Lessons in variation with case studies from health care, including Bristol and Shipman

Mohammed A. Mohammed
K. K. Cheng
Andrew Rouse
Tom Marshall
Departments of Surgery and Public Health and Epidemiology
University of Birmingham

Abstract

During the last century, manufacturing industry has achieved great success in improving quality of its products. A critical factor in this success story has been the use of Walter A Shewhart's pioneering (c 1920's) work in the economic control of variation, which culminated in the development of a simple yet powerful graphical method known as the control chart. The control chart, classifies variation as common cause or special cause and thus guides the user to the most appropriate action to effect improvement. Using six case studies, including Bristol and Shipman, this paper clearly demonstrates a central role for Shewhart's approach in turning the rhetoric of clinical governance into a reality.

Introduction

During the last century, manufacturing industry has achieved great success in improving quality of its products. In industry, the definition of quality is "on target with minimum variation"[1]. Reducing variation is also a core concern in clinical governance[2], however there are fundamental and profound differences between the ways in which the NHS and industry make sense of variation.

This paper begins with an illustration of the industrial approach to understanding and controlling variation followed by application of this approach to health care, using six clinical governance case studies, viz: Bristol cardiac surgery, Dr Shipman, in-vitro fertilisation (IVF) treatment, mortality following fracture neck of femur, neonatal deaths and prevalence of coronary heart disease (CHD) in primary care.

Two kinds of variation

Consider a process such as writing a signature. Five of MAM's signatures are shown in the left of Figure 1. Although these signatures were produced under the same conditions and by the same process, they are not identical. On the other hand, although they show variation, the variation is controlled within "limits". They are all recognisably the same signature. This kind of variation suggests a stable process produced the signatures.

In healthcare we use three basic approaches to make sense of variation: standard setting, league tables and hypothesis testing. Were we to compare the five signatures to a standard, some could fall below the standard. We could rank the signatures from best to worst and create a league table. A statistical test might identify one signature as being significantly different from the others. Each of these approaches focuses attention on those signatures that fail the test. This is wrong. From the point of view of the underlying process of writing signatures, all these signatures are identical. No signature is better or worse than the others. If we want to reduce the variation between the signatures, we must change the way we write all signatures, not just the ones that fail an arbitrary test.

Now consider the signature on the right. It is clearly different from the others. A casual look suggests that there must be a special reason why this is so. If we want to address this kind of variation, we need to identify this special cause and prevent it from interacting with an otherwise stable process. (In this case, the signature is a forgery!)

This approach categorises variation according to the action needed to reduce it.

Common cause variation is intrinsic to the process. To reduce common cause variation we need to act on the process. Special cause variation is the result of factors extrinsic to the process. To reduce special cause variation we need to identify and act on the special causes.

The originator of these fundamental concepts was a physicist and engineer - Walter A Shewhart[3]. His pioneering work at Bell laboratories in the 1920's, successfully brought together the disciplines of statistics, engineering and economics leading to the accolade "Father of modern quality control"[4].

Shewhart devised a simple graphical method, the control chart, for discriminating between the two sources of variation, thereby guiding the user to take appropriate action. The control chart has three lines. The central line is the mean and the upper and lower lines are termed control limits. Control limits represent the limits of common cause variation. Shewhart[3], with the aid of mathematical theory, empirical evidence and practical concerns advocated the use of limits set at three standard deviations around the mean. A data point that falls outside these control limits (or unusual patterns on the control chart) suggest a special cause.

Variation can not be eliminated

Shewhart illustrated his concepts with a data set obtained from an experiment in which almost everything possible was done to obtain perfect results (ie no variation) - Millikan's Nobel Prize winning measurements of the charge of an electron[3]. Despite Millikan's best efforts, there was considerable variation in his measurements of the charge of an electron. However, as the control chart (Figure 2) of Millikan's data shows, all measurements fall within the upper and lower control limits suggesting that his experiment was stable. To suggest to Millikan that some of his measurements were better than others, or some fell below an acceptable standard, would be absurd. To award him a gold star for some of these values and to punish him for other values would be equally absurd. Stable processes exhibit common cause variation, which is best reduced by action on the underlying process.

