

DEVELOPING RESEARCH AND PRACTICE

Control, compare and communicate: designing control charts to summarise efficiently data from multiple quality indicators

B Guthrie, T Love, T Fahey, A Morris, F Sullivan

Qual Saf Health Care 2005;14:450–454. doi: 10.1136/qshc.2005.014456

Summarising the complex data generated by multiple cross sectional quality indicators in a way that patients, clinicians, managers and policymakers find useful is challenging. A common approach is aggregation to create summary measures such as star ratings and balanced score cards, but these may conceal the detail needed to focus quality improvement. We propose an alternative way of summarising and presenting multiple quality indicators, suitable for use for quality improvement and governance. This paper discusses (1) control charts for repeated measurements of single processes as used in industrial statistical process control (SPC); (2) control charts for cross sectional comparison of many institutions for a single quality indicator (rarely used in industry but commonly proposed for health care); and (3) small multiple graphics which combine control chart signal extraction with efficient graphical presentations for multiple indicators.

outliers, an explicit choice needs to be made about where to set confidence or control limits.

This paper focuses particularly on the use of control chart techniques when applied to data from multiple quality indicators. As examples it uses Scottish immunisation data,¹⁰ data extracted from a comprehensive, population based, externally validated diabetes register in the Tayside region of Scotland,¹¹ and data from one practice collected as part of the requirements of the new UK general practice contract.

LEAGUE TABLES

League tables are ubiquitous in health care and other sectors for interpreting cross sectional quality data comparing different institutions. Figure 1 shows a league table for the proportion of patients with type 2 diabetes whose glycated haemoglobin is $\leq 7.4\%$ in the 12 months before 31 December 2003 in all general practices in the Tayside region of Scotland. Practices are plotted in ascending order of measured performance and the horizontal line shows the regional mean (55.6%) for comparison.

Such a league table has the advantages of familiarity and ease of interpretation, but the disadvantage of implying considerable variation between practices and overemphasising the ranking and the ends of the distribution (the “best” and “worst”).³ However, ranks are statistically highly unreliable, and most of the variation implied cannot be distinguished from chance.^{8,12} More sophisticated versions follow biostatistical convention by including 95% confidence intervals around each practice point to allow a test of whether the practice is different from the mean, but are less commonly used (fig 2). Although they more appropriately account for chance variation, they require some prior knowledge or instruction to interpret.

In the UK and internationally, comparison of the quality of care in different healthcare institutions is increasingly common, but there is uncertainty about the most effective way to present the data to clinicians, managers, or patients.^{1–3} The way that data are presented is known to affect interpretation of treatment effectiveness and risk,⁴ and there is some evidence that the form of graphic chosen influences interpretation of quality data by the public^{5,6} and professionals.⁷

This paper focuses on comparative analysis of primary care performance data from a quality improvement perspective, where measures are treated as *indicators* that prompt further investigation rather than being used to make definitive judgements that a practice is “good” or “bad”.

A number of general considerations are applicable, irrespective of the form of graphical analysis used. Firstly, the quality measures themselves have to be valid and reliable.^{8,9} Secondly, careful consideration should be given to what the appropriate comparator is—which may be locality, regional and national means, or comparisons against practices serving similar populations in terms of deprivation, urban/rural, or age distribution (as a way of incorporating some case-mix adjustment). Finally, where statistical techniques are used to identify

CONTROL CHARTS FOR LONGITUDINAL DATA

Control charts are a tool developed for industrial statistical process control (SPC) where they are used to examine the performance of a single process over time. All processes are considered subject to *common cause variation*, which is the sum of all random events influencing the process being measured. Such variation is predictable within a range defined by statistical theory and requires no intervention. *Special* or *assignable causes* are specific non-random disturbances

See end of article for authors' affiliations

Correspondence to:
Dr B Guthrie, Department of Community Health Sciences, University of Dundee, MacKenzie Building, Kirsty Semple Way, Dundee DD2 4BF, UK; b.guthrie@chs.dundee.ac.uk

