A collection of key articles

This special collection reflects the depth and breadth of content from BMJ Quality & Safety, which encourages innovation and creative thinking to improve the quality of health care and the science of improvement.
A collection of key articles

1. Identification of doctors at risk of recurrent complaints: a national study of healthcare complaints in Australia
   Marie M Bismark, Matthew J Spittal, Lyle C Gurrin, Michael Ward, David M Studdert

2. The global burden of unsafe medical care: analytic modelling of observational studies
   Ashish K Jha, Itziar Larizgoitia, Carmen Audera-Lopez, Nittita Prasopa-Plaizier, Hugh Waters, David W Bates

   Michael J Taylor, Chris McNicholas, Chris Nicolay, Ara Darzi, Derek Bell, Julie E Reed

4. Assessing adverse events among home care clients in three Canadian provinces using chart review
   Régis Blais, Nancy A Sears, Diane Doran, G Ross Baker, Marilyn Macdonald, Lori Mitchell, Stéphane Thaïès

5. ‘Care left undone’ during nursing shifts: associations with workload and perceived quality of care
   Jane E Ball, Trevor Murrells, Anne Marie Rafferty, Elizabeth Morrow, Peter Griffiths

6. Allocating scarce resources in real-time to reduce heart failure readmissions: a prospective, controlled study
   Ruben Amarasingham, Parag C Patel, Kathleen Toto, Lauren L Nelson, Timothy S Swanson, Billy J Moore, Bin Xie, Song Zhang, Kristin S Alvarez, Ying Ma, Mark H Drazner, Usha Kollipara, Ethan A Halm

7. Culture and behaviour in the English National Health Service: overview of lessons from a large multimethod study
   Mary Dixon-Woods, Richard Baker, Kathryn Charles, Jeremy Dawson, Gabi Jerzembek, Graham Martin, Imelda McCarthy, Loma McKee, Joel Minion, Piotr Ozieranski, Janet Willars, Patricia Wilkie, Michael West

qualitysafety bmj.com
ABSTRACT

Objectives (1) To determine the distribution of formal patient complaints across Australia’s medical workforce and (2) to identify characteristics of doctors at high risk of incurring recurrent complaints.

Methods We assembled a national sample of all 18 907 formal patient complaints filed against doctors with health service ombudsmen (‘Commissions’) in Australia over an 11-year period. We analysed the distribution of complaints among practicing doctors. We then used recurrent-event survival analysis to identify characteristics of doctors at high risk of recurrent complaints, and to estimate each individual doctor’s risk of incurring future complaints.

Results The distribution of complaints among doctors was highly skewed: 3% of Australia’s medical workforce accounted for 49% of complaints and 1% accounted for a quarter of complaints. Short-term risks of recurrence varied significantly among doctors: there was a strong dose-response relationship with number of previous complaints and significant differences by doctor specialty and sex. At the practitioner level, risks varied widely, from doctors with <10% risk of further complaints within 2 years to doctors with >80% risk.

Conclusions A small group of doctors accounts for half of all patient complaints lodged with Australian Commissions. It is feasible to predict which doctors are at high risk of incurring more complaints in the near future. Widespread use of this approach to identify high-risk doctors and target quality improvement efforts coupled with effective interventions, could help reduce adverse events and patient dissatisfaction in health systems.

INTRODUCTION

To many doctors who are sued or complained against, the event seems random. At the population level, however, there are patterns. Previous studies have compared doctors who experienced multiple malpractice claims,1–5 complaints,6 7 and disciplinary actions8–10 with doctors who experienced few or none, and identified differences in the sex, age and specialty profile of the two groups. Such research helps to explain medico-legal risk retrospectively, but does not provide practical guidance for identifying risks prospectively. Clinical leaders, risk managers, liability insurers and regulators all lack reliable methods for systematically determining which doctors should be targeted for assistance and preventive action before they acquire troubling track records. Consequently, the medico-legal enterprise remains reactive, dealing primarily with the aftermath of adverse events and behaviours that lead to costly disputes.

The conventional wisdom is that future medico-legal events cannot be predicted at the doctor level with acceptable levels of accuracy.11 12 Numerous studies have tried,13–25 most with limited success. This body of research has two important shortcomings. First, only a few studies15 17 21 report a method for predicting medico-legal risk that is potentially replicable, and these methods are statistically complex. The practical consequence is that regulators and liability insurers today have no clear way of estimating risk at the practitioner level, and doing so is not a standard part of risk management practice.

Second, no study to date has found a way to deal well with temporal aspects of risk, such as the evolving nature of doctors’ medico-legal event histories, which can be crucial information in assembling a risk profile. Previous claims and complaints have been identified as an important predictor of future events, but only in analyses that specify this variable crudely—usually by ‘freezing’ a doctor’s
track record at a specific point to estimate a ‘one-time’ effect. This approach is out of step with how claims and complaints are managed. The frontline challenges are to determine how a practitioner’s risk profile changes over time as new information (including new events) comes to hand; when support or intervention measures to prevent further events are warranted; and how strong those measures should be. A risk prediction method that helped to address these questions would have considerable potential for boosting the contribution of medico-legal institutions to quality improvement.

We assembled a national sample of nearly 19,000 formal healthcare complaints lodged against doctors in Australia between 2000 and 2011. We then used a time-to-event method of analysis to determine characteristics of doctors posed to incur recurrent complaints, and to estimate each practitioner’s risk of recurrence at specific time points. The study had two main goals: to identify predictors of complaint-prone doctors in Australia, and to develop a robust and useful method for forecasting medico-legal risk.

**METHODS**

**Setting**

Health service commissions (Commissions) are statutory agencies established in each of Australia’s six states and two territories. Commissions have responsibility for receiving and resolving patient complaints about the quality of healthcare services. Patients or their advocates must initiate complaints in writing, but the process is free and legal representation is optional.

Table 1 compares the jurisdiction and functions of Commissions to those of the two other agencies that handle medico-legal matters in Australia—civil courts and the Medical Board of Australia.

### Table 1: Jurisdiction and functions of key agencies with responsibility for medico-legal matters in Australia

<table>
<thead>
<tr>
<th>Civil courts</th>
<th>Health complaints commissions</th>
<th>Medical Board of Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cases handled</strong></td>
<td>Negligence claims</td>
<td>Patient complaints</td>
</tr>
<tr>
<td><strong>Jurisdictional focus</strong></td>
<td>Substandard care causing patient harm</td>
<td>Low-quality care</td>
</tr>
<tr>
<td><strong>Procedures used</strong></td>
<td>Out-of-court negotiation</td>
<td>Early resolution</td>
</tr>
<tr>
<td></td>
<td>Alternative forms of dispute resolution (eg, mediation, arbitration)</td>
<td>Conciliation</td>
</tr>
<tr>
<td></td>
<td>Trials before judges</td>
<td>Investigation</td>
</tr>
<tr>
<td><strong>Remedies</strong></td>
<td>Monetary damages</td>
<td>Communication (eg, facilitate apology or explanation)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Restoration (eg, facilitate provision of further treatment, fee forgiveness, monetary settlement)</td>
</tr>
</tbody>
</table>

*Typically, such sanctions are imposed by external administrative tribunals in proceedings initiated by the Medical Board of Australia.

Outside of the clinic or hospital in which care is received, Commissions are the primary avenue of redress for patients dissatisfied with the quality of care they have received. Plaintiffs’ lawyers in Australia will rarely take on cases unless they have first proceeded through Commission processes (although the vast majority of complaints do not become negligence claims). At least 10 other Organisation for Economic Co-operation and Development (OECD) countries—including Austria, Finland, Israel, New Zealand and the UK—have similar bodies. In the UK, the closest analogue is the Parliamentary and Health Service Ombudsman.

Commissions in all Australian states and territories except South Australia participated in the study. These seven jurisdictions have 21 million residents and 90% of the nation’s 88,000 registered doctors. The study was approved by the ethics committee at the University of Melbourne.

### Data

Between May 2011 and February 2012 we collected data on-site at Commission offices in each participating state and territory. Complaints against doctors were identified by querying the Commissions’ administrative data systems. The filing period of interest spanned 12 years and differed slightly by jurisdiction: 2000–2011 for the Australian Capital Territory, the Northern Territory, Queensland, Tasmania and Victoria; 2000–2010 for Western Australia; and 2006–2011 for New South Wales.

All Commissions record the names of persons and institutions that are the subject of complaints, as well as the filing date, the nature of the complaint, the type of health professional named and their practice location. Although all Commissions recorded doctors’ clinical specialty, the quality of this variable was mixed. Doctors’ age and sex were not routinely
collected. We therefore supplemented the Commissions’ administrative data with data from another source.

AMPCo Direct, a subsidiary of the Australian Medical Association, maintains a comprehensive list of doctors in Australia, including information on their sex, date of birth, specialty and subspecialty, and practice location. We purchased the AMPCo Direct database and matched doctors listed in it to doctors named in the complaints databases. The matching method is described in an online supplementary appendix.

Variables
We coded specialty into 13 categories, based on those promulgated by the Medical Board of Australia.29 Doctors’ principal practice address was classified as urban or rural, based on the location of its postcode within a standard geographic classification system.30 The nature of concerns raised in complaints was sorted into 20 broad ‘issue’ categories. Commissions run dispute resolution processes; they generally do not rule on the merit of complaints, nor make findings for or against parties, so it was not possible to include a variable indicating how meritorious complaints were.

Statistical analysis

Distributional analysis
We plotted the cumulative distribution of complaints among two populations of doctors: (1) all unique doctors named in complaints and (2) all practicing doctors in the seven jurisdictions under study (ie, regardless of whether they had been named in complaints). The size of this second population was based on the number of doctors in employment in 2006,31 the median study year. Because certain classes of complaints do not name doctors individually (eg, complaints arising in public hospitals in several of the study jurisdictions), we adjusted the proportions in the distributional calculations to ensure the numbers (number of complaints) matched the denominators (size of the ‘exposed’ segment of the medical workforce). Details are provided in the online supplementary appendix.

Multivariable survival analysis
We used multivariable survival analysis to identify predictors of doctors’ risks of recurrent complaints. Specifically, we used an Anderson–Gill model12 in which the time-scale ran from time from first event (ie, a doctor’s earliest complaint) and allowed each doctor in the sample to accrue multiple complaints over the period of observation. The outcome variable was the occurrence of a complaint against a doctor, conditional on the doctor having been named in an earlier complaint. The covariates were the number of prior complaints a doctor had experienced, jurisdiction, and the doctor’s specialty, age, sex and principal practice location.

The number of prior complaints was specified as a time-varying covariate. Age was also time-varying in the sense that we allowed doctors to move into higher age categories, commensurate with their age at the time of the complaint. We fit cluster-adjusted robust SEs to account for doctors who experienced repeated complaints over time.

Details of model selection and specification are described in the online supplementary appendix. All statistical analyses were conducted using Stata 12.1.

Risk predictions
To estimate doctors’ risks of experiencing complaints over time, we plotted adjusted failure curves.33 34 Details of the statistical techniques used to create these curves are provided in the online supplementary appendix. We also plotted failure curves showing the predicted risk of recurrent complaints for several individual doctors. Values for all failure curves were computed using coefficients from the main multivariable model, and hence, derived from the survivor function, S(t).

Sensitivity analysis
We tested the robustness of estimates from the main multivariable analysis by rerunning the analysis on a subsample of complaints (n=10 010) with issue codes suggestive of relatively serious concerns (namely, poor clinical care, breach of conditions, rough or painful treatment and sexual contact or relationship).

RESULTS

Characteristics of complained-against doctors and complaints
The study sample consisted of 18 907 complaints against 11 148 doctors. Sixty-one percent of the complaints addressed clinical aspects of care, most commonly concerns with treatment (41%), diagnosis (16%) and medications (8%) (table 2). Nearly one quarter of complaints addressed communication issues, including concerns with the attitude or manner of doctors (15%), and the quality or amount of information provided (6%).

Seventy-nine percent of the doctors named in complaints were male, 47% were general practitioners and 14% were surgeons (table 3). Examples of several complaints are included in the online supplementary appendix.

Incidence and distribution of complaints
Doctors in the sample were complained against an average of 1.98 times (SD 2.31). The distribution was highly skewed, with a small subgroup of doctors accounting for a disproportionate share of complaints.

Figure 1 plots the cumulative distribution of complaints among doctors in six jurisdictions over a decade. (New South Wales data was not included in these plots because the complaints window there spanned only 5 years.) The curve on the left side of
the figure shows the distribution of complaints among doctors who experienced one or more complaints in the decade. Fifteen percent of doctors named in complaints accounted for 49% of all complaints, and 4% accounted for a quarter of all complaints. The curve on the right side of the figure shows the distribution of complaints across the full population of practicing doctors, not just those who experienced complaints. Three percent of all doctors accounted for 49% of all complaints, and 1% accounted for a quarter of all complaints.

**Multivariable predictors of recurrent complaints**

In multivariable analyses, the number of prior complaints doctors had experienced was a strong predictor of subsequent complaints, and a dose-response relationship was evident (table 4). Compared with doctors with one prior complaint, doctors with two complaints had nearly double the risk of recurrence (HR 1.93; 95% CI 1.79 to 2.09), and doctors with five prior complaints had six times the risk of recurrence (HR 6.16; 95% CI 5.09 to 7.46). Doctors with 10 or more prior complaints had 30 times the risk of recurrence (HR 29.56; 95% CI 19.24 to 45.41).

Risk of recurrence also varied significantly by specialty. Compared with general practitioners, plastic surgeons had twice the risk (HR 2.04; 95% CI 1.75 to 2.38), and risks were approximately 50% higher among dermatologists (HR 1.56; 95% CI 1.30 to 1.88) and obstetrician-gynecologists (HR 1.50; 95% CI 1.29 to 1.76). Anaesthetists had significantly lower risks of recurrence (HR 0.65; 95% CI 0.54 to 0.79).

Male doctors had a 40% higher risk of recurrence than their female colleagues (HR 1.36; 95% CI 1.23 to 1.50). Location of practice (urban vs rural) was not significantly associated with recurrence. Compared with doctors 35 years of age or younger, older doctors had 30–40% higher risks of recurrence; this level of heightened risk was similar through the middle-aged and older-aged groups.

**Risks of recurrence over time**

Doctors named in a third complaint had a 38% chance of being the subject of a further complaint within a year, and a 57% probability of being complained against again within 2 years (figure 2A). Doctors named in a fifth complaint had a 59% 1-year
complaint probability and a 79% 2-year complaint probability. Recurrence was virtually certain for doctors who had experienced 10 or more complaints, with 97% incurring another complaint within a year. Regardless of the number of previous complaints, doctors’ risks of further complaints increased sharply in the first 6 months following a complaint, and then declined steadily thereafter. This is evident from the steep rise and then plateauing of the curves in figure 2A (these curves plot cumulative risks over time).

The curves shown in figure 2A depict average population-level risks for selected predictors, controlling for other covariates. However, our modelling approach is fundamentally designed to predict risk at the practitioner level. Figure 2B illustrates this; it shows wide variation in risk profiles among a selection of seven doctors in the sample. Doctor A, for instance, is a 62-year-old male general practitioner who accumulated 10 complaints over 9.2 years of observation. He had a 39% risk of recurrence after his fourth complaint, a 61% risk after his fifth complaint and a 94% risk after his sixth complaint.

Sensitivity analysis
Re-estimating the main multivariable model using a subset of ‘severe’ complaints produced very similar results to the main model. The online supplementary appendix shows the full set of results.

DISCUSSION
This study of patient complaints made to the chief health-quality regulators in Australia found that the complaints clustered heavily among a small group of doctors. Approximately 3% of practicing doctors accounted for half of all complaints. The number of prior complaints doctors had experienced was a particularly strong predictor of their short-term risk of further complaints. At the practitioner level, short-term risks of recurrence varied widely, from <10% risk among low-risk doctors to >80% risk among high-risk doctors. Overall, recurrent-event survival

Table 4  Multivariable regression analysis estimating risk of recurrent complaints

<table>
<thead>
<tr>
<th>Number of prior complaints</th>
<th>HR (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (ref)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.93 (1.79 to 2.09)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3</td>
<td>3.21 (2.87 to 3.59)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4.54 (3.90 to 5.27)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6.16 (5.09 to 7.46)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8.83 (7.05 to 11.05)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>9.57 (7.40 to 12.37)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>9.49 (7.05 to 12.77)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>16.09 (11.72 to 22.10)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>10 or more</td>
<td>29.56 (19.24 to 45.41)</td>
<td></td>
</tr>
</tbody>
</table>

States and territories

<table>
<thead>
<tr>
<th>States and territories</th>
<th>HR (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (ref)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.23 (1.86 to 2.67)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3</td>
<td>2.10 (1.75 to 2.53)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.91 (1.53 to 2.37)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.86 (1.52 to 2.29)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.73 (1.37 to 2.20)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.25 (1.02 to 1.53)</td>
<td></td>
</tr>
</tbody>
</table>

Male doctor

<table>
<thead>
<tr>
<th>Male doctor</th>
<th>HR (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.36 (1.23 to 1.50)</td>
<td></td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Urban practice location

<table>
<thead>
<tr>
<th>Urban practice location</th>
<th>HR (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.98 (0.90 to 1.07)</td>
<td></td>
<td>0.65</td>
</tr>
</tbody>
</table>

Specialty of doctor

<table>
<thead>
<tr>
<th>Specialty of doctor</th>
<th>HR (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plastic surgery</td>
<td>2.04 (1.75 to 2.38)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Dermatology</td>
<td>1.56 (1.30 to 1.88)</td>
<td></td>
</tr>
<tr>
<td>Obstetrics and gynaecology</td>
<td>1.50 (1.29 to 1.76)</td>
<td></td>
</tr>
<tr>
<td>General surgery</td>
<td>1.45 (1.17 to 1.80)</td>
<td></td>
</tr>
<tr>
<td>Orthopaedic surgery</td>
<td>1.32 (1.20 to 1.44)</td>
<td></td>
</tr>
<tr>
<td>Other surgery</td>
<td>1.30 (1.19 to 1.43)</td>
<td></td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>1.19 (1.02 to 1.40)</td>
<td></td>
</tr>
<tr>
<td>Psychiatry</td>
<td>1.15 (1.02 to 1.29)</td>
<td></td>
</tr>
<tr>
<td>General practice (ref.)</td>
<td>1.00 (ref.)</td>
<td></td>
</tr>
<tr>
<td>Internal medicine</td>
<td>0.93 (0.80 to 1.09)</td>
<td></td>
</tr>
<tr>
<td>Radiology</td>
<td>0.89 (0.34 to 2.37)</td>
<td></td>
</tr>
<tr>
<td>Anaesthesia</td>
<td>0.65 (0.54 to 0.79)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.65 (0.51 to 0.82)</td>
<td></td>
</tr>
</tbody>
</table>

Age of doctor

<table>
<thead>
<tr>
<th>Age of doctor</th>
<th>HR (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;35 years</td>
<td>1.00 (ref.)</td>
<td></td>
</tr>
<tr>
<td>36–45 years</td>
<td>1.31 (1.13 to 1.51)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>46–55 years</td>
<td>1.40 (1.22 to 1.62)</td>
<td></td>
</tr>
<tr>
<td>56–65 years</td>
<td>1.43 (1.23 to 1.67)</td>
<td></td>
</tr>
</tbody>
</table>

| Gamma         | –0.21 (–0.23 to –0.19) |         |

*Analysis based on 14 986 index complaints against 8749 doctors, and 6237 subsequent complaints.
analysis showed considerable promise as a statistical approach for flagging complaint-prone doctors early in their complaints trajectory, using only a few simple descriptive characteristics.

Our study used a national sample to examine the distribution and predictors of medico-legal events. Patients treated in healthcare facilities throughout seven states and territories were eligible to file complaints with a Commission about the quality of the care they received. Previous studies of claims and complaints risk have tended to focus on pools of doctors covered by a single liability insurer or a few hospitals.

The extent to which complaints were concentrated in a small group of doctors was striking, consistent with other studies of complaints7 19 22 and claims.18 This highly skewed distribution of medico-legal events among doctors has several implications. The obvious one is that there is a pressing need for interventions that address the behaviour of doctors who are chronically complained or claimed against. Medical boards in Australia and elsewhere already address conduct, competence and health concerns with certain practitioners, but these efforts may fall short. Our study identifies a target population within which systematic deployment of interventions to improve performance35 36 might be manageable: less than 500 doctors accounted for 25% of all complaints that named doctors in the decade under study. Immediate steps to improve, guide or constrain the care being provided by these ‘high-risk’ practitioners could be a very cost-effective way to advance quality and safety, and produce measurable benefits at the system level.

A more sobering implication of the clustering phenomenon is that remediation activities targeted at doctors who have attracted many complaints, while critical, come too late. Complaints are best understood as sentinel events, and complainants as representatives of much larger groups of harmed or dissatisfied patients.37–39 By the time multiple complaints have accrued, substantial damage to quality of care is likely to have occurred already. The clustering of medico-legal events highlights the huge gains that would be put in reach by a capability to identify early doctors who are on course to incur multiple complaints.

Our approach is ripe for replication, not only by hospitals and regulators that hold complaints data, but within liability insurers with malpractice claims data, large hospital systems with risk management data, and medical boards and other professional bodies with data on disciplinary matters. Several distinctive aspects of our approach, descriptions of which follow, pave the way for better prediction of medico-legal risk in these settings than has been achieved to date.

Previous efforts to predict malpractice risk in liability insurance pools have included doctors with and without claims in their analyses.11 14 15 17 19 21 This approach suits a core goal in many of these studies: to explore the feasibility of ‘experience rating’ doctors’ liability insurance premiums.24 40 By contrast, our study sought to predict risk for purposes of targeting quality-improvement interventions. In this context, it is appropriate to focus on doctors who have been the subject of at least one complaint because this is the group with whom regulators have a natural point of contact and opportunities to intervene. An ancillary benefit of this ‘conditional’ approach to modelling medico-legal risk is that it enhances the ability to identify strong predictors of recurrent risk.
A key technical challenge encountered in previous studies has been how to deal with the recurrent nature of medico-legal events. The approach used by Rolph and others who have emulated his method,15 17 24 ‘fixes’ the effect of prior events in a single variable at the doctor level. The ‘weighted sum algorithm’ behind the PARS risk score, developed by Hickson and colleagues, comes from analyses regressing a sample of ‘risk management events’ on information obtained from unsolicited patient complaints.19 25 A limitation of both approaches is their static consideration of doctors’ event histories. In its application, however, the PARS algorithm adopts dynamic features (doctors risk scores can be recalculated as new complaints appear over time).

An advantage of recurrent-event survival analysis is that it permits dynamic consideration of the effect of time-varying factors in the predictive model itself. In other words, it is not necessary to rely on a snapshot taken of a doctor’s situation at a particular point in time: as risk profiles evolve—and the coefficients on the previous complaints variable in our study illustrate how dramatically this may occur—survival analysis incorporates these changes into the estimation of future risk. A related advantage of survival analysis is that it permits estimation of doctors’ risk levels at different points in time—a year after an index event, 2 years later and so on. Our analysis showed that for some predictors, particularly the number of previous complaints, doctors’ risks of additional complaints were non-linear: the risk tends to rise quickly over the several months after a complaint and then level off by the time the doctor reaches a year without further incidents. For clinical leaders, regulators and liability insurers trying to determine when in a doctor’s trajectory of events to intervene to prevent recurrence, and how aggressively, this kind of temporal information may be very informative.

Our study has several limitations. First, the generalisability of our findings and method—to other types of medico-legal events, to other types of health practitioners, and outside Australia—is unknown, and should be tested. In other medico-legal settings, it may not be possible for practitioners to accrue the large numbers of events that some doctors in our sample did. Lower ceilings on the number of prior events may reduce the predictive value of this variable. Nonetheless, our analyses showed high risks of recurrence within 2 years (>60%) among doctors with as few as four complaints.

Second, the predictors we examined were doctor-focused. Other variables—including, patient characteristics,41–44 case-type and outcomes,39 45 doctors’ ethnicity and country of training,46 47 the practice setting, and aspects of the patient-doctor relationship48—may also predict complaint risk. However, because these variables are usually more difficult to measure at the population level, their suitability for large-scale predictive modelling is questionable. Moreover, given the high predictive values obtained with the simple doctor-level variables used in our analysis, the scope to boost predictive values with the addition of other variables is limited. Finally, we used head counts of practitioners, not more sophisticated measures of doctors’ exposure to complaint risk, such as volume of patients treated or procedures conducted.

During the rise of the quality and safety movement over the last 15 years, medico-legal institutions have been largely on the sidelines. They remain essentially reactive enterprises, with workloads that focus on dealing with the fallout from care that has gone wrong. Patient safety experts regard the medico-legal system’s fixation with post hoc assessments of individual behaviour, rather than prevention and systems, as anachronistic.39 But as Rolph recognised 30 years ago,11 methods for accurately and reliably forecasting the medico-legal risk of clinicians have transformative potential because they could focus and drive prevention. Identifying and intervening early with doctors at high risk of attracting recurrent medico-legal events has considerable potential to reduce adverse events and patient dissatisfaction system-wide; it may also help those doctors avoid the vicissitudes of medico-legal processes.

Acknowledgements We thank the health services commissioners and their staff in the Australian Capital Territory, New South Wales, the Northern Territory, Queensland, South Australia, Tasmania, Victoria and Western Australia; their support, assistance and advice made this study possible. Troyen Brennan, Atul Gawande, Michelle Mello and John Rolph provided helpful comments on earlier drafts of this manuscript.

Contributors MB, MS and DS developed the study idea, collected the data and conducted the analyses; MB and DS wrote the first draft of the manuscript; MW advised on design of the study, contributed expertise in interpretation and analysis of study data, and helped revise the draft manuscript; LG contributed to design and conduct of the statistical analysis and helped revise the draft manuscript; all authors reviewed and agreed on the submitted version of the manuscript. MB, MS and DS are guarantors for the study.

Funding This study was funded by an ARC Laureate Fellowship (FL110100102 to Dr Studdert) from the Australian Research Council. The research was conducted independently from the funder. The funder had no role in the study design, collection, analysis, and interpretation of data; writing of the report, or the decision to submit the article for publication.

Competing interests All authors have completed the Unified Competing Interest form and declare that: (1) MB, MS and DS have support from the Australian Research Council (Laureate Fellowship to DS); (2) none of the authors have had a financial relationship with any organisation that may have an interest in the submitted work in the previous 3 years; (3) none of the authors’ spouses, partners or children have any financial relationships that may be relevant to the submitted work and (4) none of the authors have any non-financial interests that may be relevant to the submitted work.

Ethics approval The study was approved by the Human Research Ethics Committee at the University of Melbourne.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non-Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially.
REFERENCES


The global burden of unsafe medical care: analytic modelling of observational studies

Ashish K Jha,1 Itziar Larizgoitia,2 Carmen Audera-Lopez,2 Nittita Prasopa-Plaizier,2 Hugh Waters,3 David W Bates4

ABSTRACT

Objective To contextualise the degree of harm that comes from unsafe medical care compared with individual health conditions using the global burden of disease (GBD), a metric to determine how much suffering is caused by individual diseases.

Design Analytic modelling of observational studies investigating unsafe medical care in countries’ inpatient care settings, stratified by national income, to identify incidence of seven adverse events for GBD modelling. Observational studies were generated through a comprehensive search of over 16 000 articles written in English after 1976, of which over 4000 were appropriate for full-text review.

