Jeremy Dawson, research fellow, Work & Organisational Psychology Group, Aston Business School, Aston University.

Bryony Dean Franklin, professor of medication safety, centre for medication safety and service quality, Department of Practice and Policy, the School of Pharmacy, University of London; Imperial College Healthcare NHS Trust, London.

M Clare Derrington, public health physician, independent contractor.

Mary Dixon-Woods, professor of medical sociology, Department of Health Sciences, University of Leicester.

Maisoon Ghaleb, lecturer in pharmacy practice/patient safety, Department of Practice and Policy, School of Pharmacy, University of London; School of Pharmacy, University of Hertfordshire.

Alan Girling, senior research fellow, School of Health and Population Science, University of Birmingham.

Karla Hemming, senior research fellow, School of Health and Population Science, University of Birmingham.

Amit Kotecha, registrar, Royal Orthopaedic Hospital, Birmingham.

Richard Lilford, professor of clinical epidemiology, School of Health and Population Science, University of Birmingham.

Thirumalai Naicker, honorary research associate, School of Health and Population Science, University of Birmingham.

Ugochi Nwulu, research associate, School of Health and Population Science, University of Birmingham.

Gavin Rudge, data scientist, School of Health and Population Science, University of Birmingham.

Foreword

The Health Foundation is an independent charity that aims to improve the quality of healthcare across the UK. We are here to inspire and create the space for people, teams, organisations and systems to make lasting improvements to health services.

In 2006, we launched the second phase of the Safer Patients Initiative (SPI), a large-scale intervention and the first major programme addressing patient safety in the UK. We set up the initiative to test ways of improving patient safety on an organisation-wide basis within 20 hospitals in across the UK. The participating trusts undertook improvement in leadership and four clinical areas. They had two stretch aims: a 30% reduction in adverse events and a 15% reduction in mortality over a 20-month timescale. In addition, trusts had specific goals relating to a range of process and intermediate outcomes measures.

In 2006, we also appointed a consortium led by the University of Birmingham to undertake an evaluation of the second phase of SPI (the same team evaluated the first phase). The evaluation sought to assess the wider organisational impact of SPI and so looked beyond the pilot populations of the clinical interventions. It measured the average effect of the programme across a range of practices, based on the starting assumption that SPI would transform organisation-wide approaches to patient safety.

The evaluation reports that the intervention did heighten managerial awareness of and commitment to patient safety. It also created organisational understanding about how to implement safety improvement efforts. Case note review found that many aspects of evidence based medical and peri-operative care were good at baseline (over 90% on some criteria), leaving little room for improvement. Overall, a significant additive effect of SPI on the measures included in the study was not detected.

A rising tide in patient safety

The evaluators consider possible explanations for the absence of an additional effect of the programme, including a 'rising tide'

phenomenon, where improvements in patient safety were driven by common forces across the NHS.

We believe that SPI was part of that rising tide that has placed safety firmly on wider policy and professional agendas. Throughout SPI and since, we have been committed to being at the forefront of work to accelerate the UK-wide patient safety agenda, shape the debate and develop learning on the challenges of building a sustainable culture of patient safety.

Our work has had an impact on the development of national patient safety initiatives in each of the four UK Countries.

- In 2006, the English Department of Health publication, *Safety First,* identified the Health Foundation as one of the organisations that had played a significant role in patient safety at national level. It recommended that a national patient safety campaign be established and that it should be 'in keeping with the approach already successfully used by organisations such as the Health Foundation and Institute for Healthcare Improvement. The programme should be specifically designed to engage and inform frontline staff and should enable staff to take ownership and harness the opportunity to influence the national patient safety agenda.'
- In Scotland, a report from the Scottish Government in 2007 (*Better Health, Better Care: Action Plan*) said that the Scottish Patient Safety Alliance will 'build upon the successes of the current SPI which is already improving safety standards in NHS Ayrshire and Arran, NHS Dumfries and Galloway and NHS Tayside.'
- In Northern Ireland, a proposal in 2007 to develop national indicators for safe and effective care drew on the work of the three Trusts involved in SPI; and a report by Northern Ireland's Chief Medical Officer, in 2008, cited working with the Health Foundation as enabling Northern Ireland to adopt internationally recognised best practice in tackling healthcare-associated infections.
- In Wales, a report in 2007 to the Welsh Assembly, *Minimising Healthcare Associated Infections in NHS Trusts in Wales*, includes examples of good practice from SPI site (phase one) Conwy and Denbighshire NHS Trust.

We have led and contributed actively to the national debate. In a speech to the 2008 Patient Safety Congress, Prime Minister Gordon Brown referred to the influence that SPI has had on the patient safety agenda. In 2009, we made a submission to the Health Select

Committee's Inquiry into patient safety and in the Government's response to the consultation it said:

> *'In the Committee's views SPI, The Health Foundation's important work in applying carefully researched methodology for improving safety performance, were welcomed. We also value the contribution The Health Foundation is making as a member of the National Patient Safety Forum and the NQB, and in particular its major contribution with the NPSA and the NHS III in supporting the national initiative for improving safety in England'*

More recently, the 2011 Department of Health's White Paper consultation response cites our contribution, highlighting the Health Foundation as being a leading and influential organisation in patient safety.

Taking all of these impacts together, we believe that we contributed to wider policy changes and were instrumental in creating the rising tide of policy and professional forces.

Evaluation's contribution to the science of improvement

The evaluations of SPI phase one and two make valuable contributions to the literature and debate about the role of the collaborative model in improving quality. Hulscher et al.'s (2009) systematic review of collaboratives (available on the Health Foundation's website: www.health.org.uk) identified ten published controlled evaluations of collaboratives – three show positive effects, two show null effects and five had mixed effects. The review concludes that the evidence of impact of collaboratives is positive but limited and the effects cannot be predicted with great certainty.

Hulscher et al. caution against over-claiming what collaboratives can achieve. What is critical, therefore, to the design of a collaborative is the development of an explicit programme theory and organisational theory of change. This will help to clarify whether the proposed dose of intervention is likely to result in a localised or systemic intervention; determine whether there is a sufficiently specified plan for vertical and horizontal spread, to allow the work to move from project status to becoming embedded in mainstream structures; and make clear the strategy for clinical engagement.

With hindsight, more could have been done in SPI at the outset to develop and critically examine the underlying programme theory, and then ensure that the proposed evaluation design reflected this.

As the evaluators remark in this report:

> *'In that case a more focused and less ambitious intervention, and somewhat narrower evaluation, might have ensued.'*

We think there is value in greater integration between the science of improvement and evaluation methods. We welcome closer collaboration between leaders in these areas to develop the science of evaluating improvement initiatives. From such collaboration will come the rigorously derived knowledge urgently required to bring about organisation-wide improvement in patient care across the health system.

Dr. Dale Webb
Director of Evaluation & Strategy
The Health Foundation

Contents

Figures and tables

Abbreviations

AHR	Alcohol hand rub
APACHE II	Acute Physiology and Chronic Health Evaluation II
BNF	British National Formulary
BTS	British Thoracic Society
C. diff	*Clostridium difficile*
CI	Confidence intervals
CMP	Case Mix Programme
COPD	Chronic obstructive pulmonary disease
CURB	Confusion urea respiratory rate blood pressure
DVT	Deep vein thrombosis
HCAI	Healthcare Associated Infections
HES	Health Episode Statistics
HPA	Health Protection Society
ICC	Intra-class correlation coefficients
ICNARC	Intensive Care National Audit and Research Centre
ICU	Intensive Care Unit
IHI	Institute for Healthcare Improvement
LSOA	Lower level super output areas
MRSA	Methicillin-resistant *Staphylococcus aureus*
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NOSEC	National Observational Study to Evaluate the Clean Your Hands campaign
OR	Odds ratios
PCA	Patient controlled analgesia
SE	Standard errors
SPI	Safer Patients Initiative
SPI1	Pilot phase hospitals of the Safer Patients Initiative
SPI2	Second phase hospitals of the Safer Patients Initiative

Executive summary

Objectives

To evaluate the second phase of the Health Foundation's Safer Patients Initiative (SPI), a large scale multiple component intervention intended to improve the safety of hospital care.

Setting and participants

Nine NHS hospitals in England participating in phase two of the Health Foundation's Safer Patients Initiative (SPI2) and nine matched English control hospitals.

Intervention

The second phase of a multi-component intervention mentored by the US Institute for Healthcare Improvement (IHI), with an investment from the Health Foundation of approximately £270,000 per hospital. It was delivered over 20 months and focused on improving the reliability of specific front-line care processes within designated clinical areas and engaging senior leaders to change the culture of the organisation. The intervention is fully described in the *Safer Patients Initiative: phase one* evaluation report.

Design and outcomes

A controlled evaluation comprising of five linked sub-studies:

- Before and after assessment of attitudes of front-line staff using a structured postal survey in both control and SPI2 hospitals.
- Case note review of the hospital records of high-risk patients in medical wards treated before and after the intervention in both control and SPI2 hospitals. Quality of care was measured by two teams who were independent of the hospitals – one assessed

quality against specific standards (explicit review of acute medical care), and the other undertook holistic assessments (implicit review of acute medical care).
- Explicit case note reviews of high-risk perioperative care patients against specific standards, carried out by a third independent team.
- Indirect evaluation of hand hygiene by measuring used hygiene consumables from trend data already collected to compare the matched controls with the SPI2 hospitals.
- Measurement of outcomes: adverse events and mortality among high-risk patients admitted to medical wards; hospital-wide mortality; intensive care unit (ICU) outcomes; hospital-acquired infection rates and patient satisfaction. Comparisons were made of control hospitals versus the SPI2 hospitals at baseline and over time.

Results

Only one dimension of the staff survey changed significantly (in favour of control hospitals). Measurements of vital signs and use of risk scoring improved markedly over time, but did so similarly in both control and SPI2 hospitals. Many aspects of evidence-based medical and perioperative care were good at baseline, leaving little room for improvement.

There was a marked improvement in use of hand-washing materials and a dramatic decrease in hospital-acquired infections across all hospitals. A significant additive effect of the SPI on the measures included in the study was not detected.

Conclusion

Many aspects of care are already good or improving across the NHS, suggesting considerable gains in quality across the board. These improvements might be due to policy activities, including some with features similar to the SPI, and the emergence of professional consensus on some clinical processes.

An additional effect of a large-scale organisational intervention (SPI) was not detected. It is possible that any effect was too small to detect, that the null additive effect was due to sub-optimal implementation, or that there may be longer-term additive effects that take longer to surface.

Chapter 1

Introduction

The first phase of the Health Foundation's Safer Patients Initiative (SPI1) programme involved four UK hospitals that were selected to take part in an organisational intervention to transform organisational approaches to delivering safer care designed by the Institute of Healthcare Improvement (IHI) and implemented in 2004.[1]

To build on the experience and learning from this first phase, a second phase of the intervention, known as the Safer Patients Initiative: phase two (SPI2), was rolled out from March 2007 to September 2008 inclusive. SPI2 included a further 20 UK hospitals (10 in England and 10 in the other countries of the UK) that were selected following a process similar to that used for SPI1.

The second phase of the intervention remained much the same as SPI1 intervention. For a full description and rationale for end-points used please see our report on phase one, *Evidence: Safer Patients Initiative phase one*, where these are described in full.

The programme was again mentored by the IHI. It was designed to strengthen the organisations generically, while putting in place specific front-line activities, such as the introduction of early warning score systems (EWSS) to improve the management of acutely sick patients, the use of ventilator bundles to reduce ventilator-acquired pneumonia in intensive care and the introduction of a surgical bundle of evidence-based standards to reduce surgical complications.

There were five main differences between SPI1 and SPI2 in the overall management of the programme based on experiences gleaned from SPI1 sites:

– The hospitals were required to work with a partner organisation (a buddy system) and encouraged to hold regular meetings between the lead implementation teams (10–12 people) from each site. By using this system it was envisaged that sites would

support each other, share the burden and provide support in quickly achieving the goals of the intervention.
- There was a longer period between dissemination of the preparatory materials (December 2006) and the first kick-off session where the various teams came together with IHI to share experiences (March 2007). This gave sites more time for planning and developing the intervention and to obtain a baseline measurement in the safety climate survey.
- The financial package was smaller than in the case of SPI1; a mean of £270,000 per site rather than £775,000.
- There were four learning sessions as with SPI1, but an additional reliability and capability workshop was provided.
- SPI2 sought a 15% reduction in mortality rates; this was not an explicit SPI1 aim.

Specific aspects of the intervention also changed:
- the reduction of adverse event target was revised from 50% to 30% as it was felt that this was a more achievable yet aspirational target
- removal of the routine use of beta blockers in the surgical bundle as this clinical standard was contentious in the UK.

1.1 Selection of participating sites

As with the selection of the SPI1 sites, SPI2 sites were selected through a competitive bidding process. A similar format to the phase one selection was followed with initial applications reviewed by an international panel with expertise in patient safety, organisational change and improvement methodology. Applications were assessed against the following criteria:

- leadership commitment
- capacity and capability
- openness, transparency and communication
- collaboration.

The short-listed sites were subject to an on-site assessment and the final 20 sites were chosen by a selection board.

Chapter 2

Methods

This evaluation was conducted with ethical approval and its methods were similar to those used for the evaluation of SPI1. The SPI2 evaluation used a series of linked sub-studies to address generic outcomes (that might be expected to improve if a general strengthening of organisational systems in relation to patient safety occurred) and specific outcomes (that were targeted specifically by SPI interventions).

2.1 Framework for the evaluation

All of the quantitative studies undertaken in the SPI1 evaluation were replicated in SPI2, but no qualitative elements (senior staff interviews and ethnographic study on the wards) were collected. The following SPI1 studies were repeated:

- Staff survey
- Explicit case note review of patients with acute respiratory disease to:
  - audit care against explicit standards
  - measurement of error rates implicitly (holistic case note review)
  - measurement of adverse events (preventable and non-preventable)
  - measurement of mortality among patients included in the case note reviews
- Patient survey.

The quantitative collection of processes and outcomes data was expanded to include:

- Case note review of surgical case notes to measure compliance with a bundle of standards for perioperative care
- ICU outcome data to provide evidence relevant to the effectiveness of the critical care bundles
- Consumption of alcohol hand rub (AHR) and soap in hospital trusts, along with measures of *Clostridium difficile* (C. diff) and Methicillin-resistant *Staphylococcus aureus* (MRSA) infection

rates to provide evidence on measures to reduce healthcare associated infections (HCAI)
- Overall hospital mortality rates in adult patients, standardised for sex and age.

The complete list of sub-studies for the evaluation are summarised in table 2.1.

Each sub-study was based on before and after comparisons in both control and SPI2 sites. The use of both the before and after observations across control and SPI2 sites enables rates of change to be compared across control and SPI2 hospitals.

2.2 Control and SPI sites

We focused on the ten English SPI2 hospitals so that we could take advantage of routinely collected data in England. Although the hospitals worked in pairs, each hospital formed a unit of analysis for the statistical power calculation and for the evaluation.

One of the ten SPI2 hospitals declined to participate in the evaluation leaving nine available for study. Nine SPI2 matched control sites were selected using the following criteria:

- Only non-specialist acute hospitals in England were considered.
- Control and SPI2 hospitals should have a similar directorate structure (as described in the NHS national staff survey).
- The hospitals should have the same foundation or non-foundation status (to gain foundation status a hospital must satisfy the government that it has the management capacity to warrant greater operational autonomy).
- Hospitals should be similarly located in either urban or rural settings.
- Once these criteria were satisfied, the hospital with the most similar size (usually within 1000 staff) to the SPI2 hospital was selected as the control hospital.
- If a trust had more than one hospital, quantitative data collection was focused on the largest hospital with an ICU.

Although nine control and nine SPI2 sites agreed to participate in the evaluation, we were also required to obtain further consent for each sub-study. In some instances this was not granted.

In addition, certain hospitals did not participate in specific routine data collection exercises, while others failed to supply case notes for specific analysis. It is for these reasons that discrepancies exist in the number of sites agreeing to participate in the evaluation and the number included in each sub-study. Full details are provided in the results section of each sub-study.

Table 2.1: Summary of sub-studies comprising the evaluation of SPI2

Sub-study	Purpose	Location	Data collection	Analysis
1) Staff survey	Measure effects of SPI2 on staff morale, culture and opinion	Control and SPI2 hospitals	Validated structured questionnaire before and after intervention phase of SPI2	1) Comparison of control versus SPI2 hospitals: a. At baseline b. Over time i.e. difference in difference 2) Comparisons within control and SPI2 cohorts
2) Quality of care: acute medical care	Measure effects of SPI on the quality of care being delivered using independent case note reviews in acute medical care (both explicit and holistic).	Control and SPI2 hospitals	Before and after intervention phase of SPI2	1) Comparison of control versus SPI2 hospitals: a. At baseline b. Over time i.e. difference in difference (epochs 1+2 vs. epoch 3)*
3) Quality of care: perioperative care	Measure effects of SPI on the quality of care using independent case note reviews in perioperative care (explicit).	Control and SPI2 hospitals	Before and after intervention phase of SPI2	Comparison of control versus SPI2 hospitals: a. At baseline b. Over time i.e. difference in difference (epoch 2 vs. epoch 3)*
4) Clinical process measures	Indirect measure of hand hygiene by counting used hygiene consumables.	Control and SPI2 hospitals	Trend data collected as part of the National Observation Study of Effectiveness of the national Clean Your Hands Campaign study	1) Comparison of control versus SPI2 hospitals: a. At baseline b. Over time i.e. difference in difference
5) Outcomes	Measure effects of SPI on: a. Adverse events among acute medical care case notes reviewed b. Mortality among acute medical care case notes reviewed c. Hospital wide mortality d. ICU outcomes e. HCAI rates f. Patient satisfaction	Control and SPI2 hospitals	Before and after study using: a) and b) case notes c) d) and e) routine data f) validated structured questionnaire	1) Comparison of SPI2 versus control hospitals: a. At baseline b. Over time i.e. difference in difference 2) Comparisons within SPI2 and control cohorts 3) Measurement of reliability and learning/ fatigue effects 4) Hospital wide mortality rates not included in original protocol

*Sub-studies involving case note review that overlapped with SPI1 have two pre-intervention phases (epochs 1+2), while sub-studies specific to SPI2 have only one pre-intervention phase (epoch 2). In all cases epoch 3 is the post-intervention phase.

2.3 Sub-study 1: Staff surveys

All hospitals in England participate in the national staff survey, a yearly survey run by the Care Quality Commission (formerly the Healthcare Commission).

All nine control sites and nine SPI2 sites were included in both the 2006 and 2008 national staff surveys, conducted between October and December in each of these years, and so data from these surveys were used to test for effects of the intervention.

Questionnaires were sent to a simple random sample of 850 staff in each hospital trust, as this is the standard methodology employed in the survey. A sample size of 850 is such that an average 60% response rate – around 500 responses per site – would yield 95% confidence intervals of no greater than 10% for all scores within a single organisation.

The detail of the survey methods is not repeated here but is available from the staff survey website (www.nhsstaffsurveys.com).

Approximately 28 survey items are regularly collected on behalf of the Care Quality Commission (although the precise number has varied from year to year according to the content of the questionnaires).

Of these, 13 items (table 2.2) were identified at the start of the evaluation as being of likely relevance to the SPI programme. This was either because they reflect safety issues directly or because they relate to working practices known from research to be linked to safety and health outcomes. Eleven of these scores were the same as those used in the SPI1 evaluation. A further two that were clearly relevant to the SPI programme, but had not been available at the earlier evaluation period, were also included.

Details of these questions and how they are calculated can be found in appendix 1.[2;3]

Differences between the control and SPI2 hospitals, in terms of changes between the two survey periods, were tested using a generalised linear mixed model with SPI2/control and survey period as fixed factors (with interaction), and hospital as a random factor.

Table 2.2: Staff survey items deemed relevant to the SPI

1.	Well-structured appraisals[2;3]
2.	Working in well-structured teams[4]
3.	Witnessing potentially harmful errors or near misses in previous month
4.	Suffering work-related injury
5.	Suffering work-related stress
6.	Experiencing physical violence from patients/relatives
7.	Intention to leave
8.	Job satisfaction
9.	Quality of work-life balance
10.	Support from supervisors
11.	Organisational climate[5]
12.	Fairness and effectiveness of incident reporting procedures*
13.	Availability of hand-washing materials*

* These scores were not included in the SPI1 evaluation.

In order to control for known differences between groups of staff, the following background factors were included as covariates in the models:

- age
- sex
- ethnic background (white or other)
- occupational group (nursing/midwifery, medical/dental, allied health professional/scientific & technical, admin/clerical, general management, maintenance/ancillary, or other)
- length of service
- management status (line manager or not).

A statistical correlation for multiple observations was not applied but the confidence intervals were set at 0.99 ($p<0.01$).

2.4 Sub-study 2: Error rates/quality of care – acute medical care

Case note selection criteria

Patients over the age of 65 with acute respiratory disease admitted to acute medical wards were selected as the focus for study for the following reasons:

– Improving recognition and response to acute deterioration in a patient's condition was a specific SPI target, and patients admitted with acute respiratory disease are at high risk of such deterioration[6;7]
– A number of specific evidence-based guidelines exist for this condition
– There is a high incidence of co-morbidities in people aged over 65, making this a high-risk population (as confirmed in the evaluation of SPI1) where the opportunity for error is high and hence where there should be headroom for improvement.

The areas of review included both those specifically targeted by the SPI, and those that might plausibly be expected to improve if an overall shift in organisational systems and culture related to patient safety had occurred.

Case note assembly (and statistical power calculation)

We collected case notes from both the nine control and nine SPI2 hospitals from time periods that both preceded (epochs 1 and 2) and followed (epoch 3) the SPI2 intervention period. The pre-implementation observations were spread over two epochs (epoch 1, October 2003 to March 2004 and epoch 2, October 2006 to March 2007) so that the sites participating in the SPI2 evaluation could also serve as controls for the preceding SPI1 evaluation. Epoch 3 (October 2008 to March 2009) was therefore the post-SPI2 period. The temporal change between epochs 1 and 2 was included as a fixed effect in the statistical models. Each six-month time period was made to correspond across the calendar to control for seasonal effects.

We aimed to analyse, using review against explicit criteria, 15 case notes from each control and SPI2 hospital per epoch (810 in total). This would give 80% power to detect effects summarised in table 2.3. For example, for a standard (such as measurement of respiratory rate at least six hourly) with a baseline compliance of 70%, the study is powered to detect an SPI associated improvement to 83% compliance, or a deterioration to 55%.