Accepted for publication 11 September 2005

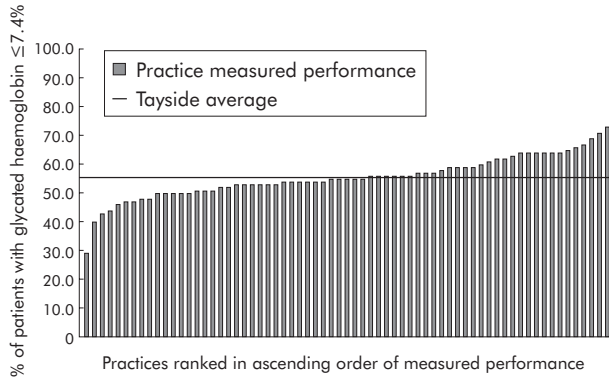


Figure 1 Simple league table and regional mean for glycated haemoglobin control in type 2 diabetes in Tayside practices.

which control charts are designed to identify to allow intervention to remove them.^{13 14}

Figure 3 shows a longitudinal control chart¹³ for the percentage of Scottish children with completed primary tetanus immunisation at the age of 1 year.¹⁰ The central horizontal line shows the mean for June 1996 to December 2000 (95.2%). Following US industrial convention, control and warning limits set at three and two standard deviations from this mean are shown by the two outer and inner lines.¹⁴ There are a number of potential signals of special cause variation, several of which are present here (box 1).

Firstly, immunisation rates were lower than expected in the four quarters from March 2001. Secondly, more recent performance has improved (10 successive points plot above the historical mean). Thirdly, there is a seasonal pattern in the early part of the time series. Investigating the reasons why special cause variation is present may identify remediable problems or identify examples of good practice that can be generalised. For example, lower rates in 2001 were probably caused by a vaccine shortage¹⁵ (the data for 2 year olds shows no dip, consistent with some delayed immunisation). Equally, if recent improvement is predominately in a few Health Boards, then there may be generalisable lessons for other areas.

The attraction of SPC methods for longitudinal data is that they are statistically informed and rigorous, but pragmatic with a long history of use in other settings. Users are not required to understand the underlying statistics because the chart summarises complex data by signalling likely special cause variation to prompt appropriate further investigation or action.

Box 1 Commonly used signals of special cause variation in longitudinal control charts¹³

Signals based on plotting a single new point

- Any single point outside a control limit

Signals based on patterns of two or more plotted points

- Two out of three consecutive points between a warning limit and a control limit
- Eight or more consecutive points on one side of the mean
- Eight or more consecutive points in a continually ascending or descending run
- Any unusual or non-random pattern

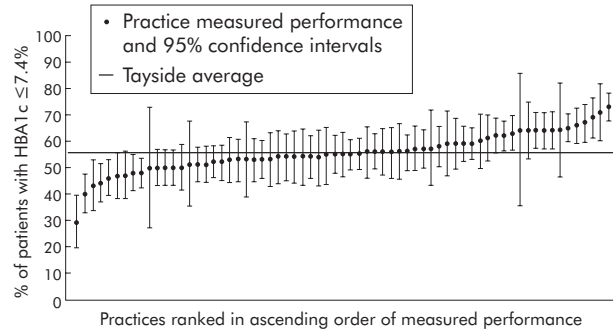


Figure 2 League table and 95% confidence intervals for glycated haemoglobin control in type 2 diabetes in Tayside practices.

CROSS SECTIONAL CONTROL CHARTS FOR SINGLE MEASURES

Unlike industrial uses where the focus is on longitudinal measurement of single settings,^{13 14} healthcare quality data analysis is usually of cross sectional data from many settings. Control charts for cross sectional data have been proposed,^{3 16} and fig 4 uses a funnel plot design to analyse the same diabetes data as in fig 1.¹⁷ The horizontal line shows the regional mean, with exact 95% warning limits and 99% control limits plotted around it. Practices plotting outside the control limits are considered to show special cause variation requiring further investigation or action. However, because each practice only contributes one plotted point, the only extractable signal that depends on patterns of two or more points is that the chart may give a visual indication of systematic differences in quality between smaller and larger practices.¹⁷

