Routine mortality monitoring for detecting mass murder in UK general practice:
test of effectiveness using modelling

Bruce Guthrie, Tom Love, Rebecca Kaye, Margaret MacLeod and Jim Chalmers

ABSTRACT

Background
The Shipman Inquiry recommended mortality rate monitoring if it could be 'shown to be workable' in detecting a future mass murderer in general practice.

Aim
To examine the effectiveness of cumulative sum (CUSUM) charts, cross-sectional Shewhart charts, and exponentially-weighted, moving-average control charts in mortality monitoring at practice level.

Design of study
Analysis of Scottish routine general practice data combined with estimation of control chart effectiveness in detecting a 'murderer' in a simulated dataset.

Method
Practice stability was calculated from routine data to determine feasible lengths of monitoring. A simulated dataset of 405 000 'patients' was created, registered with 75 'practices' whose underlying mortality rates varied with the same distribution as case-mix-adjusted mortality in all Scottish practices. The sensitivity of each chart to detect five and 10 excess deaths was examined in repeated simulations. The sensitivity of control charts to excess deaths in simulated data, and the number of alarm signals when control charts were applied to routine data were estimated.

Results
Practice instability limited the length of monitoring and modelling was consequently restricted to a 3-year period. Monitoring mortality over 3 years, CUSUM charts were most sensitive but only reliably achieved >50% successful detection for 10 excess deaths per year and generated multiple false alarms (>15%).

Conclusion
At best, mortality monitoring can act as a backstop to detect a particularly prolific serial killer when other means of detection have failed. Policy should focus on changes likely to improve detection of individual murders, such as reform of death certification and the coroner system.

Keywords
family practice; homicide; outcome and process assessment (health care); quality assurance, health care; regulation.

INTRODUCTION

The UK GP Harold Shipman murdered over 200 patients during his 23-year career (Figure 1).[1,2] Although unusual in the number he killed, there are other examples of serial killing by health professionals across the world.[3] The Shipman Inquiry that followed his conviction recommended a package of reforms.[4] Reforms of the NHS complaints systems, the control and monitoring of opiate drugs, and regulation and revalidation of doctors have been implemented to some degree.[5] Two other recommended reforms were particularly directed at increasing the chance of deterring and detecting a future murderer. These are the closer scrutiny of individual deaths by a reformed death certification and coroner system,[6] and the use of routine mortality monitoring in general practice. The Shipman Inquiry proposed that the latter should be 'seriously considered' as part of the wider package of reforms.[4] However, at the time of writing, only mortality monitoring seems likely to be implemented soon, with the Chief Medical Officer recommending that the NHS should further develop and pilot: 'a national

B Guthrie, PhD, MRCGP, professor of primary care medicine; T Love, BA, MPH, CSO research fellow, Division of Community Health Sciences, University of Dundee, Dundee; R Kaye, BSc, statistician; M MacLeod, BSc, PhD, quality improvement programme manager; J Chalmers, MRCGP, FFPH, consultant in public health medicine, Information and Statistics Division, NHS National Services Scotland, Edinburgh, Scotland.

