Measuring harm and informing quality improvement in the Welsh NHS: the longitudinal Welsh national adverse events study

Sharon Mayor,[1]* Elizabeth Baines,[2] Charles Vincent,[3] Annette Lankshear,[4] Adrian Edwards,[1] Mansel Aylward,[5] Helen Hogan,[6] Paul Harper,[7] Jan Davies,[8] Ameet Mamtora,[9] Emily Brockbank[10] and Jonathon Gray[11,12]

[1]Division of Population Medicine, Cardiff University, Cardiff, UK
[2]Quality and Safety, NHS Wales Delivery Unit, Pencoed, UK
[3]Department of Psychology, University of Oxford, Oxford, UK
[4]School of Healthcare Sciences, Cardiff University, Cardiff, UK
[5]Public Health Wales, Cardiff, UK
[6]Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK
[7]School of Mathematics, Cardiff University, Cardiff, UK
[8]Quality and Safety, Welsh Government, Cardiff, UK
[9]Cardiff and Vale University Health Board, Cardiff, UK
[10]Aneurin Bevan University Health Board, Newport, UK
[11]Centre for Healthcare Improvement and Innovation, Counties Manukau District Health Board, Auckland, New Zealand
[12]Healthcare Improvement and Innovation, Victoria University of Wellington, Wellington, New Zealand

*Corresponding author

Declared competing interests of authors: Charles Vincent reports personal fees from Cardiff University during the conduct of the study, personal fees from the Swiss Federal Office of Health, personal fees from haélò, personal fees from Wiley-Blackwell, personal fees from Healthcare at Home and grants from The Health Foundation outside the submitted work. Jonathon Gray reports other from Counties Manukau District Health Board and other from Victoria University of Wellington, Wellington, New Zealand, outside the submitted work.

Published February 2017
DOI: 10.3310/hsdr05090

Scientific summary

Measuring harm in Welsh NHS hospitals

Health Services and Delivery Research 2017; Vol. 5: No. 9
DOI: 10.3310/hsdr05090

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background and aims

Despite global activity over the past 15 years to improve patient safety, the measurement of adverse event (AE) rates remains challenging. Although many approaches have been used, and are in use, the dual-stage retrospective review of medical records has been the methodology of choice in large-scale studies. The first stage involves the completion of criterion-referenced screening review form 1 (RF1), which is undertaken by nurses or experienced clinical governance facilitators, and the second, undertaken by doctors, is completion of the modular review form 2 (MRF2). The Global Trigger Tool (GTT), developed by the US Institute for Healthcare Improvement, is another widely used retrospective review of medical records, which uses a time-limited pragmatic approach. This tool requires the random selection and limited case note review of 20 inpatient records per month per organisation. At the commencement of this study in 2011, Welsh health boards had experience in using this tool, but its implementation was not regulated and commitment to the methodology and the outcomes generated was highly variable across the Welsh health boards.

Study aims and objectives

This study aimed to obtain definitive longitudinal data on harm in NHS Wales hospitals and to compare the performance of the GTT with the two-stage retrospective review process using findings to improve an approach to ongoing surveillance of harm in the Welsh NHS.

The specific aims over the course of the project were as follows:

1. to gain an in-depth understanding of the nature and number of AEs occurring in the Welsh population admitted to hospital over a 4-year period (2011–15), comparing the use of the retrospective two-stage process and GTT reviews
2. to compare the scale and scope of health-care-related harm identified by the retrospective two-stage process and GTT reviews
3. to develop a robust ongoing measurement system for harm
4. to embed this harm surveillance in organisations and to determine the organisational response.

Methods

The study was undertaken in two discrete data generation phases separated by a transition period to undertake interim analysis informing phase 1. The second phase was followed by a final process evaluation examining the impact of the study and its output data on the NHS health board recipients. The study sample comprised two acute hospital sites from each of the six Welsh health boards. A sample size of 5000 records was calculated based on the incidence of harm found in previous studies.

Phase 1

In phase 1, we undertook a two-stage retrospective review of 20 adult patients per month with an inpatient stay of > 24 hours who had been identified using a random system generator and whose notes had been pulled for routine GTT review by health board staff. The study team then undertook a two-stage retrospective review process. We added a harm determination phase at the end of the nurse screening process and asked the nurses to rate their confidence in their decision on a 1–5 scale. Positive notes from phase 1 were sent for a second-stage review by research physicians, who used the MRF2 to judge the occurrence, nature, preventability and consequences of the referred event.

Analysis of phase 1 data using Stata version 14 (StataCorp LP, College Station, TX, USA) followed the conventional GTT method in which the aggregated data are converted into a harm rate and then compared with a baseline. We used the same methodology to compare aggregate rates in the two methodologies. An unweighted kappa coefficient was used to assess inter-rater reliability with corresponding 95% confidence interval (CI) and the percentage of records that were concordant or discordant.

Transition phase

During the transition phase, a number of factors influenced the planning and execution of phase 2, including:

- the decision of NHS Wales to cease use of the GTT
- a mandate in Wales that every inpatient death must be assessed by review of the inpatient episode of care, currently using the seven questions of the Universal Mortality Review (UMR) tool
- the need of the NHS for an operationally manageable tool using a one-step process.