Although their interpretation is less intuitive than the league table, most professionals can use them appropriately with minimal instruction.⁷ Control charts avoid the problems of ranking in league tables, and clearly indicate that most of the variation between practices is what would be expected by chance or common cause variation. They have been used where a single quality indicator is assumed adequately to capture the overall quality of care for an episode, particularly in the analysis of surgical mortality.¹⁸⁻²⁰

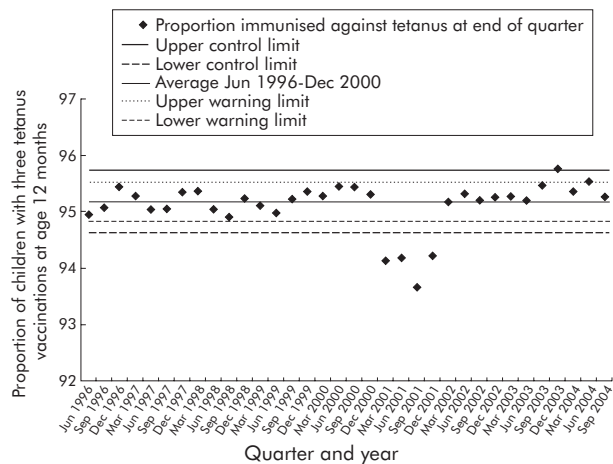


Figure 3 Control chart for percentage of 1 year old children in Scotland with three tetanus vaccinations (p chart for proportional data with varying denominator).

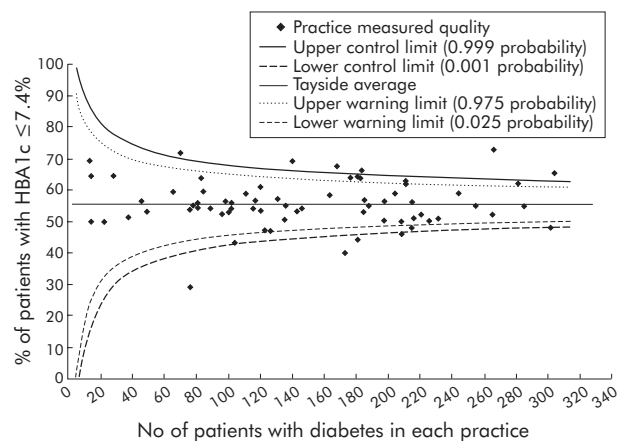


Figure 4 Cross sectional control chart for percentage of patients with type 2 diabetes with HBA1c $\leq 7.4\%$ in Tayside practices.

SMALL MULTIPLE GRAPHICS BASED ON CONTROL CHART SIGNALS

The assumption that single measures are adequate proxies for overall care is less sustainable for chronic diseases where multiple quality indicators are applicable.²¹ The new GMS contract includes 65 indicators for 10 conditions that could be compared using cross sectional control charts like that in fig 3.²² However, 65 separate charts would not facilitate the detection of patterns across measures, and occasional false positive signals are inevitable given the multiple comparisons being made. One approach to this problem is to aggregate data into a smaller number of measures, as happens with “star ratings” for hospital and primary care trusts in England and Wales.²³ However, the hidden assumptions underlying the construction of such aggregates (including how different measures are weighted) make them relatively opaque to users.²⁴

An alternative is to create forms of data presentation that facilitate the detection of patterns in the original data structure. An attractive concept is that of small multiple graphics commonly used in the consumer press which are “inevitably comparative, deftly multivariate, shrunken high density graphics, ... efficient in interpretation, often narrative in content.”²⁵ (page 175)

In the context of healthcare quality indicators, each cross sectional control chart can be reduced to a set of varying strength signals of evidence of special cause variation. Figure 5 shows an example intended to facilitate comparison of neighbouring practices for a locality based quality improver. A colour version can be found in the online supplement available at www.qshc.com/supplemental. It displays comparative data for 13 indicators of the quality of type 2 diabetes care in 14 practices in one locality. Each dot encodes the control chart signal for one indicator in one practice, where the comparator is the overall Tayside mean and control and warning limits are defined with exact 99% and 99.9% probability. Examining the columns can identify practices outlying on multiple indicators where there are likely to be systematic factors affecting quality for better or worse (for example, practices 4 and 10). Single signals may still be deserving of attention but are less likely to be meaningful given the multiple comparisons being made (for example, practices 12 and 3). Examining the rows can identify indicators where there may be a more global problem across the locality compared with the area mean (for example, the pattern seen for foot examination may relate to access to podiatry within this locality).