These calculations are appropriate for analysis in binary data where each patient is associated with a single opportunity for error. However, the power available to analyse prescribing errors will tend to be considerably greater than that in table 2.3 since the typical patient is associated with more than one medication order and thus has several opportunities for error. That said, some actions, such as use of blood culture in patients who may have blood stream infection, were contingent (did not apply to the whole sample) and less power would be available in such cases.

Table 2.3: Detectable effect sizes, at 5% significance and 80% power, for a sample with 135 case notes in each epoch at the intervention sites and 135 case notes in each epoch at the control sites

The assumed analysis adjusts for unexplained variation between hospitals.

Baseline proportion	Modified proportions detectable with 80% power	
0.05	0.14	0.00
0.10	0.21	0.02
0.15	0.27	0.05
0.20	0.34	0.09
0.25	0.39	0.13
0.30	0.45	0.17
0.35	0.50	0.21
0.40	0.56	0.25
0.45	0.61	0.30
0.50	0.65	0.35
0.55	0.70	0.39
0.60	0.75	0.44
0.65	0.79	0.50
0.70	0.83	0.55
0.75	0.87	0.61
0.80	0.91	0.66
0.85	0.95	0.73
0.90	0.98	0.79
0.95	1.00	0.86

Patients over 65 years of age and admitted with acute respiratory disease, primarily community-acquired pneumonia, exacerbation of chronic obstructive pulmonary disease (COPD) or acute asthma were included in the study (for rationale see case note selection criteria, p 7). The case notes from the first two or three patients who fulfilled the eligibility criteria were selected from each hospital in each month from each epoch.

For each case note, the admission of interest was photocopied and anonymised (with respect to the patient's name, hospital name and year of admission) by medical-record clerks in each hospital. Photocopied notes were despatched to Birmingham before being

Box 2.1: Components of an ideal respiratory history

- Duration of presenting symptoms
- Normal (pre-morbid) exercise tolerance
- Presence/absence of shortness of breath
- Presence/absence of orthopnoea
- Presence/absence of cough
- Whether or not cough was productive (if present)
- Smoking history taken
- Presence/absence haemoptysis
- Whether or not chest pain was present
- Occupation/previous occupation
- Pet ownership

sent to reviewers. In Birmingham, anonymisation was quality-assured, the notes were digitised and the year of admission was removed so that reviewers would be blinded to the epoch from which the case notes originated.

We audited the quality of anonymisation by asking the reviewer in the explicit review (see explicit case note review below) to note if the hospital of origin, the year of origin and the patient name had been recognised by the reviewer.

Explicit case note review

We developed a set of explicit criteria to define medical care for respiratory patients with reference to British Thoracic Society (BTS) guidelines,[8;9] the British National Formulary (BNF) (versions 53, 54 and 56 – the editions that covered the study period[10–12]) and expert opinion (consultant respiratory physicians from a teaching and a general hospital – see acknowledgements).

The areas of review and source of guidelines were:

- Quality of medical history-taking. Eleven items (box 2.1) were identified, using expert opinion, as constituting the ideal history for a patient admitted with acute respiratory disease

Table 2.4: Vital signs that should be recorded

	Admission	6 and 12 hours later
Temperature	✓	✓
Respiratory rate	✓	✓
Cyanosis/oxygen saturation	✓	–
Presence of confusion/mental state (new onset)	✓	–
Pulse	✓	✓
Blood pressure	✓	–
Oxygen saturation	–	✓

- Proportion of routine investigations (urea and electrolytes, chest x-ray and full blood count) ordered within six hours of a patient's admission (expert opinion – see above)
- Observations and signs of patient deterioration. The completeness with which patients vital signs were recorded (table 2.4) was evaluated on admission and then for the first and subsequent 6 hour time periods (BTS). Vital sign data that were recorded in the case notes constituted the numerator, while all vital signs that should have been recorded constituted the denominator
- Appropriate clinical response for abnormal vital signs was measured (table 2.5) (BTS)
- Investigating features of good care for specific classes of patients by:
  - Calculating the CURB score to determine the severity of community acquired pneumonia and hence appropriate antibiotic selection (box 2.2) (BTS, BNF)
  - Use of intravenous steroids for patients with acute exacerbations of asthma and COPD (BTS)
  - Measurement of peak flow in asthma patients (expert opinion)
  - To exclude hypercapnia in COPD patients, by performing arterial blood gases, before prescribing/administering oxygen (BTS).

Table 2.5: Appropriate clinical response for abnormal observations

Abnormal vital sign	Appropriate clinical response
Oxygen saturation <90, at any time	One of: Full blood gases within 2 hours Given oxygen if not on oxygen Doctor called or transferred to ICU if on oxygen
Blood pressure systolic <90	Both of: At least next six hours, hourly observations Blood culture
Sputum present	Sputum culture
Respiratory rate >20 at any time after admission	One of: Given oxygen (if not on oxygen) Doctor called (if on oxygen)
Temperature over 38° C – any episode	Blood culture
Failure to improve by 48 hours or subsequent deterioration	One of: Review by consultant Repeat chest x-ray White cell counted/repeated Appropriate addition of further antibiotics

Box 2.2: Assessment of severity of community acquired pneumonia using the CURB score

CURB score

Confusion: new mental confusion (defined as an Abbreviated Mental Test score of 8 or less)
Urea: raised >7 mmol/l
Respiratory rate: raised > 30/min
Blood pressure: low blood pressure (systolic blood pressure <90 mm Hg , diastolic blood pressure < 60 mm Hg).

Interpretation of CURB score

- Patients who have two or more 'core' adverse prognostic features are at high risk of death and should be managed as having severe pneumonia
- Patients who display one 'core' adverse prognostic feature are at increased risk of death. The decision to treat such patients as having severe or non-severe pneumonia is a matter of clinical judgement, preferably from an experienced clinician. This decision can be assisted by considering 'pre-existing' and 'additional' adverse prognostic features.

Influence on antibiotic therapy

Non-severe community-acquired pneumonia
Most patients can be adequately treated with oral antibiotics. Combined oral therapy with amoxicillin and a macrolide (erythromycin or clarithromycin) is preferred for patients who require hospital admission for clinical reasons. When oral treatment is contraindicated, recommended parenteral choices include intravenous ampicillin or benzylpenicillin, together with erythromycin or clarithromycin.

Severe community acquired pneumonia
Patients with severe pneumonia should be treated immediately after diagnosis with parenteral antibiotics. An intravenous combination of a broad spectrum b-lactamase stable antibiotic such as co-amoxiclav or a second generation (e.g. cefuroxime) or third generation (e.g. cefotaxime or ceftriaxone) cephalosporin together with a macrolide (e.g. clarithromycin or erythromycin) is preferred.

Rates of prescribing errors. The following definition was used:

> *'A clinically meaningful prescribing error occurs when, as a result of a prescribing decision or prescription writing process, there is an unintentional significant reduction in the probability of treatment being timely and effective or increase in the risk of harm when compared with generally accepted practice.'*[13]

Errors were identified using a previously developed pro forma.[14] SPI1 had identified reductions in the number of adverse effects related to anticoagulant therapy as a key aim (see Outcomes, below), so prescribing error in this area was investigated as a sub-category (as listed in section 2.8 of the BNF).

Finally, medicines reconciliation on admission was also a target of the SPI. We therefore examined failures to continue to prescribe medicines on the transition from primary to secondary care where no explanation for this was recorded in the notes.

All case notes were reviewed by a single reviewer (Maisoon Ghaleb) over the period November 2006 to November 2009. Ideally reviews would be conducted in a random sequence once all records had

been collected. This was not possible due to the time taken to collect the case notes and the reporting requirements of the evaluation. Therefore, to control for any learning or fatigue (or both) effect on the part of the reviewer, the case notes were scrambled to ensure that the notes were not reviewed entirely in series and in particular, so that the same hospitals and epochs were not examined in series.

Generalised linear mixed models were used to analyse the effect of the SPI intervention. Within all models, pre-intervention levels were estimated by pooling data from the first two epochs and post-intervention levels were estimated using data from the third epoch. Fixed effects were included:

- for differences in pre-intervention levels between control and SPI2 hospitals (baseline comparisons)
- for temporal changes between epochs 1 and 2 across all hospitals.
- the temporal change experienced in the control hospitals between the pre-intervention period (i.e. epochs 1 and 2 pooled together) and the post-intervention period (epoch 3)
- the effect of the SPI, interpreted as the difference between the temporal changes pre/post intervention experienced in the control and SPI2 hospitals.

Adjustment for the patient-level covariates, age and sex was included in all analyses. Cubic polynomials at the time of review were used to adjust for learning/fatigue effects in the review process and were included in all analyses save that for mortality. Binary observations were modelled using mixed effects logistic regressions with a random component for variation between hospitals. Medication errors (per recorded prescription) were analysed with population-averaged negative binomial models with grouping by hospital, fitted using generalised estimating equations.

Where the data were insufficient to support a full analysis as described here, the hospital effects were excluded from the model leading to logistic regression analyses (for binary data) and negative binomial regression models (for prescribing errors.) The calculations were performed in STATA 11.0. Statistical significance is claimed for p-values less than 0.01, and 99% confidence intervals are used throughout.

Holistic case note review

In addition to the explicit review, each case note was evaluated holistically (implicit review) by a specialist in general medicine (M Clare Derrington). M Clare Derrington has considerable experience in case note review and has investigated hospitals who were outliers on hospital mortality statistics.[15] To measure inter-observer reliability, a subset (n=74) was independently re-evaluated by an

Box 2.3: Definitions of error and adverse events

Error:

Undesirable event in healthcare management which could have led to harm, or did so, but which did not impact on duration of admission or lead to disability at discharge.

A failure to complete a planned action as it was intended or to adopt an incorrect plan.

Adverse Event:

Unintended injury or complication.

Prolonged admission, disability at discharge or death.

Caused by healthcare management rather than the disease process.

Poor outcomes, some of which are the result of preventable actions or poor plans.

experienced trainee in

Table 2.6: Classification of errors and adverse events

Category	Nature of the problem
Diagnosis/Assessment admission error	– failure to diagnose promptly/correctly – failure to assess patient's overall condition adequately (including comorbidities)
Hospital-acquired infection	– hospital-acquired infection
Technical/management	– technical problem relating to a procedure – problem in management/monitoring (including nursing and other professional care)
Medication/maintenance/test results	– failure to give correct/monitor the effect of medication – failure to maintain correct hydration/electrolytes – failure to follow up abnormal test
Clinical reasoning	– obvious failure of clinical reasoning
Discharge information	– information needed by GP not transferred at discharge for whatever reason

Note that a particular error/event could be assigned to more than one category. For example, a test result showing severe hyperthyroidism was ignored and this error could be classified under 'Medication/Maintenance/Test results' and 'Discharge information'.

respiratory medicine (Thirumalai Naicker). Using expert clinical judgement, an overall quality score was assigned, graded on a scale from one (unsatisfactory, an error had occurred) to 10 (very best care).

A specific score for each of three stages of care – admission, management and pre-discharge – was also allocated on a scale from one (unsatisfactory) to six (excellent care).

Reviewers recorded errors and adverse events using the definitions found in box 2.3.[16–20] The number of errors and adverse events (of all

types, not just those relating to medication) were recorded for each patient. It was possible for a patient to have more than one error or adverse event.

The results are presented as average numbers of errors or adverse events per 100 patients. Average ratings and average numbers of adverse events and errors were calculated for both control and intervention groups. Adverse events and errors were further classified by broad categories (table 2.6), and adverse events were also categorised into four levels of preventability: definitely preventable; preventable on balance of probabilities; not preventable on the balance of probabilities; and definitely not preventable.

A mixed modelling approach was used to test for differences in changes in outcomes between epochs 1 and 2, and epoch 3.

Random effects were included to allow for within hospital correlation, using an exchangeable correlation structure. Covariates included:

- binary variable 'after' indicating whether the observation was before or after the intervention period
- binary variable 'intervention' indicating whether the hospital was a control or SPI2 hospital
- binary variable 'epoch 1 (or 2)' indicating whether the observation was from the pre-intervention phase
- an interaction between 'after' and 'intervention', to evaluate the estimated difference in change between the control and SPI2 hospitals (between epoch 3 and the average of the pre-intervention epochs).

All models were adjusted for age and sex of patients.

For the adverse events and errors, inter-observer reliability was assessed comparing errors and adverse events identified by both reviewers, using the Kappa statistic.

2.5 Sub-study 3: Error rates/quality of care – perioperative care

Case note selection

Patients undergoing major surgical operations of two types (total hip replacement and open colectomy) were selected for the following reasons:

- improving perioperative care was a specific SPI2 target
- specific guidelines apply to this group of patients
- it was believed that compliance with the guidelines was poor.

We developed a set of explicit criteria for perioperative care using clinical guidelines from IHI[21], British Orthopaedic Association[22] and the National Institute for Health and Clinical Excellence (NICE).[23;24] The areas of review were as follows:

- Administration of prophylactic antibiotics prior to inclusion.
- The use of prophylactic deep vein thrombosis (DVT) treatment (unless contraindicated), which included pharmacological intervention (unfractionated or low molecular weight heparins) and/or mechanical interventions, such as anti-thromboembolism stockings, foot pumps and sequential compression devices.
- Intra-operative temperature monitoring (on at least one occasion).
- The use of advanced methods of pain control (epidural anaesthesia and/or patient controlled analgesia) for post-operative pain control. It was decided to look at the types of anaesthesia administered, as there is evidence that using neuraxial blocks (spinal and epidural) with sedation only or in combination with a general anaesthetic helps with early post-operative pain control and recovery. Likewise there is evidence to support the use of patient controlled analgesia (PCA). Our quality criterion was that at least one of the modalities (neuraxial block or PCA) should be used.

Within the SPI intervention, the IHI advocated the removal of hair by clipping (not shaving); as this standard is not routinely recorded, this was not included as a process measure for the evaluation.

Case note assembly

Again, notes were selected from nine control and nine SPI2 hospitals. In this case there was a single pre-intervention epoch (corresponding to epoch 2, that is October 2006 to March 2007) for comparison with the post-intervention epoch (corresponding to epoch 3, that is October 2008 to March 2009).

The intention was to analyse 10 case notes from each epoch (five of each surgical operation type) to yield a total sample of 360. To control for seasonal effects the case notes were spread across each time period (approximately two per month).

The anonymisation procedures used in the sub-study dealing with the management of the acutely sick respiratory patients was followed (see section Case note assembly (and statistical power calculation), p 8).

All case notes were reviewed by a single medically trained reviewer (Ugochi Nwulu) over a period from November 2009 to January 2010. The first 20 cases were read jointly by Ugochi Nwulu and Richard Lilfordand each one was discussed for training purposes.

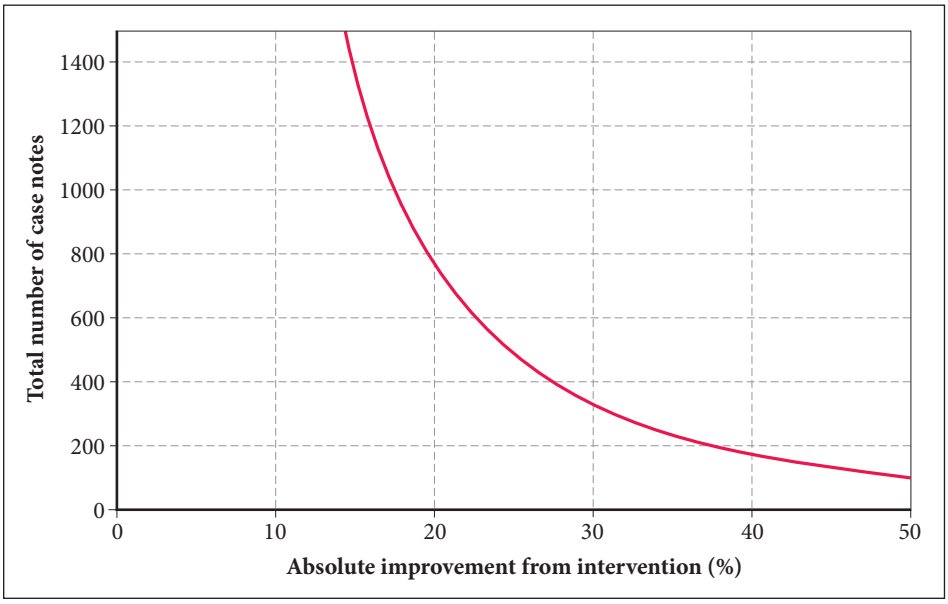


Figure 2.1: Sample sizes for 80% power (at 5% significance)

The notes were partially scrambled over epochs to assess, and if necessary control, for learning/fatigue effects. Inter-rater agreement was measured using 27 case notes reviewed by a second reviewer (Amit Kotecha), a surgical trainee.

Sample size calculation

We performed the sample size calculation after analysing results for 42 case notes. We found high compliance (>90%) with the venous thrombo-prophylaxis and antibiotic criteria such that there was little headroom for post intervention improvement.

We therefore based the calculation on intra-operative temperature monitoring where compliance was about 40% at baseline (that is, there was plenty of room for improvement in response to SPI).

Table 2.7: Sample sizes for 80% power (at 5% significance)

Effect Size (%)	Total number of cases needed for 80% power
15	1,364
20	764
22.5	600
25	484
30	328
35	236

Assuming that control hospitals experience an improvement from 40% to 50% compliance over the study period, our sample (n=360) is sufficient to detect an additional 25% to 30% improvement in association with SPI at 80% power, see figure 2.1 and table 2.7.

2.6 Sub-study 4: Indirect measure of hand hygiene

Improvement in hand hygiene was a specific aim of the SPI intervention.

In the UK there has also been a national initiative to improve hand hygiene amongst acute hospital employees – the Clean Your Hands campaign.[25]

This initiative consisted of actions to make AHR available at the bedside, monthly updated posters on wards and a patient empowerment component to encourage patients to ask staff to clean their hands.

The campaign was rolled out in England and Wales between December 2004 and June 2005 and continues to date. Since hand hygiene is also an SPI target we tested the hypothesis that SPI would have an additive effect.

The success of this campaign was measured by the National Observational Study to Evaluate the Clean Your Hands campaign (NOSEC).[26] As part of their study, monthly data from NHS Logistics for soap and AHR consumption (litres) was collected as an indirect measure of hand hygiene compliance. Data were available on a monthly basis for the period July 2004 to September 2008. This spanned a before period (July 2004 to February 2007) and a period concurrent with the intervention (March 2007 to September 2008). To adjust for potential variations in consumption due to hospital size, these data, which were available at hospital trust level and were expressed as a rate (in litres) per 1,000 bed occupied days.

Bed occupancy days were based on yearly averages spanning financial years.[27]

Population averaged (marginal) models were used to used to assess the effects of the intervention on soap and AHR consumption. To allow for decays in correlations (within hospitals) over time, an auto-regressive (AR 3) correlation structure was included. Model fits were compared between log and identity scales, and results presented here are based on the identity scale (as this allows estimation of difference in change).

Covariates within the models included an indicator variable denoting intervention or control hospital and time as a continuous variable (from one to maximum number of temporal observations available). The effect of time was modelled as a polynomial function (cubic) as there was an indication that changes in rates were non-linear.

Finally, a fixed effect interaction between time and intervention allowed assessment of whether the change in rates of infection differed between control and SPI2 hospitals.

Both models were fitted in STATA using the GEE population averaged class of models. For the before and after comparisons, estimates of differences in differences (as estimated by the GEE models) are presented along with 99% confidence intervals. For the temporal models, smoothed estimates of outcomes over the study period are presented in graphical format, along with p-values for tests of significant differences in changes between control and SPI2 hospitals.

Models were weighted with a suitably appropriate denominator – either number of events or standard deviation of outcome for summary data.

2.7 Sub-study 5: Outcomes

Adverse events detected in acute medical case notes

SPI2 aimed to make a 30% reduction[28;29] in the total number of adverse events. The incidence of patient harm caused by medication was measured as part of the explicit review.

The holistic review also measured adverse events both overall and by degree of preventability. In addition, each death was re-analysed by a second reviewer (blind to epoch and group), who had been trained in anaesthesia and public health, and who had experience as a reviewer of deaths for the National Confidential Enquiry into Perioperative Deaths (CL).

This study of deaths was not included in the original protocol and was added as a further quality control procedure after completion of the data collection.

Rates of mortality among acute medical care patients

We compared mortality rates across pre and post-intervention epochs, among patients whose case notes were selected for review. This was because this was feasible and, arguably, a higher signal to noise ratio would be expected among this group, which not only was especially well placed to benefit from specific SPI interventions, but also tends to have high mortality.

Hospital-wide mortality

This analysis was not part of the original protocol and was added at a later stage. The standardised mortality rates were derived from discharge information captured by Hospital Episode Statistics (HES).

The analysis included the discharge episodes of all patients aged 15 and over where the patient classification was coded as one. This excluded day cases, regular attendees for recurrent treatments such as dialysis and chemotherapy, or patients attending to give birth.

The purpose of the exclusions was to reduce the extent to which the denominator of discharged patients was inflated with low-risk episodes in those units having large day-case suites or maternity units. All in-year discharges were analysed and the rates of those discharged dead were directly standardised within sex and quinary age groups using a reference population of total discharges in each age and sex group.

We used HES records for intervention and control hospitals for financial years 2002/03 to 2008/09 inclusive.

ICU: Mortality, morbidity and length of stay

To provide information relevant to the effectiveness of the critical care bundles, we accessed data from the Case Mix Programme (CMP)[30] – a comparative audit run by the Intensive Care National Audit and Research Centre (ICNARC).

This programme collects patient outcomes from adult, general critical care units (intensive care and combined intensive care/high dependency units) covering England, Wales and Northern Ireland. Critical care units volunteered to join and collect standardised datasets (case mix, patient outcome and activity data) on patients admitted to their unit. These data are submitted to ICNARC for validation and analyses.

Data for the ICUs for all the study hospitals were available on a monthly basis for six months prior to the SPI (from October 2006 to March 2007) and for six months after the intervention (from October 2008 to March 2009).

Mortality data were available on the observed numbers of deaths and the risk-adjusted number of deaths, both of which were used to calculate observed to expected mortality ratios. Information was also available on the mean length of stay in the unit, along with standard deviation.

Finally, data were available on the mean risk prediction scores: the APACHE II score31 and the ICNARC score32 for patients admitted directly from a ward (along with standard deviation).

For data on intensive care outcomes, a mixed modelling population averaged approach was again used to provide information relevant to the effects of the intervention. However, since these data were only available for a single six-month period prior to the intervention, and for a single six-month period after the intervention (continuous time series data throughout the study period were not available), these data were modelled using a simple difference of difference model (that is, not including time as a continuous variable and not including an auto-regressive component).