On the basis of phase 1 data, we identified the criteria that were the most sensitive and specific markers of AEs in current use and, furthermore, refined others to identify acts of omission to improve both the identification of a wide range of events and the face validity of the tool with a range of health-care professionals. The new screening criteria were included in a new tool, the Harm2 tool, which was developed for the routine surveillance of AEs in practice. The Harm2 tool is a hybrid of the one-stage GTT process and a condensed RF1 and MRF2 screening and measurement structure. We then trained a team of research nurses in using the Harm2 tool to undertake medical record assessment across study sites.

Phase 2

The monthly sample of reviewed notes included 20 random discharges and 10 admissions of patients who died from each study site using explicit case note selection criteria. During the analysis we assessed both intermethod (RF1/MRF2 vs. Harm2) and inter-rater reliability.

We determined the percentage of patients experiencing an AE and the severity and level of preventability of the incident. Univariate logistic regression assessed the relationship between demographic and chronic disease states and the subsequent development of AEs. Descriptive statistics were used to describe and compare the nature and severity of AEs in the random and deceased patient cohorts. We tested the Harm2 tool against the two-stage process (intermethod reliability), and an unweighted kappa coefficient was used to assess inter-rater reliability.

Service evaluation and impact assessment

From a constructivist perspective and using semistructured interviews, we explored the review process with the RF1 reviewers and, from the perspective of key NHS informants, the value and use of the data by health boards. Interview data were anonymised and thematically analysed. When available, documentary evidence was used to confirm the narrative.

Findings

Phase 1

In total, 4536 episodes of care from 11 of the 12 invited hospitals were reviewed, with the number of samples from individual hospitals ranging from 174 to 560. The mean age of patients was 64.1 years (95% CI 63.48 to 64.66 years). A total of 1430 (32.6%) inpatient episodes of care had at least one identified positive screening criterion for potential AEs (with a range seen across individual organisations of between 23% and 43%) and, of these, 821 inpatient episodes (18.1%) were referred for in-depth physician review and 790 were completed across the study sites (96.2%). At least one AE was determined in 10.3% of all episodes of care (95% CI 9.4% to 11.2%). In total, 73.1% of patients with AEs needed additional

treatment and/or days in hospital, 16% were discharged with a significant impairment to their functional status and in 9% of cases the AE may have been associated in some way with the patient dying during the inpatient episode. Physician assessment of problems in care in the referred sample revealed that failure in clinical monitoring and management was identified in one-third of all events. Failure to manage and prevent infection (29%), problems with an operation or procedure (21%) and problems in the prescribing, administration or monitoring of drugs and fluids (18.7%) were identified as common problems in care.

We also found the following.

- There was a statistically significant increase in positive criteria with increasing age up to 85 years.
- The percentage of harm identified by non-physicians was 10.4%, which was comparable with that determined by physicians, of 10.3%.
- Although there is a high level of agreement on harm identified in the professional groups, differences in the type of events identified are recognised.
- There was a statistically significant association between mean length of hospital stay and risk of experiencing an AE ($p < 0.0001$).
- The mean rate of AEs using the GTT methodology, across NHS Wales, was lower than the rate reported through the two-stage retrospective review process (mean 9.0%, 95% CI 8.82% to 9.18%).
- We could profile the risk of AEs in individual organisations by producing signatures of harm for each NHS site.

As part of the process of quality assuring the data generated from the RF1 reviews, we began to see distinct patterns of harm emerging from individual organisations. We tested these patterns with site leads by ranking the top five frequently occurring criteria that were linked to AEs for each organisation and presenting these as their 'organisational signatures' of harm. Assessment of signatures of harm along with the AE rate allows for (1) examination of how big an individual issue is within a health-care system, (2) the triangulation of harm data with other organisational data such as incident reports, (3) a mechanism for awareness raising around specific organisational issues or interventions and (4) a potential mechanism for quality improvement evaluation.

Phase 2

Twelve NHS organisations participated in this phase between May 2014 and September 2015, during which time we reviewed 4396 records from randomly selected discharges ($n = 3352$) and randomly selected deceased discharges ($n = 1044$) samples. In the randomly selected discharges sample, at least one AE requiring intervention or subsequent readmission was determined in 11.3% of all episodes of care (95% CI 10.22% to 12.40%) and 59.6% of AEs were deemed to be preventable (95% CI 55.29% to 63.91%). In the deceased sample, at least one AE was determined in 315 (30.1%) episodes of care studied (95% CI 28.13% to 33.8%2) and 61.7% of AEs identified were deemed to be preventable (95% CI 57.49% to 65.91%).

Patient factors associated with AEs in the randomly selected discharges sample included being aged over 85 years and having peripheral vascular disease (PVD), hemiplegia or dementia. No association with chronic disease was identified in the deceased group, but AEs occurred twice as frequently in emergency admissions and were associated with increased length of stay.