Address for correspondence
Professor Bruce Guthrie, Tayside Centre for General Practice, Community Health Sciences, University of Dundee, Mackenzie Building, Kirsty Semple Way, Dundee DD2 4BF. E-mail: b.guthrie@chs.dundee.ac.uk
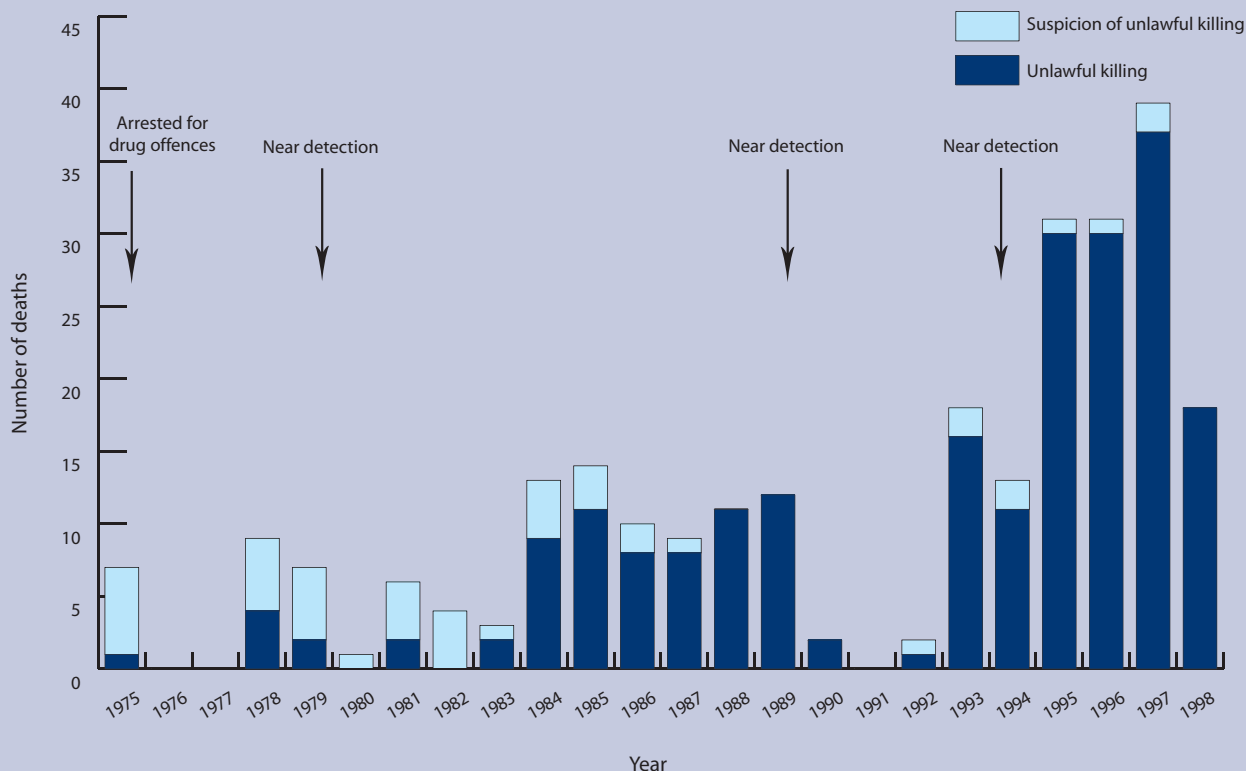
How this fits in

Previous work examining GP-specific mortality rates has shown that Harold Shipman would have been identified as having a high mortality rate, but the shift to practice-based registration in 2004 adds new uncertainty to the effectiveness of mortality monitoring to detect a murderous GP. For practice-level monitoring intended to detect an individual murderer, instability of practices limits the length of time that monitoring can be used for up to 3 years at best, since the assumption of any monitoring system should be that a murderer will move practice to avoid detection. Over 3 years, none of the control charts examined would reliably detect any murderer except one with Shipman's *modus operandi* (murder at home) and his mid-career rate of killing (10 patients annually). Mortality monitoring cannot substitute for reforms intended to ensure a proper account of every death.

system for death monitoring as part of a wider clinical quality assurance framework'.[5]

Since Shipman's conviction, two systems of mortality monitoring in UK general practice have been examined (Appendix 1). Aylin *et al* retrospectively applied a cumulative sum (CUSUM) chart monitoring system to general practice mortality data for 1993–2000.[7,8] Altogether, 3.3% of GPs were identified as having higher than expected mortality rate, including Harold Shipman. Notably, Shipman was not the most extreme outlier, despite murdering at least 142 people in this period. No cause for concern was found on investigation of the others.[4,9]

Based on simulation results, the study concluded that after 7 years' monitoring, CUSUM charts would reliably detect a GP murdering 10 patients per year (estimated successful detection rate after 7 years' monitoring; $eSDR_7$) 82–97% depending on where the alarm threshold was set). However, monitoring individual GPs' mortality rates is not feasible since the shift to practice-based registration in 2004. At practice level, the $eSDR_7$ was only 41–75%.[8]

The Northern Ireland pilot monitored practice mortality rates using cross-sectional Shewhart charts for 5 years' aggregated case-mix-adjusted mortality data.[10,11] Altogether, 15.8% of practices were identified as having higher than expected mortality. Investigation was based on a quality-improvement model, and so was open and collaborative.[12] After investigation, unmeasured case-mix heterogeneity was considered to be the cause of all high-mortality alarms.[10] The ability of the system to detect a murderer could not be examined, because there was no known murderer in the dataset.