Covariates within the model included an indicator variable denoting control or SPI2 hospital, and an indicator variable denoting before or after the intervention. Correlations within hospitals were incorporated using an exchangeable correlation structure. Adjustment was made for the morbidity covariates, mean APACHE II score and mean ICNARC physiology score.

Finally, a fixed effect interaction between intervention and before/after period allowed assessment of whether the change in outcomes between the before and after period differed between control and SPI2 hospitals.

All models were fitted in STATA using the GEE population averaged class of models. For the before and after comparisons, estimates of differences in differences (as estimated by the GEE models) are presented along with 99% confidence intervals.

Full results from fitted GEE models are provided in appendix 4.

C. diff and MRSA infection rates

Several components of the SPI intervention are related to infection control. We obtained the numbers of all C. diff and MRSA bacteraemia associated diarrhoea in the study sites from the Health Protection Agency (HPA), which collects mandatory HCAI data from all acute trusts in England and Wales.

The C. diff and MRSA data relate to both community and hospital-based infections (that is, they include cases diagnosed within the first 48 hours of stay) in patients older than 65 years.

C. diff data were available quarterly for the period January 2004 to June 2009. MRSA data were available from April 2001 to September 2009. These data therefore spanned a pre-intervention period (April 2001 or January 2004 to March 2007), a period concurrent with the intervention (April 2007 to September 2008) and a post-intervention period (October 2008 to June 2009 or September 2009).

To adjust for potential variations in numbers of cases due to hospital size, these data were expressed as a rate per 1,000 bed occupancy days for C. diff infections and as a rate per 100,000 bed occupancy days for the MRSA infections. Bed occupancy days were based on yearly averages spanning financial years.

Population averaged (marginal) models were used to assess the effects of the intervention on rates of C. diff and MRSA infections. To allow for decays in correlations (within hospitals) over time, an auto-regressive (AR 3) correlation structure was included.

Model fits were compared between log and identity scales, and results presented here are based on the identity scale (as this allows estimation of difference in change).

Covariates within the models included an indicator variable denoting control or SPI2 hospital, and time as a continuous variable (from one to maximum number of temporal observations available). The effect of time was modelled as a polynomial function (cubic) as there was an indication that changes in rates were non-linear.

Finally, a fixed effect interaction between time and intervention allowed assessment of whether the change in rates of infection differed between control and SPI2 hospitals.

Both models were fitted in STATA using the GEE population averaged class of models. For the before and after comparisons, estimates of differences in differences (as estimated by the GEE models) are presented along with 99% confidence intervals. For the temporal models, smoothed estimates of outcomes over the study period are presented in graphical format, along with p-values for tests of significant differences in changes between control and SPI2 hospitals.

Full results from fitted GEE models are provided in appendix 4.

Patient surveys

Since quality of care and avoidance of adverse events are important to patients, improvements in practice might plausibly affect patients' views of their care. Their views were assessed by means of a patient survey.

All English hospitals participate in the Care Quality Commission's National NHS Acute Inpatient Survey in England. The detail of this methodology is available from www.nhssurveys.com

Data were collected in October to December 2006 (pre-intervention) and October to December 2008 (post-intervention). Methods similar to those for the staff survey were used in the analysis, except that the control variables included were sex, age, length of stay and whether the admission was emergency or elective.

Five scores (table 2.8) were identified for analysis: three overall satisfaction scores and two related to cleanliness. The details of these scores can be found in appendix 2.

Table 2.8: Patient survey questions deemed relevant to the SPI

1.	Overall, how would you rate the care you received?
2.	How would you rate how well the doctors and nurses worked together?
3.	Overall, did you feel you were treated with respect and dignity while you were in the hospital?
4.	In your opinion, how clean was the hospital room or ward that you were in?
5.	How clean were the toilets and bathrooms that you used in hospital?

Chapter 3

Results

3.1 Sub-study 1: Staff surveys

In the nine SPI2 hospitals, the overall response rate for the first, before, survey was 53% (3,957 of 7,402 valid questionnaires returned).

This rate remained the same for the second, after, survey (3940/7448). In the nine control hospitals, the response rates were 50% (3,634/7,301) and 49% (3,616/7,424) respectively.

Table 3.1 shows the changes in both control and SPI2 hospitals on each of the 13 scores identified, along with the differences between the groups in these changes (with associated 99% confidence intervals).

Comparison with control hospitals is important because national changes in the NHS over this period resulted in generally more positive scores from the second survey than from the first.[34]

Only one of the 13 scores (organisational climate) shows a statistically significant ($p<0.01$) change over time between the control hospitals and SPI2 hospitals. Organisational climate, which refers to extent of positive feeling within the organisation relating to communication, staff involvement, innovation and patient care, was significantly lower in the control hospitals than the SPI2 hospitals at baseline (2.79 versus 2.91 on a scale where 1 is very negative and 5 is very positive).

Thus, although the increase in this score in control hospitals was higher than in SPI2 hospitals (0.08 compared with 0.01), the score was still higher in SPI2 hospitals at the second survey. The effect size for this difference in change between the control and SPI2 hospitals after covariates are taken into account was modest, at 0.07 points on a five point scale where there was a range at baseline of 0.55 points between hospitals.

Table 3.1: Staff survey scores in control and SPI2 hospitals at the two periods*

	Control hospitals					SPI2 hospitals					Range at baseline	Difference in change (99% CI)	p-value
	N	Survey 1 score (SE)	N	Survey 2 score (SE)	Absolute % change	N	Survey 1 score (SE)	N	Survey 2 score (SE)	Absolute % change			
% staff having well-structured appraisals within previous 12 months[2;3]	3477	28 (1)	3429	28 (1)	-1	3783	28 (1)	3734	26 (1)	-2	20–39	3 (-3, 9)	0.191
% staff working in well-structured teams[4]	3498	36 (1)	3408	37 (1)	1	3781	38 (1)	3747	38 (1)	0	32–42	4 (-4, 12)	0.205
% staff witnessing potentially harmful errors or near misses in previous month	*3602*	*37 (1)*	*3532*	*33 (1)*	*-4*	*3918*	*41 (1)*	*3851*	*40 (1)*	*-1*	*32–47*	*-4 (-10, 3)*	*0.167*
% staff suffering work-related injury in previous 12 months	*3524*	*19 (1)*	*3490*	*16 (1)*	*-3*	*3848*	*19 (1)*	*3796*	*18 (1)*	*-1*	*16–23*	*-2 (-5, 2)*	*0.182*
% staff suffering work-related stress in previous 12 months	*3575*	*33 (1)*	*3532*	*27 (1)*	*-6*	*3882*	*32 (1)*	*3842*	*27 (1)*	*-6*	*26–40*	*-1 (-6, 5)*	*0.670*
% staff experiencing physical violence from patients/relatives in previous 12 months	*3598*	*11 (1)*	*3536*	*11 (1)*	*-1*	*3884*	*11 (1)*	*3849*	*11 (1)*	*0*	*7–16*	*-1 (-3, 3)*	*0.645*
Intention to leave[5]	*3557*	*3.26 (0.02)*	*3544*	*3.40 (0.02)*	*0.14*	*3880*	*3.31 (0.01)*	*3865*	*3.42 (0.01)*	*0.11*	*3.07–3.50*	*-0.04 (-0.12, 0.04)*	*0.198*
Staff job satisfaction[5]	3593	3.34 (0.01)	3568	3.44 (0.01)	0.10	3902	3.40 (0.01)	3898	3.49 (0.01)	0.09	3.23–3.50	-0.02 (-0.08, 0.04)	0.422
Quality of work-life balance[5]	3568	2.77 (0.02)	3536	2.56 (0.02)	-0.22	3868	2.68 (0.02)	3857	2.51 (0.02)	-0.17	2.46–2.97	0.05 (-0.04, 0.14)	0.142
Support from supervisors[5]	3583	3.39 (0.02)	3551	3.56 (0.02)	0.17	3894	3.43 (0.01)	3869	3.61 (0.01)	0.18	3.22–3.53	0.00 (-0.08, 0.07)	0.889
Organisational climate[5;33]	3578	2.79 (0.01)	3551	2.87 (0.01)	0.08	3861	2.91 (0.01)	3886	2.92 (0.01)	0.01	2.52–3.07	-0.07 (-0.14, 0.00)	0.009
† Fairness and effectiveness of incident reporting procedures[5]	3555	3.36 (0.01)	3487	3.41 (0.01)	0.05	3861	3.41 (0.01)	3803	3.45 (0.01)	0.04	3.27–3.54	-0.01 (-0.05, 0.04)	0.664
† Availability of hand-washing materials[5]	2939	4.58 (0.01)	3126	4.75 (0.01)	0.17	3231	4.51 (0.01)	3418	4.67 (0.01)	0.16	4.32–4.72	-0.01 (-0.07, 0.04)	0.587

* The first six of these scores were percentages, simply reflecting the percentage of respondents who answered 'yes' to a single question or a set of questions. The other seven are on a scale of one to five, and are based on the mean of between three and six questions, each of which was scored between one and five for each respondent. For six of these seven scores, the higher the score the better, although for intention to leave, lower scores are better. To aid interpretation, scores where a lower value is better are shown in *italics*. Range at baseline indicates the range of scores across SPI and control hospitals in the first survey to give some context for the level of change shown. The difference in change and corresponding confidence interval does not necessarily reflect the difference in absolute change because of the inclusion of covariates in the models tested.

† These scores were not included in the SPI1 evaluation.

3.2 Sub-study 2: Error rates/quality of care – acute medical care

Explicit review

The intended sample size of 405 from the SPI2 hospitals was not met – 347 case notes were reviewed. These case notes were split approximately equally across the epochs – 116 from epoch 1, 117 from epoch 2 and 114 from epoch 3. Control hospitals yielded 355 case notes out of the intended sample size of 405: 120 from epoch 1, 123 from epoch 2 and 112 from epoch 3.

History taking (tables 3.2a and 3.2b)

Baseline comparisons showed no significant differences between control and SPI2 hospitals. An effect of SPI was not apparent and was not statistically significant for any of the outcomes measured.

For two items (exercise tolerance and occupation) measured in relation to history taking, there was significant evidence of an improvement overtime in both control and SPI2 hospitals (see table 3.2b). There was some evidence of a reviewer learning/fatigue effect for exercise tolerance ($p<0.001$), chest pain ($p=0.010$) and occupation ($p=0.001$).

Several of the questions were asked less often for older patients. Age was a significant predictor for items 3, 6 and 7 ($p\leq0.001$ in all cases), typically reducing the odds of the question being asked by about 5% per year of age.

Vital signs (tables 3.3a and 3.3b)

There is no significant evidence for an effect associated with SPI. However, compliance in taking patient observations at six and 12 hours after admission also improved in both groups of hospitals when epochs 1 and 2 are compared to epoch 3.

This was most evident for respiratory rate where practice continued to improve across all three epochs. In addition, improvement took place between the first two epochs on these and most of the other six and 12 hour items ($p<0.010$ for all items except for six hour pulse, for which $p=0.016$).

Appropriate clinical response (tables 3.4a and 3.4b)

The data are sparse, and formal analysis was possible for only three items (see table 3.4b). No significant conclusions were indicated.

Steroids and antibiotics – compliance with standards (tables 3.5a and 3.5b)

There is no significant evidence that the SPI had an effect. Use of the CURB score (a clinical prediction rule for predicting mortality from community-acquired pneumonia and infection at any site) has improved significantly over time (OR=7.3; 1.4 – 37.7), though from a very low base, and differences were not statistically significant between control and SPI2 hospitals.

A negative age-effect ($p<0.001$) was apparent for item four yielding a reduction in odds of compliance of about 6% per year of age. There is a reviewer learning effect ($p=0.002$) for item 2 (oxygen prescription for COPD).

Prescribing errors (tables 3.6a and 3.6b)

A reviewer learning/fatigue effect was significant ($p=0.009$) in the review of prescribing errors, with a decreasing rate of error detection with time of review; this was allowed for in the analysis. No significant time effects for SPI arm, time or SPI were detected (table 3.6b).

Anti-coagulant prescribing errors (table 3.7)

A total of 10 errors were recorded. Six occurred in SPI2 hospitals before the introduction of the intervention, the other four in control hospitals in epoch 3. The breakdown is shown in table 3.7, but no further analysis was possible.

Reconciliation errors (table 3.8a and 3.8b)

The results can be found in tables 3.8a and 3.8b. Again, there is no significant evidence that the SPI has an effect ($p=0.914$).

Table 3.2a: Medical history taking (% of patients asked)

	Control hospitals						SPI2 hospitals					
	Epoch 1		Epoch 2		Epoch 3		Epoch 1		Epoch 2		Epoch 3	
No. of patients	236		240		381		380		381		380	
	%	SE	%	SE	%	SE	%	SE	%	SE	%	SE
1. Duration of presenting symptom	92.5	2.4	91.1	2.6	95.5	2.0	96.6	1.7	98.3	1.2	99.1	0.9
2. Normal exercise tolerance	26.5	4.1	31.7	4.2	38.4	4.6	38.8	4.5	38.3	4.6	33.9	4.5
3. Presence/absence shortness of breath	88.3	2.9	91.1	2.6	88.4	3.0	91.4	2.6	93.2	2.3	92.0	2.6
4. Presence/absence orthopnoea	23.3	3.9	28.1	4.1	17.0	3.6	32.8	4.4	29.3	4.2	18.0	3.7
5. Presence/absence cough	88.3	2.9	89.4	2.8	86.6	3.2	91.4	2.6	91.5	2.6	83.9	3.5
6. If cough, was it productive	78.3	3.8	84.6	3.3	77.7	4.0	87.1	3.1	88.0	3.0	76.8	4.0
7. Smoking history taken	73.9	4.0	81.3	3.5	66.1	4.5	77.6	3.9	79.5	3.7	74.1	4.2
8. Presence/absence of haemoptysis	22.2	3.9	28.1	4.1	16.1	3.5	25.2	4.1	23.3	3.9	26.1	4.2
9. Chest pain (of any type)	68.1	4.3	71.5	4.1	54.5	4.7	54.3	4.6	65.5	4.4	59.8	4.7
10. Occupation/previous occupation	44.4	4.6	37.7	4.4	53.6	4.7	34.8	4.5	38.5	4.5	38.4	4.6
11. Pets	3.4	1.7	3.3	1.6	0.9	0.9	1.7	1.2	2.6	1.5	6.3	2.3
% over all items		55.7		58.2		54.1		57.5		59.0		57.4

Entries are percentages with binomial standard errors.

Darker shaded areas relate to post-intervention epochs.

Table 3.2b: Medical history taking – differences between control and SPI2 hospitals, changes over time and the effect of SPI2

	Baseline comparisons		Changes in controls		Effect of SPI2	
	OR (99% CI)	p-value	OR (99% CI)	p-value	OR (99% CI)	p-value
1. Duration of presenting symptom	3.2 (0.7, 14.0)	0.040	1.6 (0.4, 7.3)	0.391	1.7 (0.07, 40.3)	0.672
2. Normal exercise tolerance	1.4 (0.8, 2.4)	0.125	2.2 (1.1, 4.4)	0.005	0.7 (0.3, 1.7)	0.312
3. Presence/absence shortness of breath	1.3 (0.5, 3.5)	0.480	0.8 (0.3, 2.3)	0.539	1.3 (0.3, 5.7)	0.701
4. Presence/absence orthopnoea	1.3 (0.7, 2.5)	0.330	0.6 (0.3, 1.5)	0.159	0.9 (0.3, 2.6)	0.749
5. Presence/absence cough	1.2 (0.5, 2.9)	0.506	0.7 (0.2, 1.8)	0.286	0.7 (0.2, 2.4)	0.407
6. If cough, was it productive	1.4 (0.7, 2.9)	0.208	0.7 (0.3, 1.6)	0.307	0.7 (0.2, 2.1)	0.418
7. Smoking history taken	1.1 (0.5, 2.1)	0.841	0.6 (0.3, 1.2)	0.061	1.5 (0.5, 4.0)	0.313
8. Presence/absence of haemoptysis†	0.9 (0.4, 1.9)	0.686	0.6 (0.2, 1.4)	0.106	2.2 (0.7, 6.5)	0.061
9. Chest pain (of any type)	0.6 (0.4, 1.1)	0.041	0.7 (0.4,1.4)	0.193	2.1 (0.9, 5.2)	0.028
10. Occupation/previous occupation	0.9 (0.5, 1.7)	0.696	2.0 (1.0, 4.0)	0.010	0.6 (0.3, 1.5)	0.178
11. Pets	0.9 (0.2, 4.7)	0.872	0.3 (0.02, 5.6)	0.299	8.3 (0.3, 210.0)	0.093

† Denotes items with significant ($P < 0.010$) between hospital variation within the arms of the study.

Table 3.3a: Vital signs – percentage compliance with standards

	Control hospitals						SPI2 hospitals					
	Epoch 1		Epoch 2		Epoch 3		Epoch 1		Epoch 2		Epoch 3	
	%	SE	%	SE	%	SE	%	SE	%	SE	%	SE
On admission:												
Temperature	96.7	1.6	99.2	0.8	99.1	0.9	99.1	0.9	99.1	0.9	96.5	1.7
Respiratory rate	95.8	1.8	99.2	0.8	100.0	0.0	96.5	1.7	98.3	1.2	100.0	0.0
Cyanosis/oxygen saturation	98.3	1.2	98.4	1.1	100.0	0.0	99.1	0.9	99.1	0.9	100.0	0.0
Confusion/mental state	53.3	4.6	71.5	4.1	74.1	4.2	62.6	4.5	57.3	4.6	80.7	3.7
Pulse	98.3	1.2	99.2	0.8	100.0	0.0	99.1	0.9	99.1	0.9	100.0	0.0
Blood pressure	98.3	1.2	99.2	0.8	100.0	0.0	99.1	0.9	99.1	0.9	100.0	0.0
At six hours:												
Temperature	61.7	4.5	69.9	4.2	69.6	4.4	63.2	4.5	77.8	3.9	68.1	4.4
Respiratory rate	40.8	4.5	69.1	4.2	72.3	4.2	47.4	4.7	76.1	4.0	77.9	3.9
Pulses	69.2	4.2	73.2	4.0	75.0	4.1	64.9	4.5	81.2	3.6	79.6	3.8
Oxygen saturation	61.7	4.5	71.5	4.1	74.1	4.2	60.5	4.6	78.6	3.8	79.6	3.8
At 12 hours:												
Temperature	58.3	4.5	70.7	4.1	68.8	4.4	58.8	4.6	69.8	4.3	72.6	4.2
Respiratory rate	35.0	4.4	69.9	4.2	73.2	4.2	44.7	4.7	67.5	4.3	78.8	3.9
Pulse	63.3	4.4	76.4	3.8	75.0	4.1	59.6	4.6	70.9	4.2	79.6	3.8
Oxygen saturation	54.2	4.6	75.6	3.9	74.1	4.2	57.0	4.7	70.9	4.2	79.6	3.8
Routine investigations:												
U & E	99.2	0.8	98.4	1.1	99.1	0.9	100.0	0.0	99.1	0.9	100.0	0.0
Chest X-ray	96.7	1.6	97.6	1.4	97.3	1.5	96.5	1.7	98.3	1.2	100.0	0.0
Full blood count	98.3	1.2	97.6	1.4	99.1	0.9	99.1	0.9	99.1	0.9	100.0	0.0

Darker shaded areas relate to post-intervention epochs.

Table 3.3b: Vital signs – differences between control and SPI2 hospitals, changes over time and the effect of SPI2

	Baseline comparisons		Changes in controls		Effect of SPI2	
	OR (99% CI)	p-value	OR (99% CI)	p-value	OR (99% CI)	p-value
On admission:						
Temperature	2.2 (0.2, 21.1)	0.381	0.7 (0.02, 24.0)	0.823	0.1 (0.002, 4.1)	0.108
Respiratory rate	0.7 (0.1, 3.9)	0.617	–	–	–	–
Cyanosis/oxygen saturation	1.6 (0.1, 18.2)	0.605	–	–	–	–
Confusion/mental state	0.9 (0.5, 1.7)	0.674	1.8 (0.8,3.7)	0.045	1.7 (0.6, 4.5)	0.187
Pulse	1.1 (0.1, 14.9)	0.942	–	–	–	–
Blood pressure	1.1 (0.1, 14.9)	0.942	–	–	–	–
At six hours:						
Temperature	1.3 (0.7, 2.4)	0.323	1.4 (0.7, 2.8)	0.239	0.8 (0.3, 1.9)	0.457
Respiratory rate	1.3 (0.7, 2.5)	0.281	2.1 (1.0, 4.3)	0.010	1.0 (0.4, 2.8)	0.907
Pulse	1.1 (0.6, 2.1)	0.604	1.3 (0.6, 2.8)	0.327	1.2 (0.4, 3.3)	0.662
Oxygen saturation	1.2 (0.7, 2.2)	0.433	1.4 (0.7, 3.0)	0.223	1.2 (0.4, 3.1)	0.703
At 12 hours:						
Temperature	1.0 (0.6, 1.8)	0.934	1.2 (0.6, 2.4)	0.583	1.2 (0.5, 2.9)	0.685
Respiratory rate	1.2 (0.6, 2.3)	0.524	2.4 (1.1, 5.0)	0.002	1.2 (0.4, 3.1)	0.713
Pulse	0.8 (0.5, 1.4)	0.394	1.2 (0.6, 2.5)	0.510	1.5 (0.6, 4.1)	0.268
Oxygen saturation	1.0 (0.6, 1.7)	0.953	1.4 (0.7, 2.9)	0.231	1.4 (0.5, 3.6)	0.430
Routine investigations:						
U & E	0.9 (0.03, 28.8)	0.944	0.6 (0.01, 27.7)	0.762	–	–
Chest X–ray	1.1 (0.2, 5.1)	0.904	0.7 (0.1, 5.6)	0.641	–	–
Full blood count	1.6 (0.2, 16.9)	0.609	1.7 (0.1, 40.4)	0.663	–	–

No items showed significant variation between hospitals within arms.

Blanks are associated with 100% compliance in table 3.3b, for which logistic regression analysis is impossible.