Results The incidence, clinical outcomes, demographics and costs for each of the seven adverse events were collected from each publication when available. We used disability-adjusted life years (DALYs) lost as a standardised metric to measure morbidity and mortality due to specific adverse events. We estimate that there are 421 million hospitalisations in the world annually, and approximately 42.7 million adverse events. These adverse events result in 23 million DALYs lost per year. Approximately two-thirds of all adverse events, and the DALYs lost from them, occurred in low-income and middle-income countries.

Conclusions This study provides early evidence that adverse events due to medical care represent a major source of morbidity and mortality globally. Though suffering related to the lack of access to care in many countries remains, these findings suggest the importance of critically evaluating the quality and safety of the care provided once a person accesses health services. While further refinements of the estimates are needed, these data should be a call to global health policymakers to make patient safety an international priority.

INTRODUCTION

Most efforts to improve healthcare globally have focused on improving care for diseases that cause substantial morbidity and mortality, hoping to increase access to lifesaving therapies for the world’s population. These efforts have begun to pay off, with increasing access to antimalarial drugs and antiretroviral therapies for patients with HIV.1 However, institutionalising these gains will require focus on healthcare systems and their ability to deliver safe, effective care. This will be especially important in low-income and middle-income countries that are growing economically, and are looking to improve their health systems to care for a growing population.2

One lens through which to examine the functioning of healthcare systems is that of patient safety. Unsafe medical care—where patients are harmed by the medical care designed to help them—can have wide-ranging consequences.3 Adverse events, or injuries as a result of medical care, lead to direct harm and waste, and have spillover effects on patient confidence in the healthcare system.4 Many policymakers have primarily considered patient safety as an issue for high-income countries, where most of the population has access to basic healthcare. Indeed, estimates suggest that tens of thousands of citizens are injured, or die, due to medical errors in these countries.5 While the lack of access to basic healthcare services in many countries remains a clear policy challenge in the context of health systems strengthening, the extent to which people face suffering due to unsafe care once accessing medical services is less obvious. In other words, the extent to which unsafe medical care—or adverse events resulting from medical care—is a problem for developing and transitional countries, once a person accesses these health services, is not well known.
WHO undertook the challenge of estimating the global burden of unsafe care as an essential step to guide global actions in strengthening health systems. The global burden of disease (GBD) is a standard metric used by policymakers throughout the globe to determine how much suffering is caused by individual diseases. Its application has been more recently expanded to examine events like road accidents and other public health dangers. The GBD uses disability-adjusted life-years (DALYs) lost to quantify the morbidity and mortality associated with individual conditions and injuries. Understanding the GBD of unsafe medical care would be helpful to quantify the degree to which the world’s population encounters harm from unsafe healthcare interventions, allowing policymakers to better compare the DALYs lost from unsafe medical care to other causes of human suffering. Such data would allow policymakers to better prioritise the interventions likely to improve care and health for the world’s citizens.

Therefore, in this study, we sought to answer three questions: first, what is the global burden of unsafe medical care? Second, to what extent does the issue of unsafe medical care affect low-income and middle-income countries (LMICs) compared with high-income countries (HICs)? And third, are there certain types of adverse events resulting from unsafe medical care that are particularly harmful that policymakers can target in order to eliminate unnecessary suffering?

**METHODS**

**Definition of terms**

For the purposes of this analysis, we consider adverse events as unsafe experiences in an inpatient hospital setting and are thereby contingent on people having access to these medical services. We then explore the ‘clinical outcomes’ (eg, the proportion of patients who die, the proportion who have an injury and the duration of an injury) of these adverse events in order to quantify the burden of these adverse events or unsafe medical practices on human suffering.

**Identifying types of adverse events**

In July 2007, WHO’s Patient Safety programme convened a panel of international experts to discuss priorities for research on patient safety. The committee identified 20 topics that were of importance to patient safety, including structural factors, process of care and outcomes of unsafe care. Of these twenty, twelve adverse events were candidates for estimating the GBD of unsafe medical care. After consultation with the WHO committee, and an exhaustive literature review, we excluded five of the 12 outcomes due to severe data limitations (eg, substandard or counterfeit drugs, unsafe blood products, unsafe injections, medical devices and surgical errors, although we captured some of the injuries from surgical care in venous thromboembolism or nosocomial pneumonia). Due to the inadequacy of these data on adverse events in the ambulatory care setting, we elected to focus only on inpatient adverse events. As such, the final set of seven outcomes or types of adverse events used for the analyses were: (1) adverse drug events (ADEs), (2) catheter-related urinary tract infections (CR-UTIs), (3) catheter-related blood stream infections (BSIs), (4) nosocomial pneumonia, (5) venous thromboembolisms (VTEs), (6) falls and (7) decubitus (pressure) ulcers. Hospitalisations resulting from these adverse events occurring to outpatients were excluded. Additionally, we excluded hospitalisations due to childbirth, as we had little information about adverse events among these hospitalisations.

**Data sources**

We used two primary sources of data: a literature review and findings from recent WHO-commissioned epidemiologic studies. First, the literature review strategy, as detailed in the online supplementary methodological appendix, was designed to be a comprehensive examination of both peer-reviewed and non-peer-reviewed studies that focused on the seven aforementioned adverse events of interest, the clinical features of the patients who were injured (eg, age), and their outcomes. For this analysis, we relied upon two separate literature reviews; the first was conducted in late 2007 through early 2008, and it was then repeated in 2011. We supplemented the literature review with discussions with international experts in each topic area to ensure that key studies had been identified. The outcome of our literature review, including the specific studies that contributed incidence data for our models, is reported in the online supplementary methodological appendix.

The second data source for this study came from epidemiologic studies that were commissioned by WHO, which aimed to estimate the scale to which inpatient adverse events harmed patients. These studies have previously been described. In brief, they consisted of the identification of adverse events by a two-stage medical record review: initial screening by nurses or junior physicians using 18 explicit screening criteria followed by a review by a senior physician for determination of the adverse event, its preventability and the resulting disability. The studies were conducted in 26 hospitals across eight low-income and middle-income countries in the Eastern Mediterranean and North African regions, and 35 hospitals across five countries in Latin America. These studies also provided incidence data for our models.

**Global burden of disease model**

The GBD, run by WHO, uses disability-adjusted life years (DALYs) to measure morbidity and mortality due to a specific condition. The GBD DALYs model requires several key inputs: the number of people...
affected, the age at which they are affected, and the clinical consequence of the adverse events, including the type of disability encountered (ie, clinical outcomes). Due to the paucity of data, we used a single average age per event, as opposed to standard GBD calculations done by age group and sex. The details of the model, including the formulae used, are detailed in the online supplementary methodological appendix. For all our modelling approaches, we estimated each input separately for high-income versus low-income or middle-income (LMIC) countries (as defined by the World Bank).

Identifying inputs for the GBD model

Incidence of adverse events: We estimated the incidence of each of our seven adverse events in a hospitalised population based on reported data from the literature review and epidemiologic studies described earlier. Given that there was a range of estimates for both HICs as well as for LMICs, we generally took the median incidence for each category as our ‘best estimate’ but allowed the entire range of incidence estimates in the Monte Carlo models (see analysis below).

Number of adverse events: In order to calculate the number of patients harmed due to adverse events after accessing medical services, we needed to estimate the number of hospitalisations that occur globally. To our knowledge, there is no single source where such data are available. Consequently, we used data from WHO, the World Bank, the Organisation for Economic Co-operation and Development (OECD), and others, including the Centers for Disease Control and Prevention in the USA to create these estimates. We used the median as our ‘best estimate’ for the number of hospitalisations, but allowed the modelling to take into account the entire range of data identified. For each of our seven outcomes or adverse events of interest, we multiplied the number of hospitalisations by the incidence to estimate the number of adverse events that occurred.

Demographics and outcomes of adverse events: We used data from the literature review to estimate demographics (eg, age and gender) of patients injured from an adverse event as well as their clinical outcomes (eg, the proportion of patients who typically die, the proportion that would suffer a long-term and a short-term injury, the duration of that injury, and the proportion that would suffer only minor injuries but no sustained disability). The distribution of age at the time of acquiring the condition and the outcomes for these injuries are shown in the online supplementary methodological appendix.

Calculating DALYs: To calculate DALYs, this required that we apply disability weights for the injuries or harms that are attributable to the seven adverse events explored in this analysis. We used WHO’s GBD reports to identify disability weights for injuries when available; when not available, we identified the closest analogue or clinical condition for which there were disability weights available (see online supplementary methodological appendix). As is standard in these models, we assumed that the life expectancy was 81.3 years based on model life-table West Level 26, which has a life expectancy at birth of 82.5 for females and 80.1 for males.

Analysis

Our primary analytic approach was to build a Monte Carlo simulation model with 1000 simulations for each of the seven adverse events within LMICs, and then separately for HICs. In these models, the best estimate was assumed to be the midpoint of the range with a triangular distribution. Therefore, we had 14 sets of Monte Carlo models (one for each of the seven adverse events for HICs and for LMICs). These models yielded the best overall aggregate estimate of the global burden of harm resultant from these adverse events. Moreover, the models produced distributions for each of the input variables, as well as each of the output variables (see online supplementary methodological appendix). Analyses were performed using SAS V9.2.

RESULTS

We estimated that there were 117.8 million hospitalisations among the approximately 1.1 billion citizens in HICs in 2009, while there were 203.1 million hospitalisations among the 5.5 billion citizens of the LMICs. Hospitalisation rates for HICs were higher (mean 10.8 per 100 citizens per year) compared with those for LMICs (mean 3.7 per 100 citizens per year; see table 1).

Table 1 Hospitalisation rates in high, middle and low income countries

<table>
<thead>
<tr>
<th></th>
<th>High-income countries</th>
<th>Low-income and middle-income countries</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospitalisation rates*</td>
<td>10.8 (8.6 to 13.2)</td>
<td>3.7 (2.0 to 6.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>Mean (95% CI)</td>
<td>1 056 300 000</td>
<td>5 554 000 000</td>
<td>6 610 300 000</td>
</tr>
<tr>
<td>Total population</td>
<td>117.8 M (94.3 M—143.4 M)</td>
<td>203.1 M (121.9 M—312.2 M)</td>
<td>421 M (225.5 M—616.3 M)</td>
</tr>
</tbody>
</table>

*Rates are per 100 citizens per year; M, million
We found large variations in the reported incidence of adverse events both within HICs and LMICs (see table 2). Of the seven adverse events analysed in the inpatient hospital setting, the most common type in HICs was adverse drug events with an incidence rate of 5.0% (CI 2.7% to 7.2%) while the most common in LMICs was venous thromboembolism (incidence rate of 3.0%, CI 1.0% to 4.8%). We found comparable incidence between HICs and LMICs of three types of adverse events: catheter-related blood stream infections, venous thromboembolism, and decubitus ulcers. We found lower rates of adverse drug events in LMICs compared with HICs (2.9% vs 5.0%) and nosocomial pneumonia (0.4% vs 0.8%), while rates of two types of adverse events (catheter-related urinary tract infection (CR-UTI) and falls) were higher in LMICs compared with HICs (see table 2). Based on these incidence data, we estimated that of every 100 hospitalisations, there were approximately 14.2 of these adverse events in HICs and 12.7 in LMICs. The age at which these adverse events occurred was generally 8 to 19 years higher in HIC (see online supplementary methodological appendix).

We estimate that there are approximately 16.8 million injuries annually due to these adverse events among hospitalised patients in HICs. LMICs, which have five times the population of HICs, experienced approximately 50% more adverse events (25.9 million, see table 3). The number of adverse events varied substantially depending on the type of event examined, and the estimates for each type of adverse event corresponded with wide confidence intervals. For instance, we estimated that there were approximately 1.4 million (95% CI 0.8 million to 2.0 million) catheter-related urinary tract infections among HICs, while there was a substantially higher number and rate in LMICs (4.1 million, 95% CI 0.5 million to 9.2 million).

Based on these findings, we estimate that there were 22.6 million DALY’s lost due to these adverse events in 2009 (table 4). The number of DALY’s lost were more than twice as high in LMICs (15.5 million) as they were in HICs (7.2 million). The biggest source of lost DALY’s appeared to be venous thromboembolism (5.4 million DALY’s in LMICs, 95% CI 1.1 million to 11.7 million) and 2.3 million in HICs (95% CI 1.1 million to 3.9 million). Although the underlying numbers of several of the infections were much smaller, they caused a comparable number of DALY’s lost, often because the clinical outcomes were poor when these infections occurred (table 4).

For most of the adverse events explored in this study, the primary source of DALY’s lost was premature death: 78.6% of all adverse events in HICs and 80.7% in LMICs. The proportion of DALY’s lost due to short-term or long-term disability (as opposed to premature death) ranged from as little as 0.8% (catheter-related UTI) to a high of 32.9% (falls in the hospital) in all countries (table 4). While premature death constituted the primary source of DALY’s lost, disability (both short and long term) were generally more common than death itself (see data presented in online supplementary methodological appendix).

**DISCUSSION**

Injuries secondary to adverse events from unsafe care present significant challenges to health systems across the globe. We projected a collective 22.6 million injuries annually due to these adverse events...
DALYs lost due to adverse events experienced by the world’s hospitalised population. Compared with other conditions, the combination of these seven adverse events alone estimated in this study rank as the 20th leading cause of morbidity and mortality for the world’s population. It is unlikely that these are ‘new’ previously undiscovered DALYs, but rather that they are captured within the injuries and deaths attributed to other conditions such as cardiovascular disease. We suspect that these DALYs resulting from unsafe medical care may be one of the reasons why patients are disabled or die from these other conditions.

While lack of access to healthcare, especially hospital care, is clearly a major source of ill health and poor outcomes, especially in low-income countries, our work focuses on the safety of care once a person has accessed the medical resources available to them. We are unaware of any prior effort to examine the global burden of unsafe care across multiple types of adverse events. WHO estimates that the global burden of unsafe injection practices was over 9.2 million DALYs lost per year in the year 2000 alone. If we had included those estimates, the resultant GBD from unsafe care would have been over 33 million DALYs, placing it as the 14th leading cause of morbidity and mortality in the world, comparable to the burden from tuberculosis or malaria. Including adverse events that were not possible to include in this study due to data limitations, such as unsafe surgery, harm due to counterfeit drugs, unsafe childbirth and unsafe blood use, as well as safety issues with ambulatory care, would further raise these estimates substantially. A recent systematic review found that healthcare-associated infections are ubiquitous and occur at much higher rates in low-income countries than in HICs. Although these investigators did not calculate the GBD of these infections, their data underscore and support our findings that adverse events once reaching a hospital setting are common and likely cause unnecessary suffering across the globe.

These findings should prompt policymakers across the globe to invest further into systematic data collection, as well as programmes to measure and improve the safety of the healthcare systems. While the lack of access to care presents substantial harm, it is important to maintain high standards for safety and quality within the healthcare systems that we subject patients to across the globe. Unsafe medical care may even

Table 4  Disability-adjusted life-years (DALYs) lost and source of the DALYs, in 2009

<table>
<thead>
<tr>
<th></th>
<th>DALYs*</th>
<th>Short-term disability (%)</th>
<th>Long-term disability (%)</th>
<th>Premature death (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High-income countries</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catheter-related UTI</td>
<td>402 (214–620)</td>
<td>2.2</td>
<td>0.1</td>
<td>97.7</td>
</tr>
<tr>
<td>Adverse drug events</td>
<td>779 (350–1332)</td>
<td>5.7</td>
<td>0.3</td>
<td>94.0</td>
</tr>
<tr>
<td>Falls in the hospital</td>
<td>27 (6–51)</td>
<td>27.5</td>
<td>6.0</td>
<td>66.5</td>
</tr>
<tr>
<td>CR blood stream infections</td>
<td>1126 (328–2088)</td>
<td>3.0</td>
<td>0.2</td>
<td>96.8</td>
</tr>
<tr>
<td>Nosocomial pneumonia</td>
<td>2545 (1673–3703)</td>
<td>1.4</td>
<td>0.0</td>
<td>98.5</td>
</tr>
<tr>
<td>Decubitus ulcers</td>
<td>134 (58–268)</td>
<td>5.9</td>
<td>4.4</td>
<td>89.8</td>
</tr>
<tr>
<td>Venous thromboembolisms</td>
<td>2282 (1054–3855)</td>
<td>28.2</td>
<td>7.4</td>
<td>64.4</td>
</tr>
<tr>
<td>Total</td>
<td>7208 (5371–9271)</td>
<td>15.7</td>
<td>5.7</td>
<td>78.6</td>
</tr>
<tr>
<td><strong>Low-income and middle-income countries</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catheter-related UTI</td>
<td>3420 (450–8012)</td>
<td>0.7</td>
<td>0.0</td>
<td>99.4</td>
</tr>
<tr>
<td>Adverse drug events</td>
<td>1435 (126–3453)</td>
<td>2.3</td>
<td>0.1</td>
<td>97.6</td>
</tr>
<tr>
<td>Falls in the hospital</td>
<td>76 (6–169)</td>
<td>26.9</td>
<td>5.9</td>
<td>67.2</td>
</tr>
<tr>
<td>CR blood stream infections</td>
<td>2150 (958–4065)</td>
<td>3.0</td>
<td>0.2</td>
<td>96.8</td>
</tr>
<tr>
<td>Nosocomial pneumonia</td>
<td>2674 (996–5403)</td>
<td>1.4</td>
<td>0.0</td>
<td>98.5</td>
</tr>
<tr>
<td>Decubitus ulcers</td>
<td>291 (104–652)</td>
<td>30.0</td>
<td>5.6</td>
<td>64.4</td>
</tr>
<tr>
<td>Venous thromboembolisms</td>
<td>5399 (1126–11 730)</td>
<td>26.8</td>
<td>7.0</td>
<td>66.1</td>
</tr>
<tr>
<td>Total</td>
<td>15 454 (9009–23 607)</td>
<td>14.1</td>
<td>5.2</td>
<td>80.7</td>
</tr>
<tr>
<td><strong>Total (combined)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catheter-related UTI</td>
<td>3822 (844–8412)</td>
<td>0.8</td>
<td>0.0</td>
<td>99.4</td>
</tr>
<tr>
<td>Adverse drug events</td>
<td>2214 (807–4274)</td>
<td>3.4</td>
<td>0.2</td>
<td>96.7</td>
</tr>
<tr>
<td>Falls in the hospital</td>
<td>103 (29–199)</td>
<td>27.0</td>
<td>5.9</td>
<td>68.1</td>
</tr>
<tr>
<td>CR blood stream infections</td>
<td>3276 (1752–5379)</td>
<td>3.0</td>
<td>0.2</td>
<td>98.2</td>
</tr>
<tr>
<td>Nosocomial pneumonia</td>
<td>5219 (3226–8120)</td>
<td>1.4</td>
<td>0.0</td>
<td>99.1</td>
</tr>
<tr>
<td>Decubitus ulcers</td>
<td>426 (209–804)</td>
<td>13.8</td>
<td>4.8</td>
<td>82.7</td>
</tr>
<tr>
<td>Venous thromboembolisms</td>
<td>7681 (3115–14 034)</td>
<td>27.3</td>
<td>7.1</td>
<td>70.7</td>
</tr>
<tr>
<td>Total</td>
<td>22 644 (15 899–30 979)</td>
<td>14.4</td>
<td>5.3</td>
<td>80.2</td>
</tr>
</tbody>
</table>

*All DALY numbers are in thousands; DALYs, disability-adjusted life years.
lead patients, especially in low-income countries, to opt out of using the formal healthcare system, thereby making unsafe care a potentially significant barrier to access for many of the world’s poor. Such a phenomenon would suggest that the distinction between access and quality (or in this case, safety) may not be so clear. Finally, other costs of unsafe care, such as increased consumption of resources due to prolonged stay and extra care—and loss of wages and productivity—are important, and would benefit from further investigation.

Limitations
This work has important limitations. The primary one is the lack of availability of high-quality data. Although there are nearly five times as many people living in LMICs, the number of adverse events we calculated was only 50% higher, primarily due to the lower hospitalisation rates and the poor quality of data sources in LMICs (including medical records). These poor quality data sources lead to undercounting of adverse events that are often not recorded. Nevertheless, the number of DALYs per event was substantially higher, likely due to a combination of a younger age at which these events occur, and the worse outcomes that often result. While this limitation may raise concerns about the validity of our findings, we used data from a large number of sources, and reassuringly found a consistent rate of adverse events.

The paucity of data also limited our ability to run calculations per age group and sex, leading us to calculate average estimates. Additionally, the data limited our analysis to reporting the aggregate harm resulting from total adverse events as opposed to preventable adverse events. While estimating preventable harm would be valuable, there is even greater uncertainty about how much harm is preventable at any given time, and as technology and clinical care changes, the proportion of adverse events that are preventable, likely will, as well. While our estimates are imprecise, we believe that as more data on adverse events become available, WHO will be able to refine these estimates and track them over time.

Second, while there are several high-quality studies, few use standardised definitions or approaches to identifying adverse events. Therefore, the data we relied on all used slightly different approaches and likely lead to some degree of imprecision.

Third, we elected to use the same life expectancy value for all individuals, and although this has been controversial, it is the standard approach used by WHO. Had we chosen different life expectancies for different countries, we likely would have estimated a lower number of DALYs lost, especially for low-income and middle-income countries.

Fourth, we excluded publications not written in English, which may have affected the precision of our estimates. Nearly all major epidemiologic studies of adverse events from HICs over the past decade have found that they occur in 8–15% of hospitalised patients. Data from LMICs suggest that the rates are even higher. Further, we excluded studies that were clearly of low quality, including those that used non-standard methods (such as convenience samples), or had unclear denominators, or extremely small sample sizes. Whether and to what degree these exclusions biased our findings is unclear.

Fifth, as described above, key inpatient adverse events that the WHO Committee on Patient Safety viewed as important were excluded due to data limitations (eg, unsafe childbirth), leading us to underestimate the true burden of harm from unsafe medical care. Also, we excluded adverse events in the ambulatory setting, which recent data suggest are a major source of harm.

Finally, we lacked disability weights specifically designed for the injuries we examined attributable to our seven adverse events. However, most of the injuries did have clinically analogous events for which there were disability weights. In other words, we identified ‘proxy’ conditions for each adverse event, usually choosing diseases that affected the same organ system with a generally similar level of severity. We attempted to use the most conservative disability weight in the model, though we recognise that our efforts at matching are imperfect. For example, for catheter-related infections, we used the proxy condition of endocarditis which has a disability weight range from 0.17 to 0.32. This is more fully described in the online supplementary methodological appendix. WHO has a well designed and rigorous process for creating disability weights, and the potential impact of these results will likely spur them to create specific disability weights for these injuries.

Although our estimates are quite conservative, they still represent a relatively wide range of possible outcomes because of inadequate data. Poor quality data on health systems, especially on adverse events, hampered our ability to effectively calculate the number of DALYs lost due to unsafe care, especially within LMICs. Even in HICs, these data are not routinely measured and made publicly available,16 hampering not only our ability to calculate their consequences, but also limiting the ability of clinical leaders and policymakers to track the potential impact of policies designed to increase the safety of healthcare. As LMICs prosper economically, it is hopeful that citizens will have greater access to medical services, and more encounters with the healthcare system. Without concomitant improvements in the safety of health systems, the number of injuries will likely grow.

CONCLUSION
Using a conservative approach, we estimated that there are at least 43 million injuries each year due to medical care, and that nearly 23 million DALYs are
lost as a consequence. A large majority of these injuries and harm occur in developing and transitional countries—and these numbers will likely grow. Given the magnitude of these effects, our findings suggest that to improve the health of the world’s citizens, we will need to improve access to care and also to invest substantial focus on improving the safety of the healthcare systems that people access worldwide. When patients are sick, they should not be further harmed by unsafe care. This should be a major policy emphasis for all nations.

**Acknowledgements** We are grateful to Dr Colin Mathers of WHO for his thoughtful input into our work. We are also grateful to Kathyana Aparicio of the Patient Safety Programme of WHO for her valuable insights, and Dr Victor Rosenthal of the International Nosocomial Infection Program for sharing data from his programme.

**Contributors** AKJ and DWB contributed to the conception and design of the project. HW and AKJ analysed the data and all authors contributed to the interpretation of the findings. AKJ drafted the manuscript and all authors contributed substantially to critical revisions of the manuscript. All authors approved the final version of the manuscript.

**Funding** WHO Patient Safety Programme funded the project. All final decisions about the topics studied, the analytic approach, and the manuscript was made by the authors. This paper represents the views of the authors and not necessarily those of WHO.

**Competing interests** None.

**Provenance and peer review** Not commissioned; internally peer reviewed.

**REFERENCES**


Systematic review of the application of the plan–do–study–act method to improve quality in healthcare

Michael J Taylor,1,2 Chris McNicholas,2 Chris Nicolay,1 Ara Darzi,1 Derek Bell,2 Julie E Reed2

ABSTRACT
Background Plan–do–study–act (PDSA) cycles provide a structure for iterative testing of changes to improve quality of systems. The method is widely accepted in healthcare improvement; however there is little overarching evaluation of how the method is applied. This paper proposes a theoretical framework for assessing the quality of application of PDSA cycles and explores the consistency with which the method has been applied in peer-reviewed literature against this framework.

Methods NHS Evidence and Cochrane databases were searched by three independent reviewers. Empirical studies were included that reported application of the PDSA method in healthcare. Application of PDSA cycles was assessed against key features of the method, including documentation characteristics, use of iterative cycles, prediction-based testing of change, initial small-scale testing and use of data over time.

Results 73 of 409 individual articles identified met the inclusion criteria. Of the 73 articles, 47 documented PDSA cycles in sufficient detail for full analysis against the whole framework. Many of these studies reported application of the PDSA method that failed to accord with primary features of the method. Less than 20% (14/73) fully documented the application of a sequence of iterative cycles. Furthermore, a lack of adherence to the notion of small-scale change is apparent and only 15% (7/47) reported the use of quantitative data at monthly or more frequent data intervals to inform progression of cycles.

Discussion To progress the development of the science of improvement, a greater understanding of the use of improvement methods, including PDSA, is essential to draw reliable conclusions about their effectiveness. This would be supported by the development of systematic and rigorous standards for the application and reporting of PDSAs.

INTRODUCTION
Delivering improvements in the quality and safety of healthcare remains an international challenge. In recent years, quality improvement (QI) methods such as plan–do–study–act (PDSA) cycles have been used in an attempt to drive such improvements. The method is widely used in healthcare improvement; however there is little overarching evaluation of how the method is applied. This paper proposes a theoretical framework for assessing the quality of application of PDSA cycles and explores the consistency with which PDSA cycle application against this framework as documented in peer-reviewed literature.

Use of PDSA cycles in healthcare
Despite increased investment in research into the improvement of healthcare, evidence of effective QI interventions remains mixed, with many systematic reviews concluding that such interventions are only effective in specific settings.1–4 To make sense of these findings, it is necessary to understand that delivering improvements in healthcare requires the alteration of processes within complex social systems that change over time in predictable and unpredictable ways.5 Research findings highlight the influential effect that local context can have on the success of an intervention5–7 and, as such, ‘single-bullet’ interventions are not anticipated to deliver consistent improvements. Instead, effective interventions need to be complex and multi-faceted8–11 and developed iteratively to adapt to the local context and respond to unforeseen obstacles and unintended effects.12 13 Finding effective QI methods to support iterative development to test and evaluate
interventions to care is essential for delivery of high-quality and high-value care in a financially constrained environment.