Both studies concluded that mortality monitoring was feasible, and would detect future murderers. However, during a pilot to examine implementation issues of the Northern Ireland system in Scotland, doubts arose about this conclusion for a number of reasons.

Firstly, murder detection requires that murderers cannot avoid monitoring. In the Northern Ireland

Figure 1. Shipman Inquiry findings of unlawful killings by Harold Shipman 1975–1998.[1]

pilot, 22.5% of practices merged or split over a 5-year period, and were therefore not included in monitoring. Even if a practice is in continual existence, a murderer could still avoid detection by changing practice frequently enough that the excess mortality attributable to them was not detected. Although it is potentially dangerous to design a monitoring system based on one case, Shipman himself appeared to change practice in 1991 following a near detection (Figure 1).[4] The key point is that to be effective, any monitoring system for murder detection has to include all or nearly all potential murderers.

Secondly, the total number of alarms generated by each system is critical to their likely success. In the Shipman dataset, 3.3% of GPs crossed the chosen signalling threshold,[7] compared to 15.8% of practices in Northern Ireland.[10] Investigation of alarms in a mortality-monitoring system to detect murderers has to be confidential and forensic so as not to prejudice any future police investigation.[4,13,14] Minimising false alarms is important to avoid wasting resources and the potential harmful effects of such investigations on the innocent.

Finally, it is uncertain which control chart is preferable in terms of maximising detection and minimising false alarms. Although the comparative power of CUSUMs and longitudinal Shewhart charts is well known,[15] cross-sectional Shewhart charts have little industrial use and their performance in this circumstance is less certain.

The aim of this study is to examine the effectiveness of routine mortality monitoring to detect murderers in general practice, to inform decisions over implementation. As mass murder is rare, this question cannot be examined using real data. Under these circumstances, modelling using a combination of real data and simulation informed by real data can inform decision making. Therefore, the analysis uses NHS Scotland routine data to examine coverage (the proportion of practices and GPs being monitored consistently over time) and define the parameters of a simulated dataset. This information helps define the parameters of a simulated dataset, that is used to examine rates of successful detection of 'murderers' by different control charts.

METHOD

Coverage of practices and GPs over different time periods

Practice and GP codes were extracted from the NHS Scotland GP registration list on 31 December 2005,[16] and the proportion of practices in existence over periods of 1, 3, 5, and 10 years measured. For practices in existence for the whole of each period, the proportion of practices where the same GPs

were practising together at the beginning and end of each period was calculated. These results were used to determine the number of years of cumulative data that could be reasonably included in monitoring. The assumption was that any murderer would actively seek to avoid detection by changing practice and, therefore, that the monitoring system should be tested for periods over which nearly all practices would be included in monitoring, and where the GPs working in those practices were reasonably static.

Detection of excess deaths in the simulated dataset

The simulated dataset consisted of 405 000 'patients' registered with 75 'practices', each with a list size of 5400 (representing a notional Scottish health board where all practices have Scottish mean list size). Each 'practice' was assigned a mean total and out-of-institution mortality rate for each of the 3 years. The assigned rate was the actual Scottish mean death rate for financial years 2001–2002 to 2003–2004 (~1.1% for total mortality, ~0.3% for out-of-institution mortality), varied by a normally distributed random factor to model unmeasured case-mix heterogeneity. The variability introduced was designed to reflect actual variation in mortality rates in routine Scottish practice data after adjustment for patient age, sex, and deprivation.