Table 3.4a: Appropriate clinical response

	Control hospitals									SPI2 hospitals								
	Epoch 1			Epoch 2			Epoch 3			Epoch 1			Epoch 2			Epoch 3		
	N	%	SE	N	%	SE	N	%	SE	N	%	SE	N	%	SE	N	%	SE
Oxygen saturation <90 at any time:																		
Full blood gases within 2 hours	13	61.5	13.5	10	50.0	15.8	0	–	–	2	50.0	35.4	10	70.0	14.5	4	25.0	21.7
Given oxygen (if not on oxygen)	12	66.7	13.6	7	57.1	18.7	1	0.0	0.0	4	75.0	21.7	9	77.8	13.9	2	50.0	35.4
Doctor called or transferred to ICU (if on oxygen)	8	25.0	15.3	6	50.0	20.4	0	–	–	2	50.0	35.4	5	80.0	17.9	2	50.0	35.4
Blood pressure systolic <90:																		
At least next six hours, hourly observations	7	28.6	17.0	8	25.0	15.3	8	50.0	17.7	4	50.0	25.0	6	16.7	15.2	2	100.0	0.0
Blood culture	4	50.0	25.0	5	40.0	21.2	8	37.5	17.1	4	25.0	21.7	5	80.0	17.9	2	100.0	0.0
Sputum Present																		
Sputum culture	70	41.4	5.9	72	48.6	5.9	69	24.6	5.2	71	36.6	5.8	78	46.2	5.7	62	29.0	5.8
Respiratory rate >20 at any time after admission:																		
Given oxygen (if not on oxygen)	3	0.0	0.0	0	–	–	0	–	–	2	0.0	0.0	1	0.0	0.0	0	–	–
Doctor called (if on oxygen)	5	0.0	0.0	1	100.0	0.0	0	–	–	3	0.0	0.0	2	0.0	0.0	3	0.0	0.0
Temperature over 38°C – any episode:																		
If yes, blood culture	16	68.8	11.6	14	71.4	12.0	15	73.3	11.4	19	73.7	10.1	25	76.0	8.5	13	61.5	13.5
Failure to improve by 48 hours or subsequent deterioration:																		
Review by consultant	11	100.0	0.0	12	100.0	0.0	10	100.0	0.0	9	100.0	0.0	10	100.0	0.0	3	100.0	0.0
Repeat chest X-ray	10	100.0	0.0	9	100.0	0.0	9	100.0	0.0	8	100.0	0.0	8	100.0	0.0	3	100.0	0.0
White cell counted/repeated	11	100.0	0.0	12	100.0	0.0	11	100.0	0.0	8	100.0	0.0	10	100.0	0.0	3	100.0	0.0
Appropriate addition of further antibiotics	9	100.0	0.0	5	100.0	0.0	8	75.0	15.3	7	85.7	13.2	6	100.0	0.0	2	50.0	35.4
Follow up:																		
Arrange follow up?	45	71.1	6.7	47	61.7	7.1	38	42.1	8.0	49	59.2	7.0	52	63.5	6.6	44	38.6	7.3

Note: The columns headed N represent the opportunities for error. The opportunities vary within categories, e.g. the reviewer may judge that it would have been inappropriate to call a doctor, or move a patient to ICU despite falling oxygen saturation, e.g. because death was expected.

Entries are error rates as percentages of N, with binomial standard errors.

Darker shaded areas relate to post-intervention epochs.

Table 3.4b: Appropriate clinical response – difference between control and SPI2 hospitals, changes over time and the effect of SPI2 (formal analyses for three items only, because of sparse data for other items.)

	Baseline comparisons		Changes in controls		Effect of SPI2	
	OR (99% CI)	p-value	OR (99% CI)	p-value	OR (99% CI)	p-value
Sputum present:						
Sputum culture	0.8 (0.4, 1.6)	0.411	0.5 (0.2, 1.2)	0.040	1.7 (0.5, 6.0)	0.250
Temperature >38°C:						
If yes, blood culture	1.0 (0.2, 4.5)	0.969	0.9 (0.1, 7.1)	0.874	0.6 (0.04, 9.6)	0.636
Appropriate follow-up:						
Clinical review arranged if appropriate	0.7 (0.3, 1.7)	0.343	0.3 (0.1, 1.0)	0.009	1.2 (0.3, 5.4)	0.698

Table 3.5a: Use of steroids and antibiotics, CURB score and other standards applicable to specific cases – compliance with standards

	Control hospitals									SPI2 hospitals								
	Epoch 1			Epoch 2			Epoch 3			Epoch 1			Epoch 2			Epoch 3		
	N	%	SE	N	%	SE	N	%	SE	N	%	SE	N	%	SE	N	%	SE
Asthma or COPD given steroids within 24 hrs	70	84.3	4.4	63	91.8	3.5	56	92.9	3.5	59	91.5	3.7	74	93.2	2.9	53	94.3	3.2
COPD: appropriate prescription of oxygen	30	33.3	8.7	20	55.0	11.1	7	57.1	18.1	19	57.9	11.3	21	47.6	10.9	7	85.7	13.2
Peak flow record	10	80.0	12.6	11	63.6	14.5	5	40.0	21.9	24	79.2	8.3	18	94.4	5.4	8	75.0	15.3
Severity of pneumonia patients recorded in notes?	52	73.1	6.2	68	70.6	5.6	57	77.2	5.6	49	77.6	5.9	45	77.8	6.3	60	70.0	6.0
CURB score recorded in notes?	52	1.9	1.9	67	22.4	5.1	56	21.4	5.5	50	2.0	2.0	44	25.0	6.1	60	41.7	6.4
Was appropriate antibiotic treatment given?	51	94.1	3.3	68	92.6	3.2	53	96.2	2.6	49	91.8	4.0	42	100.0	0.0	55	94.5	3.1

Darker shaded areas relate to post-intervention epochs.

Table 3.5b: Steroids and antibiotics, CURB score and other standards applicable to specific cases – differences between control and SPI2 hospitals, changes over time and the effect of SPI2

	Baseline comparisons		Changes in controls		Effect of SPI2	
	OR (99% CI)	p-value	OR (99% CI)	p-value	OR (99% CI)	p-value
Asthma or COPD given steroids within 24 hrs	1.8 (0.6, 5.7)	0.183	0.9 (0.2, 4.8)	0.813	0.6 (0.05, 6.8)	0.568
COPD: appropriate prescription of oxygen	1.7 (0.1, 19.2)	0.585	0.1 (0.001, 4.0)	0.092	1.0 (0.005, 220.5)	0.985
Peak flow record	1.1 (0.03, 40.9)	0.954	0.1 (0.001, 13.5)	0.255	29.7 (0.1, 15943)	0.165
Severity of pneumonia patients recorded in notes?	0.9 (0.3, 3.2)	0.829	0.9 (0.3, 3.1)	0.821	0.7 (0.1, 3.0)	0.478
CURB score recorded in notes?	1.4 (0.4, 4.9)	0.453	7.3 (1.4, 37.7)	0.002	2.1 (0.4, 11.1)	0.236
Was appropriate antibiotic treatment given?	1.4 (0.2, 10.5)	0.676	1.5 (0.1, 15.7)	0.665	0.5 (0.02, 10.0)	0.519

No items showed significant variation between hospitals within arms.

Table 3.6a: Prescribing errors*

	Control hospitals						SPI2 hospitals					
	Epoch 1		Epoch 2		Epoch 3		Epoch 1		Epoch 2		Epoch 3	
No. of patients†	120		122		112		113		117		114	
No. of prescriptions	2953		3269		2871		2529		2938		2656	
Prescriptions per patient	24.6		26.8		25.6		22.4		25.1		23.3	
Errors:												
Total	345		298		216		251		266		167	
	Rate	SE	Rate	SE	Rate	SE	Rate	SE	Rate	SE	Rate	SE
Unadjusted rates:												
Error rate per prescription	0.120	0.017	0.092	0.013	0.076	0.012	0.106	0.015	0.088	0.013	0.060	0.010
Rates adjusted for date of review:												
Overall rate (all errors)	0.103	0.017	0.086	0.013	0.089	0.015	0.098	0.015	0.084	0.013	0.073	0.014
By five most prevalent stages of the drug use process:												
Need for drug therapy	0.006	0.002	0.018	0.003	0.018	0.004	0.014	0.003	0.011	0.002	0.015	0.002
Selection of dose	0.057	0.009	0.031	0.005	0.032	0.006	0.039	0.006	0.039	0.006	0.030	0.006
Selection of drug	0.002	0.001	0.002	0.001	0.001	0.001	0.004	0.001	0.002	0.001	0.001	0.001
Selection of formulation	0.010	0.003	0.005	0.001	0.005	0.002	0.007	0.002	0.008	0.002	0.009	0.003
Provide information needed for supply	0.029	0.006	0.029	0.005	0.035	0.007	0.032	0.006	0.026	0.005	0.025	0.006

* A breakdown of error types, including failure to reconcile a patient's previous medicines with prescription on admission (a particular focus of the SPI) are given in the full report – there were no significant SPI effects.

† The number of patients are those with medication charts available for review.

Darker shaded areas relate to post-intervention epochs.

Table 3.6b: Prescribing errors – differences between control and SPI2 hospitals, changes over time and the effect of SPI2

	Baseline comparisons		Changes in controls		Effect of SPI2	
	Rate ratio (99% CI)	p-value	Rate ratio (99% CI)	p-value	Rate ratio (99% CI)	p-value
Overall rate (all errors)	1.0 (0.6, 1.5)	0.860	0.9 (0.6, 1.4)	0.662	0.9 (0.5, 1.5)	0.444
By five most prevalent stages of the drug use process:						
Need for drug therapy	1.0 (0.6, 1.6)	0.825	1.5 (0.7, 2.9)	0.157	0.8 (0.3, 2.1)	0.626
Selection of dose	0.9 (0.6, 1.4)	0.595	0.8 (0.5, 1.3)	0.166	1.0 (0.5, 2.0)	0.982
Selection of drug	1.2 (0.5, 2.9)	0.670	0.7 (0.1, 4.2)	0.643	0.7 (0.05, 9.3)	0.687
Selection of formulation	1.1 (0.6, 2.0)	0.788	0.8 (0.3, 2.0)	0.506	1.6 (0.5, 5.3)	0.277
Provide information needed for supply	1.0 (0.6, 1.8)	0.842	1.1 (0.7, 1.9)	0.556	0.7 (0.3, 1.5)	0.220

Table 3.7: Anti-coagulant prescribing errors

	Control hospitals			SPI2 hospitals		
	Epoch 1	Epoch 2	Epoch 3	Epoch 1	Epoch 2	Epoch 3
No. of patients	28	42	62	31	64	74
No. of prescriptions	43	61	92	54	92	99
No. of errors	0	0	4	1	5	0

Darker shaded areas relate to post-intervention epochs.

Table 3.8a: Reconciliation errors at admission

	Control hospitals			SPI2 hospitals		
	Epoch 1	Epoch 2	Epoch 3	Epoch 1	Epoch 2	Epoch 3
No. of patients	28	42	62	31	64	74
No. of admissions	120	122	112	113	117	114
Admissions with reconciliation errors:						
N	8	14	10	7	8	6
% (SE)	6.7 (2.3)	11.5 (2.9)	8.9 (2.7)	6.2 (2.3)	6.8 (2.3)	5.3 (2.1)
Mean no. of errors when error is present (SE)	1.4 (0.3)	2.1 (0.3)	2.2 (0.6)	2.9 (1.1)	2.3 (0.7)	2.3 (0.6)

Darker shaded areas relate to post-intervention epochs.

Table 3.8b: Reconciliation errors at admission – differences between control and SPI2, changes over time and the effect of SPI2

	Baseline comparisons		Changes in controls		Effect of SPI2	
					Ratio of temporal changes	
	OR (99% CI)	p-value	OR (99% CI)	p-value	OR (99% CI)	p-value
Admissions with reconciliation errors	1.2 (0.3, 4.8)	0.727	1.5 (0.6, 3.8)	0.292	0.9 (0.3, 3.2)	0.914

Odds-ratios (OR) derive from a logistic model with random effects for hospitals, adjusted for the date of review.

Table 3.9: Holistic review: changes in ratings and numbers of adverse events and errors between control and SPI2 hospitals (standard errors in parenthesis)

	Control hospitals			SPI2 hospitals			Difference in change (99% CIs)*
	Epoch 1	Epoch 2	Epoch 3	Epoch 1	Epoch 2	Epoch 3	
No. of patients	126	126	114	117	120	122	
Quality ratings:							
Admission rating†	4.76 (0.13)	4.94 (0.12)	4.97 (0.10)	5.03 (0.10)	4.93 (0.11)	4.87 (0.10)	-0.26 (-0.77, 0.24)
Management rating†	3.98 (0.17)	4.18 (0.17)	4.29 (0.16)	4.35 (0.16)	4.03 (0.17)	4.25 (0.16)	-0.18 (-0.92, 0.56)
Pre-discharge rating†	4.13 (0.16)	4.25 (0.14)	4.32 (0.13)	4.28 (0.15)	4.16 (0.15)	4.25 (0.14)	-0.10 (-0.74, 0.54)
Overall care rating‡	7.42 (0.13)	7.62 (0.12)	7.77 (0.11)	7.72 (0.11)	7.46 (0.12)	7.47 (0.11)	-0.41 (-0.94, 0.11)
Errors/Adverse Events							Rate ratios
No. errorsΦ	52.4 (5.6)	39.7 (5.2)	30.7 (5.3)	35.9 (4.9)	45.0 (5.7)	38.5 (5.0)	1.47 (0.74, 2.90)
No. adverse eventsΦ	4.76 (2.21)	3.97 (1.74)	3.51 (1.73)	0.85 (0.85)	5.00 (1.99)	0 (–)	

* Differences in changes are estimated from a mixed effects model (see methods for details) and represent a difference in change between epoch 3 and epochs 1 and 2.

† Score scale: 1 (below best practice) to 6 (excellent care)

‡ Score scale: 1 (unsatisfactory) to 10 (very best care)

Φ The numbers of errors and numbers of adverse events are per 100 patients (patients could experience more than one error and more than one adverse event) .

Darker shaded areas relate to post-intervention epochs.

Table 3.10: Rates per 100 patients of errors identified by broad category of error (standard errors are in parenthesis)

	Control hospitals			SPI2 hospitals			Rate ratio (99% CIs)*
	Epoch 1	Epoch 2	Epoch 3	Epoch 1	Epoch 2	Epoch 3	
No. of patients	126	126	114	117	120	122	
No. of errors	67	50	36	44	54	47	
Diagnosis/assessment/ admission error	63.49 (7.18)	42.86 (6.00)	36.84 (6.74)	44.44 (6.91)	55.00 (7.28)	46.72 (6.60)	1.34 (0.72, 2.51)
Hospital-acquired infection	0	0	0.87 (0.87)	0	0	0	Not estimable
Technical/management	10.32 (2.72)	9.52 (2.63)	9.65 (2.78)	4.27 (1.88)	8.33 (2.53)	5.74 (2.11)	0.94 (0.21, 4.28)
Medication/ maintenance/follow-up	24.60 (4.46)	16.67 (3.52)	8.77 (2.66)	22.22 (4.22)	16.67 (3.80)	17.21 (3.43)	2.13 (0.69, 6.53)
Clinical reasoning	36.50 (4.30)	27.78 (4.00)	20.18 (3.78)	24.79 (4.01)	29.17 (4.17)	27.87 (4.08)	1.65 (0.72, 3.77)
Discharge information	12.70 (2.98)	14.29 (3.13)	9.65 (2.78)	11.11 (2.92)	16.67 (3.42)	13.93 (3.15)	1.43 (0.44, 4.68)

* Differences in changes are estimated from a mixed effects model (see methods for details) and represent a difference in change between epoch 3 and epochs 1 and 2.

Errors can be of multiple categories.

Darker shaded areas relate to post-intervention epochs.

Implicit (holistic) case note review

The sample

In the nine SPI2 hospitals, 359 case notes were holistically reviewed (roughly equally divided between the nine hospitals). For the nine control hospitals, 366 cases notes were holistically reviewed (again roughly equally divided between the nine hospitals).

For the control and SPI2 hospitals, roughly equal numbers of cases notes were reviewed from each of the three epochs (243 cases notes were reviewed from epoch 1; 246 from epoch 2; and 236 from epoch 3). This means that a total of 489 cases notes were reviewed from the pre-intervention period and 236 cases notes were reviewed from the post-intervention period. A small number of case notes analysed by explicit review did not get included in the holistic review, and vice versa, due to logistical problems and time constraints.

For this reason the homology between the two sets of notes is not complete. For example, there were 31 deaths among the explicit case notes reviewed, and 30 among the implicit case notes.

Reliability

In total, 74 case notes were reviewed by two reviewers. Measures of reliability between the two holistic reviewers were, as expected for holistic reviews, low[35] (ICCs were 0.05 (99% CI: -0.25, 0.34) for admission rating; 0.05 (99% CI: -0.25,0.34) for the management rating; 0.37 (99% CI: 0.08,0.60) for the pre-discharge care rating; and 0.31 (99% CI: 0.02, 0.56) for the overall care rating).

The main reviewer tended to assign higher average ratings with more variability, whereas the second reviewer tended to assign lower average ratings with less variability.

The errors and adverse events identified by the two reviewers had small Kappas (0.08 and 0.00 respectively).

Quality of care

The average quality of care scores during epoch 1 with standard errors (SE) for admission, management and pre-discharge ratings were 4.89 (SE 0.08), 4.15 (SE 0.12) and 4.20 (SE 0.12) respectively on a scale of one (below best practise) to six (excellent care); and the average score for overall care was 7.56 (SE 0.09), on a scale of one (unsatisfactory) to 10 (very best care).

During epoch 1, all of the four quality of care ratings were higher in the SPI2 hospitals compared with the control hospitals (table 3.9), although not significantly so. However, during both epoch 2 and epoch 3, all four quality of care ratings were higher in the control hospitals compared to the SPI2 hospitals (although, not significantly so).

In the control hospitals, all ratings tended to increase with time. Whereas in the SPI2 hospitals, all ratings decreased between epoch 1 and epoch 3 (although once again, not significantly so). However, differences in changes across control and SPI2 hospitals were not significant for any of the four ratings (table 3.9).

Errors

Over all hospitals and all epochs, the average number of errors observed was 41 (SE 2.17) per 100 patients, which equates to approximately one error in every 2.5 case notes reviewed.

In the control hospitals, the average number of errors per 100 patients decreased over the three epochs from 52.4 (SE 5.6) errors per 100 patients in the first epoch to 30.7 (SE 5.3) in the third epoch (table 3.10). Whereas, in the SPI2 hospitals, the average number of errors per 100 patients was relatively stable over epochs: from 35.9 (SE 4.9) in the first epoch to 38.5 (SE 5.0) in the third.

Again, differences in changes in the average number of errors before and after the intervention across control and SPI2 hospitals were not significant (rate ratio 1.47; 0.74-0.90).

A total of 153 errors were identified in the control hospitals and 145 errors identified in the SPI2 hospitals (table 3.10). The most frequent categories of errors related to diagnosis, assessment or admission, or were errors relating to poor clinical reasoning.

Errors relating to both these types were more frequent in the control hospitals in epoch 1, but were less frequent during epochs 2 and 3. Rates of other errors also differed between control and SPI2 hospitals and between epoch 1 and epoch 2, although no differences in changes were significant.

Table 3.11a: Reviewer agreement in the perioperative case note review

	Appropriate pain relief	Prophylactic antibiotics	Temperature monitored	DVT prophylaxis
% Agreement	85%	93%	59%	96%
Kappa*	0.46	–	0.24	–

*Blank entries for Kappa indicate that one reviewer put all cases in the same category.

Table 3.11b: Rates of compliance with perioperative care standards

	Control hospitals				SPI hospitals			
	Pre-intervention		Post-intervention		Pre-intervention		Post-intervention	
No. of patients	51		43		79		69	
	%	SE	%	SE	%	SE	%	SE
Advanced method of pain relief*	94.0	3.4	94.9	3.6	85.3	4.1	82.5	4.8
Perioperative antibiotic given†	94.1	3.3	100.0	–	97.5	1.8	97.1	2.0
Temperature monitored‡	16.0	5.2	30.2	7.1	29.1	5.1	41.2	6.0
Appropriate DVT prophylaxis†Φ	100.0	–	100.0	–	98.7	1.3	100.0	–

* Hospital staff identified 15 cases with contraindications to this standard, all of which were corroborated by the reviewers. The data relates to the 227 eligible patients.

† Logistic regression impossible because 100% in one cell.

‡ Evidence of heterogeneity between hospitals at baseline.

Φ Three cases had contraindications yielding a denominator of 238. It was withheld in only two cases where no contraindications were present but wrongly administered in two cases where there was a contraindication.

Darker shaded areas relate to post-intervention epochs.

Table 3.11c: Perioperative review: changes in the level of compliance between SPI2 and control hospitals and the effect of SPI

	Baseline comparisons (SPI/control)		Changes in controls (Epoch 2/epoch 1)		Effect of SPI	
	OR (99% CI)	p-value	OR (99% CI)	p-value	OR (99% CI)	p-value
Advanced method of pain relief	0.3 (0.03, 2.6)	0.148	1.0 (0.1, 17.2)	0.978	0.6 (0.03, 18.4)	0.820
Perioperative antibiotic given	0.8 (0.06, 11.5)	0.862	–	–	–	–
Temperature monitored*	1.8 (0.5, 6.5)	0.227	1.8 (0.4, 7.6)	0.279	0.9 (0.1, 5.2)	

*Temperature monitoring is subject to significant (P=0.010) variation between hospitals within the arms of the study.

Table 3.12: Soap and AHR consumption – median and inter-quartile ranges for control and intervention hospitals, pre- and post-intervention period

	Control hospitals		SPI2 hospitals	
	Pre-intervention	Post-intervention	Pre-intervention	Post-intervention
Total hospital consumption rates per 1,000 bed days*:				
Soap	43 (33,54)	63 (35,86)	49 (30,64)	75 (5,102)
AHR	34 (12,45)	56 (45,67)	39 (28,74)	60 (42,96)

* Before period is July 2004 to February 2007 and after (during) period is March 2007 to September 2008; units are litres per 1,000 bed days.

Darker shaded areas relate to post-intervention epochs.

3.3 Sub-study 3: Error rates/quality of care – perioperative care

Sample, reviewer reliability and headline message

We fell short of the target number of 360 case notes and were able to retrieve 242 notes. At total of 127 came from admissions for total hip replacements and 115 from admissions for open colectomies. A second reviewer examined 27 case notes.

Percentage agreement and Kappa statistics are given in table 3.11a. These figures indicate low agreement on whether the temperature had been monitored (59%). For all other items the reviewers agreed on at least 85% of the cases.

No significant SPI effects were observed for any of the four clinical standards examined and the before/after comparison if anything, leaned towards the control hospitals. The hospitals were similar at baseline except with respect to intra-operative temperature monitoring where controls had more headroom for improvement.

The results relating to the individual criteria are given in table 3.11b and the outcomes of the mixed effects logistic regressions are given in table 3.11c.

Pain relief

Hospital staff identified contraindications to either epidural or self-administered analgesia in 15 of 242 cases. The existence of the contraindication was confirmed by the reviewers in all of these 15 cases, with an additional contraindication in a patient identified by one of the reviewers.