PDSA cycles provide one such method for structuring iterative development of change, either as a standalone method or as part of wider QI approaches, such as the Model for Improvement (MFI), Total Quality Management, Continuous QI, Lean, Six Sigma or ‘Quality Improvement Collaboratives’.28 Despite increased use of QI methods, the evidence base for their effectiveness is poor and under-theorised.15–17 PDSA cycles are often a central component of QI initiatives, however few formal objective evaluations of their effectiveness or application have been carried out.18 Some PDSA approaches have been demonstrated to result in significant improvements in care and patient outcomes,19 while others have demonstrated no improvement at all.20–22

Although at the surface level these results appear disheartening for those involved in QI, there is a need to explore the extent to which the PDSA method has been successfully deployed to draw conclusions from these studies. Rather than see the PDSA method as a ‘black box’ of QI,25 it is important to understand that the use of PDSA cycles is, itself, a complex intervention made up of a series of interdependent steps and key principles that inform its application24 25 and that this application is also affected by local context.26 To interpret the results regarding the outcome(s) from the application of PDSA cycles (eg, whether processes or outcomes of care improved) and gauge the effectiveness of the method, it is necessary to understand how the method has been applied.

No formal criteria for evaluating the application or reporting of PDSA cycles currently exist. It is only in recent years, through SQUIRE guidelines, that frameworks for publication have been developed that explicitly consider description of PDSA application.27 28 We consider that such criteria are necessary to support and assess the effective application of PDSA cycles and to increase their legitimacy as a scientific method for improvement. We revisited the origins and theory of the method to develop a theoretical framework to evaluate the application of the method.

### The origins and theory of PDSA cycles

The PDSA method originates from industry and Walter Shewhart and Edward Deming’s articulation of iterative processes which eventually became known as the four stages of PDSA.25 PDCA (plan–do–check–act) terminology was developed following Deming’s early teaching in Japan.29 The terms PDSA and PDCA are often used interchangeably in reference to the method. This distinction is rarely referred to in the literature and for the purpose of this article we consider PDSA and PDCA but refer to the methodologies generally as ‘PDSA’ cycles unless otherwise stated.

Users of the PDSA method follow a prescribed four-stage cyclic learning approach to adapt changes aimed at improvement. In the ‘plan’ stage a change aimed at improvement is identified, the ‘do’ stage sees this change tested, the ‘study’ stage examines the success of the change and the ‘act’ stage identifies adaptations and next steps to inform a new cycle. The MFI30 and FOCUS13 (see figure 1) frameworks have been developed to precede the use of PDSA and PDCA cycles respectively (table 1).

In comparison to more traditional healthcare research methods (such as randomised controlled trials in which the intervention is determined in advance and variation is attempted to be eliminated or controlled for), the PDSA cycle presents a pragmatic scientific method for testing changes in complex systems.32 The four stages mirror the scientific experimental method33 of formulating a hypothesis, collecting data to test this hypothesis, analysing and interpreting the results and making inferences to iterate the hypothesis.

The pragmatic principles of PDSA cycles promote the use of a small-scale, iterative approach to test interventions, as this enables rapid assessment and provides flexibility to adapt the change according to feedback to ensure fit-for-purpose solutions are developed.10 12 13 Starting with small-scale tests provides users with freedom to act and learn; minimising risk to patients, the organisation and resources required and providing the opportunity to build evidence for change and engage stakeholders as confidence in the intervention increases.

In line with the scientific experimental method, the PDSA cycle promotes prediction of the outcome of a test of change and subsequent measurement over time (quantitative or qualitative) to assess the impact of an intervention on the process or outcomes of interest. Thus, learning is primarily achieved through intervention experiments designed to test a change. In recognition of working in complex settings with inherent variability, measurement of data over time helps understand natural variation in a system, increase awareness of other factors influencing processes or outcomes, and understand the impact of an intervention.

As with all scientific methods, documentation of each stage of the PDSA cycle is important to support scientific quality, local learning and reflection and to ensure knowledge is captured to support organisational memory and transferability of learning to other settings.

This review examines the application of PDSA cycles as determined by these principle features of the PDSA method described above. We recognise that a number of health and research related contextual factors may affect application of the method but these factors are beyond the scope of this review. The review intends to improve the understanding of

---

**Information sources**

The following databases were searched for articles: Embase (1980 to present), Health Business Elite (1985 to present), British Nursing Index (BNI; 1985 to present), Allied and Complementary Medicine Database (AMED; 1985 to present), and CINAHL (1985 to present). Systematic review

**Methods**

A theoretical framework was constructed by comparing and discussing articles around the use of PDSA cycles to facilitate iterative development of the PDSA method in line with the theoretical framework. Taking into account the development of the method and terminology, the search excluded from the study selection.

**Data collection process and study selection**

Data were collected regarding application of the PDSA method in line with the literature informed criteria and there were no restrictions on study design.

**Search**

The following databases were searched for articles: Embase (1980 to present), Health Business Elite (1985 to present), British Nursing Index (BNI; 1985 to present), Allied and Complementary Medicine Database (AMED; 1985 to present), British Nursing Index (BNI; 1985 to present), Allied and Complementary Medicine Database (AMED; 1985 to present), and CINAHL (1985 to present).

**Selection criteria**

To be included in this review, articles needed to be published in peer-reviewed journals and use PDSA cycles to facilitate iterative development of the PDSA method.
whether the PDSA method is being used and reported in line with the literature informed criteria and therefore inform the interpretation of studies that have used PDSA cycles to facilitate iterative development of an intervention.

METHODS
A systematic narrative review was conducted in accordance to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.

Search
The search was designed to identify peer-reviewed publications describing empirical studies that applied the PDSA method. Taking into account the development of the method and terminology, the search terms used were ‘PDSA’, ‘PDCA’, ‘Deming Cycle’, ‘Deming Circle’, ‘Deming Wheel’ and ‘Shewhart Cycle’. No year of publication restrictions were imposed.

Information sources
The following databases were searched for articles: Allied and Complementary Medicine Database (AMED; 1985 to present), British Nursing Index (BNI; 1985 to present), Cumulative Index to Nursing and Allied Health Literature (CINAHL; 1981 to present), Embase (1980 to present), Health Business Elite (EMBESCO Publishing, Ipswich, Massachusetts, USA), the Health Management Information Consortium (HMIC), MEDLINE from PubMed (1950 to present) and PsychINFO (1806 to present) using the NHS Evidence online library (REF), and the Cochrane Database of Systematic Reviews. The last search date was 25 September 2012.

Data collection process and study selection
Data were collected and tabulated independently by MJT, CM and CN in a manner guided by the Cochrane Handbook. Eligibility was decided independently, in a standardised manner and disagreements were resolved by consensus. If an abstract was not available from the database, the full-text reference was accessed.

Inclusion criteria for articles were as follows: published in peer-reviewed journal; describes PDSA method being applied to improve quality in a healthcare setting; published in English. Editorial letters, conference abstracts, opinion and audit articles were excluded from the study selection.

Data items
A theoretical framework was constructed by compartmentalising the key features of the PDSA method into observable variables for evaluation (table 2). This framework was developed in accordance with recommendations for PDSA use cited in the literature, describing the origins and theory of the method. Face validity of the framework was achieved through discussion among authors, with QI facilitators and at local research meetings.

Data were collected regarding application of the PDSA method in line with the theoretical framework. Other data collected included first author, year of publication, country, area of healthcare, use of PDSA or PDCA terminology, and use of MFI or FOCUS as
supporting frameworks. Ratios were used to analyse the results regarding the majority of variables, and mean scores regarding data associated with length of study, length of PDSA cycle and sample size were also used for analysis. Data were analysed independently by MJT and CM. Discrepancies (which occurred in less than 3% of data items) were resolved by consensus.

**Risk of bias in individual studies**

The present review aimed to assess the reported application of the PDSA method and the results of individual studies were not analysed in this review.

**RESULTS**

**Study selection**

A search of the databases yielded 942 articles. After removal of duplicates, 409 remained; 216 and 120 articles demonstrated sequential cycle use, 8 solely applied the PDSA method as part of an empirical study and/or understand variation? 33 articles described cycles that used PDSA stages in the following cycle may seek to modify, expand, adopt or follow. Depending on the knowledge gained from a PDSA cycle, the following cycle may seek to modify, expand, adopt or abandon a change that was tested.

**Table 2  Theoretical framework based on key features of the plan–do–study–act (PDSA) cycle method**

<table>
<thead>
<tr>
<th>Feature of PDSA</th>
<th>Description of feature</th>
<th>How this was measured</th>
</tr>
</thead>
</table>
| **Iterative cycles** | To achieve an iterative approach, multiple PDSA cycles must occur. Lessons learned from one cycle link and inform cycles that follow. Depending on the knowledge gained from a PDSA cycle, the following cycle may seek to modify, expand, adopt or abandon a change that was tested | ▶ Were multiple cycles used?  
▶ Were multiple cycles linked to one another (ie, does the ‘act’ stage of one cycle inform the ‘plan’ stage of the cycle that follows)?  
▶ Were statistics used to test the effect of changes while providing feedback? |
| **Prediction-based test of change** | A prediction of the outcome of a change is developed in the ‘plan’ stage of a cycle. This change is then tested and examined by comparison of results with the prediction | ▶ Was a change tested?  
▶ Was an explicit prediction articulated? |
| **Small-scale testing** | As certainty of success of a test of change is not guaranteed, PDSAs start small in scale and build in scale as confidence grows. This allows the change to be adapted according to feedback, minimises risk and facilitates rapid change and learning | ▶ Sample size per cycle?  
▶ Temporal duration of cycles?  
▶ Number of changes tested per cycle?  
▶ Did sequential cycles increase scale of testing? |
| **Use of data over time** | Data over time increases understanding regarding the variation inherent in a complex healthcare system. Use of data over time is necessary to understand the impact of a change on the process or outcome of interest | ▶ Was data collected over time?  
▶ Were statistics used to test the effect of changes and/or understand variation?  
▶ Did sequential cycles increase scale of testing? |
| **Documentation** | Documentation is crucial to support local learning and transferability of learning to other settings | ▶ How thoroughly was the application of the PDSA method detailed in the reports?  
▶ Was each stage of the PDSA cycles documented? |
were further discarded following review of abstracts and full texts, respectively. Excluded articles did not apply the PDSA method as part of an empirical study or coincidently used the acronyms PDSA or PDCA for different terms, or were abstracts for conferences or poster presentations. A total of 73 articles met the inclusion criteria and were included in the review (see figure 2).

**General study characteristics**

Country of study
The retrieved articles describe studies conducted in the USA (n=46), the UK (n=13), Canada (n=3) Australia (n=3), the Netherlands (n=2) and one each from six other countries (see online supplementary appendix A for complete synthesis of results).

Healthcare discipline to which method was applied
This varied across acute and community care and clinical and organisational settings. The most common settings were those of pain management and surgery (six articles each).

Method terminology
Of the 73 articles identified, 42 articles used ‘PDSA’ as terminology and 31 referred to the method as ‘PDCA’. Eight of these reported using the MFI. Thirty-one articles used ‘PDCA’ terminology, with 20 using the preceding FOCUS framework. One article described use of FOCUS and MFI. Over time there was an increase in the prevalence of PDSA use with PDCA use diminishing (see online supplementary figure S1). The earliest reported use of PDCA and PDSA in healthcare was 1993 and 2000, respectively.

**Documentation**
The following four categories were used to describe the extent to which cycles were documented in articles (n=73): no detail of cycles (n=16); themes of cycles (but no additional details) (n=8); details of individual cycles, but not of stages within cycles (n=8); details of cycles including separated information on stages of cycles (n=41).

Analysis of articles against the developed framework was dependent on the extent to which the application of PDSA cycles was reported. Articles that provided no details of cycles or only themes of cycles were insufficient for full review and excluded for analysis against all features. Articles that provided further details of cycles completed (n=49) were included for analysis against the remaining four features of the framework. A full breakdown of findings can be viewed in online supplementary appendix B.

**Application of method**

**Iterative cycles (n=49)**
Fourteen articles described a sequence of iterative cycles (two or more cycles with lessons learned from one cycle linking and informing a subsequent cycle), 33 described isolated cycles that are not linked, and 2 articles described cycles that used PDSA stages in the incorrect order (in one article, one plan, one do, two checks and three acts were described, PDACACA\(^{35}\); a further study did not report use of a ‘check’ stage; PDA\(^ {36}\)) and are excluded from further review. Of the 33 articles that described non-iterative cycles, 29 reported a single cycle being used, and 4 described multiple, isolated (non-sequential) cycles. Although future actions are often suggested in articles that reported a single cycle, only three explicitly mentioned the possibility of further cycles taking place. A total of 13.6% (3/22) of PDCA studies described the application of iterative cycles compared with 44% (11/25) of PDSA studies describing the application of iterative cycles (see figure 3).

**Prediction-based testing of change (n=47)**
The aims of the cycles adhered to one of two themes: tests of a change; and collection or review of data without a change made. Of the 33 articles with single cycles, 30 aimed to test a change while 3 used the PDSA method to collect or review data. Of the 14 articles demonstrating sequential cycle use, 8 solely used their cycles to test change while 5 began with a cycle collecting or reviewing data followed by cycles testing change. One article described a mixture of cycles testing changes and cycles that involved collection/review of data. Four of the 47 studies contained an explicit prediction regarding the outcome of a
Systematic review

Figure 3 Iterative nature of cycles for all articles and split by plan–do–check–act and plan–do–study–act terminology.

change; all 4 aimed to test a change (see online supplementary table S1).

Small-scale testing (n=47)
Scale was assessed in three ways: sample size, duration and complexity. Sample size refers to quantity of observations used to measure the change; duration refers to the length of PDQA cycle application; and complexity refers to the quantity of changes administered per cycle.

Sample size
Patient data, staff data and case data were used as samples within PDQA cycles. Twenty-seven articles reported a sample size from at least one of their cycles. Twenty-one of these were isolated cycle studies with sample size ranging from 7 to 2079 (mean=323.33, SD=533.60). The remaining six studies reporting individual cycle sample sizes used iterative cycles; the sample size of the first cycles of these ranged from 1 to 34 (mean=16.73, SD=11.47). Two of these studies described the use of incremental sample sizes across cycles, three used non-incremental sample sizes across cycles, and one changed the type of sample. Of the eight iterative cycle articles that did not report individual cycle sample sizes, two did not differentiate sample sizes between cycles and instead gave an overall sample for the chain of cycles and six did not report sample size.

Duration
Reported study duration of isolated cycles ranged from 2 weeks to 5 years (mean=11.91 months, SD=12.81). Only five articles describing iterative cycles explicitly reported individual cycle duration. Individual cycle duration could be estimated from the total duration of the PDQA cycle chain and the number of cycles conducted, resulting in approximate cycle lengths ranging from three cycles in 1 day to one cycle in 16 months (mean=5.41 months, SD=4.80, see online supplementary figure S2). The total PDQA cycle duration for series of iterative cycles (first to last cycle of one chain) ranged from 1 day to 4 years (mean=20.38, SD=20.39 months).

Complexity
Twenty-two articles reported more than one change being tested within a single cycle. Of the articles describing iterative cycles, 42% administered more than one change per cycle compared with 48% of the articles describing non-iterative PDSA cycles.

Data over time (n=47)
All studies used a form of qualitative and quantitative data to assess cycles. Studies were categorised according to four types of reporting quantitative data: regular (n=15), three or more data points with consistent time intervals; non-regular (n=16), before and after or per PDQA cycle; single data point (n=8), a single data point after PDQA cycle(s); and no quantitative data reported (n=8). Of the 13 articles that used regular data, only 7 used monthly or more frequent data intervals (see online supplementary figure S3 for full frequency of regular quantitative data reporting).

No studies reported using statistical process control to analyse data collected from PDQA cycles. Eleven included analysis of data using inferential statistical tests (five of these studies collected isolated data, six involved continuous data collection).

Of the eight articles that did not report any quantitative data, two reported that quantitative analyses had taken place but did not present the findings and six described the use of qualitative feedback only (one non-regular, five single data point). Qualitative data were gathered through a range of mechanisms from informal staff or patient feedback to structured focus groups.

DISCUSSION
PDQA cycles offer a supporting mechanism for iterative development and scientific testing of improvements in complex healthcare systems. A review of the historic development and rationale behind PDQA cycles has informed the development of a theoretical framework to guide the evaluation of PDQA cycles against use of iterative cycles, initial small-scale testing, prediction-based testing of change, use of data over time and documentation.

Using these criteria to assess peer-reviewed publications of PDQA cycles demonstrates an inconsistent approach to the application and reporting of PDQA cycles and a lack of adherence to key principals of the method. Only 2/73 articles demonstrated compliance with criteria in all five principles. Assessment of compliance was problematic due to the marked variation in reporting of this method, which reflects a lack of standardised reporting requirements for the PDQA method.

From the articles that reported details of PDQA cycles it was possible to ascertain that variation is
inherent not just in reporting standards, but in the conduct of the method, implying that the key principles of the PDSA method are frequently not followed. Less than 20% (14/73) of reviewed articles reported the conduct of iterative cycles of change, and of these, only 15% (2/14) used initial small-scale tests with increasing scale as confidence in the intervention developed. These results suggest that the full benefits of the PDSA method would probably not have been realised in these studies. Without an iterative approach, learning from one cycle is not used to inform the next cycle, and therefore it is unlikely that interventions will be adapted and optimised for use in a particular setting. Furthermore, large-scale cycles risk significant resource investment in an intervention that has not been tested and optimised within that environment and risk producing ‘false’ negatives.

Only 14% (7/47) of articles reported use of regular data over time at monthly or more frequent intervals, indicating a lack of understanding around the use of the PDSA method to track change within a ‘live’ system, and limiting the ability to interpret the results from the study. Cycles that included an explicit prediction of outcomes were reported in only 9% (4/47) of articles, suggesting that PDSA cycles were not used as learning cycles to test and revise theory-based predictions.

Overall these results demonstrate poor compliance with key principles of the PDSA method, suggesting that it is not being used optimally. The increasing trend in using PDSA (as opposed to ‘PDCA’) cycles in recent years, however, does seem to have been accompanied by an increase in compliance with some key principles, such as use of iterative cycles. Deming was cautious over the use of the ‘PDCA’ terminology and warned it referred to an explicitly different process, referring to a quality control circle for dealing with faults within a system, rather than the PDSA process, which was intended for iterative learning and improvement of a product or a process. This subtle difference in terminologies may help to explain the better compliance with key methodological principles in studies that refer to the method as ‘PDSA’.

One of the articles identified in the search included comments by the authors that the PDSA method should be ‘more realistically represented’, as ineffective cycles can be ‘abandoned’ early on, making it needless to go through all four stages in each iteration. These comments may provide insight into an important potential misunderstanding of the PDSA methodology. Ineffective changes will result in learning, which is a fundamental principle behind a PDSA cycle. However minor this abandoned trial may have been, it can still be usefully described as a PDSA cycle. A minor intervention may be planned (P) and put into practice (D). A barrier may be encountered (S), resulting in a decision being made to retract the intervention, and to do something differently (A).

The theoretical framework presented in this paper highlights the complexity of PDSA cycles and the underpinning knowledge required for correct application. The considerable variation in application observed in the reported literature suggests that caution should be taken in interpreting results from evaluations in which PDSAs are used in a controlled setting and as a ‘black box’ of QI. This review did not compare the effectiveness of use to reported outcomes and therefore this study does not conclude whether better application of the PDSA method results in better outcomes, but instead draws on theoretical principles of PDSAs to rationalise why this would be expected. Prospective mechanistic studies exploring the effective application of the method as well as study outcomes would be of greater use in drawing conclusions regarding the effectiveness of the method. The framework presented in this paper could act as a good starting point for such studies.

The fact that only peer-reviewed publications were assessed in this study means that results may be affected by publication bias. This is anticipated both in terms of what is accepted for publication but also the level and type of detail that is requested and allowed in typical publications (eg, before and after studies are more common than presenting data over time and this may make these types of studies easier to publish). Though QI work may be easier to publish now through recent changes in publication guidelines, possible publication outlets continue to be relatively limited.

To support systematic reporting and encourage appropriate usage, we suggest that reporting guidelines be produced for users of the PDSA method to increase transparency as to the issues that were encountered and how they were resolved. While PDSA is analogous to a scientific method, it appears to be rarely used or reported with scientific rigour, which in turn, inhibits perceptions of PDSA as a scientific method. Such guidelines are essential to increase the scientific legitimacy of the PDSA method as well as to improve scientific rigour or application and reporting. Although the SQUIRE guidelines make reference to the potential use of PDSA cycles, further support to users and teachers, and publication of this improvement method seems necessary. Consistent reporting of PDSA structure would allow meta-evaluation and systematic reviews to further build the knowledge of how to use such methods effectively and the principles to apply to increase chances of success.

It is clear from these findings that there is much room for improvement in the application and use of the PDSA method. Previous studies have discussed the influence of different context factors on the use of QI methods, such as motivation, data support infrastructure and leadership. Understanding how high-quality usage can be promoted and supported needs to become the focus of further research if such
QI methods are going to be used effectively in mainstream healthcare.

CONCLUSIONS
There is varied application and reporting of PDSAs and lack of compliance with the principles that underpin its design as a pragmatic scientific method. The varied practice compromises its effectiveness as a method for improvement and cautions against studies that view QI or PDSA as a ‘black box’ intervention.

There is an urgent need for greater scientific rigour in the application and reporting of these methods to advance the understanding of the science of improvement and efficacy of the PDSA method. The PDSA method should be applied with greater consistency and with greater accordance to guidelines provided by founders and commentators.23 30 44 45

Acknowledgements The authors would like to thank Dr Thomas Woodcock for his valuable input into the theoretical framework and data analysis.

Contributors All listed authors qualify for authorship based on making one or more of the substantial contributions to the intellectual content: conceptual design (MJT, CM, CN, DB, AD and JR), acquisition of data (MJT, CM and CN) and/or analysis and interpretation of data (MJT, CM, CN and JR). Furthermore all authors participated in drafting the manuscript (MJT, CM, CN, DB, AD and JR) and critical revision of the manuscript for important intellectual content (MJT, CM, CN, DB, AD and JR).

Disclaimer This article presents independent research commissioned by the National Institute for Health Research (NIHR) under the Collaborations for Leadership in Applied Health Research and Care (CLAHRC) programme for North West London. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests The authors declare no conflict of interest.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

REFERENCES


Assessing adverse events among home care clients in three Canadian provinces using chart review

Régis Blais, Nancy A Sears, Diane Doran, G Ross Baker, Marilyn Macdonald, Lori Mitchell, Stéphane Thalès

ABSTRACT

Objectives The objectives of this study were to document the incidence rate and types of adverse events (AEs) among home care (HC) clients in Canada; identify factors contributing to these AEs; and determine to what extent evidence of completion of incident reports were documented in charts where AEs were found.

Methods This was a retrospective cohort study based on expert chart review of a random sample of 1200 charts of clients discharged in fiscal year 2009–2010 from publicly funded HC programmes in Manitoba, Quebec and Nova Scotia, Canada.

Results The results show that 4.2% (95% CI 3.0% to 5.4%) of HC patients discharged in a 12-month period experienced an AE. Adjusting to account for clients with lengths of stay in HC of less than 1 year, the AE incidence rate per client-year was 10.1% (95% CI 8.4% to 11.8%); 56% of AEs were judged preventable. The most frequent AEs were injuries from falls, wound infections, psychosocial, behavioural or mental health problems and adverse outcomes from medication errors. More comorbid conditions (OR 1.15; 95% CI 1.05 to 1.26) and a lower instrumental activities of daily living score (OR 1.54; 95% CI 1.16 to 2.04) were associated with a higher risk of experiencing an AE. Clients’ decisions or actions contributed to 48.4% of AEs, informal caregivers 20.4% of AEs, and healthcare personnel 46.2% of AEs. Only 17.3% of charts with an AE contained documentation that indicated an incident report was completed, while 4.8% of charts without an AE had such documentation.

Conclusions Client safety is an important issue in HC, as it is in institutionalised care. HC includes the planned delivery of self-care by clients and care provision by family, friends and other individuals often described as ‘informal’ caregivers. As clients and these caregivers can contribute to the occurrence of AEs, their involvement in the delivery of healthcare interventions at home must be considered when planning strategies to improve HC safety.

INTRODUCTION

Problems of client safety have been well investigated in acute care hospital settings, and, to a lesser extent, in long-term care, emergency room and primary care.

However, only a few studies have used a similar approach to safety issues among home care (HC) clients. The studies of HC clients are limited with regard to small sample size population studied (eg, long-stay clients) and types of harm (eg, when only physical harm is examined).

Client safety is usually assessed by measuring the incidence of adverse events (AEs). An AE has been defined as ‘an event that results in unintended harm to the client by an act of commission or omission rather than by the underlying disease or condition of the client’.

HC includes the provision of healthcare interventions to clients of all ages (birth to extreme old age), for the purposes of providing curative, supportive, palliative and rehabilitation care for acute and longer term illnesses and conditions. HC differs from the hospital setting in terms of the nature of formal service provision, the inclusion of clients and family members in the direct provision of healthcare interventions, and the characteristics of the client receiving care.

The aging of the population with concomitant increases in the prevalence of chronic conditions, combined with the need for healthcare authorities to find models of care that are more sustainable than institutionalised care, makes it likely that use of HC services will increase considerably in the coming years. For example, in Canada there was a 51%
increase in the number of HC clients from 1997 to 2007, with over 900,000 individuals receiving HC services in 2007. Improving the safety of HC becomes more crucial given this rising utilisation and intensity of services provided in the home. One of the first Canadian HC client safety studies reported a 5.5% annual incidence rate of AEs in a sample of 400 Winnipeg HC clients, of which injurious falls accounted for nearly half (46%). Two recent studies, one conducted in the USA and one in Canada, found that 13% of HC clients experienced an AE. The two Canadian studies were limited in sample size, and each involved only one jurisdiction. The US study was based on clinical-administrative data not specifically designed to assess client safety. Clinical-administrative databases constitute rich and cost-efficient sources of information on clients and services provided, but they have limitations in terms of the detail they can provide regarding the nature of AEs and contributing causes. Incident reports filed by healthcare staff are another means of documenting problems of quality but they have been found to under-report safety events, at least in institutional settings.

One more characteristic that distinguishes HC from hospital care, and has an impact on the calculation of the rate of AEs is the duration of care. The common way to calculate an AE rate in acute care hospitals is to assess the proportion of patients who experienced an AE, without necessarily taking into consideration the length of stay, which does not vary considerably across patients. However, since HC can be as short as a few days or lasts several years, it is more important to consider the length of ‘exposure’: the risk of experiencing an AE is likely to be higher in clients receiving HC for a full year than in those who were cared for only a few days. However, past HC studies did not fully take that reality into account while calculating AE rates.

This study attempted to fill these gaps by using data from client health records (or charts) to assess AEs and calculate an AE incidence rate per client-year.

The objectives of this study were to document the incidence rate and types of AEs among HC clients in three Canadian provinces in different regions of the country: Western Canada (Manitoba), Central Canada (Quebec) and Eastern Canada (Nova Scotia); identify the factors contributing to AEs among HC clients; and determine to what extent evidence of completion of incident reports was documented in charts where AEs were found.

METHODS

Definitions

The following definition of an AE for HC clients was adapted from that used in international studies of hospital AEs and a recent Canadian study of HC AEs: an unintended physical, mental or social injury, harm or complication that results in disability, death or increased use of healthcare resources, and that is caused by health care rather than by the client’s underlying disease process. In the context of HC, ‘healthcare’ includes three sources of care: (paid) healthcare providers, (unpaid) informal caregivers and self-care (care provided by the clients themselves). An AE can be caused by an act of commission (e.g., inadequate treatment plan or poorly executed treatments) or of omission (e.g., missed diagnosis, failure to treat).