Using these individual 'practice' rates, a chance experiment was conducted for every 'patient', with each assigned to 'survive' or 'die', repeated 1000 times for each year. Excess deaths annually were then added to one practice, representing the presence of a murderer. Five and 10 excess deaths were chosen, modelling Shipman's early and mid-career rate of killing (Figure 1). The 'murderer' was placed in practices at the upper and lower quartile of unmeasured case-mix heterogeneity.

The ability to detect excess deaths compared to those expected if each practice had the Scottish mean mortality rate was determined. Three control charts with widespread use in industrial or healthcare monitoring were examined: normal log-likelihood

Table 1. Practice and GP stability for 1049 Scottish practices in existence on 31 December 2005, n (%).

Years of monitoring	Practices continually open over that period	Practices open over that period AND with no change in GP	Practices open over that period AND with no change in GP principals, salaried GPs or GP retainers
1	1037 (98.9)	866 (82.6)	807 (76.9)
3	1006 (95.9)	627 (59.8)	423 (40.3)
5	963 (91.8)	432 (41.1)	308 (29.4)
10	888 (84.7)	220 (21.0)	165 (15.7)

Table 2. Control chart sensitivity in the simulated dataset, for practices at the lower (first number) and upper (second number) quartile of case-mix-adjusted mortality.

| | Sensitivity, % | | | |
| | Out-of-institution mortality | | Total mortality | |
	10 excess deaths/year	5 excess deaths/year	10 excess deaths/year	5 excess deaths/year
Shewhart chart				
95.0% control limit	81.8–100	7.9–92.5	8.7–91.7	0.9–64.2
99.0% control limit	54.7–100	2.2–78.8	1.9–76.4	0.0–38.9
99.9% control limit	23.1–98.4	0.4–54.8	0.2–49.9	0.0–17.1
CUSUM[a]				
$k = 1, h = 3$	95.1–100	44.3–95.9	53.2–95.4	26.6–85.8
$k = 1, h = 5$	88.8–99.6	27.3–90.7	43.5–93.0	19.3–80.6
$k = 2, h = 3$	93.2–100	41.6–93.8	50.8–94.1	26.3–84.2
$k = 2, h = 5$	84.7–99.4	25.0–88.1	41.6–91.6	18.6–78.6
EWMA[b]				
$\lambda = 0.2, L = 2$	68.6–74.2	34.7–61.1	15.0–35.8	4.1–27.1
$\lambda = 0.2, L = 3$	53.6–58.1	21.7–46.2	10.1–24.9	2.4–18.7

[a]*In this dataset, 'k of 1' represents a CUSUM chart designed to detect ~4 excess deaths annually, and a k of 2 one designed to detect ~8. 'h' is the threshold triggering an alarm (a larger h means the CUSUM statistic must be more extreme to trigger an alarm).*
[b]*Exponentially-weighted, moving-average (EWMA) charts use 2 consecutive years' data. λ is the weight given to the previous year's data, and L is the width of the control limit in standard deviations. An L of 2 ≈ 95%, and an L of 3 ≈ 99.8% control limits. CUSUM = cumulative sum.*

cumulative sum (CUSUM) charts as used in the analysis of the Shipman dataset;[7,8] cross-sectional control charts based on a funnel plot design with exact control limits,[17] similar to the Shewhart charts used in the Northern Ireland pilot;[10] and exponentially-weighted, moving-average (EWMA) charts.[13] The performance of each chart was examined using a range of chart parameters and detection thresholds.

Numbers of alarms requiring investigation
To examine how many alarms would require

Table 3. Number of practices needing investigation for out-of-institution mortality monitoring in 998 Scottish practices 2001–2002 to 2003–2004.

	Number (%) of practices needing investigation, $n = 998$
Shewhart chart	
95.0% control limit	31 (3.1)
99.0% control limit	6 (0.6)
99.9% control limit	2 (0.2)
CUSUM	
$k = 1, h = 3$	282 (28.2)
$k = 1, h = 5$	169 (16.9)
$k = 2, h = 3$	261 (26.2)
$k = 2, h = 5$	149 (14.9)
EWMA	
$\lambda = 0.2, L = 2$	284 (28.5)
$\lambda = 0.2, L = 3$	206 (20.6)

CUSUM = cumulative sum. EWMA = exponentially-weighted, moving-average. See Table 2 footnote for information on symbols.

investigation, the same control charts were applied to actual Scottish practice out-of-institution mortality data for 2001–2002 to 2003–2004 and the number of practices that signalled an alarm was counted.