Thus, 226 patients were eligible for modern analgesic methods and 199 (88%) received such care. There was little room for improvement and there were no differences between control and SPI2 hospitals at either baseline, or over time.

Prophylactic antibiotics

These were given in 235 of 242 cases (97%). While the breakdown across arms and epochs is summarised in table 3.11c, the full logistic regression analysis was not feasible because of the 100% compliance in the control hospitals at epoch 2.

Temperature monitoring

There was marked but non-significant increase in compliance over epochs in both control and SPI2 hospitals with little difference in rate of improvement (OR 1.8; 0.4-7.6). There is evidence of heterogeneity between hospitals.

DVT prophylaxis

Anticoagulation prophylaxis was given in 239 of the 242 cases (99%). Two of these 239 were contraindicated for prophylaxis. It was correctly withheld in one further contraindicated case, and in two cases where no contraindications were recorded.

3.4 Sub-study 4: Indirect measure of hand hygiene

Data available

Data on soap and AHR (in litres) were available for nine and eight of the control trusts and for seven and six of the SPI2 trusts respectively.

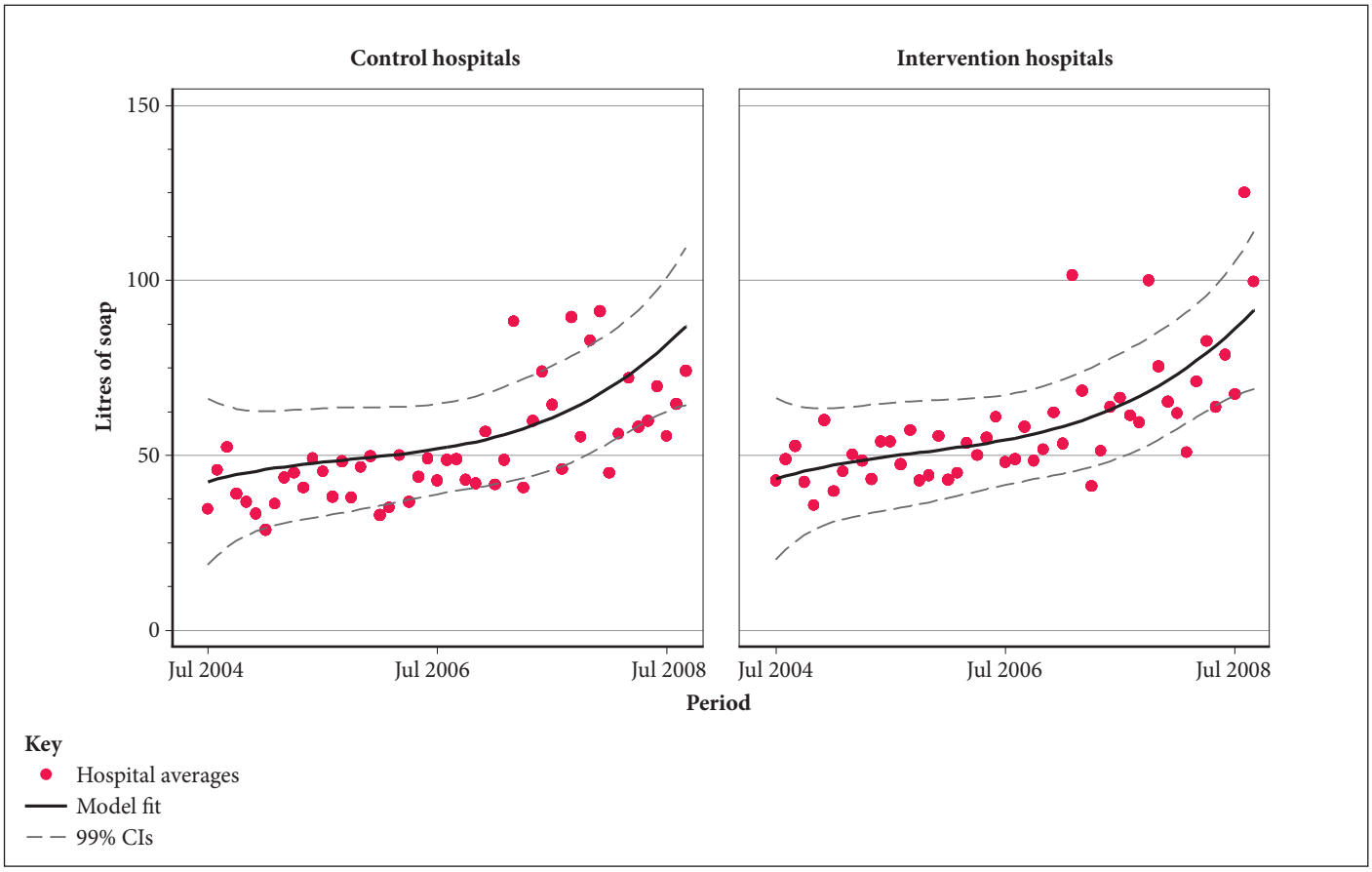


Figure 3.1: Rate of soap consumption per 1,000 bed days over time in control and SPI2 hospitals

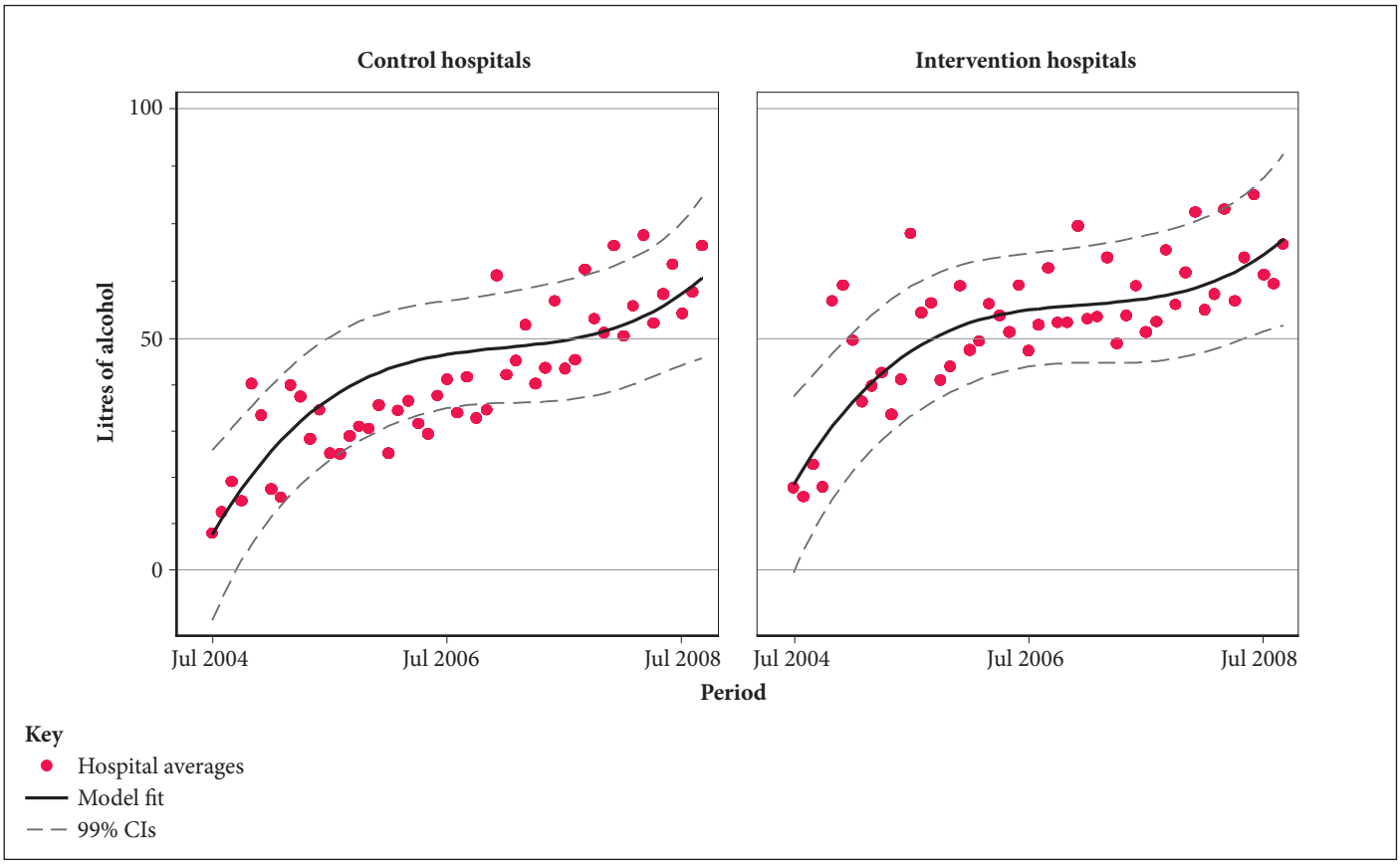


Figure 3.2: Rate of AHR consumption per 1,000 bed days over time in control and SPI2 hospitals

Soap and AHR consumption

The median rate of soap consumption over all hospitals and all time periods was 50 litres per 1,000 bed days (IQR: 32, 71) and the median rate of AHR consumption was 44 litres per 1,000 bed days (IQR: 29, 61). Averaging over all time periods (July 2004 to September 2008) the median rate of soap and AHR consumption was higher in the SPI2 hospitals compared to the control hospitals: the median rate of soap consumption in the SPI2 hospitals was 53 litres (IQR: 30, 79) compared to 46 litres (IQR: 34, 65) in the control hospitals; and the median rate of AHR consumption was 49 litres (IQR: 31, 79) compared to 43 in the control hospitals (IQR: 34 ,65).

Rates of both soap and AHR consumption increased in both control and SPI2 hospitals over the study period (table 3.12). For example, in the control hospitals the median rate of soap consumption increased from 43 litres (IQR: 32, 54) in the period before the intervention to 63 litres (IQR: 35, 86) in the period during the intervention; and in the SPI2 hospitals this rate similarly increased from 49 litres (IQR: 30, 64) to 71 litres (IQR: 5, 102). Smoothed estimates of rates of increase of consumption of both products, as estimated by the GEE population averaged model, are presented in figures 3.1 and 3.2.

The rate of increase in rates of consumption of both soap and AHR (that is, the difference of the differences) were similar between control and SPI2 hospitals and were not significant (p=0.760 and p=0.889 respectively, appendix 4, table A2), reflecting the fact that rates of consumption of both products were higher in the SPI2 hospitals throughout the study, and not only after the intervention phase.

3.5 Sub-study 5: Outcomes

Adverse events among patients on acute medical wards

Over all hospitals and all epochs, the main reviewer identified 22 adverse events among the 725 case notes and the average number of adverse events observed was 3.03 per 100 patients.

In the control hospitals, the average number of adverse events per 100 patients decreased over the three epochs from 4.76 (SE 2.21) adverse events per 100 patients in the first epoch, to 3.51 (SE 1.73) in the third epoch. In contrast, in the SPI2 hospitals, the average number of adverse events per 100 patients increased between the first and second epoch from 0.85 (SE 0.85) to 5.00 (SE 1.99); and decreased to zero in the third epoch. Again, differences in changes in numbers of adverse events across control and SPI2 hospitals were not significant (rate ratio=1.47; 0.74 – 2.90).

Classifications by type of adverse event are presented in table 3.13. Small numbers of identified adverse events preclude informative comparisons.

The principal reviewer identified strong or certain evidence of preventability in four of the 22 adverse events (that is, 0.5% of cases overall). None of these four adverse events was fatal and all occurred in the pre-intervention epochs (itemised in table 11 of the SPI1 paper).[1] However, the second reviewer found two preventable deaths (both among control hospitals) in the third epoch, one due to brachycardia in a patient with hypokalaemica, and another due to delay in diagnosis of femoral artery thrombosis. She also found three preventable deaths in earlier epochs.

A further case where the probability of a causal link was less than 50% was also identified again in the control group. Due to such small numbers of adverse events being assessed as preventable, these percentages were not analysed between control and SPI2 hospitals.

Table 3.13: Rates (per 100 patients) of adverse events among patients admitted with acute respiratory disease

	Control hospitals			SPI2 hospitals			Difference in change (99% CIs)*
	Epoch 1	Epoch 2	Epoch 3	Epoch 1	Epoch 2	Epoch 3	
No. of patients	126	126	114	117	120	122	
No. of errors	6	5	4	1	6	0	
Diagnosis/assessment/ admission error	3.97 (1.75)	1.59 (1.12)	2.63 (1.95)	0.85 (0.85)	3.33 (1.65)	0	-1.98 (-8.18, 4.23)
Hospital-acquired infection	2.38 (1.36)	1.59 (1.12)	1.75 (1.24)	0	2.50 (1.43)	0	-1.03 (-5.78, 3.74)
Technical/management	0.79 (0.79)	1.59 (1.12)	0.88 (0.88)	0	0.83 (0.83)	0	-0.10 (-3.48, 3.28)
Medication/ maintenance/follow-up	0	0	0	0.85 (0.85)	1.67 (1.17)	0	-1.27 (-3.88, 1.33)
Clinical reasoning	0	0	0	0.85 (0.85)	0	0	-0.42 (-1.94, 1.09)
Discharge information	0.79 (0.79)	0	0	0.85 (0.85)	0	0	-0.02 (-2.17, 2.11)

* Differences in changes are estimated from a mixed effects model (see methods for details) and represent a difference in change between epoch 3 and epochs 1 and 2.

Errors can be of multiple categories.

Darker shaded areas relate to post-intervention epochs.

They serve to shed light on mortality estimates however. A breakdown of deaths by level of preventability and reviewer is given in table 3.14.

Three medication related adverse events were found on holistic review. At around 0.004% (3/725), this is also a somewhat lower rate than reported elsewhere.[19]

Mortality among acute medical care patients

Crude mortality was higher in the control hospitals than in the SPI2 hospitals (OR 0.7; 0.2-2.1) (Table 3.15a), but neither this, nor any other effect – including that of the SPI – was significant at the pre-determined 1% level after adjustment for age of patient (OR 0.3; 0.068-1.4) (although the result was just significant [p=0.043] at the 5% level).

Sex and number of co-morbidities were also included as patient-level covariates, though only age was significant (p<0.001). The mortality rate increased by 10.3% (CI 6.8%-15.1%) per year of patient age.

Hospital-wide mortality

Over time, the general trend of hospital-wide mortality is downwards in both control and SPI2 hospitals (figure 3.3). Using the standard deviations supplied, there appears to be no simple functional relationship consistent with the data.

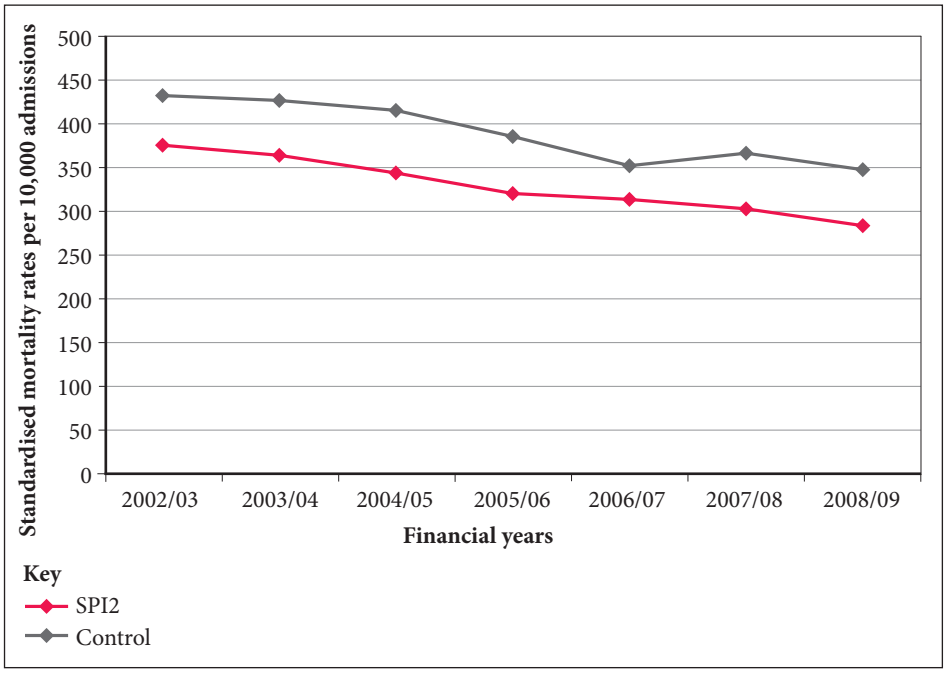


Figure 3.3: Hospital directly age sex standardised mortality rates per 10,000 admissions, all medical specialties, controls and SPI2, 2002/3 – 2008/9

Table 3.14: Preventable deaths in acute medical wards across the study epochs*

Epoch	Control					Intervention				
	No. of deaths within holistic review	Preventable deaths: ≥50% 1st reviewer	Preventable deaths: ≥50% 2nd reviewer	Preventable deaths: <50% 1st reviewer	Preventable deaths: <50% 2nd reviewer	No. of deaths within holistic review	Preventable deaths: ≥50% 1st reviewer	Preventable deaths: ≥50% 2nd reviewer	Preventable deaths: <50% 1st reviewer	Preventable deaths: <50% 2nd reviewer
1	17	0	1†	2	0	9	0	0	0	0
2	24	0	1†	1	1	11	0	1†	0	1
3	23	0	2	2	2	7	0	0	0	0

*Preventable deaths <50%: substandard practice was present that could have led to death but the probability that it did so was less than 50%
Preventable deaths ≥50%: substandard practice led to death on the balance of probabilities.

† These deaths (both associated with CO2 retention in patients denied non-invasive ventilation – one of whom was given 60% oxygen) are not included in table 11 of SPI1 evaluation.

Darker shaded areas relate to post-intervention epochs.

Table 3.15a: Mortality among acute medical care patients whose case notes were reviewed

	Control hospitals			SPI2 hospitals		
	Epoch 1	**Epoch 2**	**Epoch 3**	**Epoch 1**	**Epoch 2**	**Epoch 3**
No. of patients	120	123	112	116	117	114
Deaths	18	24	24	9	15	7
1 % Mortality (SE)	15.0 (3.3)	19.5 (3.6)	21.4 (3.9)	7.8 (2.5)	12.8 (3.1)	6.1 (2.3)
2 Age: mean (SD)	77.6 (7.7)	81.1 (7.9)	79.6 (8.0)	77.7 (7.6)	78.1 (7.1)	80.6 (7.8)
3 % Female	63.3	53.7	53.6	53.4	50.4	52.6
4 Co-morbidities: mean	2.9	3.1	2.6	2.8	3.0	2.9

Darker shaded areas relate to post-intervention epochs.

Table 3.15b: The effect of SPI2 on the mortality among acute medical care patients

	Baseline comparisons		Changes in controls		Effect of SPI2	
	Odds ratio (99% CI)	**p-value**	**Odds ratio (99% CI)**	**p-value**	**Odds ratio (99% CI)**	**p-value**
Mortality (adjusted for age, sex, number of co-morbidities)	0.7 (0.2, 2.1)	0.391	1.4 (0.6, 3.1)	0.320	0.3 (0.08, 1.4)	0.043

Odds-ratios (OR) derive from a logistic model with random effects for hospitals, adjusted for the date of review.

Furthermore, the difference between control and SPI2 hospitals is not constant over time, whether measured on the natural scale or the log scale (the latter represents a relative measure).

However, calibration using between hospital information may disturb these conclusions – for example, it is conceivable that the data are consistent with a constant temporal difference, when assessed against standard deviations that incorporate an allowance for variation between hospitals within the arms of the study.

We investigated the baseline differences in mortality in control verses SPI2 hospitals by considering the possibility that the control hospitals served a more deprived area. We obtained a distribution of income deprivation scores from the neighbourhoods of all admitted patients for control and intervention hospitals.

The neighbourhoods used were Lower Level Super Output Areas (LSOA) which are fairly homogenous areas, each containing around 1,600 residents offering a good granularity of measurement for deprivation and other social and environmental variables. Each LSOA in England has an income deprivation score calculated as part of the Indices of Multiple Deprivation 2007.

The score is effectively a proportion of people in a neighbourhood who live in a household with less than 60% of the national median income and/or are in receipt of one of a number of means-tested welfare benefits.

We took the median and upper and lower quartile scores for all admitted patients in both control and SPI2 hospitals for all years. On aggregate the median income scores for both control and SPI2 were very similar (0.12 and 0.13 respectively). However the variation of medians and quartile values within the two groups were markedly different, the SPI2 group appearing to be much more heterogeneous (figure 3.4).

We thus failed to account for the difference between control and SPI2 hospitals in baseline mortality. The mortality in SPI2 hospitals did indeed improve by the 15% target, but similar improvement was evident among controls.

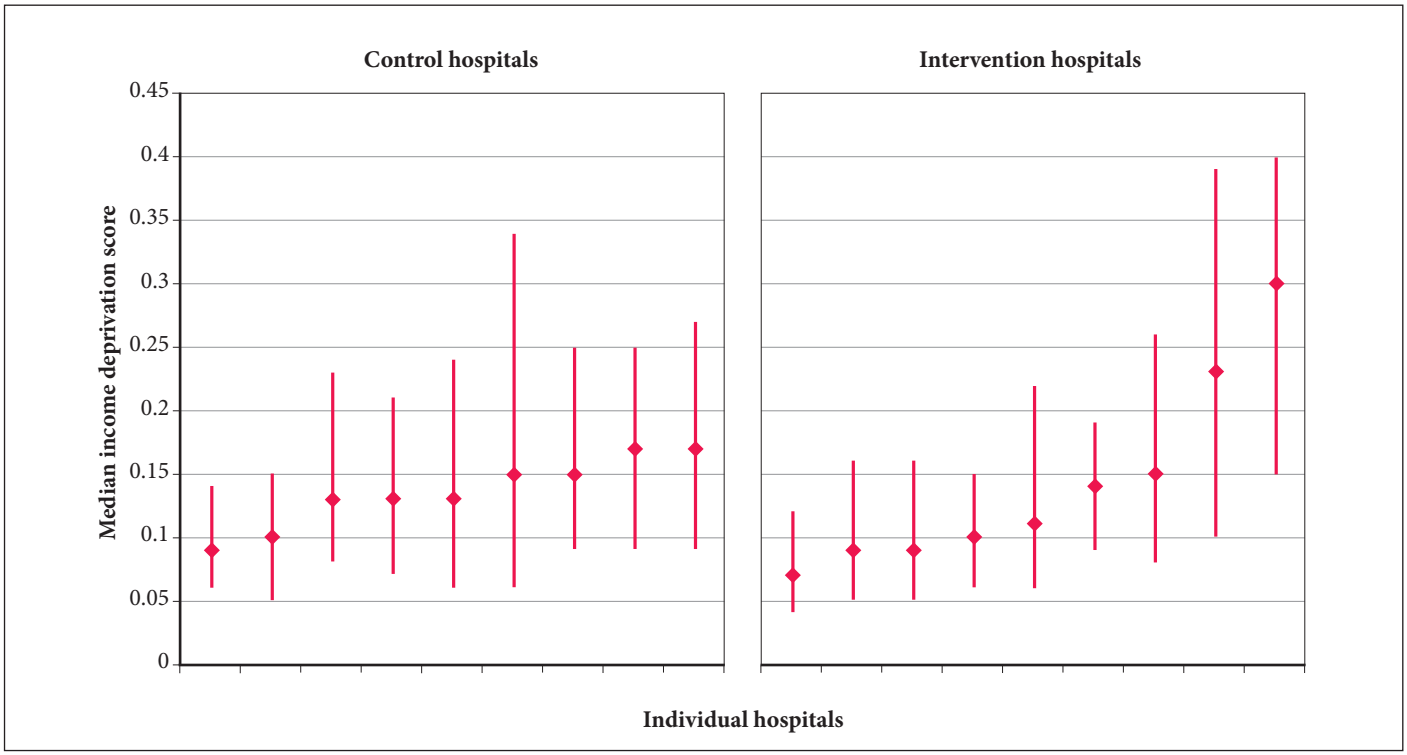


Figure 3.4: Median income deprivation scores of control and SPI2 hospitals

ICU: Mortality, morbidity and length of stay

Data available

Data on mortality, length of stay and several other outcome measures for ICUs were available for 16 hospitals, eight of which were control hospitals and eight of which were SPI2 hospitals.