Figure 6 takes a similar form but is designed to facilitate comparison of indicators within one practice. It compares related indicators across disease areas. Examining the columns allows comparisons of diseases across common indicators (for example, quality of care for diabetes in this practice appears generally better than that for other conditions). Examining the rows compares indicators shared across disease areas (for example, the chart suggests a potential quality problem across the diagnostic indicators). In principle, for a single practice all 65 of the continuous indicators in the new GMS contract could be summarised on a single sheet.

The key advantage over multiple graphs for single measures is that it is easier to detect patterns in the data to help users interpret complex sets of indicators that are related to each other. Although we have used control chart techniques to signal that a practice is different from average, the same kind of graphic could be constructed using signals from league tables with varying width of confidence interval.

CONCLUSIONS

Single measure control charts are more statistically robust than simple league tables, but neither is ideal where quality

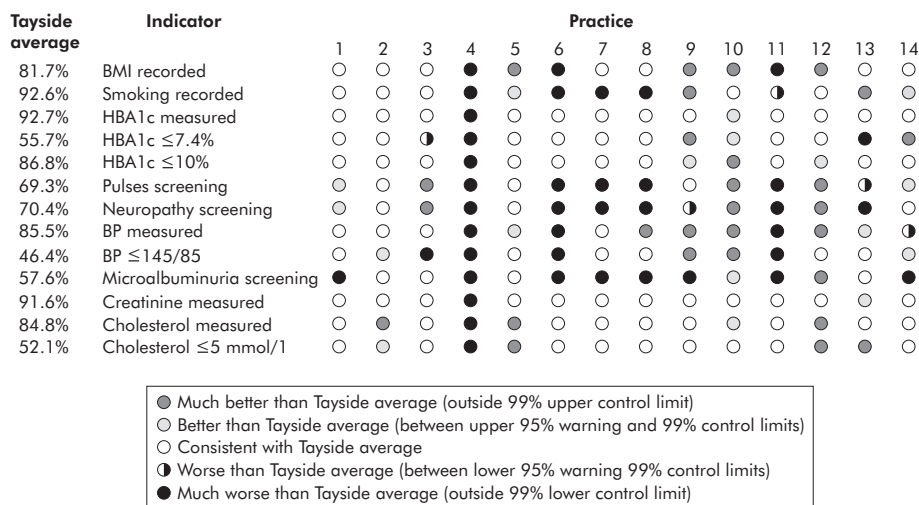


Figure 5 Control chart signals for 13 measures in 17 practices in one locality (practices compared with Tayside regional mean).

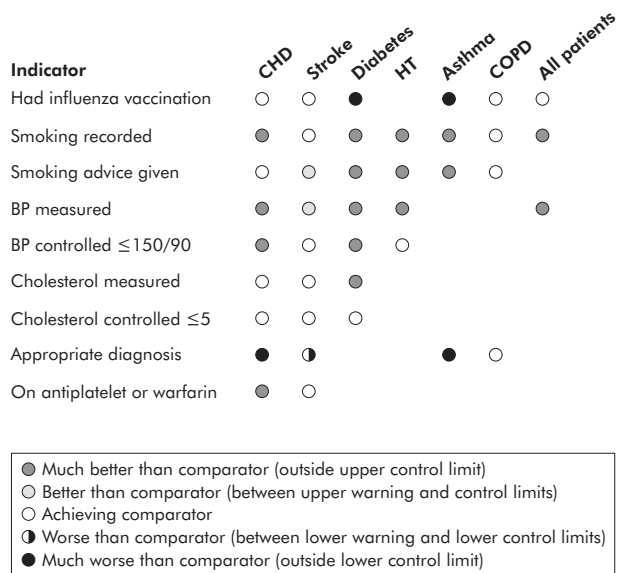


Figure 6 Selected new GMS contract indicators for one practice compared against arbitrary standard.

measurement requires multiple indicators. In contrast, small multiple graphics are an efficient tool for screening data from multiple quality indicators to prompt reflection, further investigation, or action. They embody the statistically informed pragmatism of longitudinal control charts to facilitate detection of meaningful patterns in complex data to allow users to hypothesise and investigate causes. This form of control chart has strong face value but shares key uncertainties with league tables and single measure control charts.