Study design and setting

This was a retrospective cohort study based on a review of charts of clients receiving publicly funded HC services in Manitoba, Quebec and Nova Scotia. In Manitoba, data were examined for clients receiving services under the responsibility of the Winnipeg Regional Health Authority, the largest city and major population centre for the province. In Quebec, publicly funded HC is provided by Health and Social Services Centers (CSSS). Due to budgetary limitations, only CSSSs in regions located within 260 km from Montreal (the provincial research coordination centre) were included in the sampling frame; this includes 10 of the 18 health regions in the province comprising about 90% of the Quebec population. Because Montreal is the most populous health region, five CSSSs were randomly selected out of its 12 CSSSs. In addition, five regions were randomly selected from the nine remaining regions and one CSSS was randomly selected from each of those five regions, for a total of 10 CSSSs in Quebec. In Nova Scotia, clients who received services provided by the Victorian Order of Nurses (VON) in the Halifax and Cape Breton Island (Sydney) regions were eligible for inclusion; these regions include the majority (57%) of the population of the province. All selected sites agreed to participate in the study and to provide access to client charts. Ethics approvals were obtained from the University of Manitoba, the Winnipeg Regional Health Authority, the University of Montreal, Dalhousie University, VON and each Quebec participating CSSS that required it.

Case selection and sample

Using a 95% confidence level and an estimated AE incidence of 13% (the highest rate reported in recent studies), a sample of 1200 charts provides a margin of error of 1.84% (Roasoft Inc sample size calculator) and allows for the identification of a sufficient number of AEs and the analysis of risk factors. Consistent with this parameter, the target study sample was 1200 cases, including 300 from the Winnipeg region, 600 from Quebec (60 from each of the 10 CSSSs) and 300 from Nova Scotia (150 from each of the two regions). In each of the study sites a random sample of clients who were discharged from the publicly funded HC programme in fiscal year 2007, with over 900,000 individuals receiving HC services in 2007. Improving the safety of HC becomes more crucial given this rising utilisation and intensity of services provided in the home. One of the first Canadian HC client safety studies reported a 5.5% annual incidence rate of AEs in a sample of 400 Winnipeg HC clients, of which injurious falls accounted for nearly half (46%). Two recent studies, one conducted in the USA and one in Canada, found that 13% of HC clients experienced an AE. The two Canadian studies were limited in sample size, and each involved only one jurisdiction. The US study was based on clinical-administrative data not specifically designed to assess client safety. Clinical-administrative databases constitute rich and cost-efficient sources of information on clients and services provided, but they have limitations in terms of the detail they can provide regarding the nature of AEs and contributing causes. Incident reports filed by healthcare staff are another means of documenting problems of quality but they have been found to under-report safety events, at least in institutional settings.

One more characteristic that distinguishes HC from hospital care, and has an impact on the calculation of the rate of AEs is the duration of care. The common way to calculate an AE rate in acute care hospitals is to assess the proportion of patients who experienced an AE, without necessarily taking into consideration the length of stay, which does not vary considerably across patients. However, since HC can be as short as a few days or lasts several years, it is more important to consider the length of ‘exposure’: the risk of experiencing an AE is likely to be higher in clients receiving HC for a full year than in those who were cared for only a few days. However, past HC studies did not fully take that reality into account while calculating AE rates.

This study attempted to fill these gaps by using data from client health records (or charts) to assess AEs and calculate an AE incidence rate per client-year.

The objectives of this study were to document the incidence rate and types of AEs among HC clients in three Canadian provinces in different regions of the country: Western Canada (Manitoba), Central Canada (Quebec) and Eastern Canada (Nova Scotia); identify the factors contributing to AEs among HC clients; and determine to what extent evidence of completion of incident reports was documented in charts where AEs were found.

METHODS

Definitions

The following definition of an AE for HC clients was adapted from that used in international studies of hospital AEs and a recent Canadian study of HC AEs: an unintended physical, mental or social injury, harm or complication that results in disability, death or increased use of healthcare resources, and that is caused by health care rather than by the client’s underlying disease process. In the context of HC, ‘healthcare’ includes three sources of care: (paid) healthcare providers, (unpaid) informal caregivers and self-care (care provided by the clients themselves). An AE can be caused by an act of commission (e.g., inadequate treatment plan or poorly executed treatments) or of omission (e.g., missed diagnosis, failure to treat).

Study design and setting

This was a retrospective cohort study based on a review of charts of clients receiving publicly funded HC services in Manitoba, Quebec and Nova Scotia. In Manitoba, data were examined for clients receiving services under the responsibility of the Winnipeg Regional Health Authority, the largest city and major population centre for the province. In Quebec, publicly funded HC is provided by Health and Social Services Centers (CSSS). Due to budgetary limitations, only CSSSs in regions located within 260 km from Montreal (the provincial research coordination centre) were included in the sampling frame; this includes 10 of the 18 health regions in the province comprising about 90% of the Quebec population. Because Montreal is the most populous health region, five CSSSs were randomly selected out of its 12 CSSSs. In addition, five regions were randomly selected from the nine remaining regions and one CSSS was randomly selected from each of those five regions, for a total of 10 CSSSs in Quebec. In Nova Scotia, clients who received services provided by the Victorian Order of Nurses (VON) in the Halifax and Cape Breton Island (Sydney) regions were eligible for inclusion; these regions include the majority (57%) of the population of the province. All selected sites agreed to participate in the study and to provide access to client charts. Ethics approvals were obtained from the University of Manitoba, the Winnipeg Regional Health Authority, the University of Montreal, Dalhousie University, VON and each Quebec participating CSSS that required it.

Case selection and sample

Using a 95% confidence level and an estimated AE incidence of 13% (the highest rate reported in recent studies), a sample of 1200 charts provides a margin of error of 1.84% (Roasoft Inc sample size calculator) and allows for the identification of a sufficient number of AEs and the analysis of risk factors. Consistent with this parameter, the target study sample was 1200 cases, including 300 from the Winnipeg region, 600 from Quebec (60 from each of the 10 CSSSs) and 300 from Nova Scotia (150 from each of the two regions). In each of the study sites a random sample of clients who were discharged from the publicly funded HC programme in fiscal year
2009–2010 (1 April 2009–31 March 2010) was selected. If a client was discharged more than once in the period, the first discharge (considered the index admission) was selected. Exclusion criteria were discharged outside of the study range, cases in which no HC services were initiated, and/or where the HC services provided were not directed towards the client (eg, services were a direct support to a family caregiver). Charts were randomly selected in each province until the required numbers were obtained. In total, 1200 valid charts were reviewed. The sampling cascade is shown in figure 1.

Data collection
The format and approach to charting across the sample sites varied, but generally all charts included administrative data consistent with client identification; social data describing living situations; clinical data including healthcare history, diagnoses and medication use; and clients’ physical, cognitive and interpersonal functional statuses (table 1 and see online supplementary appendix table 1). The charts were specific to the clients’ HC and were compiled and maintained by HC case managers or nurses; charts were not physician records or cross-sector healthcare electronic records. Entries were made on a daily, weekly, monthly or other schedule depending on the protocols required by the HC programme.

Selected charts were reviewed using an adaptation for HC of the standard method developed for hospital chart review,†‡§ a method used in the recent study in Ontario Canada.6 Chart review forms, adapted from Sears et al6 were computerised and installed on portable computers. Chart review was carried out on the premises of each HC agency. The review process was performed in two stages. In the first stage, each selected HC chart was assessed by one member of a team of trained nurses for the presence of one or more of 24 screening criteria potentially sensitive to the occurrence of an AE (table 2) and used in the Ontario study.6 As reported by Sears et al6 these criteria were adapted from those applied in hospital AE studies, modified and validated through an expert Delphi process, and then tested for validity. Nurse reviewers also recorded information on the clients’ demographics, functional status (activities of daily

---

**Figure 1** Sampling cascade. *Reasons for ineligibility were dates outside of study range; no home care episode; or home care service request was for informal caregiver.
living (ADL); instrumental ADL (IADL)) and comorbid conditions (presence/absence). Ordinal ADL scores (1–4) were calculated as the sums of the scores on each item (1=independent; 2=with difficulty; 3=assistance; 4=dependent) divided by the number of items. IADL scores (1–4) were calculated the same way. A higher score thus meant a higher level of functional impairment.

In the second stage, charts that were positive for at least one screening criterion were reviewed by one member of a team of trained physicians experienced in HC. An AE was identified when a physician reviewer determined that all three AE criteria were met: there was an injury and the client experienced disability, death or increased use of services, and it was likely caused by healthcare (ie, the causation rating was at least four: more than a 50% likelihood of being caused by healthcare). To make this determination, physician reviewers first assessed whether the client suffered any unintended injury, harm or complication. If there was no injury, the review process stopped. If there was an injury reviewers determined if the injury resulted in disability, death or increased use of healthcare services (eg, intensified or prolonged

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No</th>
<th>Yes</th>
<th>Total</th>
<th>p Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>71.20</td>
<td>75.93</td>
<td>71.52</td>
<td>0.016</td>
</tr>
<tr>
<td>Difficulty with communication later on during index admission (%)†</td>
<td>14.46</td>
<td>23.75</td>
<td>15.10</td>
<td>0.025</td>
</tr>
<tr>
<td>Difficulty with communication at intake or later on during index admission (%)</td>
<td>15.35</td>
<td>24.05</td>
<td>15.94</td>
<td>0.041</td>
</tr>
<tr>
<td>Cognitive patterns problems (out of 3)§</td>
<td>0.65</td>
<td>1.12</td>
<td>0.68</td>
<td>0.004</td>
</tr>
<tr>
<td>Communication problems (out of 4)</td>
<td>0.71</td>
<td>0.96</td>
<td>0.72</td>
<td>0.017</td>
</tr>
<tr>
<td>Mood and behaviour problems (out of 2)</td>
<td>0.31</td>
<td>0.49</td>
<td>0.32</td>
<td>0.022</td>
</tr>
<tr>
<td>IADL (1 independent, 4 dependent)</td>
<td>2.35</td>
<td>2.82</td>
<td>2.39</td>
<td>0.000</td>
</tr>
<tr>
<td>IADL (% independent)</td>
<td>43.9</td>
<td>27.4</td>
<td>42.6</td>
<td>0.000</td>
</tr>
<tr>
<td>ADL (1 independent, 4 dependent)</td>
<td>1.63</td>
<td>1.86</td>
<td>1.65</td>
<td>0.015</td>
</tr>
<tr>
<td>ADL (% independent)</td>
<td>68.5</td>
<td>59.3</td>
<td>67.8</td>
<td>0.011</td>
</tr>
<tr>
<td>Client functional status has deteriorated during admission (%)</td>
<td>24.88</td>
<td>56.79</td>
<td>27.11</td>
<td>0.000</td>
</tr>
<tr>
<td>Pain (%)</td>
<td>55.57</td>
<td>75.38</td>
<td>57.01</td>
<td>0.002</td>
</tr>
<tr>
<td>History of falls (%)</td>
<td>30.00</td>
<td>43.84</td>
<td>31.09</td>
<td>0.014</td>
</tr>
<tr>
<td>Medications: number of prescriptions and over-the-counter meds:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>33</td>
<td>2</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>1–3</td>
<td>154</td>
<td>9</td>
<td>163</td>
<td></td>
</tr>
<tr>
<td>4–9</td>
<td>532</td>
<td>36</td>
<td>568</td>
<td></td>
</tr>
<tr>
<td>&gt;9</td>
<td>267</td>
<td>34</td>
<td>301</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>986</td>
<td>81</td>
<td>1067</td>
<td>0.015</td>
</tr>
<tr>
<td>Number of comorbid conditions (out of 32)</td>
<td>3.94</td>
<td>5.56</td>
<td>4.05</td>
<td>0.000</td>
</tr>
<tr>
<td>Cardiac and vascular disease (out of 7)</td>
<td>1.37</td>
<td>2.01</td>
<td>1.41</td>
<td>0.000</td>
</tr>
<tr>
<td>Transient ischaemic attack</td>
<td>0.08</td>
<td>0.15</td>
<td>0.08</td>
<td>0.032</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>0.21</td>
<td>0.40</td>
<td>0.22</td>
<td>0.000</td>
</tr>
<tr>
<td>Hypertension</td>
<td>0.50</td>
<td>0.67</td>
<td>0.51</td>
<td>0.004</td>
</tr>
<tr>
<td>Neurological</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other dementia</td>
<td>0.07</td>
<td>0.17</td>
<td>0.08</td>
<td>0.002</td>
</tr>
<tr>
<td>Infections (out of 5)</td>
<td>0.25</td>
<td>0.53</td>
<td>0.27</td>
<td>0.001</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>0.07</td>
<td>0.14</td>
<td>0.07</td>
<td>0.018</td>
</tr>
<tr>
<td>Urinary tract infection</td>
<td>0.06</td>
<td>0.20</td>
<td>0.07</td>
<td>0.000</td>
</tr>
<tr>
<td>Other infection</td>
<td>0.11</td>
<td>0.20</td>
<td>0.12</td>
<td>0.026</td>
</tr>
<tr>
<td>Diabetes</td>
<td>0.26</td>
<td>0.38</td>
<td>0.26</td>
<td>0.012</td>
</tr>
<tr>
<td>Respiratory (out of 2)</td>
<td>0.18</td>
<td>0.33</td>
<td>0.19</td>
<td>0.020</td>
</tr>
<tr>
<td>Emphysema/chronic obstructive lung disease</td>
<td>0.12</td>
<td>0.25</td>
<td>0.13</td>
<td>0.001</td>
</tr>
<tr>
<td>Other comorbid conditions not noted above</td>
<td>0.59</td>
<td>0.70</td>
<td>0.60</td>
<td>0.045</td>
</tr>
</tbody>
</table>

*p Value indicates significant difference between cases with and without adverse event, based on χ² test for categorical variables and t test for continuous variables. None of the p values correct for multiple comparisons.

†This table presents only the statistically significant variables. See online supplementary appendix table 1 for all tested variables.

‡Indicates the percentage of clients with this characteristic.

§Indicates the average number of problems or conditions in this category.

Table 1  Characteristics of clients with or without AE (selected variables)†
HC, additional treatment, medical consultations, emergency room visit, hospital admission). Finally, the physician reviewers used the six-point scale (1 = virtually no evidence of healthcare causation; 2 = slight to modest evidence of healthcare causation; 3 = healthcare causation not likely (less than 50/50, but ‘close call’); 4 = healthcare causation more likely (more than 50/50, but ‘close call’); 5 = moderate to strong evidence of healthcare causation; 6 = virtually certain evidence of healthcare causation), employed in previous studies,14–16 to determine the extent to which healthcare (whether by paid healthcare providers, informal caregivers or the client), rather than the clients’ disease processes, was responsible for the injury. Physician reviewers also judged the preventability of each AE using the 6-point scale (1 = virtually unpreventable; 2 = slight to modest preventability; 3 = preventability not quite likely (less than 50/50, but ‘close call’); 4 = preventability more than likely (more than 50/50, but ‘close call’); 5 = strongly preventable; 6 = virtually certain for preventability) also used in previous studies.14–16

After each day of chart review, nurses and physician reviewers transferred data to a secure web depot at the coordinating research centre at the University of Montreal. At both stages of the review process, interrater reliability was also assessed on a random sample of 10% of the charts using the κ statistic.16

### Table 2 Screening criteria in the stage 1 review

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Charts with criterion</th>
<th>N</th>
<th>%†</th>
<th>Adverse event charts with criterion</th>
<th>N</th>
<th>%§</th>
<th>χ² †</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Unplanned admission (including readmission) to home care within the</td>
<td>57</td>
<td>4.8</td>
<td></td>
<td></td>
<td>13</td>
<td>16.0</td>
<td>**</td>
</tr>
<tr>
<td>6 months after discharge from index admission</td>
<td>2  Request for admission (denied or wait-listed) to home care within the</td>
<td>3</td>
<td>0.3</td>
<td></td>
<td>0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>6 months after discharge from index admission</td>
<td>3  Recognised actual or potential environmental risks</td>
<td>108</td>
<td>9.0</td>
<td></td>
<td>12</td>
<td>14.8</td>
<td></td>
</tr>
<tr>
<td>4  Recognised actual or potential risks related to client behaviour</td>
<td>115</td>
<td>9.6</td>
<td></td>
<td></td>
<td>19</td>
<td>23.5</td>
<td>**</td>
</tr>
<tr>
<td>5  Inappropriate/inaccurate home care or service provider assessment of client</td>
<td>11</td>
<td>0.9</td>
<td></td>
<td></td>
<td>0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>6  New problem/diagnosis noted during index admission</td>
<td>243</td>
<td>20.3</td>
<td></td>
<td></td>
<td>38</td>
<td>46.9</td>
<td>**</td>
</tr>
<tr>
<td>7  Client injury, harm, trauma or complication during home care admission</td>
<td>182</td>
<td>15.2</td>
<td></td>
<td></td>
<td>34</td>
<td>42.0</td>
<td>**</td>
</tr>
<tr>
<td>8  Unplanned assessment/treatment by primary care provider during index admission</td>
<td>67</td>
<td>5.6</td>
<td></td>
<td></td>
<td>17</td>
<td>21.0</td>
<td>**</td>
</tr>
<tr>
<td>9  Unplanned visit to hospital emergency department during index admission</td>
<td>279</td>
<td>23.3</td>
<td></td>
<td></td>
<td>49</td>
<td>60.5</td>
<td></td>
</tr>
<tr>
<td>10 Unplanned admission to acute care hospital during index admission</td>
<td>242</td>
<td>20.2</td>
<td></td>
<td></td>
<td>31</td>
<td>38.3</td>
<td>**</td>
</tr>
<tr>
<td>11 Unplanned admission/request for admission to long-term care facility</td>
<td>45</td>
<td>3.8</td>
<td></td>
<td></td>
<td>3</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>12 Adverse drug reaction during index admission</td>
<td>7</td>
<td>0.6</td>
<td></td>
<td></td>
<td>4</td>
<td>4.9</td>
<td>**</td>
</tr>
<tr>
<td>13 Acquired infection/sepsis</td>
<td>115</td>
<td>9.6</td>
<td></td>
<td></td>
<td>32</td>
<td>39.5</td>
<td>**</td>
</tr>
<tr>
<td>14 Development of neurological deficit not present on admission but present at the time of discharge from the index home care stay</td>
<td>8</td>
<td>0.7</td>
<td></td>
<td></td>
<td>2</td>
<td>2.5</td>
<td>*</td>
</tr>
<tr>
<td>15 Emotional or psycho-social problem in patient or informal caregiver</td>
<td>44</td>
<td>3.7</td>
<td></td>
<td></td>
<td>9</td>
<td>11.1</td>
<td>**</td>
</tr>
<tr>
<td>16 Unexpected death:</td>
<td>37</td>
<td>3.1</td>
<td></td>
<td></td>
<td>2</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>17 Other client complications, for example, AMI, CVA, PE, DVT, etc</td>
<td>101</td>
<td>8.4</td>
<td></td>
<td></td>
<td>12</td>
<td>14.8</td>
<td>*</td>
</tr>
<tr>
<td>18 Expected family/informal caregiver availability for client assistance not realised</td>
<td>17</td>
<td>1.4</td>
<td></td>
<td></td>
<td>4</td>
<td>4.9</td>
<td>**</td>
</tr>
<tr>
<td>19 Dissatisfaction with care documented in the client record and/or evidence of complaint lodged</td>
<td>17</td>
<td>1.4</td>
<td></td>
<td></td>
<td>1</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>20 Adverse event reported by a caregiver</td>
<td>50</td>
<td>4.2</td>
<td></td>
<td></td>
<td>8</td>
<td>9.9</td>
<td>**</td>
</tr>
<tr>
<td>21 Documentation or correspondence indicating litigation, either contemplated or actual</td>
<td>1</td>
<td>0.1</td>
<td></td>
<td></td>
<td>0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>22 Inappropriate discharge/inadequate discharge plan for index admission</td>
<td>2</td>
<td>0.2</td>
<td></td>
<td></td>
<td>0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>23 Unplanned admission to any hospital within the 6 months after discharge from index admission</td>
<td>21</td>
<td>1.8</td>
<td></td>
<td></td>
<td>4</td>
<td>4.9</td>
<td>**</td>
</tr>
<tr>
<td>24 Any other undesirable outcomes not covered above</td>
<td>7</td>
<td>0.6</td>
<td></td>
<td></td>
<td>0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

Mean number of criteria per chart 3.40 3.63

* p<0.05; ** p<0.01.
† χ² indicates the association between the presence of a criterion and the risk of an adverse event.
‡ Out of 1200 fully audited charts.
§ Out of 81 charts with adverse events.
AMI, acute myocardial infarction; CVA, cerebrovascular accident (stroke); DVT, deep vein thrombosis; PE, pulmonary embolus.
Agreement between nurses for the first stage of the review process (presence of a screening criterion) was moderate ($\kappa = 0.58$). Agreement between physicians for the second stage of the review ranged from fair for the determination of whether an injury had occurred ($\kappa = 0.35$) to excellent for the determination of whether the injury had a consequence ($\kappa = 1.0$).\textsuperscript{16}

Reviewers looked for AEs that occurred during the HC index admission and that were detected during either the index or during subsequent HC admissions over the 6-month period after discharge from the index admission. Since HC length of stay can be very long, they also limited identification of AEs to the 12 months preceding discharge from the index admission. Only information present in the HC charts could be used; information in hospital charts or elsewhere was not available unless mentioned in the HC charts.

Data analysis

Descriptive analysis was used to document client characteristics and AEs. The AE rate was calculated as the proportion of charts in which at least one AE was found (as done in other studies) and as the incidence rate per client-year (number of AEs over the total length of stay of all clients divided by 365); both rates were weighted for the sampling strategy. The sampling weights were based on inverse probability of being included in the sample for each province (Manitoba, Quebec and Nova Scotia). For Quebec and Nova Scotia, the weights were based on a stratified sampling by region. For Manitoba, the weights were based on a simple random sample.\textsuperscript{1}

Bivariate analysis and forward stepwise multivariate logistic regression were used to identify factors associated with the risk of having an AE. In addition to age and sex, the following variables that were significant in bivariate analysis were tested in the multivariate regression: number of comorbid conditions; number of medications; length of stay; ADL score (1–4); IADL score (1–4); communication; mood; and cognitive problems.

RESULTS

Of the 1200 charts that were reviewed by nurses, 518 (43.2%) were positive for at least one screening criterion. Two-thirds of criteria were found to be associated ($p<0.05$) with finding an AE in the chart (table 2). The five most prevalent screening criteria identified in charts with an AE were: unplanned visit to hospital emergency department during index admission (n=49, 60.5%); new problem/diagnosis noted during index admission (n=38, 46.9%); client injury, harm, trauma or complication during HC admission (n=34, 42%); acquired infection/sepsis (n=32, 39.5%); and unplanned admission to acute care hospital during index admission (n=31, 38.3%).

The results show that 417 of the 518 clients were found to have experienced 715 injuries (first criterion of the AE definition); 409 of the 417 clients injured had resulting disability, death or increased use of healthcare resources (second criterion of the definition). The third AE criterion (caused by healthcare rather than by the client’s underlying disease process) was identified 93 times (ie, there were 93 AEs) across 81 clients. Most (71 of 81) clients had only one AE; nine clients experienced two AEs and one client experienced four AEs. After weighting for the sampling strategy, the overall AE rate (proportion of clients with AEs=81/1200) was 4.2% (95% CI 3.0% to 5.4%); 56% (n=52) of the 93 AEs were judged preventable by the physician reviewers. Adjusting the analysis to account for clients with lengths of stay on HC of less than 1 year, the AE incidence rate per client-year was 10.1% (95% CI 8.4% to 11.8%).

Table 3 presents the types of AEs suffered by clients. The most frequent AE was an injurious fall (n=16, 17.2%), but medication was involved in 21.5% (n=20) of AEs, whether it was related to an injurious fall (n=4) or another type of AE. Wound infections (n=13, 14%) and psychosocial, behavioural or mental health problems (n=11, 11.8%) were also frequent.

In terms of potential impact, 91.4% (n=85) of AEs were associated with an increased use of healthcare resources, 68.8% (n=64) with client disability and 7.5% (n=7) with death. More than one impact for each AE was possible. Healthcare personnel were

<table>
<thead>
<tr>
<th>Table 3 Types of adverse events (injuries)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adverse event</strong></td>
</tr>
<tr>
<td>Fall injury</td>
</tr>
<tr>
<td>Wound infection</td>
</tr>
<tr>
<td>Psychosocial, behavioural, mental problem</td>
</tr>
<tr>
<td>Medication problem (adverse drug reaction)*</td>
</tr>
<tr>
<td>Pressure ulcer</td>
</tr>
<tr>
<td>Other wound problem</td>
</tr>
<tr>
<td>Non-wound infection</td>
</tr>
<tr>
<td>Syncope or seizure</td>
</tr>
<tr>
<td>Delayed wound healing</td>
</tr>
<tr>
<td>Shortness of breath</td>
</tr>
<tr>
<td>Skin tear or laceration</td>
</tr>
<tr>
<td>Hypo/hyperglycaemia</td>
</tr>
<tr>
<td>Gastrointestinal problem</td>
</tr>
<tr>
<td>Intravenous site problem</td>
</tr>
<tr>
<td>Fracture</td>
</tr>
<tr>
<td>Bleeding—minor</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

*In addition to the six cases of medication problems that correspond to direct adverse drug reactions, there were 14 cases among the other listed injuries in which medication was involved, including four injurious falls, for a total of 20 (21.5%) adverse event positive cases.

judged by physician reviewers to have contributed to the occurrence of the AE (a rate of 4 or higher on the six-point causation scale) in 46.2% (n=43) of AEs, informal caregivers in 20.4% (n=19) of AEs and clients in 48.4% (n=45) of AEs. Two or three of those types of people were involved in 14% (n=13) of AEs.

Some characteristics of clients who experienced an AE differed significantly compared with non-AE clients. The characteristics significantly (p≤0.05) related to AEs are age, difficulty with communication, cognitive problems, mood and behaviour problems, IADL and ADL scores, functional status, pain, history of falls, number of medications and number of comorbid conditions (Table 1 presents the statistically significant variables; online supplementary appendix table 1 presents all variables tested). When these significant characteristics were tested together in a logistic regression analysis, only two remained statistically significant: the number of comorbid conditions which increased the risk of experiencing an AE by 15% for each additional comorbid condition a client had (OR 1.15; 95% CI 1.05 to 1.26); and the IADL score when the risk of having an AE increased by 54% with each increase in level of the four-point IADL score (ie, as the client becomes more dependent) (OR 1.54; 95% CI 1.16 to 2.04). The final model is shown in Table 4; its strength is reflected by the C statistic (0.672; 95% CI 0.613 to 0.732) and the goodness of fit (p=0.659).17

Finally, in the first stage of review, nurse reviewers checked if there was indication in the chart that an incident report was filled out by healthcare professionals. Only 17.3% (n=14) of the 81 charts in which an AE was found contained documentation of an incident report, while 4.8% (n=54) of charts without an AE did. More specifically, for the most common types of AEs, documentation regarding incident report generation was present in 6.2% (n=1) of charts with a fall injury, 7.7% (n=1) of charts with a wound infection, 27.3% (n=3) of charts with a psychosocial, behavioural or mental health problem and 33.3% (n=2) of charts in which a medication problem was found.