RESULTS
Coverage of practices and GPs by length of time monitoring occurs
Practice stability in all Scottish practices in existence on 31 December 2005 is shown in Table 1. Although monitoring is at practice level, the intention of monitoring for murder detection is to identify individual killers. Although practices are reasonably stable over 5 years, there is much greater instability of individual clinicians composing those practices. Therefore, it was decided to model 3 years as the longest period over which a monitoring system could reliably accumulate data at practice level with the intention of detecting a murderous individual. The assumption made was that any mass murderer knowing that practice mortality was being monitored would seek to avoid detection by changing practice. Even over 3 years, only 60% of practices have had no change of GP principal, and only 40% no change in any GP.

Detection of excess deaths in the simulated dataset
Table 2 shows successful detection rates in the simulated dataset for the three control charts, for five and 10 excess deaths annually for 3 years. The range shown is for a murderer in a practice with an underlying mortality rate at the lower (first number) and upper (second number) quartiles of case-mix-adjusted heterogeneity. No chart reliably detected excess deaths under all circumstances. EWMA charts were generally insensitive. Shewhart charts were only highly sensitive for 10 excess deaths at home annually (Shipman's *modus operandi* and mid-career-murder rate). CUSUM charts were the most sensitive, but did not achieve consistently >50% detection rates in a lower-quartile practice, except for 10 excess deaths per year.

Numbers of alarms requiring investigation
Shewhart charts produced few alarms requiring investigation, whereas both CUSUM and EWMA charts produced alarms for between 15% and 28% of practices (Table 3).

DISCUSSION
Summary of main findings
Coverage of mortality monitoring at practice level is significantly limited by the relative instability of the individual clinicians who make up practices. Reflecting the incomplete case-mix adjustment

possible with existing routine data, the sensitivity of the three control charts examined was poor, except for the detection of a murderer with Shipman's mid-career rate of killing (10 excess deaths at home annually). All detection systems generated large numbers of false signals.

Strengths and limitations of the study

The key strength of this study is that it has used simulation based on routine general practice data to examine carefully a problem for which there is no existing, high-quality dataset that can be used for analysis.

The analysis has several important assumptions and limitations. Firstly, it has only examined mortality monitoring for murder detection. Mortality monitoring for quality improvement might be worthwhile in its own right although this is also uncertain. It is also unclear whether it is possible to combine open, collaborative systems of investigation in quality improvement, with the confidential, forensic investigation required for murder detection.[10,11,18]

Secondly, the choice of monitoring over 3 years is to some extent arbitrary. The choice was informed by real data on Scottish practice stability, and the use of monitoring that was being examined. Although 85% of practices could be monitored over 10 years, the GPs who comprise those practices are much less stable. Even over 3 years, 40% of practices have at least one GP principal arrive or leave. The authors believe that a practice-based system for detecting an individual murderer should assume that any future mass murderer would actively change practice to avoid detection because, unlike Shipman, they would know that practice mortality rates were being monitored. For quality improvement, where the main focus is on the practice, the period of monitoring is less critical, and could be for any period that the practice exists.

Thirdly, all practices in the analysis were assumed to be average sized. Detection would be more likely in smaller practices, and less likely in larger ones. However, it cannot be assumed that a murderer would be in a single-handed or a small practice. Although Shipman killed more people after going single-handed in 1991, the Shipman Inquiry still concluded that he killed over 100 people while working in his previous seven-doctor practice.[1]

Fourthly, detection of excess deaths is assessed only for practices at the upper and lower quartile of case-mix-adjusted mortality rates. The actual detection rates of any implemented monitoring system may therefore be better or worse depending on the true underlying, case-mix-adjusted mortality rate and size of the practice that a murderer happens to work in. However, the data presented give a reasonable indication of the range in likely detection rates for a single round of monitoring.