Case Study 1 – Bristol Cardiac Surgery

A control chart (see Appendix for method) of the mortality rates data for children aged less than a year old during three epochs[5] is shown in Figure 3.

The chart for epoch 1 will be used to explain the interpretation of a control chart.

In epoch 1, the mortality rates for nine hospitals lie within the control limits: common cause variation. Action to reduce this variation must focus on the underlying process of care common to these nine hospitals. However, two hospitals (hospitals 11 and 7) are outside the control limits and this indicates that there are special causes. In hospital 11 it is important to learn why the mortality rates are high. To do this we systematically look at data collection, patient case-mix, facilities and quality of care. We must then take remedial action to help this hospital eliminate the special cause. In hospital 7, the mortality rate is low. It is important to learn why their results are better than other hospitals. If appropriate, we can use this knowledge to improve the results of all the hospitals.

In epoch 2, two hospitals (hospitals 10 and 11) show evidence of special cause variation. Hospital 10 is in need of investigation and help to eliminate the special cause. In contrast, hospital 11 has shown remarkable improvement in its results. In epoch 1 it was above the upper control limit and in epoch 2 (and subsequently) it is below the lower control limit. It is also in need of investigation. Understanding why hospital 11 has made such striking progress offers an opportunity for learning, which could help the results of all. Alternatively, it may indicate that there have been important changes in the case-mix of patients treated at hospital 11.

In epoch 3, two hospitals (hospital 1 and hospital 11) show special cause variation. Hospital 11, as in epoch 2, is below the lower control limit. Hospital 1 (Bristol Royal Infirmary) is above the upper control limit. It is in need of help to identify and eliminate the special cause for its high mortality rate.

Whilst the statistical evidence presented to the Bristol inquiry raised concern over the performance of two units[5], the control chart analysis offers much more. The Bristol enquiry focused attention on Bristol alone. Control charts suggest other units for investigation. The Bristol enquiry only identified poor performers. Control charts identify both good and bad performers. The Bristol enquiry could only criticise the poor performance. Control charts enable us to highlight opportunities for improvement by learning from centres with low mortality rates (below the lower control limit). The Bristol enquiry took action after 1995. Control charts provided a basis for action by 1987.

Case Study 2 – Dr Shipman

A control chart (Figure 4) of mortality rates of women aged 65 years and over in Thameside and Glossop 1992-1998[6] shows that during 1992 and 1994, Dr Shipman's mortality rates were within common cause variation. However, during 1993 and 1995-8 his mortality rates indicated special cause variation. Action to reduce special cause variation is to find the special cause and remove it. Subsequent legal proceedings identified that special cause as being Dr Shipman himself.

Commentators have argued that the Shipman case was not an example of poor quality of care; rather Shipman was a murderer who happened to be practicing medicine[7]. This may be so, but in Shewhart's approach murder is just one of an infinite number of special causes.

Case Study 3 – IVF Treatment

Marshall and Spiegelhalter[8] analysed case-mix adjusted live birth rate from 52 in-vitro fertilisation clinics in the UK (n=24739 treatment cycles, live birth rate range 5%-24%). They concluded that league tables were unreliable. No action point emerged from their analysis. In contrast, a control chart (Figure 5) with the upper and lower control limits divides the clinics into three groups with guidance for action:

Group A - performance above the upper control limit. Learn why their results are better than other clinics. Use this knowledge to improve the performance of all the clinics.

Group B - performance within the control limits. Make fundamental changes to the way in which IVF treatment is provided. This should be informed by lessons learned from Group A. There are no grounds for taking action in individual centres in this group.

Group C - performance below the lower control limit. Help these centres to identify and eliminate the special causes of their poor results.