DISCUSSION

This is the first study to assess AEs among HC clients across different regions of Canada using chart review. In contrast with some previous studies2 18 19 we considered all age groups, all medical conditions and all types of AEs, including psychosocial and mental-health-related AEs. We found that the proportion of clients who experienced an AE was 4.2% (95% CI 3.0% to 5.4%). This rate is much lower than the 13% found in Ontario6 and in the USA.12 However, we went beyond the calculation of a simple proportion of clients experiencing an AE: we took into consideration the ‘exposure’ time of clients, that is, the length of HC stay, and calculated an AE incidence rate per client-year. We think this measure is an improvement on previous calculations and should be applied in future studies. Our AE incidence rate per client-year was 10.1% (95% CI 8.4% to 11.8%). This is higher than the annual incidence rate of 5.5% observed in Winnipeg.7 These different rates may be due to differences in HC services, client characteristics and/or the methods used to assess the charts.6

Injurious falls were among the most common AEs found in this study and in other studies whether the methodology was based on chart review6 7 or secondary data analysis.8 We noted a particular issue with medication problems. Medication errors can cause direct and immediate adverse drug reactions (eg, rash) and we found six such cases among the 93 AEs. However, we were able to identify 14 additional cases in which a medication error contributed to another type of AE (eg, a medication caused dizziness that provoked a fall; a client refused to take prescribed medication and developed a psychotic episode). The role of medication in such AEs cannot be established by analysing secondary data alone; this underlines the need for expert review of client charts to identify the causal chain, provided sufficient information is available in the charts.

Our study shows that clients with more comorbid conditions (ie, more complex cases) and those who had a higher score for IADL, (ie, those who were more functionally vulnerable), were at greater risk of experiencing an AE. This finding is consistent with

<table>
<thead>
<tr>
<th>Variable*</th>
<th>Wald</th>
<th>p Value</th>
<th>Exp (B)</th>
<th>95% CI Low</th>
<th>95% CI High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of comorbid conditions</td>
<td>8.879</td>
<td>0.003</td>
<td>1.15</td>
<td>1.05</td>
<td>1.26</td>
</tr>
<tr>
<td>IADL score†</td>
<td>8.694</td>
<td>0.003</td>
<td>1.54</td>
<td>1.16</td>
<td>2.04</td>
</tr>
<tr>
<td>Intercept</td>
<td>101.805</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C statistic=0.672 (95% CI 0.613 to 0.732)
Goodness of fit: p=0.659

*Variables tested were age, sex, number of comorbid conditions, number of medications, length of stay, ADL score, IADL score, communication, mood and cognitive problems.
†IADL, instrumental activities of daily living. Score ranged from 1 (independent) to 4 (dependent).
the results of other studies of HC safety\(^6\)\(^ {12}\) and hospital safety.\(^3\) However, beyond clients’ conditions, being cared for at home poses particular challenges for client safety since a significant portion of care is provided by the clients themselves and their caregivers, and not by healthcare professionals. This is reflected in the physician reviewers’ determination that clients contributed to the occurrence of 48.4% of AEs, informal caregivers to 20.4% of AEs, while healthcare personnel contributed to the causation of 46.2% of AEs. This tri-partite contribution to AE causation is consistent with the findings of the Ontario study\(^6\) of contributing sources and suggests that safety in HC is dependent upon clients and informal caregivers being aware of risks and possessing the skills to mitigate those risks, as well as HC staff. The complex dynamic of care that occurs when providing service in the home may not be as easily amenable to risk reduction as care in more controlled care environments, such as hospitals. This study provides HC programmes with the evidence of where the incidence of AEs is greatest and what the consequences can be. The results can help effectively target education for staff, clients and informal caregivers to improve safety procedures and quality of care.

Our study showed that a small minority of charts (17.3%) in which an AE was found contained documentation that an incident report was filled out by healthcare professionals. This low proportion may be due to the fact that including an incident report (or a copy of it) in a client chart might not be compulsory in all HC programmes or that the incident reporting system does not allow for the linkage of incident reports with client charts. However, in at least one of the provinces studied here there is a requirement to place a copy of the incident report in the chart but the documentation of incident reports in charts with an AE was still low (14.8%). Beyond organisational policies, many reasons have been proposed to explain low incident reporting in institutionalised settings, including unclear or inadequate forms, lack of feedback on reporting, fear of personal consequences and lack of time to fill out forms.\(^{20–22}\) These same factors might apply in HC. In addition, the fact that 4.8% of cases without an AE contained evidence that an incident report was completed suggests that what is reported does not necessarily correspond to the definition of an AE that was used in this study and that HC programmes may have different criteria for generating incident reports. This finding also suggests that it would be beneficial for HC agencies to develop a national standard for what the processes for reporting AEs should be.

This study has a number of limitations. First, as with most other studies that used the same two-stage review method, charts that were not screened positive by nurses were not reviewed by physicians. It is thus possible that some of these cases may have experienced an AE, yet the list of screening criteria used was quite extensive; thus the likelihood that a chart without any of these criteria contained information indicating that there was an AE may be quite low, however human factors such as reviewer fatigue and temporal trending may affect the sensitivity with which the screening criteria were identified. Second, it is possible that information relevant to client safety may not be present in client charts either because healthcare providers are not always present with the client and hence do not observe an AE, or they are not mandated to record it in the chart. Third, although inter-rater reliability was comparable to that of other HC\(^6\) or hospital studies\(^1\)\(^ {3–5}\) using the same methodology, it was far from perfect on some dimensions. This suggests that professional judgment was not clear cut based on documentation available and given the complexity of care environment/dynamics of providing care in the home. Fourth, the fact that the number of cases with each type of AE was rather low limits the possibility of finding specific risk factors. For example, the risk factors associated with injurious falls is probably different from those associated with wound infections. Our analysis combined all AEs. A much larger sample would be needed to assess the risk factors associated with each type of AE and the specific actions that might reduce the likelihood of AEs occurring. Fifth, not all information regarding other types of healthcare received by HC clients (eg, physician and pharmacist care) was available in the charts, so it was not possible to detect all AEs that led to emergency room visits or hospital admissions. An inter-sectoral electronic client chart that records a more complete history of healthcare use (including physician visits, emergency room visits, acute care and long-term care facility admissions) would allow for a better assessment of client care safety.

This study has expanded the breadth of safety in healthcare research by providing evidence of AEs that are relevant to HC. However, HC programmes vary across Canadian provinces in many ways (eg, governance and organisation, services, including roles of professional and non professional staff, quality and accountability, conceptions of safety, etc.).\(^{11}\) Future research should document the programmes’ specific elements and attempt to link them with indicators of AEs to identify which components of care in which contexts are safer for clients. This would help programmes select actions to take to improve HC safety.

CONCLUSION

Client safety is an important issue in HC just as it is in institutionalised care. Sicker and more dependent clients are at a higher risk of experiencing an AE and special attention should be devoted to these subgroups. Moreover, unlike institutional settings, homes are not designed for healthcare and are not regulated environments; and healthcare personnel are not
present on a continuing basis. These environmental factors may contribute to risks that HC clients encounter. The important role in HC of clients themselves and of their unpaid caregivers must be taken into consideration when planning strategies to improve HC safety.

Acknowledgements We wish to thank the home care agencies that gave us access to client charts and the nurse and physician reviewers for their competent work. We also thank Caroline Slegers-Boyd (Manitoba) and Andrea Scobie (Nova Scotia) for coordinating data collection. We gratefully acknowledge Michèle Paré, Mylène Baptista, Jean-Guy Bouchard and Diane Crevier (University of Montreal) for their expert assistance with data management and analysis.

Contributors RB designed the study and data collection tools, monitored data collection, designed and supervised data analysis, and drafted and revised the paper. He is guarantor. NAS designed the study and data collection tools, analysed the data and revised the paper. DD designed the study and revised the paper. GRB designed the data collection tools and revised the paper. MM and LM designed the data collection tools, monitored data collection and revised the paper. ST designed the data collection tools, monitored data collection, analysed the data and revised the paper.

Funding This work was supported by the Canadian Patient Safety Institute; the Canadian Institutes of Health Research (CIHR) (Institutes of Health Services and Policy Research, Aging, Circulatory and Respiratory Health and Musculoskeletal Health and Arthritis); the Change Foundation; the Canadian Health Services Research Foundation (grant number HC-10-05 Doran-Blais); the Ministère de la Santé et des Services sociaux du Québec and the Nova Scotia Health Research Foundation.

Competing interests None.

Ethics approval Ethics research committee of the Faculty of Medicine of the University of Montreal. See lines 213–217 for complete ethics approval.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

REFERENCES
‘Care left undone’ during nursing shifts: associations with workload and perceived quality of care

Jane E Ball,1 Trevor Murrells,1 Anne Marie Rafferty,2 Elizabeth Morrow,1 Peter Griffiths3

ABSTRACT
Background There is strong evidence to show that lower nurse staffing levels in hospitals are associated with worse patient outcomes. One hypothesised mechanism is the omission of necessary nursing care caused by time pressure—‘missed care’.

Aim To examine the nature and prevalence of care left undone by nurses in English National Health Service hospitals and to assess whether the number of missed care episodes is associated with nurse staffing levels and nurse ratings of the quality of nursing care and patient safety environment.

Methods Cross-sectional survey of 2917 registered nurses working in 401 general medical/surgical wards in 46 general acute National Health Service hospitals in England.

Results Most nurses (86%) reported that one or more care activity had been left undone due to lack of time on their last shift. Most frequently left undone were: comforting or talking with patients (66%), educating patients (52%) and developing/updating nursing care plans (47%). The number of patients per registered nurse was significantly associated with the incidence of ‘missed care’ (p < 0.001). A mean of 7.8 activities per shift were left undone on wards that are rated as ‘failing’ on patient safety, compared with 2.4 where patient safety was rated as ‘excellent’ (p < 0.001).

Conclusions Nurses working in English hospitals report that care is frequently left undone. Care not being delivered may be the reason low nurse staffing levels adversely affect quality and safety. Hospitals could use a nurse-rated assessment of ‘missed care’ as an early warning measure to identify wards with inadequate nurse staffing.

INTRODUCTION
The National Health Service (NHS) in England, like many healthcare systems in the world, is facing intense pressure to maintain the quality and safety of care provided in hospitals at the same or less cost than in previous years.1 The quality of nursing care—and the potential for poor nursing care to do patients great harm—has been the focus of numerous recent reports in England.2 3 Poor quality care is a source of significant increased cost internationally.4 The Francis Inquiry5 examined the reasons why hundreds of patients experienced poor care at The Mid Staffordshire NHS Foundation Trust between January 2005 and March 2009. The Inquiry was instigated when hospital standardised mortality ratios (case mix adjusted mortality rates) indicated that between 400 and 1200 more patients than expected had died over a 2 year period. Numerous patient accounts were heard by the Inquiry, including negative experiences of fundamental aspects of nursing care including care such as communication, maintaining dignity, discharge planning and safety. Failure to ensure adequate nurse staffing was a central factor identified in the report.

There is clearly a need to understand the scale of potential problems in care delivery across the NHS and internationally. There is also a need to understand mechanisms which link nurse staffing to quality and safety outcomes—including our focus here—the nature and extent of care that might be being ‘left undone’.6 The purpose of this study is to describe the nature and prevalence of care left undone (as reported by nurses) and explore its association with nurse staffing levels and nurse ratings of the quality of care and patient safety environment.

BACKGROUND
The body of evidence demonstrating an association between patient outcomes and nurse staffing is substantial. A systematic
review of 102 studies concluded that increased registered nurse (RN) staffing levels are associated with lower rates of hospital related mortality and adverse patient events.7 For example, in intensive care units (ICUs) higher RN staffing was associated with lower levels of hospital related mortality; each additional full time equivalent RN per patient day corresponds to a 9% reduction in odds of death in ICUs (ORs, 0.91; 95% CI 0.86 to 0.96), which would save 5 lives per 1000 patients. Although much of the reviewed research was undertaken in hospitals in the USA similar findings have been identified in other countries, for example, Belgium,8 9 Korea,10 11 and the UK.12 Other research shows that in some cases variation in staffing levels and patient outcomes within hospitals is greater than that between hospitals, which can mask the effect of any relationship between nurse staffing and patient outcomes.13 In the USA, Needleman et al14 used a retrospective observational study to analyse 176 000 shifts. They reported that controlling for hospital and patient differences, mortality rates were significantly greater for patients receiving care on shifts where RN staffing was 8 h or more below the planned level.

Aspects of organisational climate, and in particular nursing practice environment, have also been identified as significant predictors of nursing quality and patient outcomes.15 A ‘positive work environment’ has been described as comprising factors including autonomy, positive relationships between staff, teamwork, job satisfaction and low risk of burnout.16 A series of research studies internationally have used the Practice Environment Scale (PES)15—for example, in Belgium,17 Taiwan,18 and China19—to show that practice environment characteristics have an association with a range of patient outcome measures.20 Data from the USA have been used to estimate that the increase in mortality associated with a change from good to mixed quality practice environment is greater than the change in mortality associated with a nurse caring for one more patient per shift.21 A failure to properly observe and respond to patient deterioration is also identified as being part of the causal pathway in the association between low staffing and death from treatable complications (often referred to as ‘failure to rescue’), although the staffing outcome relationship here may not be specific to nursing.22

Despite the strength of evidence for a link between nurse staffing and clinical outcomes, relatively little is known about the mechanisms through which variations in nurse staffing impact on mortality, or other patient outcomes.23 There is some evidence that unfinished care by nurses, or ‘missed care’ could be used as an indicator of overall quality,24 explaining over 40% of the variation in care quality ratings in one US study.6

AIM

The aim of the study was to use data collected from nurses working in English NHS hospitals to address the following questions:

1. What is the nature and prevalence of ‘missed care’ (ie, care that nurses regard as necessary but was left undone on their last shift due to lack of time)?
2. Is there a relationship between ward nurse staffing levels and the prevalence and type of nursing care that is left undone due to time constraints?
3. Is ‘missed care’ associated with perceptions of the overall quality of nursing care or patient safety environment of a ward?

METHODS

A cross-sectional survey design was used. The methodology of the survey in England followed a protocol established by the international RN4Cast consortium of 15 countries.25 The study focused on general medical and surgical wards in acute hospitals. Ethical approval was sought and gained (from the National Research Ethics Committee) and permissions acquired for the research to be undertaken at each hospital.

SAMPLE

In England a random stratified sample of 64 (out of a possible 341) NHS general acute hospital Trusts (the bodies managing one or more hospital) was identified to ensure mix by size, teaching status and region with a target sample size of 32 Trusts. Within each stratum the chief executive and chief nurse of Trusts were approached in a random order and invited to take part. If a Trust declined to participate then the next Trust in that stratum was approached until the quota defined by the sampling frame was fulfilled. Thirty-one of the 64 Trusts identified in the original sample agreed to take part. Within these Trusts we then took a stratified random sample of up to five general medical and five general surgical wards from each hospital operated by that Trust. Where a Trust had fewer than five wards in a given category we included all wards. Mixed medical/surgical wards were included in the medical sampling frame but analysed separately. In total 401 wards were included. Highly specialist, long-stay rehabilitation, critical care units and ICUs were excluded because of the high intensity of nursing care associated with these types of care settings.26

MEASURES

The full content of the questionnaire survey used in this study is described elsewhere and has been used extensively in previous studies of nurse staffing and patient safety.25 The questionnaire consisted of five sections presented over seven pages covering: Work Environment and Job Satisfaction, Quality and Safety, Your most recent shift, About you, and Where you work. Nurse staffing was calculated from the nurse
surveys; where nurses were asked to report the numbers of staff giving direct patient care (specifically ‘RNs’ and ‘other nursing care staff’) and the numbers of patients on the ward on the last shift they worked. From this we identified:

- patients per RN providing direct care
- patients per non-registered nursing staff (or Healthcare Support Workers (HCSW). The abbreviation HCSW is used as a variable label to refer to ‘other nursing care staff’ providing direct care who are not registered, such as Healthcare Assistants or nursing auxiliaries)
- the proportion (as a percentage) of the nursing team providing direct care that were RNs (referred to as ‘skill-mix’)

Nurses were also asked to report the number of patients requiring assistance with daily living and the number requiring frequent monitoring, to measure nursing workload intensity related to patient need.

The nurse work environment was assessed using the PES of the Nursing Work Index (revised), an internationally validated measure that has been adapted and used previously in Europe. The PES of the Nursing Work Index measures modifiable organisational factors, including managerial support for nursing, nurse participation in hospital affairs, doctor–nurse relations and promotion of care quality. Four items in the PES that related specifically to staffing and resourcing were excluded to avoid overlap (common variance) with the measures of nurse staffing used in the analysis. The mean PES Score for the remaining 28 items was used to give an overall rating of the practice environment (PES-28).

A single question asked nurses to rate the quality of care on their ward as fair, poor, good or excellent. Using an item from the Agency for Healthcare Research and Quality’s hospital survey on patient safety culture, previously validated in the UK, nurses gave their ward an overall grade on patient safety as poor, failing, acceptable, very good or excellent.

Care left undone (termed ‘missed care’ in the analyses) was assessed by asking nurses to ‘On your most recent shift, which of the following activities were necessary but left undone because you lacked the time to complete them?’ A list of 13 nursing care activities was presented and nurses asked to tick all that applied. The activities included were based on consistently recognised core components of nursing work and an existing instrument to assess ‘rationing’ of nursing care. These were:

- adequate patient surveillance
- adequate documentation of nursing care
- administering medication on time
- comfort/talk with patients
- develop or update nursing care plans/care pathways
- educating patients and/or family
- frequent changing of patient’s position
- oral hygiene
- pain management
- planning care
- preparing patients and families for discharge
- skin care
- undertaking treatments/procedures

Two measures of ‘missed care’ were derived. First, reported prevalence of any care being left undone, based on one or more of the activities having been ticked (binary measure). Second, a score indicating the volume of care left undone, by summing the number of activities ticked per person.

**DATA COLLECTION**

Questionnaires were distributed by local study coordinators to all wards in sufficient quantities for all staff identified as working on the ward to complete. Staff also had the option of completing the questionnaires online. Three reminders were sent (at approximately 2 weekly intervals): a postcard, full reminder pack, and final postcard. As the questionnaires were not distributed to named individuals, reminders were not targeted at non-responders. Data collection was from January 2010–September 2010 with the survey typically in the field at each site for 12 weeks.

**ANALYSIS**

The profile of nurses was described using summary statistics (means, SDs, frequencies and percentages). Descriptive statistics were used to describe the prevalence and nature of care left undone, and address the first research question. The average number of items missed during a shift were compared across groups (directorate, most recent shift worked, patient per RN and patients per healthcare support worker each grouped into quintiles) using analysis of variance. The proportion of shifts where nurses observed at least one item of missed care were compared across groups using the Pearson $\chi^2$ test. Associations between pairs of continuous variables were tested for statistical significance using Spearman’s correlation coefficient. Associations between a continuous and an ordinal variable were tested using the polyserial correlation coefficient. These tests were used to explore the relationship between missed care and quality and patient safety (the third research question). All the analyses mentioned so far were performed using SPSS V20.

The relationships between ‘missed care’ and other variables (staffing level and practice environment) were explored through multilevel regression models. To analyse the hierarchical cross-sectional design with nurses nested within wards, and wards within hospitals, a three-level multilevel model was fitted to the data using MLwiN, a statistical software package for fitting multilevel models using maximum likelihood estimation and Markov Chain Monte Carlo methods. Practice environment score (PES-28) and directorate (Surgical, Medical, Surgical/Medical) were ward level variables; while shift, patients per RN, patients per HCSW, patients requiring assistance with...
daily living (number of patients) and frequent monitoring (number of patients) were treated as nurse-level variables. These last two variables were included in order to control for variations in the required nursing intensity originating from variation in patient need. These independent variables were regressed onto the number of aspects of care missed (range 0–13), a global nominal dependent variable (no missed care vs one or more aspects of missed care) and 13 individual aspects of missed care.

To aid interpretation we grouped the staffing variables into quintiles. The model fitting included testing two possible interactions. The first was between PES-28 and patients per RN, to test for the potential additional effect on missed care that staffing levels and practice environment may have when taken together, compared with separately. Second, the potential interaction between RN staffing and HCSW staffing was explored to see if HCSW staffing complemented (ie, gave added value to) or substituted (ie, could be used to replace) RN staffing.

RESULTS
A total of 2917 responses were received from RNs in medical and surgical directorates (a subdivision of a hospital according to specialty). A further 73 responses from other directorates (which did not meet the criteria of general medical or surgical) were excluded from this analysis. Precise response rates cannot be calculated as we were unable to track how many questionnaires were actually distributed or how many nurses were actually working on the wards over the period studied. However, using the figures provided by hospitals of the number of RNs employed on the sampled wards, the response rate was estimated at at least 39%; it is likely to be higher due to some nurses being on annual or sick leave or some staff simply not receiving the questionnaire.

The characteristics of the sample are shown in table 1. Comparison with the profile of nurses obtained from an earlier national survey11 indicates that respondents are broadly typical of nurses working on NHS medical/surgical wards across the UK.

PREVALENCE AND NATURE OF CARE LEFT UNDONE
Across all respondents, 86% reported that on their last shift, at least 1 of the 13 care activities listed had been needed but not done due to lack of time. Nurses missed a mean of 4 items of care. The most common activities identified as missed were: comfort/talking with patients (66%), educating patients (52%) and developing or updating nursed care plans/care (47%). Pain management (7%) and treatment and procedures (11%) were least likely to be reported as missed.

Summary statistics for the number of items of missed care observed during a shift and the proportion of shifts where at least one item of missed care was observed by each independent variable are shown in table 2 based on cases with complete data from wards with two or more responding nurses.

More care was left undone on day and afternoon shifts than night shifts (p<0.001). A greater number of patients requiring assistance with daily living or frequent monitoring were associated with the number of care items left undone (Spearman’s r=0.23 and 0.18, respectively p<0.001) and having any missed care (7.66 (SD 5.76) vs 5.75(SD 5.49) t2805=6.00 p<0.001 and 3.73(SD 3.73) vs 2.87(SD 3.36) t2792=4.21, p<0.001, respectively). The better the practice environment score the fewer the care items left undone (r=-0.32, p<0.001) and higher scores were found when nurses said there was no missed care compared with some missed care (2.97 (SD 0.46) vs 2.74(SD 0.45) t2901=9.78, p<0.001).

CARE LEFT UNDONE AND STAFFING LEVELS
Staffing levels varied considerably among wards and hospitals. The average (mean) number of patients

---

Table 1 Profile of nurse respondents

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
<th>SD</th>
<th>n=</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age</td>
<td>39.6</td>
<td>10.1</td>
<td>2790</td>
</tr>
<tr>
<td>Under 25 years (&lt;25)</td>
<td>8%</td>
<td></td>
<td>221</td>
</tr>
<tr>
<td>25–34</td>
<td>25%</td>
<td></td>
<td>719</td>
</tr>
<tr>
<td>35–44</td>
<td>32%</td>
<td></td>
<td>931</td>
</tr>
<tr>
<td>45–54</td>
<td>24%</td>
<td></td>
<td>705</td>
</tr>
<tr>
<td>55 and over (55+)</td>
<td>8%</td>
<td></td>
<td>220</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>92%</td>
<td></td>
<td>2673</td>
</tr>
<tr>
<td>Male</td>
<td>8%</td>
<td></td>
<td>226</td>
</tr>
<tr>
<td>UK trained</td>
<td>83%</td>
<td></td>
<td>2413</td>
</tr>
<tr>
<td>Holds a bachelor’s degree in nursing</td>
<td>27%</td>
<td></td>
<td>785</td>
</tr>
<tr>
<td>Working hours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full time</td>
<td>76%</td>
<td></td>
<td>2218</td>
</tr>
<tr>
<td>Part time</td>
<td>22%</td>
<td></td>
<td>634</td>
</tr>
<tr>
<td>Last shift worked (and reported on)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day</td>
<td>56%</td>
<td></td>
<td>1630</td>
</tr>
<tr>
<td>Afternoon/evening</td>
<td>13%</td>
<td></td>
<td>390</td>
</tr>
<tr>
<td>Night</td>
<td>26%</td>
<td></td>
<td>769</td>
</tr>
<tr>
<td>Length of service (mean years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nursing career</td>
<td>13.8</td>
<td>10.6</td>
<td>2702</td>
</tr>
<tr>
<td>Current hospital</td>
<td>9.5</td>
<td>8.5</td>
<td>2704</td>
</tr>
<tr>
<td>Current specialty</td>
<td>7.7</td>
<td>7.0</td>
<td>2539</td>
</tr>
<tr>
<td>Current ward</td>
<td>5.8</td>
<td>5.6</td>
<td>2627</td>
</tr>
<tr>
<td>Job title</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staff nurse</td>
<td>73%</td>
<td></td>
<td>2139</td>
</tr>
<tr>
<td>Sister/charge nurse</td>
<td>24%</td>
<td></td>
<td>689</td>
</tr>
<tr>
<td>Other</td>
<td>2%</td>
<td></td>
<td>59</td>
</tr>
<tr>
<td>Directorate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical</td>
<td>47%</td>
<td></td>
<td>1384</td>
</tr>
<tr>
<td>Surgical</td>
<td>50%</td>
<td></td>
<td>1463</td>
</tr>
<tr>
<td>Medical/surgical</td>
<td>2%</td>
<td></td>
<td>70</td>
</tr>
</tbody>
</table>

BMJ Qual Saf 2013;0:1–10. doi:10.1136/bmjqs-2012-001767

---
nurses rated the practice environment as more positive when they are working on shifts with high numbers of patients per RN. The number of activities left undone was strongly related to nurses overall perceptions of the quality and safety of nursing care (polyserial correlation = 0.37, p < 0.001) and nurses overall grading of patient safety and comfort of patients (polyserial correlation = 0.37, p < 0.001). Comparing the ‘best’ and ‘worst’ practice environments (by taking the upper and lower deciles) the average number of items of care missed varied from 2.82 in the best practice environments compared with 5.61 for the poorest environments. Similarly the incidence of any care being missed (one item or more) was 79% in the best environments compared with 92% for the lowest decile.

HCSW staffing levels were not found to be associated with either the amount of missed care or the occurrence of any missed care reported by RNs (p < 0.05). Neither the interaction between PES and patients per RN (χ² = 2.738, 4 df, p = 0.602) nor between patients per RN and patients per HCSW (χ² = 21.811, 16 df, p = 0.149) were statistically significant.

RN staffing level was significantly associated with missed care for 8 of the 13 care activities (table 4, and see online supplementary appendix 1). The effect of staffing was strongest for ‘adequate patient surveillance’, ‘adequately documenting nursing care’ and ‘comforting/talking with patients’. Nurses working on shifts with the worst staffing (11.67 patients per RN) were twice as likely to report inadequate patient surveillance, when compared with those in the best staffed environments (less than 6.14 patients per RN). RN staffing level was not significantly associated with missed care in relation to frequent changing of position, administering medications on time, pain management, and preparing patients and families for discharge.

**Table 2** Summary statistics for prevalence and incidence of missed care

<table>
<thead>
<tr>
<th>No.</th>
<th>Mean</th>
<th>SD</th>
<th>%</th>
<th>(No.)</th>
</tr>
</thead>
</table>
| HCSW staffing levels were not found to be associated with either the amount of missed care or the occurrence of any missed care reported by RNs (p < 0.05). Neither the interaction between PES and patients per RN (χ² = 2.738, 4 df, p = 0.602) nor between patients per RN and patients per HCSW (χ² = 21.811, 16 df, p = 0.149) were statistically significant.

**DISCUSSION**

Most nurses working on general medical and surgical wards in this representative sample reported that some care was left undone on their last shift. Care that was frequently left undone included adequate patient surveillance, which has been hypothesised as a key mechanism explaining the association between low nurse staffing and increased mortality. The amount of care left undone was strongly related to nurses overall perceptions of the quality and safety of care.