Firstly, the paper has used regional means against which to compare practices since all are part of a regional managed clinical network. In other circumstances, locality or national comparisons may be more appropriate. Rather than seeing one comparator as “best”, the most appropriate one for the purpose in hand should be sought or multiple comparators used to test explanations for patterns detected in the data.

Secondly, it is uncertain how wide confidence intervals or control limits should be in this context. In league tables, biostatistical convention uses 95% confidence intervals because these have been found to be useful in medical research. Three standard deviation ($\approx 99.3\%$) limits are used in industry because, in this setting, they appropriately balance sensitivity (correctly identifying special causes) and specificity (avoiding potentially costly false alarms).²⁶ Routine cardiac surgery mortality monitoring uses 99.99% limits, justifying this because unmeasured case mix heterogeneity is inevitable in health care.¹⁹ Others have used narrower 99% limits,²⁰ reflecting the criticism that wider limits are overprotective of surgeons and probably too insensitive to true outliers.²⁷

We have chosen exact 99% and 99.9% limits to create signals, but what is appropriate will depend on what the quality data are used for. If the consequences of being identified as a “poor performer” are potentially extreme (closing down a hospital unit, suspending an individual), then wide control limits that prioritise specificity are more appropriate. However, if data are being used within a supportive quality improvement framework, then narrower limits that prioritise sensitivity should be used. Rather than following any particular convention, users should explicitly decide what is most appropriate for their particular purposes.

Finally, graphical tools should be designed to suit the needs of their intended audience. This paper focuses on use by professionals and managers, but other graphical analyses are likely to better suit patient and public use for choice or accountability.^{5,6} Although there are grounds for believing that well designed graphical analyses can promote quality improvement, the actual usefulness of any particular design can only be judged after implementation. We plan to pilot the implementation of these designs within the single Scottish Diabetes IT system (SCI-DC) to examine this further.

Immunisation data were supplied by the Information and Statistics Division of NHS Scotland, the GP contract data by the Ferguson Medical Practice, and the diabetes data by the DARTS/MEMO. For the latter, the authors thank Philip Thomson and Douglas Boyle for assistance in data extraction, GPs and other clinicians in Tayside, and the members of the DARTS Steering Group (D Boyle, B Brennan, K Boyle, J Broomhall, F Cargill, P Clark, A Connacher, S Cunningham, E Dow, D Dunbar, A Dutton, S Greene, K Hunter, R Jung, M Kenicer, B Kilgallon, G Leese, R Locke, T MacDonald, R McAlpine, S McKendrick, R Newton, P Slane, F Sullivan, R Walker, S Young). DARTS is supported by The Scottish Executive, Tenovus Tayside, NHS Tayside, Tayside University Hospitals NHS Trust and Tayside Primary Care NHS Trust.



Colour versions of figures 5 and 6 are shown in the online supplement available at the QSHC website www.qshc.com/supplemental.

Authors' affiliations

B Guthrie, T Love, T Fahey, A Morris, F Sullivan, Department of Community Health Sciences, University of Dundee, Dundee DD2 4BF, UK

This study was funded by the Health Foundation and the Chief Scientist's Office of the Scottish Executive Health Department.

Competing interests: none.

BG had the idea for the paper, designed the first version of the small multiples, and wrote the first draft. All authors discussed the results, contributed to development of the design, and wrote the paper.