In the 10% of the Harm2 tool records that were double reviewed, there was agreement in 336 out of 380 cases (88.4%; $\kappa = 0.50$, 95% CI 0.38 to 0.65). In 422 Harm2 reviews, the RF1 was also completed and physicians using the MRF2 reviewed 74 inpatient episodes. Agreement on the presence of AEs was evident in 58 out of 74 reviews or 78.4% ($\kappa = 0.45$, 95% CI 0.07 to 0.62), indicating a moderate level of agreement.

Service evaluation phase

We interviewed a total of 27 individuals including 14 reviewers (of whom 12 were from nursing or midwifery backgrounds) and 13 members of health boards (of whom seven were medical directors or assistant medical directors).

The review process

After training, reviewers found the Harm2 criteria clear and the tool easy to use, and monthly conference calls enhanced reliability.

Factors that facilitated reviewing included the following:

- collaborative or unified notes, which made it easier to ensure that concerns and prescribed actions were addressed
- live data, including discharge letters entered on the Welsh Clinical Portal (NHS Wales Informatics Service), that were legible and could be accessed remotely.

Factors that challenged the review process included:

- poor-quality notes, especially those of deceased patients
- notes that had been digitised, as poor scanning, lack of searchability and automatic timeouts all contributed to difficulties
- partly electronic and partly paper records, with separate records held by therapists
- a focus on a single episode of acute care, which prevented comment on the holistic care of people with long-term mental illnesses admitted to an acute facility.

The management and usefulness of the data in health boards

Independent review was thought to be useful because it was resourced and offered freedom from acculturation. However, it divorced learning from clinical practice. Owing to the small numbers of notes reviewed monthly at the health board level, the usefulness of the data was perceived as limited in the early months, although this increased as the numbers built up. The single page of output data was thought to be excellent but not easily digested, and was thus considered too complex for widespread dissemination. The key value of the data lay in the ability to compare harm rates in the randomly selected discharges sample and randomly selected deceased discharges samples, as there was a suspicion that the study's mortality review process, similar to the UMR, was picking up features of the dying process rather than harm per se. The harm signature presented to health boards contained no surprises, but was useful in validating existing findings from Datix® (Datix, London, UK) and existing clinical dashboards.

Conclusions

The rate of harm detected across NHS Wales, using both the two-stage retrospective review process and the new Harm2 tool, conforms to that found by previous studies conducted in both the UK and Europe. Chronic disease is not strongly associated with the risk of incurring AEs or complications and data do not need to be collected routinely on the presence of all comorbid conditions for improvement purposes.

With training and using a structured review process, non-physician reviewers undertake case note review efficiently and effectively, and rates of AEs, preventability and breakdown of problems in care conform to studies in which physicians undertake these classifications. The determination of preventability is, however, challenging for reviewers. In 78% of cases there is agreement between professional groups on the determination of harm events. There are, however, differences in the events identified.

The Harm2 tool performed with moderate reliability in the determination of AEs, although screening criteria can be further refined. The rate and composition of AEs are influenced by both the balance of cases selected from different specialties and whether the patient died or was discharged alive.

A pragmatic approach with a robust measurement structure offers the potential for routine screening of AEs in NHS organisations. However, stability in case note review teams is highly valued to ensure reliability and highlight local learning.

Although chronic disease is not strongly associated with the risk of incurring AEs or complications, older patients with pre-existing diagnoses of dementia, PVD and hemiplegia appear to be at greater risk of AEs, probably as a result of underlying functional and cognitive impairment. Our data suggest that organisational quality improvement efforts to improve the safety of the ward environment may reduce the levels of AEs in these groups.

The difficulties encountered by our reviewers in navigating and finding information, in both paper and digitised case notes, suggest that similar difficulties may be encountered in clinical practice.

It would be relatively straightforward to devise a means of continuing the current approach by sampling across the country on a regular basis and undertaking comparative work with incident reporting systems, but in the longer term we need to set measures of harm alongside measures of the beneficial effects of health care first at the level of populations and, then, more ambitiously, for individual patients.

Funding

Funding for this study was provided by the Health Services and Delivery Research programme of the National Institute for Health Research.

Health Services and Delivery Research

ISSN 2050-4349 (Print)

ISSN 2050-4357 (Online)

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HS&DR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hsdr. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Services and Delivery Research* journal

Reports are published in *Health Services and Delivery Research* (HS&DR) if (1) they have resulted from work for the HS&DR programme or programmes which preceded the HS&DR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

HS&DR programme

The Health Services and Delivery Research (HS&DR) programme, part of the National Institute for Health Research (NIHR), was established to fund a broad range of research. It combines the strengths and contributions of two previous NIHR research programmes: the Health Services Research (HSR) programme and the Service Delivery and Organisation (SDO) programme, which were merged in January 2012.

The HS&DR programme aims to produce rigorous and relevant evidence on the quality, access and organisation of health services including costs and outcomes, as well as research on implementation. The programme will enhance the strategic focus on research that matters to the NHS and is keen to support ambitious evaluative research to improve health services.

For more information about the HS&DR programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hsdr

This report

The research reported in this issue of the journal was funded by the HS&DR programme or one of its preceding programmes as project number 09/2001/28. The contractual start date was in February 2011. The final report began editorial review in February 2016 and was accepted for publication in June 2016. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HS&DR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).