**CARE LEFT UNDONE AND QUALITY AND PATIENT SAFETY**

There was a strong relationship between the number of items of missed care and nurses perception of quality of nursing care (polyserial correlation = −0.37, p < 0.001) and nurses overall grading of patient safety on their unit/ward (polyserial correlation = −0.40, p < 0.001) (table 5).
Table 3  Multilevel model for missed care

<table>
<thead>
<tr>
<th>Ward level variables</th>
<th>Number of items of missed care observed during a shift</th>
<th>Shifts where at least one item of missed care was observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>L95%</td>
</tr>
<tr>
<td>Directorate (χ², p value)*</td>
<td>(6.702, 0.035)</td>
<td></td>
</tr>
<tr>
<td>Surgical</td>
<td>−1.028</td>
<td>−1.902</td>
</tr>
<tr>
<td>Medical</td>
<td>−0.817</td>
<td>−1.695</td>
</tr>
<tr>
<td>Medical/surgical</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Practice Environment Scale (PES-28)</td>
<td>−2.726</td>
<td>−3.312</td>
</tr>
<tr>
<td>Shift (χ², p value)*</td>
<td>(32.545, &lt;0.001)</td>
<td></td>
</tr>
<tr>
<td>Day</td>
<td>0.866</td>
<td>0.564</td>
</tr>
<tr>
<td>Afternoon/evening</td>
<td>0.721</td>
<td>0.345</td>
</tr>
<tr>
<td>Night</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Assistance with daily living (no. patients)</td>
<td>0.049</td>
<td>0.025</td>
</tr>
<tr>
<td>Frequent monitoring(no. patients)</td>
<td>0.073</td>
<td>0.040</td>
</tr>
<tr>
<td>Nurse staffing variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients per RN (χ², p value)†</td>
<td>(36.296, &lt;0.001)</td>
<td></td>
</tr>
<tr>
<td>up to 6.13</td>
<td>−1.087</td>
<td>−1.501</td>
</tr>
<tr>
<td>6.14–7.33</td>
<td>−0.427</td>
<td>−0.839</td>
</tr>
<tr>
<td>7.40–9.25</td>
<td>−0.201</td>
<td>−0.595</td>
</tr>
<tr>
<td>9.33–11.50</td>
<td>−0.121</td>
<td>−0.488</td>
</tr>
<tr>
<td>11.67 and over</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Patients per HCSW (χ², p value)†</td>
<td>(4.51, 0.48)</td>
<td></td>
</tr>
<tr>
<td>up to 6.80</td>
<td>0.288</td>
<td>−0.098</td>
</tr>
<tr>
<td>7.00–9.25</td>
<td>0.284</td>
<td>−0.106</td>
</tr>
<tr>
<td>9.33–13.00</td>
<td>0.305</td>
<td>−0.058</td>
</tr>
<tr>
<td>13.33–17.00</td>
<td>0.261</td>
<td>−0.102</td>
</tr>
<tr>
<td>17.33 and over</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Random variance</td>
<td>σ²</td>
<td>L95%</td>
</tr>
<tr>
<td>Hospitals (n=46)</td>
<td>0.128</td>
<td>0.000</td>
</tr>
<tr>
<td>Wards (n=392)</td>
<td>0.482</td>
<td>0.231</td>
</tr>
<tr>
<td>Nurses (n=2566)</td>
<td>7.541</td>
<td>7.096</td>
</tr>
</tbody>
</table>

* Two degrees of freedom.
† Four degrees of freedom.
‡ Two degrees of freedom.

Table 4  Quality of nursing care and patient safety compared with missed care score

<table>
<thead>
<tr>
<th>Quality of nursing care delivered to patients on your unit/ward*</th>
<th>Number of items</th>
<th>Mean ± SD</th>
<th>No.</th>
<th>%</th>
<th>Mean ± SD</th>
<th>No.</th>
<th>%</th>
<th>Mean ± SD</th>
<th>No.</th>
<th>%</th>
<th>Mean ± SD</th>
<th>No.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain management</td>
<td>36</td>
<td>6</td>
<td>42</td>
<td>8</td>
<td>38</td>
<td>7</td>
<td>50</td>
<td>9</td>
<td>29</td>
<td>5</td>
<td>195</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>Treatments and procedures</td>
<td>46</td>
<td>8</td>
<td>57</td>
<td>11</td>
<td>65</td>
<td>12</td>
<td>72</td>
<td>13</td>
<td>66</td>
<td>12</td>
<td>306</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>Skin care</td>
<td>78</td>
<td>13</td>
<td>120</td>
<td>23</td>
<td>118</td>
<td>22</td>
<td>135</td>
<td>24</td>
<td>133</td>
<td>25</td>
<td>584</td>
<td>21</td>
<td>7</td>
</tr>
<tr>
<td>Administer medications on time</td>
<td>101</td>
<td>17</td>
<td>121</td>
<td>23</td>
<td>116</td>
<td>22</td>
<td>136</td>
<td>24</td>
<td>156</td>
<td>29</td>
<td>630</td>
<td>23</td>
<td>9</td>
</tr>
<tr>
<td>Planning care</td>
<td>117</td>
<td>19</td>
<td>154</td>
<td>29</td>
<td>142</td>
<td>27</td>
<td>183</td>
<td>32</td>
<td>170</td>
<td>31</td>
<td>766</td>
<td>28</td>
<td>7</td>
</tr>
<tr>
<td>Adequately document nursing care</td>
<td>134</td>
<td>22</td>
<td>190</td>
<td>36</td>
<td>199</td>
<td>38</td>
<td>204</td>
<td>36</td>
<td>193</td>
<td>36</td>
<td>920</td>
<td>33</td>
<td>3</td>
</tr>
<tr>
<td>Adequate patient surveillance</td>
<td>135</td>
<td>22</td>
<td>169</td>
<td>32</td>
<td>195</td>
<td>37</td>
<td>229</td>
<td>40</td>
<td>237</td>
<td>44</td>
<td>965</td>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>Educating patients and family</td>
<td>268</td>
<td>44</td>
<td>276</td>
<td>53</td>
<td>280</td>
<td>53</td>
<td>340</td>
<td>59</td>
<td>284</td>
<td>52</td>
<td>1448</td>
<td>52</td>
<td>5</td>
</tr>
</tbody>
</table>

In general, how would you describe the quality of nursing care...
(in this case HCSWs) act as a complement for the work of another group (in this case RNs) there is an interaction effect whereby an increase in the complement workforce increases the effect of the other group on outputs (in this case the inverse of missed care). However no such interaction was observed. While we did not study all the potential work of RNs, this finding does not support an increase in the number of HCSWs as a means of increasing the efficiency of RNs.

The desirability of increasing nurse staffing levels as a means to improve quality is contested on grounds other than cost. There is debate internationally about setting standard minimum staffing levels, but this policy is often resisted on the basis that it is inflexible and might stifle innovation in workforce planning. Evidence from this study, as elsewhere, suggests that attention should be paid to the quality of the practice environment as a potentially lower cost approach to improving the quality and efficiency of nursing work. Many of the constructs of the nursing practice environment are consistent with West et al’s high performance human resource management system including training, performance management, participation, decentralisation, involvement, use of teams and employment security which were related to lower risk adjusted mortality rates in a study in English NHS hospitals. How workforce planners can redress the balance by improving the practice environment in the face of staffing reductions and resultant lack of job security is unclear.

Table 4 Missed care by levels of registered nurse staffing

<table>
<thead>
<tr>
<th>Overall number of missed care aspects</th>
<th>Patients per registered nurse</th>
<th>Mean No. SD</th>
<th>Mean % No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>One or more aspects of missed care</td>
<td></td>
<td>3.06 2.98</td>
<td>4.13 3.17</td>
</tr>
<tr>
<td>Comfort/talk with patients</td>
<td></td>
<td>3.45 66</td>
<td>5.65 80</td>
</tr>
<tr>
<td>Educating patients and family</td>
<td></td>
<td>1.94 55</td>
<td>3.54 90</td>
</tr>
<tr>
<td>Develop or update nursing care plans/care pathways</td>
<td></td>
<td>2.53 48</td>
<td>3.88 61</td>
</tr>
<tr>
<td>Adequate patient surveillance</td>
<td></td>
<td>3.77 59</td>
<td>5.34 80</td>
</tr>
<tr>
<td>Adequately document nursing care</td>
<td></td>
<td>3.45 66</td>
<td>5.65 80</td>
</tr>
<tr>
<td>Oral hygiene</td>
<td></td>
<td>2.53 48</td>
<td>3.88 61</td>
</tr>
<tr>
<td>Frequent changing of patient position</td>
<td></td>
<td>4.19 59</td>
<td>5.34 80</td>
</tr>
<tr>
<td>Planning care</td>
<td></td>
<td>4.19 59</td>
<td>5.34 80</td>
</tr>
<tr>
<td>Administer medications on time</td>
<td></td>
<td>5.34 80</td>
<td>6.55 90</td>
</tr>
<tr>
<td>Skin care</td>
<td></td>
<td>3.88 61</td>
<td>5.12 75</td>
</tr>
<tr>
<td>Prepare patients and families for discharge</td>
<td></td>
<td>2.53 48</td>
<td>3.17 55</td>
</tr>
<tr>
<td>Treatments and procedures</td>
<td></td>
<td>2.53 48</td>
<td>3.17 55</td>
</tr>
<tr>
<td>Pain management</td>
<td></td>
<td>2.53 48</td>
<td>3.17 55</td>
</tr>
</tbody>
</table>

Table 5 Quality of nursing care and patient safety compared with missed care score

<table>
<thead>
<tr>
<th>Rating/grade</th>
<th>No.</th>
<th>Mean missed care score</th>
<th>95% CI Lower</th>
<th>95% CI Upper</th>
<th>Average (mean) no. patients per RN on day shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of nursing care delivered to patients on unit/ward*</td>
<td>Poor</td>
<td>66</td>
<td>2</td>
<td>8.08</td>
<td>7.17</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td>473</td>
<td>16</td>
<td>5.44</td>
<td>5.17</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>1455</td>
<td>50</td>
<td>4.02</td>
<td>3.88</td>
</tr>
<tr>
<td></td>
<td>Excellent</td>
<td>904</td>
<td>31</td>
<td>2.78</td>
<td>2.60</td>
</tr>
<tr>
<td>Overall grade for patient safety on unit/ward†</td>
<td>Failing</td>
<td>45</td>
<td>2</td>
<td>7.78</td>
<td>6.82</td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>146</td>
<td>5</td>
<td>6.54</td>
<td>6.03</td>
</tr>
<tr>
<td></td>
<td>Acceptable</td>
<td>870</td>
<td>30</td>
<td>4.88</td>
<td>4.69</td>
</tr>
<tr>
<td></td>
<td>Very Good</td>
<td>1321</td>
<td>46</td>
<td>3.54</td>
<td>3.39</td>
</tr>
<tr>
<td></td>
<td>Excellent</td>
<td>512</td>
<td>18</td>
<td>2.37</td>
<td>2.15</td>
</tr>
</tbody>
</table>

*Participants were asked to tick either: Poor, Fair, Good or Excellent in response to ‘In general, how would you describe the quality of nursing care delivered to patients on your unit/ward?’

†Participants were asked to tick either: Failing, Poor, Acceptable, Very Good or Excellent in response to ‘Please give your unit/ward an overall grade on patient safety’.

RN, registered nurse.
As the enquiry into excess mortality at the Mid Staffordshire NHS Trust in England critically shows the consequences of poorly informed experiments to improve the efficiency of the nursing workforce can be disastrous. Most measures used to detect problems, such as standardised mortality rates, are ‘lagging indicators’ revealing problems after they have happened. Low staffing levels and poor practice environment have already been identified as potential indicators of poor quality and are endorsed by a number of bodies, for example, the US National Quality Forum and the American Nurses Association. However missed care is a more direct indicator of quality deficiencies with a clear pathway to adverse patient outcomes and experience. Hence, missed care has the potential to be used as a leading indicator, identifying emerging problems before serious consequences occur, enabling employers, regulators or others to identify wards where workload/staffing mismatches are putting patients at risk. Further research is warranted to determine whether routine reporting on missed care can be used in this way.

Limitations
Our use of a cross-sectional survey design allows us to draw inferences about the possible nature and prevalence of missed care but a limitation of the study is that the missed care measure is generated through nurses’ accounts. The measure is therefore open to the subjective experiences of individual nurses, who may understand specific items differently (eg, ‘adequate patient surveillance’—which may vary according to ward layout) and hold different expectations and perceptions of what level and type of care is needed and whether or not it was provided. They may also have different interpretations as to the extent to which an activity was not done was due to ‘lack of time to complete’. To some extent we were able to limit variation by asking about 13 specific activities (rather than using an open-ended question about the type of activities that were missed). Other research shows that nurses’ rating of quality closely aligns to objective measures of patient outcomes. We do not know whether nurses handed over responsibility for care that they themselves might have missed at the end of a shift, or whether this care was done later by another nurse.

The measure gauges differences in nurses’ perceptions of the amount of work undone over a standard length shift, but it does not relate this to the total work required, nor does it relate to care done or undone for specific numbers of patients. Future research could usefully seek to examine in more detail whether care ‘left undone’ was unfinished, rushed or not done to a high standard, or whether it was missed entirely, and place this in a context of the total volume of care being undertaken for patients.

A further limitation is that we have taken nurses’ reports of the staffing and patients on their last shift, to produce a measure of average staffing levels. However we do not know how the grade mix of nursing staff varied (for either registered or non-registered), nor the level of temporary staff (bank or agency) that were on duty. Both of which may affect the productivity of the nursing team as a whole, and have an impact on care being left undone. Further research is needed to move beyond establishing an association between overall staffing levels and care being left undone, to explore in more detail the effect of different combinations of staff with different qualifications and experience, on the productivity of the nursing team as a whole.

CONCLUSIONS
RNs working in English NHS hospitals report that care is needed but is often not done because of insufficient time. There is a strong relationship between RN staffing levels and the prevalence of care being left undone—and, the better the practice environment the smaller the volume of care that is left undone. Greater research attention to the impact of ‘missed care’ is needed. A ‘missed care’ measure may be a useful correlate of nursing care quality, and inform staffing decisions at ward level. Further research is needed to test the measure against patient outcomes, and to support comparability between care settings and internationally.

Acknowledgements This paper draws on a study that was funded by the European Union 7th framework, and is part of the larger international RN4Cast project, led by Prof Walter Sermeus and Prof Linda Aiken, in association with the RN4Cast consortium (15 countries). The research in England was led by the National Nursing Research Unit (NNRU) at the Florence Nightingale School of Nursing and Midwifery, King’s College, London. Principal Investigators of the study are Prof Anne Marie Rafferty and Prof Peter Griffiths (formerly King’s College London, now at the University of Southampton). The ethical approval to undertake the study in England was provided by the National Research Ethics Committee. Individual hospitals (Trusts) permission and approval for taking part in the study was secured through the Integrated Research Application System. The survey of nurses was managed and administered by Employment Research Limited, Hove (ERL).

Contributors Each of the authors listed on this paper have contributed to some (or all) of: the conception and design, acquisition of data, analysis and interpretation of data. They have all been involved in drafting the article and made contributions to revising it, and have contributed to the intellectual content. Each has given their approval of the amendments made and has approved this final version submitted for publication. I declare that they have participated sufficiently in the work to take public responsibility for appropriate portions of the content. No one else who fulfils these criteria has been excluded as an author.

Funding The European Union Seventh Framework (Grant Number HEALTH-F2-2009-223468) has funded the research and the coordination of the consortium.

Competing interests None.

Ethics approval NHS REC (ref 09/H0808/69)

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non
REFERENCES

3. Health Service Ombudsman. Care and compassion. 2011, A report to the Health Service Ombudsman on ten investigations into NHS care for older people.


Allocating scarce resources in real-time to reduce heart failure readmissions: a prospective, controlled study

Ruben Amarasingham,1,2 Parag C Patel,3 Kathleen Toto,2 Lauren L Nelson,2 Timothy S Swanson,1 Billy J Moore,1 Bin Xie,1 Song Zhang,4 Kristin S Alvarez,5 Ying Ma,1 Mark H Drazner,2 Usha Kollipara,6 Ethan A Halm2

ABSTRACT
Objective To test a multidisciplinary approach to reduce heart failure (HF) readmissions that tailors the intensity of care transition intervention to the risk of the patient using a suite of electronic medical record (EMR)-enabled programmes.

Methods A prospective controlled before and after study of adult inpatients admitted with HF and two concurrent control conditions (acute myocardial infarction (AMI) and pneumonia (PNA)) was performed between 1 December 2008 and 1 December 2010 at a large urban public teaching hospital. An EMR-based software platform stratified all patients admitted with HF on a daily basis by their 30-day readmission risk using a published electronic predictive model. Patients at highest risk received an intensive set of evidence-based interventions designed to reduce readmission using existing resources. The main outcome measure was readmission for any cause and to any hospital within 30 days of discharge.

Results There were 834 HF admissions in the pre-intervention period and 913 in the post-intervention period. The unadjusted readmission rate declined from 26.2% in the pre-intervention period to 21.2% in the post-intervention period (p=0.01), a decline that persisted in adjusted analyses (adjusted OR (AOR)=0.73; 95% CI 0.58 to 0.93, p=0.01). In contrast, there was no significant change in the unadjusted and adjusted readmission rates for PNA and AMI over the same period. There were 45 fewer readmissions with 913 patients enrolled and 228 patients receiving intervention, resulting in a number needed to treat (NNT) ratio of 20.

Conclusions An EMR-enabled strategy that targeted scarce care transition resources to high-risk HF patients significantly reduced the risk-adjusted odds of readmission.

INTRODUCTION
A majority of US hospitals struggle to contain readmission rates related to heart failure (HF).1 2 Although numerous studies have found that some combination of careful discharge planning, provider coordination and intensive counselling can prevent subsequent readmissions to a hospital, success is difficult to achieve and sustain at the typical US hospital.1-4 Enrolling all patients with HF into a uniform high-intensity care transition programme (‘do everything for everyone’) may require a depth of case management resources out of reach for many institutions, particularly safety-net hospitals. It has been hypothesised that programmes to reduce readmissions may be more effective if resources are applied differentially according to a patient’s relative risk of readmission.5 6 To our knowledge, this hypothesis has not been tested.

We have previously described an electronic medical record (EMR)-enabled model (e-Model) derived from both clinical and non-clinical factors which accurately stratifies risk for 30-day readmission among patients with HF.7 Compared with other risk prediction models, the e-Model is unique in that it draws from 29 clinical, social, behavioural and utilisation factors extracted in real time from the EMR within 24 h of admission for HF.7

In concert with other EMR-based tools, the e-Model makes it possible to match the intensity of the readmission intervention to the patient’s risk of readmission on any given day. This study examines
whether a strategy targeting highest-risk HF patients by re-allocating existing hospital resources could reduce risk-adjusted rates of readmission. To control for secular trends, we compare concurrent rates of 30-day readmission for patients with acute myocardial infarction (AMI) and pneumonia (PNA), two measures that are also the subject of intense national scrutiny but for which no intervention was performed.

**METHODS**

**Setting, study design and participants**

This study was constructed as a pragmatic trial employing a prospective before and after design with concurrent controls. The study was performed at the main hospital of Parkland Health and Hospital System, a 780-bed tertiary care teaching hospital located in Dallas, Texas. The hospital serves an ethnically diverse safety-net patient population. The study was divided into two periods of equal length: the pre-intervention period (1 December 2008 to 30 November 2009) and the post-intervention period (1 December 2009 to 30 November 2010). Readmission data were collected 1 month after the end of the respective periods to ascertain the total 30-day readmission rate for that period. During the 2-year study period, detailed clinical information was collected from the hospital’s EMR platform (EPIC Systems Corporation, Verona, Wisconsin, USA) for all hospitalised patients aged ≥18 years discharged with a principal diagnosis of HF, AMI and PNA, according to the Centers for Medicare and Medicaid Services ICD-9 code groupers.8–11 The intervention described below was implemented at the end of the pre-intervention period and achieved full operational capacity immediately after going ‘live’ as the interventional staff had been trained and prepared for several weeks prior to implementation.

**Interventional overview**

Patient selection and the treatment strategy are presented in figure 1, steps A–J. Like many hospitals, a specialised but limited number of care transition and HF management resources were available for all patients with HF in the system during daytime hours Monday to Friday. This included a hospital-wide case management department and an outpatient HF programme staffed by a limited number of cardiologists, nurse practitioners, nutritionists and pharmacists who maintained other clinical responsibilities beyond follow-up of discharged HF patients. No personnel were added for the intervention with the exception of a part-time case manager, at 25% effort, to assist existing hospital case management staff with social work and care transition needs for patients with HF identified as high-risk. The intervention was thus constructed to best allocate this finite set of resources from Monday to Friday to maximally reduce 30-day readmissions associated with HF.

The intervention was driven by a suite of computerised case monitoring and coordination programmes (software programme) that: (1) extracts real-time data from the EMR (step A, figure 1); (2) identifies patients admitted with HF using free text and structured data text processing approaches (step B); (3) ranks orders and risk stratifies patients according to the e-Model within 24 h of hospitalisation (step C); and (4) sends secure electronic notifications about those at highest risk to HF personnel (step F) who then evaluate and perform an intensive bundle of coordinated inpatient and outpatient (post-discharge) tasks on eligible patients (steps F–I). The e-Model is a real-time

---

**Figure 1** Patient selection and review process.
intervention in the post-intervention period.

**Intervention components**

High-risk patients received an intensive set of evidence-based inpatient and outpatient counselling and monitoring activities drawn from well-known readmission reduction strategies that have been extensively published (steps H and I, figure 1). These included (1) detailed inpatient clinical assessment, patient education and discharge planning by a HF nurse practitioner, pharmacist, nutritionist and case manager starting early in the hospital course; (2) a follow-up telephone call from a nurse within 48 h of discharge to assess whether the patient had obtained their medication and was aware of their outpatient follow-up appointments; (3) outpatient case management (consisting of individualised care management services based on specific post-discharge needs) for 30 days; (4) a cardiology appointment with a HF specialist within 7 days of discharge and subsequent cardiology follow-up for at least 1 month; and (5) a primary care appointment scheduled according to the urgency of non-cardiac problems. While some of the intervention components such as follow-up call and outpatient care management were in place in the pre-intervention period as part of usual care, the level of intensity and involvement of the outpatient case management staff was higher for patients selected for intervention in the post-intervention period.

On each day the intervention was in effect, patients with the highest predicted risk were selected for intervention. Patients selected for intervention were encouraged to complete all intervention elements but were free to decline any. Patients with HF who were not selected for intervention received the treatment plan directed by their primary medical team as part of usual care.

A weekly review session was held to review the fidelity of the intervention and to ensure that the study protocol was faithfully carried out.

**Study outcomes and variables**

The primary outcome of this study was readmission to any hospital in the Dallas-Fort Worth metroplex for any cause within 30 days of discharge of the index HF admission. We ascertained 30-day readmission to any hospital in the North Texas region using a probabilistic matching service available through the Dallas-Fort Worth Hospital Council, a cooperative regional information-sharing initiative. Elective readmissions classified by clinical staff and coded in the EMR system were not counted as readmissions in this analysis.

In order to determine whether the intervention had any adverse mortality effects, we tracked inpatient and 30-day mortality rates for all patients in the pre- and post-intervention period for HF, AMI and PNA using a two-step process. Those patients with a documented encounter after the 30-day post-discharge period anywhere in the index health system were considered alive. Those without an encounter were subsequently classified as dead or alive within 30 days of discharge after querying the National Death Index File.

We also collected information on demographics, clinical severity and comorbid illness burden as well as several patient and neighborhood-level measures of social disadvantage included in the e-Model. Variables that were highly significant to the prognostic capability of the e-Model were also extracted from the EMR. These variables included, among others, Brain Natriuretic Peptide (BNP), troponin, creatine kinase, blood urea nitrogen and albumin and key social and behavioural readmission risk factors such as gender, marital status, payor status, number of documented home address changes within the past 12 months, history of positive urine cocaine within the past 12 months, history of missed clinic visits within the past 6 months and number of hospital admissions prior to index admissions. The software program used for this intervention allowed us to track the receipt and completion of all intervention components.

**Statistical analysis**

A quasi-experimental approach was used to assess the impact of the targeted intervention on the overall readmission rate of HF. We used a measurement framework that included all patients with a final principal discharge diagnosis of HF during the pre- and post-intervention periods. In other words, the entire HF population was used as the basis to analyse the readmission rate irrespective of whether they were classified as HF at the point of intervention by the electronic or human review process.

Readmission and mortality rates were assigned to calendar months and calculated according to methods outlined by CMS. Differences in demographics, clinical severity of illness, social risk factors and overall readmission risk between the pre- and post-intervention cohorts were tested using χ² and t tests.

Illness severity was based on the Tabak inpatient mortality model and overall readmission risk scores for
HF were calculated according to the e-Model. The Tabak score was used because it can be computed in real time using data readily available from an EMR within 24 h of admission, and was a more precise assessment of physiological status, in our view, than other measures that calculate in-hospital mortality prediction. A subset of the e-Model risk score, called the social risk score, consisted of variables in the e-Model directly related to social, utilisation and environmental risk factors. The change in unadjusted readmission rates between the pre- and post-intervention periods was analysed at the monthly level by the Wilcoxon matched pairs test. To adjust for potential differences in patient populations, a mixed logistic regression model was constructed that included demographic and hospital utilisation variables, a case mix variable, comorbidity variables and an indicator of the pre- and post-intervention periods.

The case mix variable for the HF cohort was calculated using the Tabak HF mortality score while, for the control population, the Tabak mortality risk scores for AMI or PNA were used. Patient comorbidities were captured through indicator variables for coronary artery disease, chronic kidney disease, diabetes mellitus and cancer. The intervention effect was represented by the regression coefficient of the indicator variable of the intervention period. The model also included random effects to account for potential autoregressive correlation over time and to address potential seasonal variations or other secular trends, as well as within-subject correlation for those patients who had more than one index admission during the study period. We calculated the observed to expected (OE) readmission rate ratio to measure the performance of intervention components of special interest. Expected readmission rates were calculated based on the e-Model.

For all analyses, the null hypothesis was evaluated at a two-sided significance level of 0.05. All statistical analyses were performed with STATA V10.0 (STATA Corp, College Station, Texas, USA).

Sensitivity analysis
Because patients who were admitted and discharged within a weekend or holiday were excluded from the possibility of receiving the intervention during the intervention time period, we performed a sensitivity analysis in which patients admitted and discharged during these periods were dropped from both the pre-intervention and intervention time periods. Because only the patients with the highest predicted risk on a given day received the intervention, we performed a sensitivity analysis to compare the predicted readmission risk and observed readmission rates of the next most deserving patient (defined as the patient with the highest predicted risk among the patients not receiving the intervention in each day when the intervention was in effect) and, because the intervention was implemented in December 2009, we also performed a sensitivity analysis that excluded the month of December 2009 in the post-intervention period.

RESULTS
Study population characteristics
A total of 834 patients were discharged with a principal diagnosis of HF in the pre-intervention period and 913 in the post-intervention period. The mean age of the HF population was 58 years. The pre- and post-intervention groups overall were comparable with regard to important readmission risk factors including the e-Model risk score, pro-BNP levels, length of stay and measures of prior utilisation (table 1), although there were some noticeable differences in age, gender, race and ethnicity.

After further exploratory analysis, we found that the differences in race, gender and age correlated with a noticeable increase in the absolute number of female Hispanic patients with HF in the post-intervention period compared with the pre-intervention period, with no noticeable increase in any other demographic subgroup.