Comparison with existing literature

This study's findings are broadly consistent with the practice-level analysis of Aylin *et al* that used different methods, where estimated successful detection rates after 7 years of monitoring were 41–75%.[8]

Implications for future research and practice

Although Harold Shipman was a particularly prolific murderer, he was not unique.[3] All healthcare systems have to address the detection and deterrence of serial killers working as healthcare professionals. Routine mortality monitoring to detect murderers is a rational response,[4] but the present findings highlight the limitations of monitoring at practice level when the true unit of interest is the individual practitioner. However, individual practitioner monitoring is not feasible in the UK, and even practice-based monitoring may not be feasible in countries where mandatory primary care registration is not the norm.

The large number of practices needing investigation is a significant limitation of monitoring.[8] Even an alarm rate of only a few per cent would still require up to 40 in-depth investigations annually in Scotland or an English strategic health authority. Compared to collaborative quality-improvement investigations of control chart signals, forensic examinations would have greater potential for harmful consequences to any innocents investigated, which would be expected to make quality-improvement use of the data difficult.[13] More seriously, any system that repeatedly generates false alarms over many years makes it more likely that a (presumably) rare true alarm will be ignored when it occurs.

Routine mortality monitoring in general practice is therefore only likely to signal an alarm for an extreme mass murderer like Shipman. A monitoring system that can only detect a serial killer after 30 or more people have been murdered is not 'effective' in any meaningful way, and represents a failure of other mechanisms intended to detect and deter murderers by ensuring that every death is properly accounted for.[4,19] Requiring two doctors to certify every death was recommended by the Shipman Inquiry to ensure proper accounting for every death, but would only be effective if the second certifier takes their job seriously and is open to the possibility of foul play, and if suspicions are rigorously investigated (a failing in Shipman's and other cases[3]). In Shipman's previous area of work, such a cultural change already appears to have happened,[19] but it is disappointing that recommended reforms for an enhanced coroner

COMMENTARY

Monitoring mortality: a quality improvement perspective

Guthrie *et al*'s modelling study demonstrates that monitoring mortality in general practices over a period of 3 years is unlikely to detect a murderer who kills fewer than 30 patients.[1]

Monitoring is a method for detecting and acting on signals in data for a range of problems — not just one problem. We monitor patients' temperature postoperatively not just to detect wound infection, but any infection (or even hypothermia). We would certainly miss many infections if we relied solely on temperature. Yet nobody would suggest abandoning the temperature chart. Just as the primary purpose of monitoring the patient's temperature is to improve their health, similarly the primary purpose of monitoring should be quality improvement, not detection of murder.[2–4]

Monitoring serves quality improvement by identifying unusual (special cause) variation, investigating, and learning from such a process. This means systematic investigation to identify data errors, the influence of case-mix, the resourcing, organisation, and delivery of health care. The aim is to learn why mortality might vary and take appropriate action. For mortality differences the most common explanations are unmeasured differences in case-mix.[2] The actions of individual clinicians are among the last in a series of potential special causes.

Mortality monitoring was one of the recommendations of the Shipman Inquiry.[2] Others included closer scrutiny of controlled drug prescribing, changes to the coroner system, and more rapid investigation of complaints. The focus on a single recommendation in isolation seems inappropriate.

Could mortality monitoring have a role? Guthrie *et al* observe that that 85% of practices in Scotland could be monitored for 10 years, not the 3 years that they model. The NHS's present largely *ad hoc* and haphazard system failed to detect Shipman's 200 murders. So, detecting even an excess of 30 deaths is better by comparison. However, the focus on seeking murderers is inappropriate. A better question is how often were useful lessons learned from investigating special cause variation? First lessons include correcting data errors and improving understanding of case-mix. Later lessons will be to understand how resourcing and organisation of health care might influence mortality.

Ultimately, quality improvement requires more than monitoring. It requires the trust, commitment, and cooperation of GPs. Any monitoring system that fails on these counts is unlikely to be effective in either securing quality of care or finding 'bad apples'. To start from a position of trust would be for GPs themselves to agree what indicators should be monitored and to decide how they will investigate special cause variation.