Data were supplied to ICNARC by seven control and seven SPI2 hospitals for the pre-intervention period (epoch 1) and for six control hospital and eight SPI2 hospitals post-intervention period (epoch 2) (there were some hospitals which did not provide data for both periods).

Observed to expected mortality

The median observed to expected mortality ratio over all hospitals and all time periods was 1.06 (IQR: 0.93, 1.28). Averaging over all time periods (July 2004 to September 2008), this ratio was lower in the SPI2 hospitals compared to the control hospitals: the median observed to expected mortality ratio in the SPI2 hospitals was 0.98 (IQR: 0.90, 1.15) compared to 1.18 (IQR: 1.01, 1.32) in the control hospitals.

The rate of observed-to-expected mortality increased in the control hospitals over the study period (table 3.16). For example, in the control hospitals before the intervention period, the median observed-to-expected mortality ratio was 1.14 (IQR: 0.99, 1.32), and this rate increased to 1.24 (IQR: 1.02, 1.33) in the six months after the intervention.

In the SPI2 hospitals, the observed-to-expected mortality ratio decreased over the two periods: during the first six month period the observed-to-expected mortality ratio was 1.04 (IQR: 0.90, 1.15), and during the last six month period this decreased to 0.97 (IQR: 0.90, 1.15).

At the end of the follow-up period (March 2008), the rate of observed-to-expected mortality was higher in the control hospitals. However, the adjusted difference in differences between control and SPI2 hospitals after adjustment, was not significant at the 99% level (p=0.25, appendix 4, table A3).

Median length of stay

The median length of stay was 125 hours (IQR: 96,153) over all hospitals and all time periods. Averaging over all time periods (July 2004 to September 2008) the median length of stay was lower in the SPI2 hospitals compared to the control hospitals: the median length of stay was 103 hours in the SPI2 hospitals (IQR: 82,132) compared to 146 hours in the control hospitals (IQR: 123, 183).

Based on this, control ICUs may have been dealing with a different case-mix from the SPI2 ICUs.

Length of stay increased in the control hospitals over the study period (table 3.16): during the pre-intervention period the median length of stay was 144 hours (IQR: 117, 174), and this increased to 147 hours (IQR: 126,185) in the post-intervention period.

In the SPI2 hospitals, the median length of stay remained similar between the pre and post-intervention periods: during the pre-intervention period the median length of stay was 102 (IQR: 82, 130), and during the post-intervention period the median length of stay was 103 hours (IQR: 81, 137) in the six month period October 2007 to March 2008. Once again, differences in the rate of changes in length of stay were not significant (p=0.60, appendix 4, table A3).

APACHE II and ICNARC risk prediction scores

Over all time periods and over all hospitals the median APACHE score was 20 (IQR: 17.8, 21.8) and the median ICNARC score was 22.1 (IQR: 19.5, 22.1). These scores were similar between control and SPI2 hospitals and were similar between pre and post-intervention periods (table 3.15). Tests for differences in differences were not significant (p=0.45 and p=0.16, appendix 4, table A4).

C. diff and MRSA rates

Data

Data on numbers of C. diff and MRSA cases were available for all 18 trusts.

C. diff

Over all time periods, the median C. diff infection rate was 1.14 cases per 1,000 bed occupied days (IQR: 0.77, 1.64). Averaging over all time periods, the median rate of C. diff infection was similar between the control and SPI2 hospitals: the median C. diff infection rate was 1.15 (IQR: 0.88, 1.55) in the control hospitals and 1.1 (IQR: 0.67, 1.73) in the SPI2 hospitals.

The median C. diff infection rate decreased over the study period in both the control and SPI2 hospitals (table 3.16). In the control hospitals, the median C. diff infection rate was 1.26 (IQR: 0.95, 1.67) in the period before the intervention, and this decreased to 0.77 (IQR: 0.56, 1.02) in the period after the intervention.

In the SPI2 hospitals, in the period before the intervention, the median C. diff infection rate was 1.37 (IQR: 0.65, 1.99) and this decreased to 0.66 (IQR: 0.50, 0.88) in the period after the intervention.

Differences in changes were not significant between control and SPI2 hospitals (p=0.652, appendix 4, table A1). Smoothed estimated rates of C. diff infection per 1,000 bed occupied days, by control and SPI2 hospitals, are presented in figure 3.5.

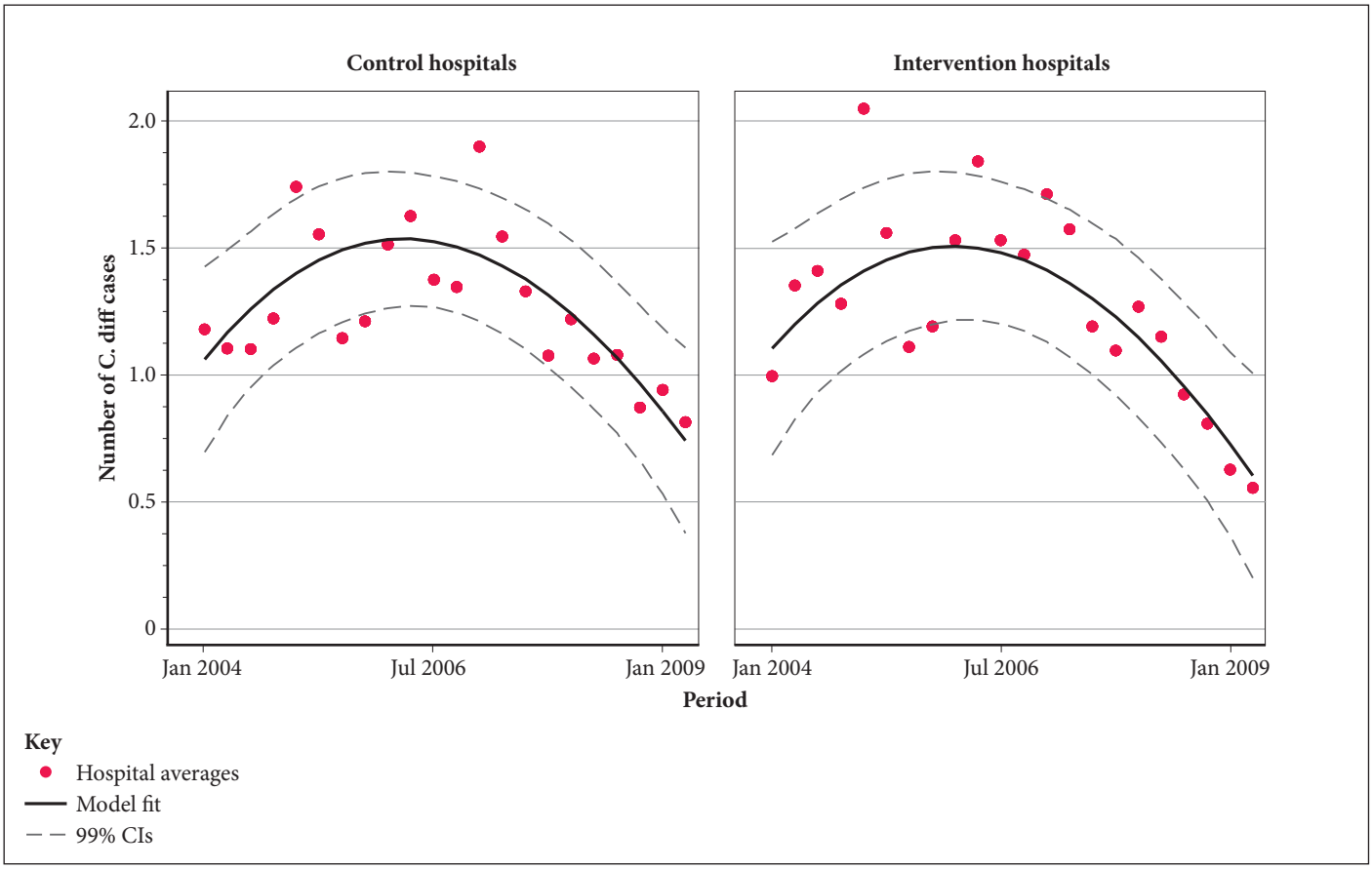


Figure 3.5: Rate of C. diff cases per 1,000 bed days in control and SPI2 hospitals

MRSA

Over all time periods, the median MRSA infection rate was 14.75 cases per 100,000 bed occupancies (IQR: 8.93, 21.98). Averaging over all time periods, the median rate of MRSA infection was similar between the control and intervention hospitals: the median MRSA infection rate was 14.87 (IQR: 9.36, 21.63) in the control hospitals and 14.58 (IQR: 8.85, 22.77) in the SPI2 hospitals.

The median MRSA infection rate decreased over the study period in both the control and SPI2 hospitals (table 3.16). In the control hospitals, the median MRSA infection rate was 17.4 (IQR: 12.01, 23.04) in the period before the intervention, and this decreased to 4.31 (IQR: 2.26, 8.18) in the period after the intervention.

In the SPI2 hospitals, in the period before the intervention, the median MRSA infection rate was 17.76 (IQR: 11.6, 24.43) and this decreased to 6.77 (IQR: 4.89, 10.65) in the period after the intervention.

Table 3.16: Intensive care outcomes and healthcare associated infection rates – median and inter-quartile ranges for control and SPI2 hospitals, pre and post-intervention period

	Control hospitals		SPI2 hospitals		Difference in difference	
	Pre-intervention	Post-intervention	Pre-intervention	Post-intervention	Change (99% CI)	p-value
Intensive and Critical Care Outcomes*						
Adjusted Mortality Ratio	1.14 (0.99,1.32)	1.24 (1.02,1.33)	1.04 (0.90,1.15)	0.97 (0.90,1.15)	0.09 (-0.11,0.29)	0.25
Mean LOS (hours)	144 (117,174)	147 (126,185)	102 (82,130)	103 (81,137)	5.86 (-22.78,34.50)	0.60
Mean APACHE II score	20.4 (17.7, 22.6)	19.0 (17.1, 20.8)	21.1 (19.1, 23.0)	20.3 (17.8, 21.8)	-0.83 (-3.63,1.98)	0.459
Mean ICNARC score	22.3 (19.5, 26.3)	20.7 (18.0, 23.5)	22.6 (21.2, 25.3)	22.2 (19.7, 25.1)	-2.26 (-6.39,1.87)	0.16
Rates of C.diff (per 1,000 bed days) and MRSA infections (per 100,000 bed days)						
C. diff†	1.26 (0.95,1.67)	0.77 (0.56,1.02)	1.37 (0.65,1.99)	0.66 (0.50,0.88)		
MRSA‡	17.41 (12.02,23.04)	4.31 (2.26,8.18)	17.76 (11.60,24.43)	6.77 (4.89,10.65)		

* Before period is October 2006 to March 2007 and after period is October 2008 to March 2009.

† Before period is April 2004 to March 2007 and after period is October 2008 to June 2009.

‡ Before period is April 2001 to March 2007 and after period is October 2008 to September 2009.

LOS: Length of Stay

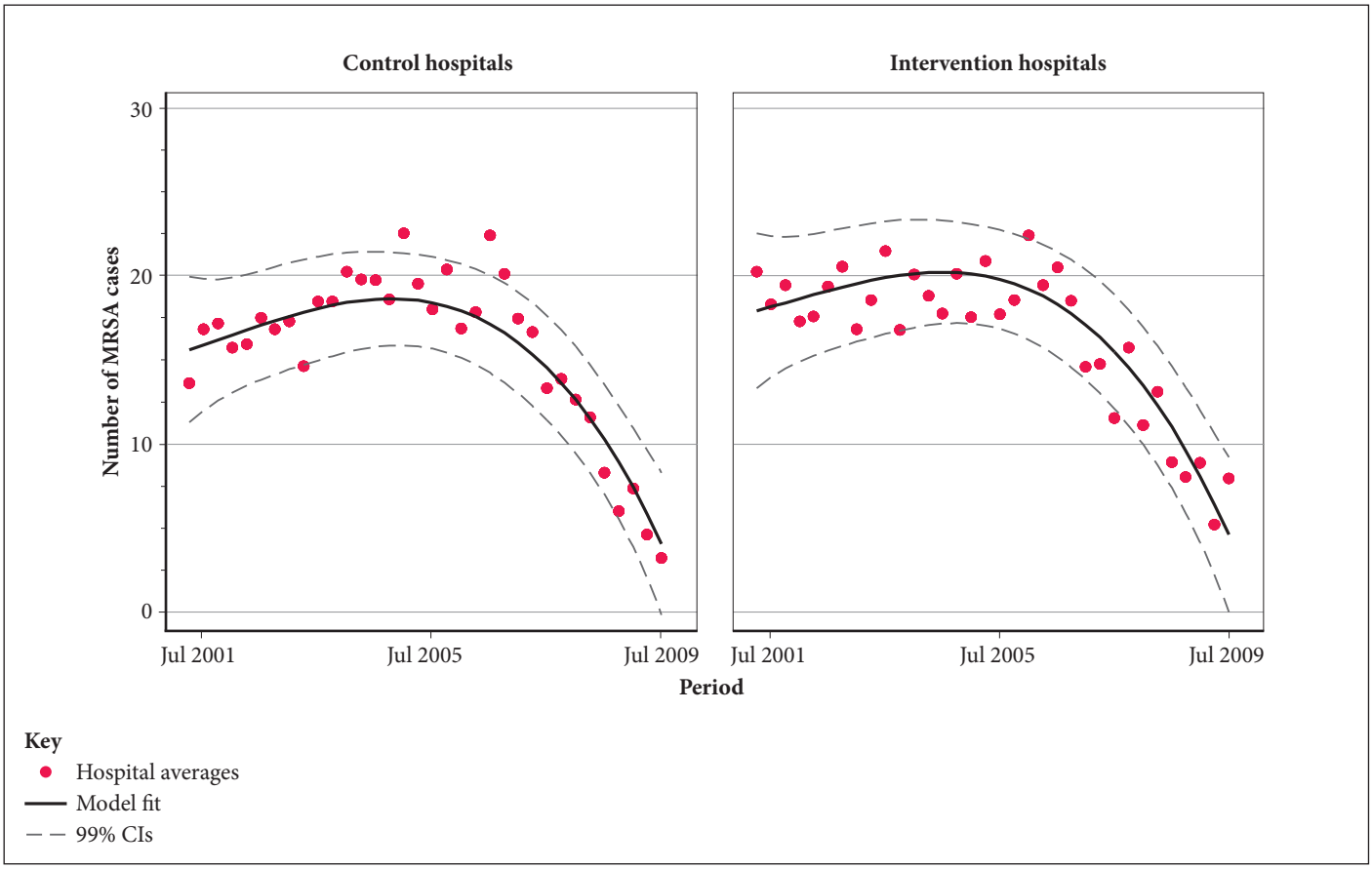


Figure 3.6: Rate of MRSA cases per 100,000 bed days in control and SPI2 hospitals

Differences in changes were not significant between control and SPI2 hospitals (p=0.693, appendix 4, table A1). Estimated smoothed rates of MRSA infection per 100,000 bed occupied days, by control and SPI2 hospitals, are presented in figure 3.6.

Patient survey

For the first survey, the overall response rate was 62% (4,328 of 7,010 valid questionnaires returned) in the nine SPI2 hospitals; for the second it was slightly lower at 55% (3,762/6,810). In the nine control hospitals, the response rates were 63% (4,62/6,791) and 57% (3,973/6,913) respectively. Table 3.17 shows the changes in both control and SPI2 hospitals on each of the five scores identified, along with the differences between the groups in these changes and associated 99% confidence intervals. All five scores improved over the study period in both the control and SPI2 hospitals. None of the five scores showed any significantly different changes between the two groups.

Table 3.17: Patient survey scores in control and SPI2 hospitals at the two periods

	Control hospitals					SPI2 hospitals					Range at baseline	Difference in change (99% CI)	p-value
	N	Survey 1 score (SE)	N	Survey 2 score (SE)	Absolute % change	N	Survey 1 score (SE)	N	Survey 2 score (SE)	Absolute % change			
% staff having well structured appraisals within previous 12 months	8046	39 (1)	7260	28 (1)	-10	6111	34 (1)	3993	27 (1)	-7	27–46	3 (-2, 8)	0.095
Overall, how would you rate the care you received?	4200	82 (0.4)	3913	85 (0.3)	4	4277	80 (0.4)	3705	84 (0.3)	4	75–87	1 (-1, 3)	0.292
Overall, did you feel you were treated with respect and dignity while you were in the hospital?	4111	78 (0.4)	3807	82 (0.4)	4	4167	76 (0.4)	3604	80 (0.4)	3	65–85	0 (-2, 2)	0.702
How would you rate how well the doctors and nurses worked together?	4182	87 (0.4)	3878	88 (0.4)	1	4220	88 (0.4)	3677	89 (0.4)	1	83–91	0 (-2, 2)	0.597
In your opinion, how clean was the hospital room or ward that you were in?	4113	75 (0.4)	3870	77 (0.4)	2	4201	77 (0.4)	3645	78 (0.4)	1	70–80	-1 (-3, 1)	0.141
How clean were the toilets and bathrooms that you used in hospital?	4141	76 (0.4)	3877	78 (0.4)	2	4220	78 (0.4)	3665	79 (0.4)	1	70–82	-1 (-3, 1)	0.204

Darker shaded areas relate to post-intervention epochs.

Chapter 4

Discussion

4.1 Non-comparative findings

There was despair in the United States at the apparent lack of progress on patient safety after the publication of two key reports in 2000.[36] Taken in the round, the data collected in this study seem to tell the story of an improving NHS.

While the staff survey shows little change between epochs, the patient survey shows improvement across all five dimensions pre-specified for our study, suggesting better patient experience. There was even an improvement in medical history taking. Hospital mortality rates are generally falling and although this may be a result of the main from improved technology and increasing proportions of people dying in the community, encouraging trends were noted in the quality of patient care.

Firstly, the baseline performance across hospitals was over 90% on many criteria relating to quality, leaving very little room for improvement. Over 90% of patients with an acute exacerbation of obstructive airways disease received steroids when indicated, and the rates of perioperative prophylaxis against venous thrombosis and wound infection approached 100%.

Secondly, where there was scope for improvement many examples of improved (and none of worsening) practice were found. Both the vigilance of monitoring vital signs on acute medical wards and the use of severity scoring has seen sharp significant increases and there was a strong upward trend in the incidence of intra-operative temperature monitoring.

Rates of hand-washing have increased (if consumption of cleansing materials is accepted as a surrogate) and the incidence of C. diff and MRSA infection has plummeted.

4.2 Control hospitals vs. SPI

Our data for SPI2, as for SPI1, suggest that it was difficult to detect an additive SPI effect. Statistically significant observations were made but not between the two groups of hospitals. In the case of the staff survey, our observations have high statistical power yet only one of the 11 dimensions examined produced a significant result. This was the same dimension (organisational climate) that was also the single dimension to yield a significant result in the evaluation of SPI1. However, in a reversal of our SPI1 evaluation results, the control hospitals improved most in the current study.

Many specific criteria reflecting the quality of care remained stable over time in both groups of hospitals, possibly reflecting a long history of quality improvement in areas such as perioperative care.

Others, such as the quality of intra-operative monitoring and recording vital signs underwent marked improvement, but did so to similar degree in both sets of hospitals.

One exception was the drop in mortality among the acute medical cases in the SPI2 hospitals and an unexplained rise in the control hospitals, such that the difference in differences would have been just significant if the $p<0.05$ threshold had been selected *a priori*.

However, this finding does not align well with either the explicit review of the quality of care or the adverse event tally observed among those same case notes – only two (or at the most three) care-related deaths were found in either group of hospitals in the post-intervention period.

Dramatic improvements in the use of hand-washing materials and in infection rates produced near mirror image results. The NHS leviathan seems responsive to the need to change in certain ways and it is hard to discern any additive effect of the SPI initiative.

Again, this corroborates the finding from the SPI1 evaluation, where improvements were noted across both control and SPI hospitals.

Overall, there is little evidence that good or improved quality and safety in participating NHS hospitals can be reliably attributed to an additive effect of the SPI.

4.3 Strengths and weaknesses

The study was based on a before and after design with contemporaneous controls. Such a design is not as strong as a cluster randomised trial. However, it is stronger than a simple before and after study of the sort that characterises most quality improvement evaluations.

One advantage of contemporaneous controls is that the groups can be compared at baseline. There were differences at baseline for some observations (most notably hospital mortality rate) but not for others.

Baseline rates on the staff and patient surveys were similar and there is little to distinguish the two groups of hospitals on the explicit reviews in either acute medical or surgical patients. For example, none of the 17 vital signs criteria differed significantly between the two groups of hospitals. Thus most of the comparisons that were made were based on end points where no material differences were evident across the groups compared.

We tested for learning/fatigue effects on the part of the reviewers. We found that this was sometimes important (especially for the tricky detection of prescribing errors where the reviewer must audit case notes against the entire formulary running to many hundreds of pages).

Where this problem was observed, we were able to allow for it in the analysis. We also tested for inter-observer agreement and while it was satisfactory with respect to explicit reviews it was poor with respect to the implicit review. This allows the reader to be discerning and treat the results of the implicit review with due caution.