There were 637 patients discharged with a principal diagnosis of either AMI or PNA in the pre-intervention period and 597 in the post-intervention period. The mean age of the concurrent control group was the same as the HF group (n=58). There was a relatively equal distribution of black, Hispanic and Caucasian patients in the control group in the pre- and post-intervention periods. As with the HF group, the mean disease-specific Tabak mortality scores were nearly identical in the pre- and post-intervention period for both conditions (18.9 vs 19.0 for AMI and 26.2 vs 26.1 for PNA).

Thirty-day readmission rates for HF
The 30-day readmission rate for HF, which included all readmissions in our region, was 23.6% over the entire study period. Of these, 26% represented readmissions to other hospitals. Table 2 shows that the unadjusted 30-day readmission rate for HF fell from 26.2% in the pre-intervention period to 21.2% in the post-intervention period (p<0.01), a relative reduction of 19%. We enrolled 913 patients and 228 received intervention in the post-intervention period, and there were 43 fewer readmissions in the post-intervention period (ie, 913 enrolled multiplied by the difference between the pre- and post-readmission rates), resulting in a number needed to treat ratio of 20.29, suggesting a fairly robust treatment effect.

The monthly paired difference between the pre- and post-intervention period also showed consistent reduction (p<0.01). After controlling for demographics, mortality risk and comorbidities in the mixed logistic model, we detected a significant reduction in
readmission rate (adjusted OR (AOR) 0.73, 95% CI 0.58 to 0.93, p=0.01) in the post-intervention period. Figure 2 shows that the monthly post-intervention readmission rate was consistently below the pre-intervention average of 26.2% in 11 of 12 months.

### Thirty-day readmission rates for AMI and PNA

The combined populations of AMI and PNA served as the concurrent control group. As shown in table 2, there was no difference in the unadjusted pre- and post-intervention readmission rates for patients hospitalised with AMI and PNA (15.5% vs 16.7%; p=0.56), and no significant differences in the readmission rate were detected by the mixed logistic regression models (AOR=1.09, 95% CI 0.80 to 1.48).

### Thirty-day mortality analyses

Since a higher 30-day death rate could theoretically lead to a lower 30-day readmission rate, we examined the mortality rates for HF pre- and post-intervention. There was no statistically significant difference in mortality in the pre- and post-intervention periods (3.2% vs 2.2%, p=0.16). It is therefore unlikely that the

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Pre-intervention (N=834)</th>
<th>Post-intervention (N=913)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e-Model (readmission risk) score, mean (SD)*</td>
<td>25.3 (0.15)</td>
<td>25.2 (0.16)</td>
<td>0.80</td>
</tr>
<tr>
<td>Demographic factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, mean (SD)</td>
<td>59.2 (11.4)</td>
<td>57.8 (10.8)</td>
<td>0.01</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>506 (60.7)</td>
<td>504 (55.2)</td>
<td>0.02</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>519 (62.4)</td>
<td>512 (56.1)</td>
<td>0.01</td>
</tr>
<tr>
<td>Hispanic</td>
<td>173 (20.7)</td>
<td>250 (27.4)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>White</td>
<td>133 (15.8)</td>
<td>136 (14.9)</td>
<td>0.59</td>
</tr>
<tr>
<td>Other</td>
<td>9 (1.1)</td>
<td>15 (1.6)</td>
<td>0.31</td>
</tr>
<tr>
<td>Single</td>
<td>605 (72.5)</td>
<td>624 (68.3)</td>
<td>0.06</td>
</tr>
<tr>
<td>Payor, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicare</td>
<td>264 (31.6)</td>
<td>254 (27.8)</td>
<td>0.08</td>
</tr>
<tr>
<td>Commercial</td>
<td>12 (1.4)</td>
<td>11 (1.2)</td>
<td>0.67</td>
</tr>
<tr>
<td>Medicaid and other</td>
<td>558 (66.9)</td>
<td>648 (71.0)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Clinical factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro-BNP, mean (SD)</td>
<td>8167 (11 564)</td>
<td>9362 (13 245)</td>
<td>0.10</td>
</tr>
<tr>
<td>Proportion with diabetes mellitus comorbidity†</td>
<td>49.3%</td>
<td>53.7%</td>
<td>0.06</td>
</tr>
<tr>
<td>Proportion with coronary artery disease comorbidity†</td>
<td>45.0%</td>
<td>38.4%</td>
<td>0.01</td>
</tr>
<tr>
<td>Proportion with chronic kidney disease comorbidity†</td>
<td>41.4%</td>
<td>42.9%</td>
<td>0.52</td>
</tr>
<tr>
<td>Utilisation or behavioural factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admitted through ED for index admission (%)</td>
<td>717 (86.0)</td>
<td>806 (88.3)</td>
<td>0.15</td>
</tr>
<tr>
<td>No. of prior inpatient admissions, mean (SD)</td>
<td>1.47 (2.5)</td>
<td>1.40 (2.2)</td>
<td>0.48</td>
</tr>
<tr>
<td>History of missed clinic visit within previous 6 months</td>
<td>96 (11.5)</td>
<td>86 (9.4)</td>
<td>0.15</td>
</tr>
<tr>
<td>History of leaving against medical advice (%)</td>
<td>20 (2.4)</td>
<td>15 (1.6)</td>
<td>0.26</td>
</tr>
<tr>
<td>Length of stay for index admission (days)</td>
<td>6.0 (6.1)</td>
<td>5.7 (5.3)</td>
<td>0.41</td>
</tr>
</tbody>
</table>

*The e-Model is an EMR-derived multivariate readmission risk model generated from the following variables: age, Tabak HF mortality score, gender, marital status, payor status, number of documented home address changes, history of positive urine cocaine, history of missed clinic visits, number of hospital admissions prior to index admissions.

†As defined by ICD-9 codes (see online supplementary appendix for list of codes).

BNP, B-type natriuretic peptide; ED, emergency department; EMR, electronic medical record; HF, heart failure.

---


---

Figure 2 Thirty-day readmission rates by month.
observed reduction in HF readmission was driven by the change in death rate.

Intervention intensity and odds of readmission
Planned intervention elements may not have been completed because of refusal by patients or logistical problems so a subgroup analysis was conducted on the 228 patients with HF who received at least one intervention element. We stratified the intervention group by the number of outpatient components received into two categories: 1–2 and ≥3. The receipt of a larger number of outpatient components was associated with a substantially lower readmission risk (table 3). Patients who received ≥3 outpatient components had an OE ratio for readmissions of 0.35 (95% CI 0.00 to 0.85; p<0.01).

Sensitivity analysis
When we excluded patients who were admitted and discharged during a weekend and holiday period in both the pre-intervention and post-intervention periods, a significant reduction in readmission remained (AOR 0.76, p=0.02), indicating that this aspect of the protocol was not responsible for the observed reduction in readmissions. Similarly, patients with AMI and PNA who were admitted and discharged during a weekend and holiday period were removed from the pre-intervention and post-intervention period in an equivalent sensitivity analysis. No change was noted in the rate of readmissions (AOR 0.97, p=0.86). For the next most deserving patients who were not selected for intervention, there was no statistically significant reduction in readmission rate, with an OE ratio of 1.03 (95% CI 0.89 to 1.17), suggesting that only patients who received the intervention experienced a reduction in readmission rates. When we excluded the month of December 2009 from the post-intervention period, the results were very similar, indicating that the intervention programme was fully functional during the first month of implementation.

DISCUSSION
In this study we found that a care transition intervention that directed largely existing resources to a smaller subgroup of patients with HF based on daily EMR-based risk stratification produced a clinically meaningful reduction in overall readmissions. By concentrating care management efforts on about one-quarter of patients with HF we were able to demonstrate a 26% relative reduction in the odds of readmission and an absolute reduction of 5.0 readmissions per 100 index admissions.

Since other institution-wide efforts to improve quality could also theoretically reduce readmissions, we examined time trends in readmission for patients with HF compared with a concurrent cohort of individuals admitted with AMI and PNA. We observed no change in the readmission rate for these two conditions, suggesting that the improvements in HF were specific to the intervention rather than institutional secular trends that would have affected overall readmission rates. While national programmes, regulations and policies designed to reduce HF readmissions nationwide could have hypothetically affected readmission rates at this institution, recent literature reveals that there has been little change nationally.1

The high-intensity care transition interventions in this study employed methods tested in several nationally emulated readmission reduction programmes and achieved a comparable reduction in readmission.6 12–16 However, in contrast to these approaches, this programme targeted approximately one-quarter of all admitted patients with HF and used predominantly existing care transition resources, suggesting that a more targeted approach is feasible and effective. Our

Table 2 Thirty-day readmission rates: pre-intervention and post-intervention periods* for all patients by HF and control groups

<table>
<thead>
<tr>
<th>Patient type</th>
<th>Pre-intervention (%)</th>
<th>Post-intervention (%)</th>
<th>Difference (95% CI)</th>
<th>p Value</th>
<th>Adjusted OR† (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF</td>
<td>All</td>
<td>26.2</td>
<td>21.2</td>
<td>5.0 (1.0 to 9.0)</td>
<td>0.01</td>
<td>0.73 (0.58 to 0.93)</td>
</tr>
<tr>
<td>AMI and PNA</td>
<td>All</td>
<td>15.5</td>
<td>16.7</td>
<td>−1.2 (−5.4 to 2.9)</td>
<td>0.56</td>
<td>1.09 (0.80 to 1.48)</td>
</tr>
</tbody>
</table>

†The risk adjustment model includes demographic variables (age, race and sex), case mix variables and indicator variables for the pre- and post-intervention periods. For HF, the case mix variable was the e-Model. For AMI and PNA, the Tabak mortality score for AMI and PNA was used. AMI, acute myocardial infarction; HF, heart failure; PNA, pneumonia.

Table 3 Outpatient intervention and odds of readmission (n=913 HF index admissions, n=228 patients receiving interventions)*

<table>
<thead>
<tr>
<th>Intervention category</th>
<th>N</th>
<th>Expected readmission rate†</th>
<th>Observed readmission rate</th>
<th>Observed/expected ratio (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outpatient intervention completion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrolled, received 1 or 2 outpatient components</td>
<td>150</td>
<td>23.8</td>
<td>17.3</td>
<td>0.73 (0.51 to 0.95)</td>
<td>0.01</td>
</tr>
<tr>
<td>Enrolled, received ≥3 outpatient components</td>
<td>27</td>
<td>21.2</td>
<td>7.4</td>
<td>0.35 (0.00 to 0.85)</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

*Intervention components include direct care and care management received from nurse practitioner, pharmacist, nutritionist and/or home visit nurse.
†Expected readmission risk as calculated by electronic readmission risk model (e-Model).
HF, heart failure.
results also provided a flexible framework that can be adapted in institutions with different levels of resource availability. Hospitals can adjust the threshold value of intervention (ie, the cut-off value or rank order of predicted risk to receive intervention) to suit their particular resource levels. Additional information—including data on the cost of implementing the software program, differences in the utilisation of resources in and outside the hospital, labour costs and the total readmission reduction opportunity available to patients in different risk groups—is needed to establish the cost-effectiveness of the intervention.

In this study we also found that, among the patients who received the intervention, those who received more outpatient components had substantially lower OE ratios for readmission than patients who did not. These findings are consistent with previous studies. Patients are often most vulnerable in the first 7–14 days after discharge when medications need to be adjusted, educational plans are best reinforced and patients are learning critical self-management skills in the context of a new after-hospital care plan. Studies have reported that rapid outpatient follow-up during this period allows for early health assessments, improves self-management and provides an opportunity to address outstanding issues before they grow into more serious events. These factors may explain, in part, the striking reduction in readmissions associated with the completion of outpatient components in this study, which placed particular emphasis on close clinic follow-up for enrolled patients.

This study has important limitations. First, we did not employ a randomised controlled trial design, nor did we think it feasible to blind providers or patients to the intervention. Other factors beyond the intervention therefore could have contributed to the observed reduction in readmissions. However, a control group of patients with AMI or PNA experienced no corresponding improvement in readmission, which would have had an impact on our findings. Nevertheless, given the abrupt reduction in readmission rates with the onset of the intervention, we believe it unlikely that the results we observed were caused by external secular trends. Fifth, the demographic characteristics of our study population might be different from some other US hospitals as our study population had a high proportion of non-white and Medicaid patients. This may affect the generalisability of our findings. We also found statistically significant differences in a number of patient characteristics during the pre- and post-intervention periods, although such differences were largely due to a net increase in Hispanic women with HF in the post-intervention period. We do not have a definitive explanation for this. However, one possible reason may be that the economic recession that occurred during the time of this study affected minority groups such as the Hispanic population and, as a result, more of these individuals were admitted to safety net hospitals similar to the institution in which the study was conducted. It is unlikely, however, that the increase in Hispanic women with HF would be the basis for the abrupt stepwise reduction in readmissions that occurred at the time of the intervention, which is more plausibly and temporally related to the onset of the intervention. Sixth, we were unable to ascertain the number of readmissions outside the region’s information-sharing initiative. Given that any underestimate would be present in both the pre- and post-intervention periods, it is unlikely that this result would have had an impact on our findings. Finally, we did not design this study as a cost-effectiveness trial. Such approaches are needed to examine the overall economic value of the resource allocation strategy presented here. Nevertheless, programmes that achieve equivalent population-level reductions in readmissions by targeting only the highest-risk patients may be implicitly cost-effective.

This study has a number of unique strengths. It is the first study of which we are aware that uses data from an EMR to stratify a patient’s risk of readmission in real time. The study also presents a new method to allocate a fixed set of often constrained hospital resources by considering an individual patient’s risk against the risk of other patients in the hospital at the same time. Finally, in contrast to many readmission intervention studies, we were able to collect readmissions to the study hospital and also to all hospitals in the Dallas-Fort Worth metroplex through a long-standing comprehensive data-sharing initiative, thus providing a more accurate assessment of the readmission rate.

CONCLUSIONS

This study provides preliminary evidence that technology platforms that allow for automated EMR data extraction, case identification and risk stratification may help potentiate the effect of known readmission
Provenance and peer review

Patient consent
Hospital System Institutional Review Board. Texas Southwestern Medical Center-Parkland Health & This study was approved by the University of

Contributors
RA, BX and YM had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Conception and design: RA, PCr, K, TT, LLN, TSS, EAH; acquisition of data: TSS, YM; analysis and interpretation of data: RA, PCr, BJM, BX, SZ, TSS, EAH; drafting of the manuscript: RA, BJM; critical revision of the manuscript for important intellectual content: RA, PCr, BJM, BX, SZ, KT, LLN, KSA, MHD, UK, EAH; statistical analysis: RA, BJM, BX, SZ; obtaining funding: RA, EAH; administrative, technical or material support: RA; supervision: RA, EAH.

Funding
The University of Texas System Patient Safety Grant Award, No. OGC 130 857 and Commonwealth Fund Grant Number 20100323.

Competing interests
None.

Ethics approval
This study was approved by the University of Texas Southwestern Medical Center-Parkland Health & Hospital System Institutional Review Board.

Patient consent
A waiver of informed consent was granted.

Provenance and peer review
Not commissioned; externally peer reviewed.

REFERENCES
6 Burke RE, Coleman EA. Interventions to decrease hospital readmissions: in particular those that emphasise intensive and early post-discharge outpatient contact.

Acknowledgements
The authors wish to thank Anita Rahman, Kashundra Foreman, Elizabeth Marchese, Adeola Gayeola, Asoham Amarasingham, Paul Mayer and Elizabeth Dwelle for their contributions to this project. We also thank the Dallas-Fort Worth Hospital Council Foundation for assisting in the collection of post-hospitalisation data through their information sharing initiative. No author has relevant financial interests in the manuscript.

Contributors
RA, BX and YM had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Conception and design: RA, PCr, K, TT, LLN, TSS, EAH; acquisition of data: TSS, YM; analysis and interpretation of data: RA, PCr, BJM, BX, SZ, TSS, EAH; drafting of the manuscript: RA, BJM; critical revision of the manuscript for important intellectual content: RA, PCr, BJM, BX, SZ, KT, LLN, KSA, MHD, UK, EAH; statistical analysis: RA, BJM, BX, SZ; obtaining funding: RA, EAH; administrative, technical or material support: RA; supervision: RA, EAH.

Funding
The University of Texas System Patient Safety Grant Award, No. OGC 130 857 and Commonwealth Fund Grant Number 20100323.

Competing interests
None.

Ethics approval
This study was approved by the University of Texas Southwestern Medical Center-Parkland Health & Hospital System Institutional Review Board.

Patient consent
A waiver of informed consent was granted.

Provenance and peer review
Not commissioned; externally peer reviewed.

REFERENCES
6 Burke RE, Coleman EA. Interventions to decrease hospital readmissions: in particular those that emphasise intensive and early post-discharge outpatient contact.

Acknowledgements
The authors wish to thank Anita Rahman, Kashundra Foreman, Elizabeth Marchese, Adeola Gayeola, Asoham Amarasingham, Paul Mayer and Elizabeth Dwelle for their contributions to this project. We also thank the Dallas-Fort Worth Hospital Council Foundation for assisting in the collection of post-hospitalisation data through their information sharing initiative. No author has relevant financial interests in the manuscript.

Contributors
RA, BX and YM had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Conception and design: RA, PCr, K, TT, LLN, TSS, EAH; acquisition of data: TSS, YM; analysis and interpretation of data: RA, PCr, BJM, BX, SZ, TSS, EAH; drafting of the manuscript: RA, BJM; critical revision of the manuscript for important intellectual content: RA, PCr, BJM, BX, SZ, KT, LLN, KSA, MHD, UK, EAH; statistical analysis: RA, BJM, BX, SZ; obtaining funding: RA, EAH; administrative, technical or material support: RA; supervision: RA, EAH.

Funding
The University of Texas System Patient Safety Grant Award, No. OGC 130 857 and Commonwealth Fund Grant Number 20100323.

Competing interests
None.

Ethics approval
This study was approved by the University of Texas Southwestern Medical Center-Parkland Health & Hospital System Institutional Review Board.

Patient consent
A waiver of informed consent was granted.

Provenance and peer review
Not commissioned; externally peer reviewed.

REFERENCES
6 Burke RE, Coleman EA. Interventions to decrease hospital readmissions: in particular those that emphasise intensive and early post-discharge outpatient contact.

Acknowledgements
The authors wish to thank Anita Rahman, Kashundra Foreman, Elizabeth Marchese, Adeola Gayeola, Asoham Amarasingham, Paul Mayer and Elizabeth Dwelle for their contributions to this project. We also thank the Dallas-Fort Worth Hospital Council Foundation for assisting in the collection of post-hospitalisation data through their information sharing initiative. No author has relevant financial interests in the manuscript.

Contributors
RA, BX and YM had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Conception and design: RA, PCr, K, TT, LLN, TSS, EAH; acquisition of data: TSS, YM; analysis and interpretation of data: RA, PCr, BJM, BX, SZ, TSS, EAH; drafting of the manuscript: RA, BJM; critical revision of the manuscript for important intellectual content: RA, PCr, BJM, BX, SZ, KT, LLN, KSA, MHD, UK, EAH; statistical analysis: RA, BJM, BX, SZ; obtaining funding: RA, EAH; administrative, technical or material support: RA; supervision: RA, EAH.

Funding
The University of Texas System Patient Safety Grant Award, No. OGC 130 857 and Commonwealth Fund Grant Number 20100323.

Competing interests
None.

Ethics approval
This study was approved by the University of Texas Southwestern Medical Center-Parkland Health & Hospital System Institutional Review Board.

Patient consent
A waiver of informed consent was granted.

Provenance and peer review
Not commissioned; externally peer reviewed.
ABSTRACT
Background Problems of quality and safety persist in health systems worldwide. We conducted a large research programme to examine culture and behaviour in the English National Health Service (NHS).
Methods Mixed-methods study involving collection and triangulation of data from multiple sources, including interviews, surveys, ethnographic case studies, board minutes and publicly available datasets. We narratively synthesised data across the studies to produce a holistic picture and in this paper present a high-level summary.
Results We found an almost universal desire to provide the best quality of care. We identified many ‘bright spots’ of excellent caring and practice and high-quality innovation across the NHS, but also considerable inconsistency. Consistent achievement of high-quality care was challenged by unclear goals, overlapping priorities that distracted attention, and compliance-oriented bureaucratised management. The institutional and regulatory environment was populated by multiple external bodies serving different but overlapping functions. Some organisations found it difficult to obtain valid insights into the quality of the care they provided. Poor organisational and information systems sometimes left staff struggling to deliver care effectively and disempowered them from initiating improvement. Good staff support and management were also highly variable, though they were fundamental to culture and were directly related to patient experience, safety and quality of care.
Conclusions Our results highlight the importance of clear, challenging goals for high-quality care. Organisations need to put the patient at the centre of all they do, get smart intelligence, focus on improving organisational systems, and nurture caring cultures by ensuring that staff feel valued, respected, engaged and supported.

INTRODUCTION
A commitment to delivering high-quality, safe healthcare has been a policy goal of governments worldwide for more than a decade, but progress in delivering on these aspirations has been modest. Patients everywhere continue to suffer avoidable harm and substandard care. England’s National Health Service (NHS) has not been immune to these problems. Despite some encouraging evidence of improvement in quality and safety, large and inexplicable variations in quality of care are evident across multiple domains and sectors of healthcare, from primary through to community and secondary care. England has also seen a number of high-profile scandals involving egregious failings in the quality and safety of individual providers. These include the case of Mid Staffordshire NHS Foundation Trust, the subject of a recently published public inquiry by Sir Robert Francis into how catastrophic failings in the quality and safety of care went undetected and uncorrected.

Francis identified the causes of organisational degradation at Mid Staffordshire as systemic; he saw the underlying faults as institutional and cultural in character. He found significant weaknesses in NHS systems for oversight, accountability and
influence for patient safety and quality of care. Central to his analysis was evidence of a large-scale failure of control and leadership at multiple levels, from what social scientists term the ‘blunt end’ of the system where decisions, policies, rules, regulations, resources and incentives are generated, through to the ‘sharp end’, often known as the ‘frontline’, where care is provided to patients. The distinction between the blunt end and the sharp end is of course a heuristic one; many within healthcare organisations function in hybrid positions as managers and practitioners. Nonetheless, it is useful to recognise how the blunt end, by shaping the environment where care is delivered, may create the ‘latent conditions’ that increase the risks of failure at the sharp end, but may equally generate organisational contexts that are conducive to providing high-quality care. Such contexts include culture: Francis blamed an ‘insidious negative culture involving a tolerance of poor standards and a disengagement from managerial and leadership responsibilities’. Culture is, of course, a term that is widely used but notoriously escapes consensual definition. Many definitions of culture (including Schein’s influential approach) nonetheless have in common an emphasis on the shared basic assumptions, norms, and values and repeated behaviours of particular groups into which new members are socialised, to the extent that culture becomes ‘the way things are done around here’.

The findings of the Francis inquiry are depressingly familiar. England is not alone in experiencing organisational crises in healthcare; examples of failures in healthcare systems have occurred as far apart as New Zealand, the USA and the Netherlands. Several demonstrate precisely the same features as Mid Staffordshire, including long incubation periods during which warning signs were discounted, poor management systems, failure to respond to patient concerns, cultures of secrecy and protectionism, fragmentation of knowledge about problems and responsibility for addressing them, and cultures of denial of uncomfortable information. An important question thus concerns the extent to which the features of the Mid Staffordshire case might be symptoms of more widespread pathologies, given that other organisations in the NHS are exposed to the same institutional and regulatory environment. In this article, we offer lessons from a large multimethod research programme on culture and behaviour related to quality and safety in the NHS.

The research programme covered a critical period between 2010, following the initial inquiry into Mid Staffordshire and the White Paper on the NHS, and 2012, when the Health and Social Care Act was passed. The programme involved a large number of substudies using different methods to seek evidence from staff and patients throughout the English NHS, from large subsamples of NHS organisations, strategic level stakeholders, teams, and patient and carer organisations, and from detailed case studies. It was thus able to provide graduated levels of focus and multiple lenses. Each of the individual substudies will be reported separately, but there is considerable value in bringing the learning from them together holistically. In this article, we provide a synthesis across the studies to draw out high-level learning about culture and behaviour in NHS organisations; what influences culture and behaviour; and what needs to change to give effect to the vision of a safe, compassionate service in which patients and their families could have trust and confidence.

METHODS
We conducted a large, mixed-method research programme involving seven separate substudies (table 1). The programme received ethical approval from an NHS Research Ethics Committee. In summary, primary data were drawn from:

- 107 interviews with key, senior level stakeholders from across the NHS and beyond;
- 197 interviews from the ‘blunt end’ (executive and board level) of NHS primary care and acute organisations through to the ‘sharp end’ (frontline clinicians) where staff care for patients;
- over 650 h of ethnographic observation in hospital wards, primary care practices, and accident and emergency units;
- 715 survey responses from patient and carer organisations;
- two focus groups and 10 interviews with patient and carer organisations;
- team process and performance data from 621 clinical teams, drawn from the acute, ambulance, mental health, primary care and community trust sectors;
- 793 sets of minutes from the meetings of 71 NHS trust boards from multiple sectors over an 18-month period, including detailed analysis of eight boards’ minutes.

We did not use a formal protocol for integrating the findings across these studies, instead deploying a more interpretive, narrative approach. We engaged in extensive discussions as a team, and identified points of convergence and updated our analytic categories as we came closer to agreement. Given the size of our datasets, we are able to provide only very limited primary data in support of our analysis in this article; our focus is on high-level messages. Further details of the methods and the data are available in a longer report.

RESULTS
Our synthesis of the findings across the substudies allowed many insights into the challenges of realising a vision of reliably safe, high-quality care across the NHS, and important learning about how improvement can best be secured.