Tom Marshall
Senior Lecturer, Department of Public Health and Epidemiology, University of Birmingham, Edgbaston, Birmingham, B15 2TT.
Mohammed A Mohammed
Senior Research Fellow, Department of Public Health and Epidemiology, University of Birmingham.

REFERENCES

1. Guthrie B, Love T, Kaye R, *et al*. Modelling the effectiveness of routine mortality monitoring in detecting mass murder in UK general practice. *Br J Gen Pract* 2008: **58(550):** 311–317. DOI: 10.3399/bjgp08X280164

2. Shipman Inquiry. *The Shipman Inquiry Fifth Report. Safeguarding patients: lessons from the past — proposals for the future.* Norwich: HMSO, 2004.

3. Mohammed MA, Booth K, Marshall D, *et al*. A practical method for monitoring general practice mortality in the UK: findings from a pilot study in a health board of Northern Ireland. *Br J Gen Pract* 2005; **55:** 670–676.

4. Chief Medical Officer. *Good doctors, safer patients: proposals to strengthen the system to assure and improve the performance of doctors and to protect the safety of patients.* London: Department of Health, 2007.

system to ensure that doctors apply due diligence to accounting for every death appear to have been rejected.[6,20] Based on this analysis, monitoring mortality rates alone is not enough and cannot substitute for other recommended reforms.[4] This study did not examine whether it is worth nationally implementing routine general practice mortality monitoring for quality improvement. Although this has some face validity, there is no strong evidence that such a system would improve the quality of care.[18] However, for the parallel aim of murder detection, then mortality monitoring could at best operate as a backstop to catch a prolific serial killer who has evaded detection by other means.

Commentary

A commentary accompanies this article:
DOI: 10.3399/bjgp08X280173

expressed

Ethical approval

The analysis used publicly available data, and no NHS ethics approval was required

Competing interests

The authors have stated that there are none.

Discuss this article

Contribute and read comments about this article on the Discussion Forum: http://www.rcgp.org.uk/bjgp-discuss

REFERENCES

1. Shipman Inquiry. *The Shipman Inquiry First Report. Death disguised.* Norwich: The Stationery Office, 2002.

2. Baker R. *Harold Shipman's clinical practice 1971–1998.* London: The Stationery Office, 2001.

3. Yorker BC, Kizer KW, Lampe P, *et al.* Serial murder by healthcare professionals. *J Forensic Sci* 2006; **51(6):** 1362–1371.

4. Shipman Inquiry. *The Shipman Inquiry Fifth Report. Safeguarding patients: lessons from the past — proposals for the future.* Norwich: HMSO, 2004.

5. Chief Medical Officer. *Good doctors, safer patients: proposals to strengthen the system to assure and improve the performance of doctors and to protect the safety of patients.* London: Department of Health, 2007.

6. Richards T. Chairwoman of Shipman Inquiry protests at lack of action. *BMJ* 2006; **332(7550):** 1111.

7. Aylin P, Best N, Marshall C. Following Shipman: a pilot system for monitoring mortality rates in primary care. *Lancet* 2003; **362(9382):** 485–491.

8. Aylin P, Best N, Bottle A, Marshall C. *Monitoring of mortality rates in primary care.* London: Imperial College of Science, Technology and Medicine, 2003.

9. Mohammed MA, Rathbone A, Myers P, *et al.* An investigation into general practitioners associated with high patient mortality flagged up through the Shipman inquiry: retrospective analysis of routine data. *BMJ* 2004; **328(7454):** 1474–1477.

10. Mohammed MA, Booth K, Marshall D, *et al.* A practical method for monitoring general practice mortality in the UK: findings from a pilot study in a health board of Northern Ireland. *Br J Gen Pract* 2005; **55(518):** 670–676.

11. Guthrie B. Can mortality monitoring in general practice be made to work? *Br J Gen Pract* 2005; **55(518):** 660–663.