Ethical approval for the study was granted by Tayside local research ethics committee

REFERENCES

- 1 **Werner RM**, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA* 2005;**293**:1239–44.
- 2 **Rand Health**. Report cards for health care: is anyone checking them? Research Briefing RB4544. Available at <http://www.rand.org/publications/RB/RB4544/index.html>, 2002.
- 3 **Adab P**, Rouse AM, Mohammed AM, et al. Performance league tables: the NHS deserves better. *BMJ* 2002;**324**:95–8.
- 4 **McGettigan P**, Sly K, O'Connell D, et al. The effects of information framing on the practices of physicians. *J Gen Intern Med* 1999;**14**:642.
- 5 **Hibbard JH**, Slovic P, Peters E, et al. Strategies for reporting health plan performance information to consumers: evidence from controlled studies. *Health Serv Res* 2002;**37**:291–313.
- 6 **Vaiana ME**, McGlynn EA. What cognitive science tells us about the design of reports for consumers. *Med Care Res Rev* 2002;**59**:3–35.
- 7 **Marshall T**, Mohammed AM, Rouse AM. A randomised trial of league tables and control charts as aids to health services decision making. *Int J Qual Health Care* 2004;**16**:309–15.
- 8 **Bird S**, Cox D, Farewell H, et al. *Performance indicators: good, bad and ugly*. London: Royal Statistical Society Working Party on Performance Monitoring in the Public Services, 2003.
- 9 **Guthrie B**, Emslie-Smith A, Morris A, et al. Quality measurement of care for people with type 2 diabetes in Tayside, Scotland: implications for the new UK general practice contract. *Br J Gen Pract* 2003;**53**:709–13.
- 10 **NHS Scotland**. *Scottish health statistics: childhood immunisations*. Available at http://www.isdscotland.org/isd/info3.jsp?pContentID=1652&p_applic=CCC&p_service=Content.Show&2005 (accessed 16 March 2005).
- 11 **Morris AD**, Boyle DIR, MacAlpine R, et al. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. *BMJ* 1997;**315**:524–8.

- 12 **Goldstein H**, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. *J R Stat Assoc A* 1996;**59**:385–443.
- 13 **Montgomery DC**. *Introduction to statistical quality control*. Chichester: John Wiley, 1991.
- 14 **Oakland JS**. *Statistical process control*. London: William Heinemann, 1986.
- 15 **Scottish Centre for Infection and Environmental Health**. DTwP vaccine shortage. *SCIEH Weekly Report* 1999;**33**:1.
- 16 **Mohammed AM**, Cheng AK, Rouse AM, et al. Bristol, Shipman and clinical governance: Shewart's forgotten lessons. *Lancet* 2001;**357**:4637–9.
- 17 **Spiegelhalter D**. Funnel plots for comparing institutional performance. *Stat Med* 2005;**23**.
- 18 **Tekkis PP**, McCulloch P, Steger AC, et al. Mortality control charts for comparing performance of surgical units: validation study using hospital mortality data. *BMJ* 2003;**326**:789.
- 19 **Keogh B**, Spiegelhalter D, Bailey A, et al. The legacy of Bristol: public disclosure of individual surgeon's results. *BMJ* 2004;**329**:454.
- 20 **Bridgewater B**. Mortality data in adult cardiac surgery for named surgeons: retrospective examination of prospectively collected data on coronary artery surgery and aortic valve replacement. *BMJ* 2005;**330**:506–10.
- 21 **Kirk SA**, Campbell SM, Kennell-Webb S, et al. Assessing the quality of care of multiple conditions in general practice: practical and methodological problems. *Qual Saf Health Care* 2003;**12**:421–7.
- 22 **NHS Confederation and British Medical Association**. *Investing in general practice: the new GMS contract*. London: British Medical Association, 2003.
- 23 **Healthcare Commission**. *2004 Performance Ratings*. Available at <http://www.chi.nhs.uk/home.asp>, 2004 (accessed 15 February 2005).
- 24 **Jacobs R**, Smith P, Goddard M. *Measuring performance: an examination of composite performance indicators*. CHE Technical Paper Series 29. York: Centre for Health Economics, University of York, 2004.
- 25 **Tufte E**. *The visual display of quantitative information*. Cheshire, Conn: Graphics Press, 1983.
- 26 **Deming WE**. *Out of the crisis*. Cambridge, MA: MIT Press, 2000.
- 27 **Blackstone EH**. Monitoring surgical performance. *J Thorac Cardiovasc Surg* 2004;**128**:807–10.