Table 1

<table>
<thead>
<tr>
<th>Study element</th>
<th>Participants and scheduling</th>
<th>Setting</th>
<th>Focus of research</th>
<th>Analytic approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stakeholder</td>
<td>107 semi-structured telephone interviews with those closely involved</td>
<td>NHS trusts, ambulance trusts, mental health trusts</td>
<td>Assessing culture and behaviour in relation to quality, change; plans to implement quality and safety improvement, enhance leadership and promote staff engagement with quality, leadership for quality</td>
<td>Analysis based on constant comparative method</td>
</tr>
<tr>
<td>Acute trusts, ambulance trusts, mental health trusts</td>
<td>650 h of observation; 197 interviews from the meetings of 71 NHS trust boards; teamwork process and performance data from 621 clinical teams</td>
<td>NHS acute trusts, ambulance trusts, mental health trusts</td>
<td>Assessing cultures of high-quality care</td>
<td>Coding within and across cases, systematically searching for where clusters of codes formed a theme</td>
</tr>
<tr>
<td>Employee focus groups and 10 interviews with patient and carer organisations</td>
<td>Observations in hospital wards, primary care practices, and accident and emergency units</td>
<td>NHS acute trusts, ambulance trusts, mental health trusts</td>
<td>Assessing views on obstacles to delivering improved patient experience</td>
<td>Analysis based on constant comparative method</td>
</tr>
<tr>
<td>NHS Trust, Primary care, ambulance, acute care and mental health trusts</td>
<td>715 survey responses from patient and carer organisations</td>
<td>NHS acute trusts, ambulance trusts, mental health trusts</td>
<td>Assessing views on obstacles to delivering improved patient experience</td>
<td>Descriptive statistics and paired sample t tests</td>
</tr>
<tr>
<td>NHS staff and patient surveys: patient satisfaction survey data</td>
<td>2007, 2009, 2011</td>
<td>NHS acute trusts, ambulance trusts, mental health trusts</td>
<td>Assessing views on obstacles to delivering improved patient experience</td>
<td>Descriptive statistics and paired sample t tests</td>
</tr>
<tr>
<td>NHS staff and patient surveys: patient satisfaction survey data</td>
<td>2007, 2009, 2011 national staff survey data</td>
<td>NHS acute trusts, ambulance trusts, mental health trusts</td>
<td>Assessing views on obstacles to delivering improved patient experience</td>
<td>Descriptive statistics and paired sample t tests</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study element</th>
<th>Participants and scheduling</th>
<th>Setting</th>
<th>Focus of research</th>
<th>Analytic approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stakeholder interviews</td>
<td>107 semi-structured telephone interviews with those closely involved in quality and safety</td>
<td>Acute trusts, ambulance trusts, mental health trusts, community trusts, foundation trusts, primary care trusts, strategic health authorities, general practices and healthcare commissioning organisations</td>
<td>Understanding of vision of high-quality and safe care; what is required to make it happen; theories of change; plans to implement quality and safety improvement, enhance leadership and promote staff engagement; views on what quality improvement means, how it could best be secured, and obstacles</td>
<td>Analysis based on constant comparative method Use of QSR NVivo 8 software</td>
</tr>
<tr>
<td>2. Ethnographic case studies: observations and interviews</td>
<td>Comparative case studies across seven purposively chosen cases 650 h of observation; 197 semi-structured interviews with executive and board-level staff and frontline staff</td>
<td>Four hospital trusts; a quality improvement collaborative; a large-scale quality improvement programme involving dozens of organisations; one primary care provider involving a chain of practices</td>
<td>Assessing culture and behaviour in relation to quality, staff engagement with quality, leadership for quality, quality improvement, practical actions for promoting cultures of high-quality care</td>
<td>Analysis based on constant comparative method Coding within and across cases, systematically searching for where clusters of codes formed a pattern Combining data from interviews across cases and stakeholders to form a single dataset Quantitative analysis—largely descriptive Open-ended responses subject to content analysis to derive themes inductively Qualitative analysis of key themes</td>
</tr>
<tr>
<td>3a. Patient and public involvement: survey</td>
<td>715 survey responses Cross-sectional</td>
<td>Patient participation groups</td>
<td>The survey consisted of 14 statements about patient experience. Open text box provided for each statement</td>
<td>Descriptive statistics and paired sample t tests</td>
</tr>
<tr>
<td>3b. Patient and public involvement: focus groups and interviews</td>
<td>Two focus groups and 10 interviews</td>
<td>Patient and carer organisations</td>
<td>Interpret the findings of the survey Assessing views on obstacles to delivering improved quality and safety and greater accountability in the NHS</td>
<td>Descriptive statistics and paired sample t tests</td>
</tr>
<tr>
<td>4a. NHS staff and patient surveys: patient satisfaction survey data</td>
<td>165 acute trusts—data from 2007, 2009, 2011</td>
<td>Acute trusts</td>
<td>Patient satisfaction came from the National Acute Inpatient Survey, using the data on patients' overall ratings of care</td>
<td>Detailed correlation analysis between staff survey and inpatient survey; multiple and multilevel regression analysis, using HR practice variables to predict engagement; regression and ordinal logistic regression analysis to predict patient satisfaction, patient mortality, staff absenteeism, staff turnover, infection rates, and Annual Health Check ratings, controlling for trust type, size and location; latent growth curve modelling to predict outcomes</td>
</tr>
<tr>
<td>4b. NHS staff and patient surveys: national staff survey data</td>
<td>309 NHS trusts from 2007, 2009, 2011 national staff survey</td>
<td>Primary care, ambulance, acute care and mental health trusts</td>
<td>Staff engagement, organisational climate, job satisfaction, manager support, job design, errors and reporting, work pressure, bullying, harassment and abuse, team working, training, appraisal, stress, teamwork</td>
<td>Descriptive statistics and paired sample t tests</td>
</tr>
<tr>
<td>4c. NHS staff and patient surveys: outcome measures</td>
<td>2005–2009 Primary care, ambulance, acute care and mental health trusts</td>
<td>Patient mortality (acute sector only) hospital standardised mortality ratio; quality of services and use of resources (Annual Health Check ratings by Healthcare Commission between 2005/2006 and 2008/2009); infection rates (MRSA) per 10000 bed days; staff absenteeism; staff turnover</td>
<td>Detailed correlation analysis between staff survey and inpatient survey; multiple and multilevel regression analysis, using HR practice variables to predict engagement; regression and ordinal logistic regression analysis to predict patient satisfaction, patient mortality, staff absenteeism, staff turnover, infection rates, and Annual Health Check ratings, controlling for trust type, size and location; latent growth curve modelling to predict outcomes</td>
<td></td>
</tr>
<tr>
<td>5. Clinical teams functioning, effectiveness and innovation</td>
<td>621 teams (4604 responses) Aston Team Performance Inventory Cross-sectional data with data on team changes collected from 388 teams (1299 individuals) 3 months later</td>
<td>51 trusts (13 acute, 17 mental health, 10 ambulance and 11 primary care trusts)</td>
<td>Team functioning: task design, team effort and skills, organisational support, resources, objectives, participation, creativity, conflict, reflexivity, task focus, leadership, satisfaction, attachment, effectiveness, inter-team relationships, innovation Leaders/external raters' evaluations of effectiveness</td>
<td>Descriptive analysis, ANOVA, regression and relative importance analysis Analysis and ratings from domain relevant experts Open-ended responses subject to content analysis to derive the themes</td>
</tr>
</tbody>
</table>
Goal setting

Virtually all those we interviewed (over 300 in total) were firmly committed to the ideal of a safe, high-quality health service for patients and to good patient experience. Many identified the values of compassion and care as at the heart of the mission for their organisations, and as their most deeply felt personal professional commitment. Our interviews, observations, surveys and documentary analysis were united in suggesting that, for organisations to succeed in delivering high-quality, safe care, they needed to have a clearly articulated vision, including explicit goals for quality and safety and a strategy for achieving them. Interviews and observations repeatedly emphasised the importance of clear goals in establishing and signposting priorities for improvement, motivating staff and ensuring resources were appropriately directed. Our survey evidence from the national patient satisfaction and national staff surveys (NSS) showed that patient satisfaction was highest in trusts that had clear goals at every level. Consistent with the findings of the Francis inquiry, boards of organisations were identified in interviews as particularly influential in setting the overall direction and demonstrating the commitment and organisational priority given to quality and safety.

But converting laudable aspirations for high-quality, safe and compassionate care into clear goals appeared challenging. Clarity about goals, how they could be achieved, and leadership for delivery were highly variable. Our questionnaire surveys of board members showed that they rarely stated clear board objectives that were challenging and measurable. The 621 frontline clinical teams we studied were generally even less clear about their objectives.

A major challenge in creating unifying visions for patient safety and quality and setting clear objectives was the range, diversity and complexity of external expectations and requirements that NHS organisations faced. Board and executive teams described an institutional and regulatory environment that was populated by external agencies and actors who served different but overlapping functions. They reported that targets, standards, incentives and measures seemed to crowd in from multiple external sources; that the same information was required many times in different formats; and that answering to so many masters and producing data for so many external audiences was costly and distracting. The proliferation of externally set priorities and the number of different agencies and actors created what we termed ‘priority thickets’—dense patches of overlapping or disjointed goals that commanded very substantial attention and resources, but did not necessarily provide clear direction or facilitate the development of clear goals, internally coherent visions or strategies linked to local priorities.

Faced with so many competing demands, some organisations tended to revert to a highly
bureaucratised form of management, characterised by proliferation of rules, procedures and forms corresponding to externally imposed demands. Many of these seemed to be motivated mostly by a need to make displays of compliance, rather than by genuine efforts to make systems safer or of better quality. Much of this activity could be characterised as defensive and reactive. It was a source of frustration throughout organisations; frontline teams complained of ‘blanket’ policies which were seen as ‘very prescriptive and not concentrated on clinical work’.

We also found considerable variability in how far organisations succeed in making their aspirations for high-quality care real: what we termed ‘bright spots’ and ‘dark spots’ were both evident, even within the same organisations. Bright spots included teams and individuals who demonstrated caring, compassion, cooperation and civility, and a commitment to learning and innovation. Direct observations found that in many settings patients were often treated with kindness and respect, systems functioned well, and staff were busy but knew what they were doing and why. Compliance with many standards of good practice, such as hygiene and equipment counting, was observed to be very good in many cases.

Though much care was of such high quality as to be inspiring, substandard care or ‘dark spots’ were also evident. Dark spots were found where staff were challenged to provide quality care, were harried or distracted, or were preoccupied with bureaucracy. Across our interviews, surveys and observations we found evidence of staff and patient concern about variability in quality of care, and a lack of confidence that care would be reliably good. Interviews and surveys with patient and carer groups suggested that patients and their carers were often concerned about quality and safety. Vulnerable patients, including older patients, young patients or those who lacked the ability to ‘speak up’, were reported to be at risk of being left to ‘fend for themselves’ or ‘being forgotten’. Our observations in clinical areas and our interviews confirmed that inconsistency was a feature of many settings. For example, many staff spoke to patients politely and with kindness, but some others were brusque, impatient or discourteous. Some senior clinical nursing staff highlighted their concern at what they saw as a tendency towards task-focused rather than person-centred care.

Further evidence of the challenges of realising a vision of consistently high-quality safe care came from our analysis of the NHS Staff Survey and inpatient survey data over the period 2007–2011 (tables 2 and 3). This suggested that there had been improvements in scores relating to quality and safety reported by patients and staff nationally between 2007 and 2009, but subsequently these improvements stalled or went into reverse. Some of the plateauing may reflect a natural maximum level being reached. For example, the percentage of staff receiving health and safety training increased from 2007 to 2009, and appears to remain relatively constant in 2011. Likewise, levels of job satisfaction increased from 2007 to a moderately high level in 2009, and then stayed approximately the same in 2011. However, measures on the staff survey relating to error and incident reporting, blame cultures and improvements following incidents, where there was headroom for improvement, appeared to have shown only very modest gains. The number of staff working paid extra hours has decreased consistently, but since 2009 the number working unpaid extra hours has increased sharply. The percentage of staff receiving training in infection control related issues increased from 2007 to 2009, but fell in 2011.

Variability in intelligence

A major challenge to achieving goals relating to quality and safety was that of ensuring that high-quality intelligence was available to organisations, teams, and individuals about how well they were doing and where the deficits and risks in organisational systems lay. NHS organisations that we studied were putting considerable time, effort and resources into data collection and monitoring systems. They typically used a combination of routinely collected data, specific data collection initiatives, and sporadic sources such as spot checks and audits. To a varying extent they also drew on feedback provided by clinical staff and patients as a means of assessing trends. However, the degree to which data collection efforts translated into actionable knowledge, and then into

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Changes in the National Staff Survey 2007–2011 (NHS trusts in England)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I have adequate materials, supplies and equipment to do my work</strong></td>
<td>3.22</td>
</tr>
<tr>
<td><strong>There are enough staff at this Trust for me to do my job properly</strong></td>
<td>2.61</td>
</tr>
<tr>
<td><strong>I do not have time to carry out all my work</strong></td>
<td>3.30</td>
</tr>
</tbody>
</table>

Results are based on the 309 NHS trusts in England with data from all 3 years shown. p values are based on paired sample t tests. Responses were on a 1–5 scale with 5 indicating greater agreement with statements; an increase of 0.10 is equivalent to 10% of respondents moving up one category of response, for example, from ‘neither agree nor disagree’ to ‘agree’.

*p<0.05; **p<0.01.

NHS, National Health Service.
effectively organisational responses, differed markedly between organisations.

Some behaviours in relation to data gathering might be described as ‘problem-sensing’; other, less positive behaviours were ‘comfort-seeking’. Problem-sensing involved actively seeking out weaknesses in organisational systems, and it made use of multiple sources of data—not just mandated measures, but also softer intelligence. Soft intelligence could be gathered in many ways, including active listening to patients and staff; informal, unannounced visits to clinical areas; and techniques such as ‘mystery shoppers’, shadowing of staff, and swapping roles for a short period. While sometimes discomfiting, this less routinely gathered knowledge enabled fresh, more penetrating insights to complement quantitative data. Senior teams displaying problem-sensing behaviours tended to be cautious about being self-congratulatory; perhaps more importantly, when they did uncover problems, they often used strategies that went beyond merely sanctioning staff at the sharp end, making more holistic efforts to strengthen their organisations and teams.

Comfort-seeking behaviours are defined here as being focused on external impression management and seeking reassurance that all was well; consequently, what was available to organisations was data, but not intelligence. Serious blind spots could arise when organisations used a very limited range of methods for gathering data, were preoccupied with demonstrating compliance with external expectations, failed to listen to negative signals from staff or lacked knowledge of the real issues at the frontline. Comfort-seeking tended to demonstrate preoccupation with positive news and results from staff, and could lead to concerns and critical comments being dismissed as ‘whining’ or disruptive behaviour. When comfort-seeking was the predominant behaviour, data collection activities were prone to being treated by sharp-end staff as wearisome and fruitless accountability exercises. Some staff reported that they felt the main purpose of much data collection was to allow individuals to be blamed if something did go wrong, not to make the system safer.

### Variability in systems

Interviews, observations and surveys showed that when staff had access to appropriate resources, perceived that staffing levels were adequate with the right skill mix, and had systems that functioned effectively, they felt that they could complete their work successfully, could explore new ways of improving quality and could develop reflective practices. This reinforced their levels of motivation and morale in a virtuous circle. But deficits in systems often obstructed and frustrated well-motivated staff in their mission to provide good care for patients. Our analysis of questionnaire reports from 621 clinical teams showed that many staff felt unable to achieve their goals for patients because of organisational factors outside their control. Observations showed that staff wasted time working with poorly designed IT systems, negotiating clinical pathways with obstructions and gaps, and battling with multiple professional groups and subsystems (e.g. pharmacy, microbiology and imaging, and many others) that did not operate in integrated ways. We also found evidence of problematic handovers between shifts, departments and teams, team conflict and a diffusion of responsibility relating to particular patients. Patient and carer groups reported discontinuities in care between institutional boundaries and even within single organisations. These ‘responsibility cords’ left patients variously ill informed, distressed and disappointed, and sometimes in danger.
Staff at the sharp end were very often aware of systems problems but felt powerless to bring about change. Changes within organisations, uncertainty about priorities, poor systems, heavy workloads and staff shortages were all blamed for staff feeling they lacked support, further reducing their motivation and morale. Given that many systems required significant improvement, it was disappointing that we found a clear trend of decreasing levels of board innovation, especially in relation to quality and safety.

We defined innovation as the intentional introduction of processes and procedures, new to the unit of adoption (team or organisation) and designed to significantly benefit the unit of adoption, staff, patients or the wider public. An analysis of board minutes from 71 NHS Trusts covering an 18-month period between January 2010 and June 2011 identified a total of 144 innovations that were implemented in organisations, representing an average of only 1–3 per organisation. More than half were focused on increasing productivity (73), with very few related to safety (14). The largest number of innovations (62) was identified between January and June 2010, followed by 56 innovations between July and December 2010. Only 26 innovations were identified in data covering the time between January and June 2011. Separately, analysis of 4976 responses to open-ended questions in our survey of 486 clinical teams identified 183 innovations over a 6-month period. This also suggested relatively low rates of innovation among frontline teams, though many of the solutions they did devise were ingenious and resourceful. The largest number of frontline staff innovations was focused on enhancing quality of patient care; fewer aimed at improving administrative effectiveness, and the smallest number concerned staff wellbeing.

Many organisations were using specific quality improvement methods to achieve change, including Plan–Do–Study–Act cycles, Collaboratives, Lean, Six Sigma, and Productive Ward (an NHS programme to support ward teams in reviewing the processes and environment used to provide patient care). Some organisations also used wider techniques to improve quality, including organisation-level campaigns. Great enthusiasm for these approaches was often reported by those leading improvement efforts, but we also sometimes observed a tendency towards uncritical or indiscriminate use, and some evidence of ‘magical thinking’ (‘this initiative will solve many problems easily and quickly’). Frontline staff who had to implement these initiatives were often not consulted or adequately informed about their purpose and implementation, and sometimes initiatives were abandoned or forgotten after a short period of intense activity. In some cases, there was insufficient acknowledgement of the effort, expertise and investment required to make such approaches work, and substantial problems with quality of data collection and interpretation.

**Culture and behaviour**

Leadership was important for setting mission, direction and tone. Our observations, interviews and surveys all emphasised the importance of high-quality management in ensuring positive, innovative and caring cultures at the sharp end of care. Some senior teams encouraged and enabled frontline teams to address challenges and to innovate, but recognised that, along with demanding personal accountability from staff, they also needed to fix systems problems that prevented staff from functioning well. A strong focus by executive and board teams on their own role in identifying and addressing systems problems was powerful in supporting cultural change that delivered benefits for patients, and our observations and interviews identified many examples of impressive gains being made by the sharp and blunt ends working together around unifying goals.

Nevertheless, an important consequence of the failure to clarify goals, to gather appropriate intelligence or to address systems deficits was the existence of frequent misalignments between the ways the blunt end and the sharp end of organisations conceived of quality and safety problems and their solutions. For sharp-end staff, threats to safety and quality were identified as weaknesses in systems, failures of reliability, suboptimal staffing, inadequate resources and poor leadership. Lack of support, appreciation and respect, and not being consulted and listened to were seen as endemic problems by staff in some organisations. In contrast, some senior managers—particularly those engaged in comfort seeking—tended to see frontline staff behaviour and culture as the cause of quality problems. In consequence, those at the blunt and those at the sharp end often did not agree on the causes of variation in quality and safety and, therefore, on how they should be addressed.

We also found substantial variation in the quality of management. Our analyses of NSS data showed that hospital standardised mortality ratios were inversely associated with positive and supportive organisational climates. Higher levels of staff engagement and health and wellbeing were associated with lower levels of mortality, as were staff reporting support from line managers, well structured appraisals (e.g. agreeing objectives, ensuring the individual feels valued, respected and supported), and opportunities to influence and contribute to improvements at work. NSS data also showed that staff perceptions of the supportiveness of their immediate managers, the extent of staff positive feeling, staff satisfaction and staff commitment were associated with other important outcomes, including patient satisfaction. In places where staff reported high work pressure, patients on the national surveys also reported too few nurses, insufficient support, and problems with information, privacy and respect. In trusts with poor staff health and wellbeing, high injury rates, and a high level of staff...
intention to quit their jobs, patients reported that they were generally less satisfied, and Care Quality Commission ratings described poorer care and poorer use of resources. These findings were consistent across trust types (primary care, ambulance, mental health, community and acute).

Also concerning was evidence that though team-working seemed well established and widespread on the surface, there was a surprising lack of clarity about team purpose, objectives, membership, leadership and performance among many teams. Our survey of 621 clinical teams demonstrated that team inputs and team processes were significantly associated with the effective provision of good-quality care, but senior managers were sometimes unable to identify teams and team leaders. When team leaders were identified, they were often confused about who their team members were. Team members themselves had low agreement about who was part of their team. Factors associated with successful teams were the effort and skills of team members, resources made available and good processes. Clarity and agreement about team objectives were key to clinical team effectiveness, along with a participative approach to decision-making that engaged all team members. Teams who regularly took time out to reflect on their objectives, how they were going about achieving these and how their performance needed to change were particularly likely to be more effective and innovative.

**DISCUSSION**

This large mixed-method study identified many ‘bright spots’ of excellent caring and practice and high-quality innovation across the NHS, but also considerable inconsistency. Though Mid Staffordshire may have been one particularly ‘dark spot’ in the NHS, organisations throughout the NHS are likely to have at least some shadows: there was little confidence that care could be relied upon to be good at all times in all parts of organisations, and we found evidence of structural and cultural threats to quality and safety. Our analysis points to how things may improve. First, organisations need to focus on developing cultures that are person centred—not just task focused—and valuing and building on the excellent care and commitment delivered by many staff throughout the NHS. This involves modelling and reinforcing values and behaviours that underpin high-quality care, patient safety and positive patient experience from the blunt end to the sharp end of the whole system.

Our work involving a very large number of organisations confirms that achieving quality and safety in NHS organisations requires a robust strategy and unifying vision. National leadership sets the tone, signals importance, legitimises, and creates accountability mechanisms. Yet the Francis public inquiry showed that a major problem for Mid Staffordshire was the large number of different agencies and bodies with a say in the NHS. This contributed to fragmentation, multiple competing pressures, ambiguity and diffusion of responsibility. Our work similarly demonstrates that proliferation of external agencies and expectations creates conflicts, distraction and confusion from the blunt to the sharp end of organisations about where resources and attention should be directed. Where incentives and external expectations conflict, compete or fail to cohere, the ability of organisations to set themselves clear, internally valued goals for achieving their aspirations is weakened.

In a distributed and complex system such as the NHS, failures are least likely when the goals are clear and uniting, and when appropriate, sensitively designed regimes of control and support are found at every level: from policymaking, through the layers of formal regulatory systems, the institutional environment, individual organisations, teams and practitioners, through to patients’ experiences. Coherence of national direction is therefore essential to avoid dispersing responsibility and accountability, and creating confusing messages and signals. As new bodies move forward, including NHS England and the Clinical Commissioning Groups, it is important that they avoid creating further competing priorities, and instead ensure focus and coherence.

National leadership needs to be matched by high-quality leadership across multiple organisational levels underpinned by clear, patient-focused goals and objectives. The role of organisational boards in securing the quality and safety of health services has become an increasing focus of academic and policy interest, not least because of evidence of the link between leadership from the top and the priority and resources given to quality and clinician engagement. But we found worrying evidence of NHS trust boards failing to set clear goals for themselves as boards and for their organisations. Goals do need to be set, and they should be limited in number (to identify priorities while avoiding the creation of priority thickets) and known not only by all board members.
but (if appropriate) more widely within their trusts. Goals should be shaped by the need to promote quality and safety, to ensure sound financial performance, and to value dignity and respect for patients. They should provide a framework for objectives at all levels of trusts, from senior management to clinical teams at the sharp end. They should be framed to encourage innovation at all levels, and quality and safety of patient care must be overriding. Not all innovations need to be grand and over-arching: fixing (apparently) small problems may result in major gains.28

The Francis public inquiry showed that discounting of warning signs of deterioration was a key feature of board and executive behaviour at Mid Staffordshire. Those in senior positions appear to have developed ‘blindsight’—a way of not seeing what was going wrong. Our work confirms the importance of high-quality intelligence (not just data) and making that intelligence actionable. But we found sobering evidence that NHS organisations are not always smart with intelligence, and need to gear more towards problem-sensing rather than comfort-seeking. At the national level, care needs to be taken to ensure that the number of measures organisations are expected to report externally is well managed,29 and that measures are aligned with local priorities, avoid imposing excessive burdens, generate accurate intelligence,30 and most of all, are useful for informing improvement locally. Thus the right intelligence needs to be gathered, interpreted correctly and fed back clearly to staff at the sharp end of care, so that they consolidate and improve their performance.31

Organisations need to be especially alert to the possibility of blind spots where they are unaware of problems. They should use multiple strategies to generate intelligence and undertake self-assessment12 of local culture and behaviours—not just rely on mandated measures—and use a range of techniques for hearing the patient’s voice and the voice and insights of those at the sharp end of care. Consistent with Francis’ findings, good management is as important as good leadership in our analysis: the wellbeing of staff is closely linked to the wellbeing of patients, and staff engagement is a key predictor of a wide range of outcomes in NHS trusts. Achieving high levels of engagement is only possible in cultures that are generally positive, when staff feel valued, respected and supported, and when relationships are good between managers, staff, teams and departments and across institutional boundaries. Staff experience frustration and conflict when asked to work in systems that do not effectively serve them, or the patients they care for; these system defects include staff shortages or inappropriate skill mix to address the needs of specific clinical areas. Our analysis suggests that improving culture, behaviour and systems requires system improvement and better communication between the blunt end and sharp end.

**Box 1 Strategies for creating positive cultures**

<table>
<thead>
<tr>
<th>Senior leaders should:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continually reinforce an inspiring vision of the work of their organisations</td>
</tr>
<tr>
<td>Promote staff health and wellbeing</td>
</tr>
<tr>
<td>Listen to staff and encourage them to be involved in decision making, problem solving and innovation at all levels</td>
</tr>
<tr>
<td>Provide staff with helpful feedback on how they are doing and celebrate good performance</td>
</tr>
<tr>
<td>Take effective, supportive action to address system problems and other challenges when improvement is needed</td>
</tr>
<tr>
<td>Develop and model excellent teamwork</td>
</tr>
<tr>
<td>Make sure that staff feel safe, supported, respected and valued at work.35</td>
</tr>
</tbody>
</table>

This needs to be sustained, intense, mutually respectful and focused on achieving a shared understanding of quality problems and joint working to put them right. Trusts can develop these cultures by using specific strategies (box 1), while recognising the complexities of trying purposefully to engineer culture.12

This article has some limitations. In the available space, we have not been able to provide full details of methods or data from this unusually large research programme. Instead we have sought to provide an overview of the key findings of the component studies. Our synthesis of findings was interpretive and narrative, and did not use a formal protocol. Others might reach somewhat dissimilar conclusions or interpretations of our data.33 However, we believe that our careful scrutiny of the data, extensive discussions, and detailed analysis of themes have enabled us to produce a powerful, robust and rich picture. Future work should assess the generalisability of these findings in other contexts, including the other countries of the UK.

**Conclusions**

This very large-scale research programme suggests that there is room for improvement in the quality and safety of care offered by the NHS, and that this improvement can build on the progress already made. Trusts need continually to refresh, reinforce and model an inspiring vision that keeps the patient at the centre. It is essential to commit to an ethic of learning and honesty,34 to work continually to improve organisational systems, and to nurture the core values of compassion, patient dignity and patient safety through high-quality leadership. This implies equal attention to systems, cultures and behaviours: setting coherent and challenging goals and monitoring progress towards them; empowering staff to provide high-quality care and providing them with the means to
achieve this through routine practice and innovation; and exemplifying and encouraging sound behaviours.

Acknowledgements We thank all of the many organisations and individuals who participated in this research programme for their generosity and support. We thank our collaborators and the advisory group for the programme.

Ethics NHS REC approval was obtained for this study (reference 10/H0406/38).

Contributors The authors listed on this manuscript meet the criteria for authorship.

Funding We thank the Department of Health Policy Research Programme (Reference No 0770017) for funding this research programme. Write up of this paper was supported by a Wellcome Trust Senior Investigator award (MDW) WT097899MA.

Competing interests None.

Ethics approval

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

Every month we scan thousands of journals then select and summarise the best healthcare improvement studies.

Signing up for the Research Scan is quick, free and will save you time getting the latest evidence.

www.health.org.uk/researchscan
We have a whole host of great features on the website from key content picked by the Editor to just being social with a tweet or two. The website carries our most up to date content so visit us today!

**Editor’s Choice**

**Key content**
The Editor highlights an important article from each issue.

**Social media**
Feeling social? Sign up and you’ll be able to join in and see all the talk about the latest issues, including offers from the journal.

**Archive**
Delve into the archive with access from 1992, vol 1 issue 1 until 2006. (Subscribers receive access to the latest archive content).

**e-alert**
An easy way to keep up to date with all the latest content. Simply sign up and we’ll email you the content every month. Perfect for a quick glance.

**Mobile**
Mobile optimised so you can access the site easily on the move.
m.qualitysafety.bmj.com

Visit us online at qualitysafety.bmj.com