Case Study 4 – Mortality after fractured hips

Todd et al[9] compared differences in mortality after fractured hip in 8 East Anglia hospitals (n=560, mortality range 5%-24%). A control chart (Figure 6) clearly shows 7 hospitals within common cause variation. Improvement at these 7 hospitals can only come from changing the underlying process of care for fractured hip patients. One hospital (number 6) had a very low mortality outside the limits of common cause variation – this mortality rate is therefore likely to have a special cause. According to Todd et al, this hospital had a well-organised multidisciplinary team, which sought early assessment and surgery, much of it performed by one surgeon followed up with early post-operative mobilisation of patients. Todd et al[9] were hesitant in recommending adoption of this hospitals' practice saying that "random variation" almost certainly plays a part in these findings. Shewhart's approach clearly shows that hospital 6 belongs to another system beyond that attributable to random and/or common cause variation. The control chart provides us with a basis for action. The model of care at hospital 6 should be more widely adopted perhaps after a randomised controlled trial. No action is not an option.

Case Study 5 – Neonatal deaths

Parry et al[10] compared mortality for nine neonatal units (n=2671 infants, mortality range 15%-28%), concluding that league tables were unreliable indicators of performance. In contrast, a control chart of the neonatal data (Figure 7) shows only common cause variation suggesting that future improvement is best sought from a fundamental change to the underlying process of care. There are no grounds for taking action on individual neonatal units.

Case Study 6 – Prevalence of coronary heart disease in primary care

The point prevalence of coronary heart disease in a primary group consisting of sixteen general practices in Birmingham was reported (private communication Birmingham

Health Authority 1999) as 9.67% (2999/3102), with wide variation (1%-38%) between practices. A control chart of this data (Figure 8) clearly identifies twelve practices within control limits indicating common cause variation. These practices should be left alone. However, five practices are outside the control limits indicating special cause variation. Two practices have much higher prevalence rates than expected from common cause variation. Special cause action needs to be taken. This should also explore the possibility of double counting. As regards the three practices below the lower control limit, they require special cause action, which should begin with a review of the data collection process.

Discussion

These case studies clearly illustrate an important role for Shewhart's approach to understanding and reducing variation. They demonstrate the simplicity and power of control charts at guiding their users towards appropriate action for improvement.

Actions based on Shewhart's approach are subject to two types of mistakes[11]. Mistake 1 is to treat an outcome resulting from a common cause as if it were a special cause. Mistake 2 is to treat an outcome resulting from a special cause as if it were a common cause. It is impossible to reduce the frequency of both errors to zero. What we can do is minimise the economic losses due to either kind of mistake. Shewhart (1931) argued that variation from stable processes lies within limits which - combining mathematical theory, empirical evidence and pragmatism - can be most usefully set at three standard deviations from the mean.

Some may consider that limits of three standard deviations are too wide a range for health care. The use of a narrower range, say two standard deviations, may seem more appealing. But there is a need for caution. Firstly as demonstrated by the electron data, stable systems can and do produce data beyond two standard deviations. So we will be guided to look for trouble more often then it actually exists (Mistake 1). Given the culture of blame in the health service we risk making matters worse, especially when the person closest to the failure is often held to be responsible. Furthermore, the case studies used here demonstrate that the three standard deviation limits are adequate to find special cause variation in practice. In any case, the sensitivity of the control chart can be increased by applying supplementary rules[1] for detecting special causes of variation based on unusual patterns in the data.

Perhaps it would be optimistic to suggest that use of control charts could prevent the recurrence of tragic and unfortunate episodes such as Bristol or Shipman. What is clear is that analysing data with an understanding of common cause and special cause variation provides the NHS with a basis to act. There is an axiom - often forgotten in the health services - that the purpose of data is action[11]. Each of the above case studies is based on data available at the time of the events. In each case, little or no action was taken at the time. Why? We believe this is largely because the current methods for understanding variation in the health service provide little or no guidance for action. One prominent advocate of Shewhart's method was so convinced of this that he wrote "Tests of significance, t-test, chi-square, are useless as inference – ie, useless for aid in prediction. Test of hypothesis has been for half a century a bristling obstruction to understanding statistical inference."[11]. At least we should seriously question the role of conventional statistical analysis in clinical governance.

Shewhart's concepts provide a sober antidote to the plague of league tables. Under stable conditions, league tables are unreliable and their guidance is equally unreliable[12]. Action based on their guidance is likely to be misguided resulting in tampering – making matters worse[11].