Source data for most end points was collected by independent researchers working across the various hospitals – we set up a supply chain of anonymised case notes for this purpose.

Certain data was collected in the participating hospitals (infection rates and data from the ICU), and this could lead to bias in the comparative study if hospital-based observers were motivated to show the SPI in a good (or bad) light. However, any bias must have affected both sets of hospitals approximately equally since the comparative results are null.

Moreover, we do not think that it is plausible that the observed dramatic reductions in infection rates across all hospitals are the result of the statutory duty to report certain infections when they are identified in the laboratory.

A particular strength of our study arises from possibilities for triangulation. Some of the observations act as a kind of internal control for others. While the funding envelope did not permit us to build qualitative studies into the design (as in SPI1), the study did provide the following internal controls:

- Findings on use of hand-washing materials and two different types of infection support the hypothesis of general improvement in this area.
- The observation that vital signs were recorded with increasing diligence, while use of risk scoring was also used more frequently supports the idea that patients at risk of deterioration are being taken more seriously.
- Mortality rates on the acute medical wards could be triangulated, not only by an audit of compliance with process standards, but also by scrutinising each death in the sample to see if it could have been caused by poor care (only two of the 30 deaths in the post-intervention period were preventable).

We wished to seek further evidence on this point by examining the incidence of unsuspected cardiac arrest crash calls, but found that this information is not yet collected in a consistent way.

The evaluation of SPI1 included qualitative observations which can provide yet a further form of internal control.

However, the study sponsor felt that theoretical saturation had already been reached in the previous evaluation. For example, ethnographic sub-studies within the SPI1 evaluation did indeed confirm that ward staff had taken the importance of close observations of sick patients increasingly to heart.

4.4 Interpretation

A large number of different observations have been made. Many of these observations relate to specific SPI objectives, such as the patient at risk of deterioration, infection control, perioperative care and intensive care. Statistically significant observations were made, but not between the two groups of hospital.

This broadly null additive effect of SPI on patient care should not, however, be translated into a conclusion that there was evidence

of no effect. While a null result can never be proven, this is a greater problem for quality initiatives, where small effect sizes may nevertheless be cost-effective, than it is for studies of clinical effectiveness.

It can, however, be translated, less problematically, into the conclusion that any effect was not large, where large is defined in terms of observed confidence limits. To put this idea in another way, our results are compatible with effects on many end points, of a magnitude that lies below the threshold that can be detected statistically in a study of this size. That said, the results will come as a disappointment to many who were involved in the intervention and who expected a rather more dramatic outcome.

Lack of a measured additive SPI effect may be explained in several ways: programme design; implementation; multiple patients; safety initiatives; and improvements may not yet be detected.

Programme design

One explanation might lie in programme design. It is possible that organisational interventions of this type are simply not highly efficacious and that alternative approaches, such as initiatives focused on professional networks, could be more powerful, as suggested in a study of motivations to change in a maternity context.[37]

Implementation

Secondly, it is possible that implementation of the SPI was not optimal, as discussed in the companion paper.[1] Looking back over the evaluations of both programmes, and following many conversations with those responsible for this and other interventions with similar aims, we suggest that the method by which vertical and horizontal spread of the SPI might have been achieved was incompletely specified.

A combination of a more explicit programme theory and organisational theory of change might have focused more attention on ensuring clinical engagement, encouraged an earlier recognition that the intervention was broad, relative to resource, and identified that effects were likely to be localised in response to a dose of intervention.

In that case, a more focused and less ambitious intervention, and somewhat narrower evaluation, might have ensued.

Multiple patient safety initiatives

A third explanation for the absence of a measured additive effect of the SPI might lie in the extent of the policy-level programmes and initiatives that were largely contemporaneous with the SPI and shared some of its goals, principles and methods, and were targeting several of the same clinical processes as the SPI.

For example, the Clean Your Hands campaign ran continuously from late 2004/05 onwards, promoting the same goal of improved hand hygiene as the SPI. Similarly, improving recognition and response to deterioration in hospitalised patients (an SPI goal) became a focus of policy attention, and guidelines on recognition and response to acutely ill patients were issued by NICE in 2007.[38]

Perhaps most significantly, several initiatives were explicitly modelled upon IHI techniques and principles, which began to have increasing impact on policy making at around the time that the SPI was launched (and it is possible that this was not a coincidence).

For example, the Department of Health's Saving Lives programme, beginning in June 2005 with a revised version in 2007,[39] included a self-assessment tool for trusts to assess their managerial and clinical performance, and a set of high impact interventions that were similar to the IHI bundles, were aimed at several clinical processes also targeted by the SPI.

In addition, the Health Act 2006 introduced new legislation on mandatory requirements on prevention and control of HCAIs.

It is further relevant that many of these policy initiatives had already been anticipated by significant consensus within professional societies and medical colleges about the appropriate measures to be adopted, and thus enjoyed considerable professional legitimacy – a crucial factor in promoting safe and effective practice.[40]

From a scientific perspective, the contemporaneous changes occurring in the control environments makes it especially difficult to isolate an additive effect of the SPI; the SPI may not have been a sufficient additional dose to generate further differences.

Detecting improvements

Finally, it is possible that any additional effects associated with SPI may simply not be detected yet. The difference between the control hospitals and the SPI hospitals was that the SPI hospitals benefited from a specific organisational intervention designed to promote the building of improvement skills into systems of care. Any SPI effect may be in the form of stickiness. SPI hospitals may potentially be

better equipped to show sustained improvements after the policy spotlight has moved elsewhere. If, however, no differences can be detected in the longer term, the role of organisational interventions of this type in promoting safety will require further examination.

4.5 Theory building

In the previous report, we put forward certain ideas that might explain the mostly null comparative results obtained in the evaluation of SPI1 (which have now been replicated in a more extensive quantitative dataset in SPI2).

These covered the scope of the intervention (the dose may have been too small), the ambitious time scale and certain features of the intervention, such that it was not fully owned by middle grade staff.

The observation that the NHS has adopted certain good practices over the same time scale as the initiative, suggests a further, rather more radical idea: the originators of SPI, along with many opinion formers in management, are working with the wrong theory.

The current theory is largely built around the concept of organisations and the pivotal role they are thought to play in driving up quality. However when it wishes to change practice generally, the NHS works with professional affiliations such as intensive care societies and medical colleges.

Research into why evidence-based guidelines were adopted or ignored in a maternity care context showed that staff were influenced almost entirely through personal/professional networks and hardly at all via the management route.[41] That is not to say that hospitals do not have an essential role to play, but the idea put forward is that this role is enabling not generative in the main. In this respect medical services (and perhaps other highly professionalised groups) may differ from many industries where the hegemony of the organisation can drive change more directly.

From our perspective the changes observed across 18 hospitals in our sample are unlikely to have resulted from concerted and simultaneous management action. This might be expected in the SPI hospitals, but it is unlikely that this would be mimicked simultaneously in the board rooms of control institutions. The idea put forward here is that health services may have learned precisely the wrong lesson by adopting certain ideas and mind-sets from managers and theorists with an industrial background.

4.6 Next steps

From the perspective of these authors there are two dangers to be avoided. The first danger is to despair and resort to nihilism. The corresponding danger is to privilege positive results over null results. Objective proof without subjective interpretations is even more difficult to come by in the evaluation of service delivery interventions than in other branches of science.

Yet while null results remain valuable, face validity is not enough. It is important to recognise that hospitals did report effects from SPI participation. These effects included heightened managerial awareness of, and commitment to, patient safety, and organisational learning about how to implement patient safety improvement efforts in the future.

The intervention did register in the hospitals even if it did not penetrate right through to the sharp end. The challenge is to build on these observed effects. The staff we interviewed theorised about the way forward.

They proposed offering more support to the middle layer of management, engaging clinical leaders at earlier stages and encouraging clinical ownership as a way of securing future success. Reducing the number of areas to be tackled and avoiding areas where there is scientific contestation or dispute about whether something is an important problem were also seen as important.

It was clear that hospitals had learned that addressing issues of legitimacy was a key task. They had identified that introducing initiatives that generated more paperwork would be unpopular among stretched ward staff, and that large scale resourcing and structural support may be needed to implement many patient safety efforts successfully.

References

1 Benning A, Ghaleb M, Suokas A, Dixon-Woods M, Dawson J, Barber N et al. (2011) Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. *BMJ,* vol 342:d195.

2 West MA, Guthrie JP, Dawson JF, Borrill CS and Carter MR (2006). Reducing patient mortality in hospitals: The role of human resource management. *J Organ Behav*, vol 27, pp983-1002.

3 Guzzo RA, Jette D and Katzell RA (1985). The effects of psychologically based intervention programs on worker productivity: a meta-analysis. *Pers Psychol*, vol 38, pp275-92.

4 Borrill CS, West MA, Shapiro D and Rees A (2006). Team working and effectiveness in health. *Brit J Healthc Manag*, vol 6, pp364-71.

5 Michie S and West M (2004). Managing people and performance: An evidence-based framework applied to health service organisations. *Int J Manag Rev*, vol 5-6, pp91-111.

6 Hillman KM, Bristow PJ, Chey T, Daffurn K, Jacques T, Norman SL et al (2001). Antecedents to hospital deaths. *Intern Med J*, vol 31, pp343-8.

7 Smith GB, Prytherch DR, Schmidt P, Featherstone PI, Knight D, Clements G et al (2006). Hospital-wide physiological surveillance-a new approach to the early identification and management of the sick patient. *Resuscitation*, vol 71, pp19-28.

8 British Thoracic Society (2001). Guidelines for the Management of Community Acquired Pneumonia in Adults. *Thorax*, vol 56, ppiv1-64.

9 British Thoracic Society (2004). *Guidelines for the Management of Community Acquired Pneumonia in Adults 2004.* [cited 2010 Jan 27]. Available at: http://www.brit-thoracic.org.uk/Portals/0/Clinical%20Information/Pneumonia/Guidelines/MACAPrevisedApr04.pdf

10 The Royal Pharmaceutical Society of Great Britain, British Medical Association (2008). *British National Formulary*, vol 56. London: Pharmaceutical Press.

11 The Royal Pharmaceutical Society of Great Britain, British Medical Association (2007). *British National Formulary*, vol 56. London: BMJ Publishing Group.

12 The Royal Pharmaceutical Society of Great Britain, British Medical Association (2007). *British National Formulary*, vol 53. London: Pharmaceutical Press.

13 Dean B, Barber N, and Schachter M. What is a prescribing error? *Qual Health Care*, vol 9, pp232-237.

14 Barber N, Franklin BD, Cornford T, Klecun E and Savage I (2006). *Safer, faster, better? Evaluating electronic prescribing. Report to the Patient Safety Research Programme* (Policy Research Programme of the Department of Health). [cited 2010 Jan 27]. Available at: http://www.haps.bham.ac.uk/publichealth/psrp/documents/PS019_Final_Report_Barber.pdf

15 Mohammed MA, Lilford RJ, Rudge G, Dawson JF, Deeks JJ, Girling A et al (2008). *Probing variations in hospital standardised mortality ratios in the West Midlands – A study commissioned by the NHS West Midlands Strategic Health Authority*. Birmingham: The University of Birmingham.

16 Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD (1995). The Quality in Australian Health Care Study. *Med J Aust*, vol 163, pp458-471.

17 Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG et al (1991). Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med*, vol 324, pp370-376.

18 Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA et al (1991). The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med*, vol 324, pp377-384.

19 Vincent C, Neale G and Woloshynowych M (2001). Adverse events in British hospitals: preliminary retrospective record review. *BMJ*, vol 322, pp517-519.

20 Woloshynowych M, Neale G and Vincent C (2003). Case record review of adverse events: a new approach. *Qual Saf Health Care*, vol 12, pp411-415.

21 Institute for Healthcare Improvement (2007). *How to guide – Prevent surgical site infections*. [cited 2010 Mar 22]. Available at: http://www.ihi.org/IHI/Topics/PatientSafety/SurgicalSiteInfections/

22 British Orthopaedic Association (2006). *Primary Total Hip Replacement: A Guide to Good Practice*. [cited 2010 Mar 22]. Available at: http://www.boa.ac.uk/site/showpublications.aspx?ID=59

23 National Institute for Health and Clinical Excellence (2008). *Clinical guideline 65: The management of inadvertent perioperative hypothermia in adults*. [cited 2010 Mar 22]. Available at: http://www.nice.org.uk/nicemedia/pdf/CG65NICEGuidance.pdf

24 National Institute for Health and Clinical Excellence (2010). *Clinical Guideline 74: Prevention and treatment of surgical site infections*. [cited 2010 Mar 22]. Available at: http://www.nice.org.uk/nicemedia/pdf/CG74NICEGuideline.pdf

25 NHS National Patient Safety Agency (2009). *NPSA Clean You Hands campaign*. [cited 2010 Feb 2]. Available at: www.npsa.nhs.uk/cleanyourhands

26 Stone S, Slade R, Fuller C, Charlett A, Cookson B, Teare L et al (2007). Early communication: does a national campaign to improve hand hygiene in the NHS work? Initial English and Welsh experience from the NOSEC study (National Observational Study to Evaluate the Clean Your Hands Campaign). *J Hosp Infect*, vol 66, pp293-296.

27 Department of Health (2009). *Beds open overnight in England*. [cited 2010 Feb 2]. Available at: http://www.dh.gov.uk/en/Publicationsandstatistics/Statistics/Performancedataandstatistics/Beds/DH_083781

28 Shirley PJ (2008). The Safer Patients Initiative: the UK experience of attempting to improve safe clinical care. *Med J Aust*, vol 189, pp414.

29 The Health Foundation (2006). *Safer Patients Initiative trusts explain how they are making hospitals safer for patients*. [cited 2010 Feb 10]. Available at: http://www.health.org.uk/news/blank_1_1.html

30 Harrison DA, Brady AR, and Rowan K (2004). Case mix, outcome and length of stay for admissions to adult, general critical care units in England, Wales and Northern Ireland: the Intensive Care National Audit & Research Centre Case Mix Programme Database. *Crit Care*, vol 8, pp99-111.

31 Harrison DA, Brady AR, Parry GJ, Carpenter JR and Rowan K (2006). Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med*, vol 34, pp1,378-1,388.

32 Harrison DA, Parry GJ, Carpenter JR, Short A, Rowan K (2007). A new risk prediction model for critical care: the Intensive Care National Audit & Research Centre (ICNARC) model. *Crit Care Med*, vol 35, pp1,091-1,098.

33 Healthcare Commission (2006). *Making sense of your staff survey data*. [cited 2009 Jan 27]. Available at: http://www.healthcarecommission.org.uk/_db/_documents/Making_sense_of_your_Staff_Survey_Data_200703214701.pdf

34 Healthcare Commission (2007). *NHS National Staff Survey 2006*. London: The Stationery Office.

35 Lilford R, Edwards A, Girling A, Hofer T, Di Tanna GL, Petty J et al (2007). Inter-rater reliability of case note audit: a systematic review. *J Health Serv Res Policy*, vol 12, pp 173-80.

36 Leape LL,.Berwick DM (2005). Five years after to err is human: what have we learned? *JAMA*, vol 293, pp2,384-2,390.

37 Wilson B, Thornton JG, Hewison J, Lilford RJ, Watt I, Braunholtz D et al (2002). The Leeds University maternity audit project. *Int J Qual Health Care*, vol 14, pp175-181.

38 National Institute for Health and Clinical Excellence (2007). *Clinical Guidelines CG50. Acutely ill patients in hospital. Recognition of and response to acute illness in adults in hospital*. [cited 2010 Mar 22]. Available at: http://www.nice.org.uk/nicemedia/pdf/CG50FullGuidelineShort.pdf

39 Department of Health (2007). *Saving Lives: reducing infection, delivering clean and safe care*. [cited 2010 Mar 22]. Available at: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_078134

40 Dixon-Woods M, Tarrant C, Willars J and Suokas A (2010). How will it work? A qualitative study of strategic stakeholders' accounts of a patient safety initiative. *Qual Saf Health Care*, vol 19, pp74-78.

41 Wilson B, Thornton JG, Hewison J, Lilford RJ, Watt I, Braunholtz D et al (2002). The Leeds University maternity audit project. *Int J Qual Health Care*, vol 14, pp175-181.

Notes from the authors

Ethics

Ethical approval for each sub-study had different considerations and hence separate applications were made for each one. Ethical approval was obtained for the staff and patient surveys from North West Multi-centre Research Ethics Committee and each site granted access to their data. The National Research Ethics Service deemed the case note review as audit/service evaluation and no further ethical approval was required. Permission was also granted from each site to access ICNARC, NOSEC and HAI data. Local research governance was followed at each site.

Contributors

Amirta Benning, Mary Dixon-Woods, Jeremy Dawson, Nick Barber and Richard Lilford designed the study and submitted the grant proposal. Richard Lilford was chief investigator. Amirta Benning, Nick Barber, Richard Lilford, Maisoon Ghaleb and Bryony Dean Franklin designed the explicit case note review pro forma and methods for the explicit case note review. Amirta Benning, Richard Lilford and Ugochi Nwulu designed the semi-structured holistic case note review pro forma and methods for data extraction. Amirta Benning and Ugochi Nwulu were responsible for the case note review collection. Maisoon Ghaleb and Byony Dean Franklin conducted the acute medicine case note review. Martin Carmalt and Thirumalai Naicker conducted the holistic case note review. Martin Carmalt and Clare Derrington carried out a separate review of deaths. Ugochi Nwulu and Maisoon Ghaleb designed the acute medicine case note review database. Gavin Rudge and Amirta Benning created the queries for data extraction. Ugochi Nwulu, Amirta Benning and Richard Lilford designed the perioperative case note review pro forma. Amirta Benning and Ugochi Nwulu designed the perioperative case note review database. Ugochi Nwulu and Amit Kotecha conducted the case note review. Alan

Girling analysed all the explicit case note review data. Gavin Rudge and Amirta Benning designed and wrote database queries for final analysis and to assess the learning effect on the case reviewers. Gavin Rudge captured processed raw mortality data and calculated hospital standardised mortality rates for hospitals in both arms, and undertook analysis of the socio-economic composition of the admitted patient populations of hospitals in the study. Karla Hemming analysed the holistic case note review data. Karla Hemming and Sopna Choudhury performed quantitative analysis of the qualitative data from the stakeholder interviews. Karla Hemming carried out analysis of the infection related data, intensive care mortality data, and hand hygiene related data. Anu Suokas carried out the ethnographic fieldwork. Jeremy Dawson was responsible for all aspects of the staff and patient surveys. All authors contributed to the final manuscript. Richard Lilford is the guarantor.

Acknowledgements

We are grateful to the following for their expertise and contribution in the development of the explicit case note review pro forma for respiratory care:

- Michael Morgan, consultant physician and head of service at University Hospitals of Leicester NHS Trust; honorary professor of respiratory medicine at the University of Leicester; chair of the executive committee, British Thoracic Society
- Martyn Partridge, professor of respiratory medicine at the National Heart and Lung Institute, Imperial College, London, based at Charing Cross Hospital
- Philip Ind, respiratory consultant and honorary senior lecturer at the NHLI, Imperial College, London.

We are also grateful to the following for their expertise in the development of the surgical explicit case note review pro forma:

- Dion Morton, professor of surgery, University Hospitals Birmingham NHS Foundation Trust
- David Thomas, consultant anaesthetist, Heart of England NHS Foundation Trust.

We would like to thank Sheldon Stone and Chris Fuller for access to data collected as part of the NOSEC study, and David Harrison and Kathy Rowan from ICNARC for access to data collected as part of the Case Mix Programme. We also thank Janet Willars for conducting stakeholder interviews and for helping to facilitate focus groups.

We would also like to thank Dale Webb, Louise Thomas and Simona Arena for their insight into the implementation of the SPI.

Finally, we are grateful to Peter Chilton for his assistance in the preparation of this report.

Funding

The Health Foundation and NPSA funded this evaluation, and it was sponsored for research governance purposes by the University of Birmingham. The study was designed independently by the researchers. The researchers acted independently but worked collaboratively with the funder. The researchers independently collected, analysed and interpreted the data. The researchers wrote this evaluation report independently. The funders were given the opportunity to provide comments prior to submission. All researchers (apart from CD) had the opportunity to access participant anonymised data.

Competing interests

None declared.

Appendix 1

Staff survey – 13 questions identified as relevant to the SPI

Six of these 13 scores are straightforward percentages:

1. **Percentage of staff having well structured appraisals** reflects the percentage of respondents who not only say that they had received an appraisal in the previous 12 months, but that this appraisal helped them improve how to do their job, helped agree clear objectives for their work, and left them feeling that their work was valued by their organisation. These aspects of appraisal have been shown to be particularly important for organisational outcomes in many sectors, including healthcare.[2;3]

2. **Percentage of staff working in well-structured teams** is the percentage of respondents who said they worked in teams, that their teams had clear objectives, that they had to work closely with team members to achieve these objectives, and that the team met regularly to discuss their effectiveness and how it could be improved. These are features of team working that have been shown to be critical for achieving high-quality team outcomes.[4]

3. **Percentage of staff witnessing potentially harmful errors or near misses in previous month** was the percentage of respondents who said they had witnessed an error or a near miss in the previous month that could have harmed either patients or staff.

4. **Percentage of staff suffering work-related injury** is the percentage of respondents who said they had suffered injury or illness as a result of moving or handling; needlestick or sharps injuries; slips, trips or falls; or exposure to dangerous substances in the previous 12 months;

5. **Percentage of staff suffering work-related stress** is the percentage of respondents who said they had suffered injury or illness as a result of work-related stress in the previous 12 months.

6. **Percentage staff experiencing physical violence from patients/ relatives** was the percentage of respondents who said they had personally experienced physical violence at work from either patients, or relatives of patients, in the previous 12 months.

Six of the other seven scores were calculated as the mean of a number of separate questionnaire items, each scored from one to five representing answers from strongly disagree through to strongly agree, or from very dissatisfied to very satisfied:

7. Intention to leave shows the extent to which employees are considering leaving their jobs. It is based on three questionnaire items.
8. Staff job satisfaction is a measure of employees' overall satisfaction with their jobs, and is based on seven items.
9. Quality of work-life balance measures the support provided by organisations for employees to maintain a good work-life balance, and is based on three items.
10. Support from supervisors is a measure of the extent to which employees feel supported by their immediate managers at work, and is based on five items.
11. Organisational climate is a measure of the overall climate, or positive feeling, within the organisation, including factors such as trust in management, communication, staff involvement in decision making and emphasis on quality. This is based on six items. Each of these scores has been shown to relate to performance outcomes, including quality of care, in healthcare organisations.[5]
12. Fairness and effectiveness of incident reporting procedures is a measure of the extent to which employees trust procedures for reporting and dealing with errors, near misses and incidents are effective and fair. This is based on seven items.