12. Lilford R, Mohammed AM, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004; **363(9415):** 1147–1154.

13. Baker R, Jones DR, Goldblatt P. Monitoring mortality rates in general practice after Shipman. *BMJ* 2003; **326(7383):** 274–276.

14. Dyer C. Investigators should be trained to 'think dirty' about cause of death, Shipman Report says. *BMJ* 2003; **327(7407):** 123.

15. Montgomery DC. *Introduction to statistical quality control.* Chichester: John Wiley and Sons, 1991.

16. Information and Statistics Division NHS in Scotland (ISD). Workforce statistics. http://www.isdscotland.org/isd/796.html (accessed 9 Apr 2008).

17. Spiegelhalter D. Funnel plots for comparing institutional performance. *Stat Med* 2005; **24(8):** 1185–1202.

18. Baker R, Sullivan E, Camossa-Stefinovich J, *et al.* Making use of mortality data to improve quality and safety in general practice: a review of current approaches. *Qua Saf Health Care* 2006; **16(2):** 84–89.

19. Tanna V. Moving on from Shipman. *BMJ* 2005; **331(7513):** 411.

20. Dyer C. Reforms no bar to new Shipman. *Guardian* 2006; 8 Feb http://www.guardian.co.uk/law/story/0,,1705035,00.html (accessed 9 Apr 2008).

Appendix 1. Reported systems for mortality monitoring in general practice.

Criteria	Retrospective examination of Shipman dataset[7,8]	Northern Ireland pilot[10,11]
Purpose or mortality monitoring	Detection of murder	Primarily, quality improvement; secondarily, detection of murder
Description of monitoring system	Total mortality for individual GPs with correction for over-dispersion was monitored for 1993–2000. Normal log-likelihood ratio CUSUM charts used for detection system. Alarm signals were subsequently investigated in two ways, one based on an open quality-improvement model,[9] one more forensic and confidential.[4] Neither found cause for concern	Five-year aggregated case-mix-adjusted total mortality for practices was monitored for 1996–2000. Cross-sectional Shewhart charts with three standard deviation control limits used for detection system. Alarm signals were openly investigated using the pyramid model of quality improvement. No cause for concern was found
Coverage — what percentage of practices/GPs/nurses were monitored?	• 11.8% of practices • 37.3% of GPs individually monitored • No nurse individually monitored • Poor data quality in 1990s meant most practices and GPs excluded	• 77.5% of practices • No individual monitoring of GPs or nurses • Nearly a quarter were excluded because of experiencing a major merger or split
Ability to detect murderers	Practice monitoring $eSDR_7$ = 74.7% ($k = 1, h = 3$)[a] $eSDR_7$ = 41.2% ($k = 1, h = 5$) GP monitoring $eSDR_7$ = 96.6% ($k = 2, h = 3$)[a] $eSDR_7$ = 82.4% ($k = 2, h = 5$)	Uncertain
False alarms (percentage of GPs /practices signalling, where later investigation concludes there is no cause for concern)	3.3% of GPs ($k = 2, h = 3$) 2.3% of GPs ($k = 2, h = 5$)	15.8% of practices Reflects quality of data available for case-mix adjustment
How long must reliable data be available for these conclusions to hold?	7 years	5 years
Comments	Conclusions rely on being able to monitor individual GPs' mortality rates which is no longer even theoretically feasible with the shift to practice-based registration in 2004	Effectiveness of mortality monitoring to detect mass murder is assumed rather than directly examines

[a]$eSDR_7$ (estimated successful detection rate after 7 years of monitoring) = proportion of truly out of control units successfully detected after 7 years. k and h are chart parameters. k relates to the size of the excess mortality the chart is tuned to detect. For the Shipman dataset, k = 1, ≈ 6 excess deaths for a system monitoring at GP level, and ≈ 13 excess deaths at practice level (for k = 2, ≈12 and 26 respectively). eSDRs shown therefore refer to charts tuned to detect Shipman in his mid-career (approximately 10 murders per year). h is the threshold that the chart statistic has to exceed to trigger an alarm (larger h means deviation from expected mortality has to be greater to signal an alarm). CUSUM = cumulative sum.