One other variable was also measured on a similar scale, but with some slight differences:

13. Availability of hand-washing materials is a measure of the extent to which hand-washing materials (hot water, soap and paper towels, or AHR) are available when needed by different groups. This was originally measured on a scale from one to four representing answers from never through to always, and then adjusted to fit a one to five scale for consistency with the other scale scores.

Appendix 2

Patient survey – five identified scores relevant to SPI

Each of these was scored between 0 and 100. The three satisfaction scores were:

1. Overall, how would you rate the care you received? (five possible responses: excellent = 100, very good = 75, good = 50, fair = 25 and poor = 0)
2. How would you rate how well the doctors and nurses worked together? (same response options)
3. Overall, did you feel you were treated with respect and dignity while you were in the hospital? (yes, always = 100; yes, sometimes = 50; and no = 0).

The two scores related to cleanliness were:

4. In your opinion, how clean was the hospital room or ward that you were in? (possible responses: very clean = 100, fairly clean = 67, not very clean = 33, and not at all clean = 0)
5. How clean were the toilets and bathrooms that you used in hospital? (same response options, plus 'I did not use a toilet or bathroom', which was excluded from the analysis).

Appendix 3

Errors and adverse events – analysis tables

Table 3.10A: Ratings and rates of adverse effects and errors: differences between SPI2 hospitals and control hospitals at baseline; and changes between epoch 3 and baseline in the control hospitals (99% CIs are in parenthesis)

	Comparisons at baseline* (1) Intervention – Control	Changes in Controls* (2) Epoch 3 – Baseline
Quality ratings:		
Admission rating†	0.12 (-0.27, 0.50)	0.11 (-0.32,0.26)
Management rating†	0.14 (-0.33, 0.61)	0.28 (-0.29, 0.84)
Pre-discharge rating†	0.00 (-0.54,0.54)	0.11 (-0.38,0.60)
Overall care rating‡	0.10 (-0.30, 0.48)	0.29 (-0.12, 0.69)
Errors/Adverse Events:		
No. errors$^{\Phi}$	-5.78 (-23.84, 12.28)	-14.35 (-32.42, 3.71)
No. adverse events$^{\Phi}$	-1.42 (-5.81, 2.97)	-1.70 (-7.37, 3.96)

* Effects are estimated from a mixed effects model (see methods for details) and represent differences at baseline (1) and the effect of time (2). Baseline refers to the average scores over epoch 1 and epoch 2.

† Score scale: one (below best practice) to six (excellent care).

‡ Score scale: one (unsatisfactory) to 10 (very best care).

$^{\Phi}$ Number of errors and number of adverse events are per 100 patients (patients could experience more than one error and more than one adverse event).

Errors can be of multiple categories.

Table 3.11A: Rates per 100 patients of errors identified by broad category of error: differences between SPI2 hospitals and control hospitals at baseline; and changes between Epoch 3 and baseline in the control hospitals (99% CIs are in parenthesis)

	Comparisons at baseline* (1) Intervention – Control	Changes in Controls* (2) Epoch 3 – Baseline
Quality ratings:		
Diagnosis/assessment/admission error	-3.28 (-27.15,20.60)	-13.08 (-36.31, 10.14)
Hospital-acquired infection	-0.00 (-0.93,0.93)	0.88 (-0.28,2.04)
Technical/management	-3.58 (-10.50, 3.34)	-1.17 (-9.66,7.31)
Medication/maintenance/follow-up	-1.08 (-11.24, 9.07)	-8.54 (-21.43, 4.35)
Clinical reasoning	-4.90 (-18.56, 8.76)	-10.93 (-24.84, 2.97)
Discharge information	0.62 (-9.43, 10.67)	-5.63 (-16.14, 4.87)

* Effects are estimated from a mixed effects model (see methods for details) and represent differences at baseline (1) and the effect of time (2). Baseline refers to the average scores over epoch 1 and epoch 2.

Errors can be of multiple categories.

Appendix 4

C. diff and MRSA – analysis tables and figures

Table A1: Fitted models for rate of C. diff (per 1,000 bed days) and MRSA infections (per 100,000 bed days)

	C. diff		MRSA	
	Coeff (se)	**p-value**	**Coeff (se)**	**p-value**
Constant	0.94 (0.22)	0.000	15.36 (2.51)	0.000
Intervention	0.05 (0.28)	0.853	2.37 (0.14)	0.420
Time	-0.13 (0.07)	0.051	0.26 (0.50)	0.601
Time^2	-0.01 (0.01)	0.264	0.01 (0.03)	0.789
Time^3	0.00 (0.00)	0.784	-0.00 (0.01)	0.208
Intervention*time	-0.01 (0.02)	0.652	-0.05 (0.14)	0.693

Table A2: Fitted models for rate of soap and AHR (litres) consumption per 1,000 bed days

	Soap		AHR	
	Coeff (SE)	**p-value**	**Coeff (SE)**	**p-value**
Constant	41.76(13.3)	0.000	3.80 (10.5)	0.708
Intervention	0.73 (13.9)	0.941	10.90 (12.2)	0.371
Time	0.73 (1.82)	0.623	3.91 (1.28)	0.002
Time^2	-0.03 (0.08)	0.657	-0.12 (0.06)	0.034
Time^3	0.00 (0.00)	0.501	0.00 (0.00)	0.065
Intervention*time	0.08 (0.44)	0.760	-0.05 (0.38)	0.889

Table A3: Fitted models for observed to expected mortality ratio (exponential scale) and mean length of stay for patients admitted to ICU

	O/E mortality		Mean LOS	
	Coeff (SE)	p-value	Coeff (SE)	p-value
Constant	1.28 (0.12)	0.000	180.4 (19.7)	0.000
Intervention	-0.14 (0.08)	0.068	-39.4 (17.2)	0.022
Before	-0.07 (0.06)	0.258	-12.9 (8.49)	0.128
Intervention before	0.09 (0.08)	0.250	5.9 (11.11)	0.598
APACHE II score	0.01 (0.01)	0.138	0.34 (1.18)	0.774
Physiology score	-0.01 (0.01)	0.015	-1.34 (0.87)	0.123

Table A4: Fitted models for APACHE II and ICNARC physiology scores for patients admitted to ICU from a ward within the hospital

	APACHE II score		ICNARC score	
	Coeff (SE)	p-value	Coeff (SE)	p-value
Constant	18.47 (0.72)	0.000	20.95 (1.00)	0.000
Intervention	1.20 (0.98)	0.225	2.32 (1.36)	0.087
Before	1.85 (0.81)	0.022	1.77 (1.19)	0.136
Intervention before	-0.83 (1.09)	0.449	-2.26 (1.60)	0.158

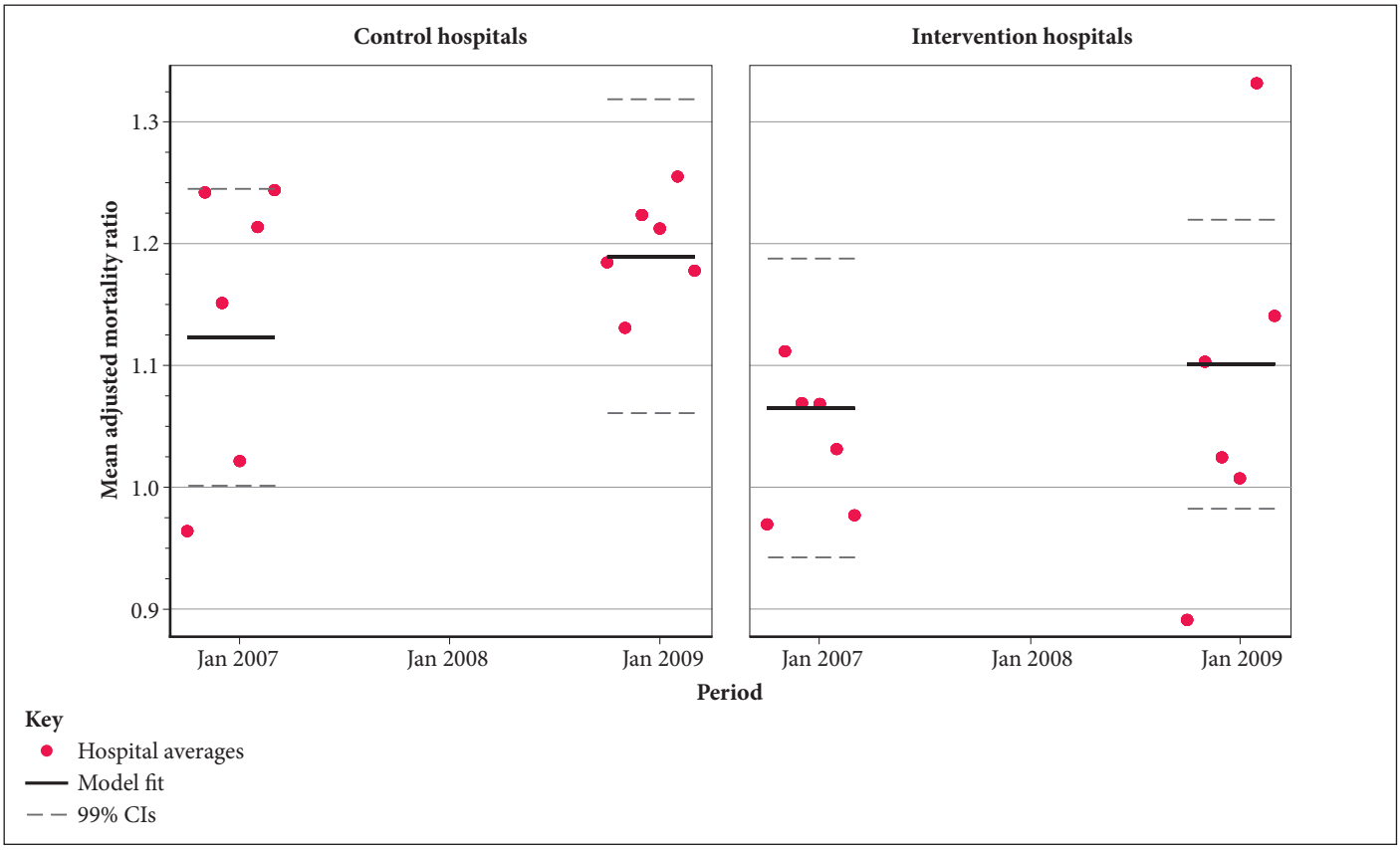


Figure A1: ICUs adjusted mortality rates

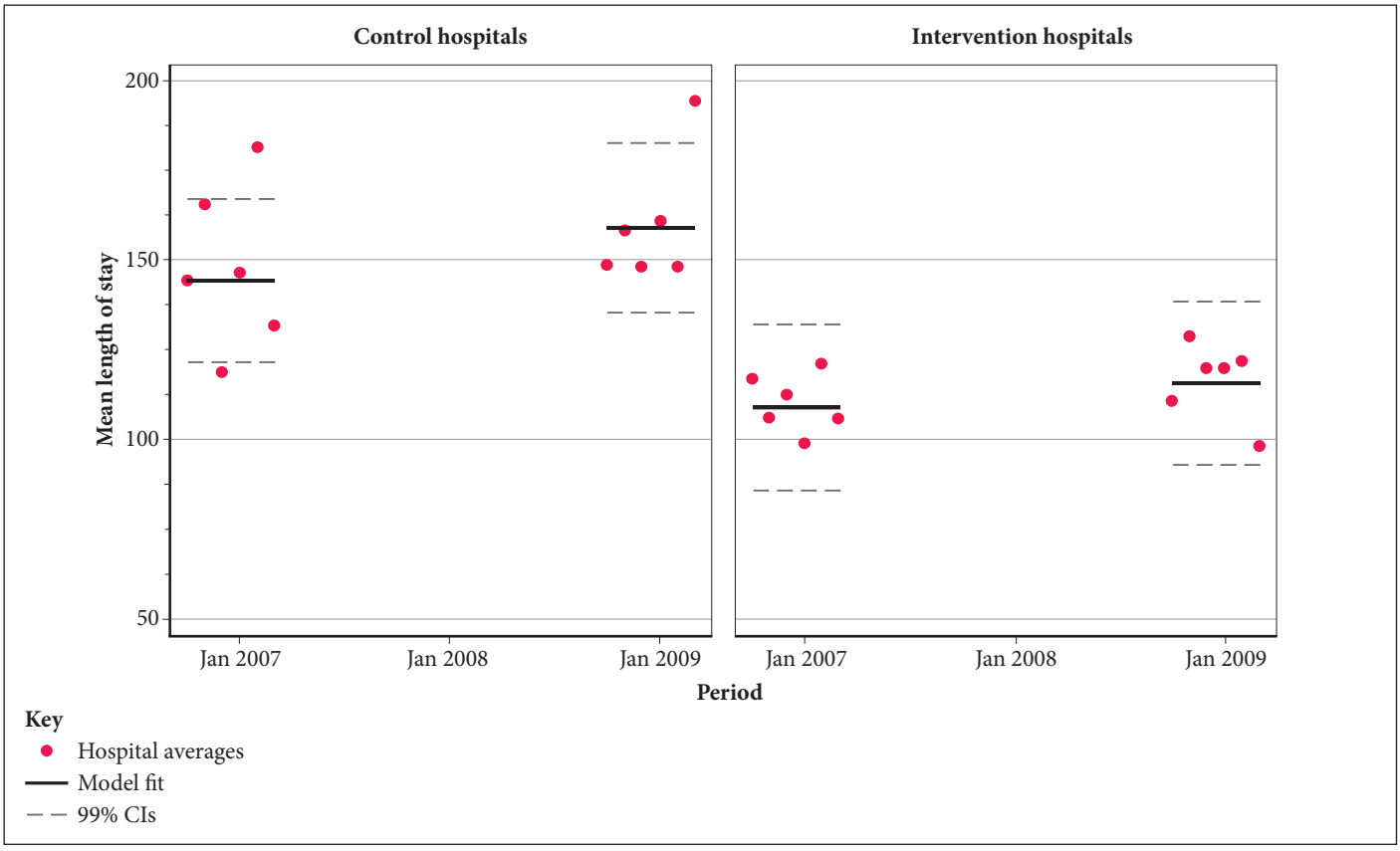


Figure A2: ICUs length of stay

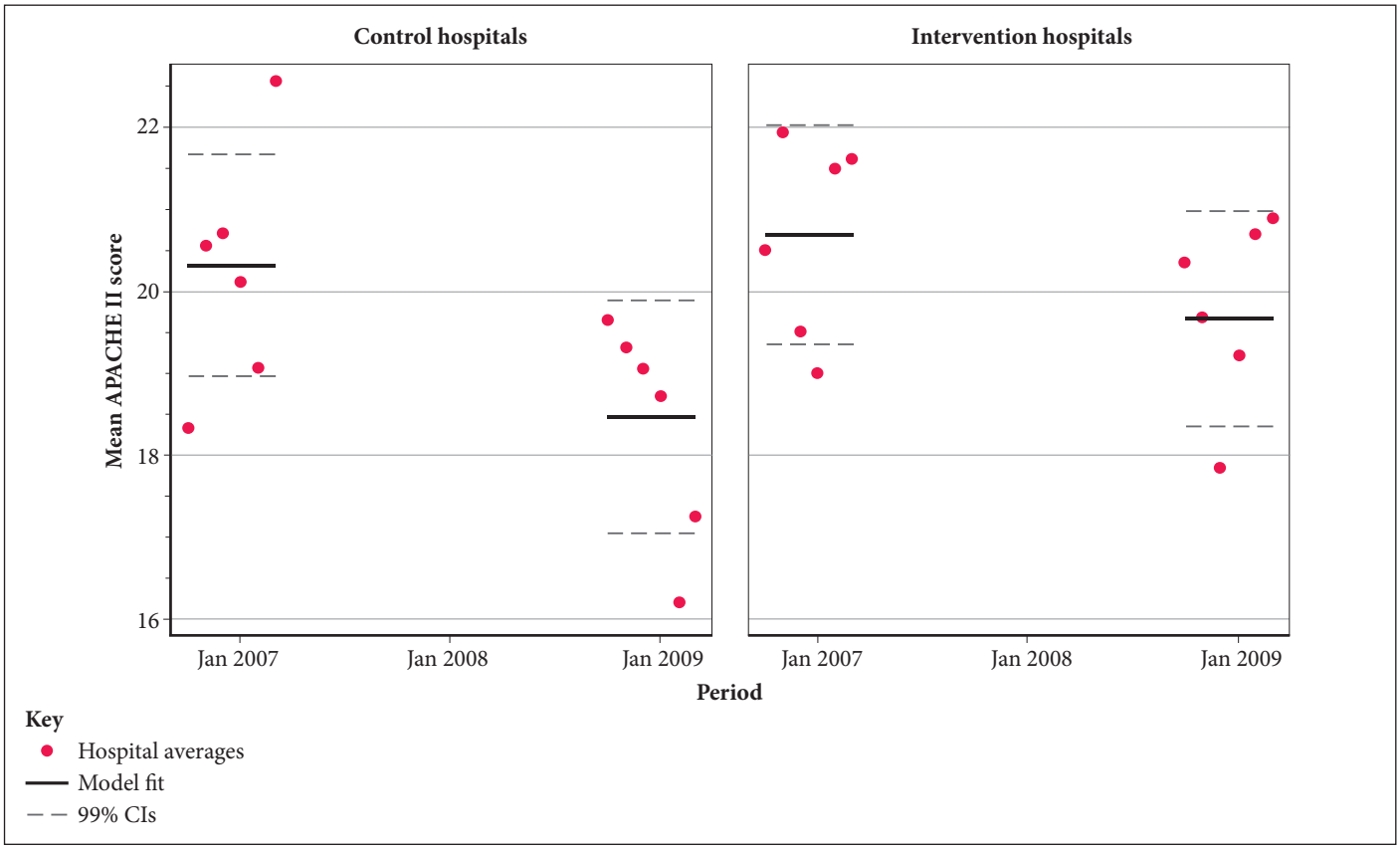


Figure A3: ICUs: APACHE II score

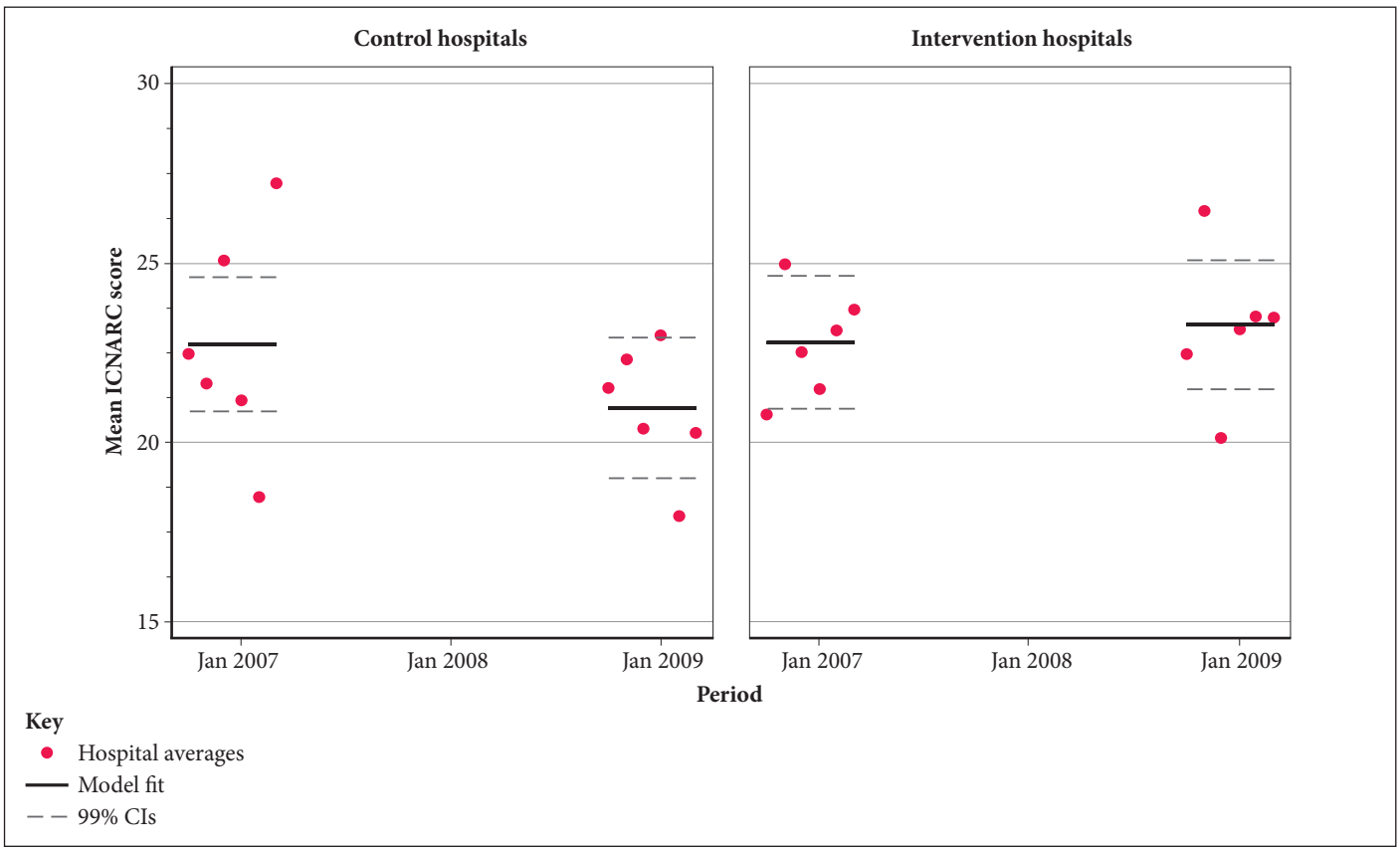


Figure A4: ICUs: ICNARC score

The Health Foundation is an independent charity working to continuously improve the quality of healthcare in the UK.

We want the UK to have a healthcare system of the highest possible quality – safe, effective, person-centred, timely, efficient and equitable. We believe to achieve this, health services need to continually improve the way they work.

We are here to inspire and create the space for people, teams, organisations and systems to make lasting improvements to health services.

Working at every level of the healthcare system, we aim to develop the technical skills, leadership, capacity, knowledge, and the will for change, that are essential for real and lasting improvement.

The Health Foundation
90 Long Acre
London WC2E 9RA
T 020 7257 8000
F 020 7257 8001
E info@health.org.uk

Registered charity number: 286967
Registered company number: 1714937

www.health.org.uk

ISBN 978-1-906461-22-5