The era of clinical governance and performance management offers immense opportunities. In the past those in possession of data may have opted for inaction or called for better data. Recent high profile cases have contributed to conditions where the tendency for action will be more frequent. The case for the control chart to guide action has been presented. Its guidance has proved immensely useful to industry over the last fifty years[1,11,14]; it is time for it to be integrated into performance management in healthcare.

MAM is keen to explore the use of these techniques in education and would be pleased to hear from interested parties.

References

1 Wheeler D. Advanced topics in statistical process control. The power of Shewhart's charts. SPC press Inc 1995. USA.

2 Scally G, Donaldson LJ. The NHS's 50 anniversary. Clinical governance and the drive for quality improvement in the new NHS in England. BMJ 1998;317:61-5.

3 Shewhart WA. Economic Control of Quality of Manufactured Product. D Van Nostrand Company New York, 1931. Reprint in 1980 by the ASQC Quality Press.

4 American Society for Quality http://www.asq.org/about/history/shewhart.html. Accessed Jan 2001.

5 Spiegelhalter DJ, Evans S, Aylin P, Murray G. Overview of statistical evidence presented to the Bristol Royal Infirmary Inquiry concerning the nature and outcomes of paediatric cardiac surgical services at Bristol relative to other specialist centres from 1984 to 1995." http://www.bristol-inquiry.org.uk/Documents/statistical%20overview%20report.pdf Also see Figure 6.1 http://www.bristol-inquiry.org.uk/Documents/fig61.pdf. Accessed Jan 2001.

6 Duce R, Jenkins R. GP probably killed more than 100. The Times. February 1 2000.

7 Anonymous. Shipman was a criminal and not a poorly performing doctor. BMJ 2000:320: 1734.

8 Marshall EC, Spiegelhalter DJ. Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates. BMJ 1998;316:1701-5.

9 Todd CJ, Palmer C, Camilleri-Ferrante C, et al. Differences in mortality after fracture of hip: the East Anglian audit. BMJ 1995;310:904-8.

10 Parry GJ, Gould CR, McCabe CJ, Tarnow-Mordi WO. Annual league tables of mortality in neonatal intensive care units: longitudinal study. BMJ 1998;316:1931-5.

11 Deming WE. The New Economics. Massachusetts Institute of Technology 1994 Second Edition.

12 Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance (with discussion). Journal of the Royal Statistical Society. Series A. 1998;159:3:385.

13 Deming WE. Sample design in business research. Wiley Classics Library Edition 1990.

14 Deming WE. Out of the crisis. Massachusetts Institute of Technology 1986.

15 Mostellor F, Tukey J. The uses and usefulness of probability paper. Jr Amer Stat Assoc. 1949;44:174-212.

Appendix

There are many different types of control charts[1]. The control charts presented with each case study were drawn using the methodology advocated by Deming[13,14] for binomial data. Such control charts are easily drawn on double square root paper (also known as probability paper) introduced in 1949 by Mostellor & Tukey[15]. Since the standard deviation on this type of paper is usefully regarded as a constant ½, the resulting control limits are straight lines above and below the mean.

Correspondence

Correspondence to:

Mohammed A Mohammed
Senior Research Fellow
Department of Public Health & Epidemiology
University of Birmingham
Edgbaston
Birmingham B15 2TT

(e-mail M.A.Mohammed@Bham.ac.uk)

Acknowledgement

This paper draws heavily on an earlier paper by the authors, published in The Lancet, Vol 357, 2001, pp.463–67.

Figure 1: "MAM" signatures: five original and one forged (right).

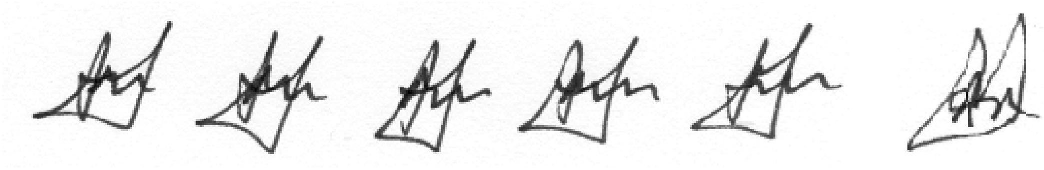

Figure 2: Shewhart control chart of Millikan's charge of an electron data.

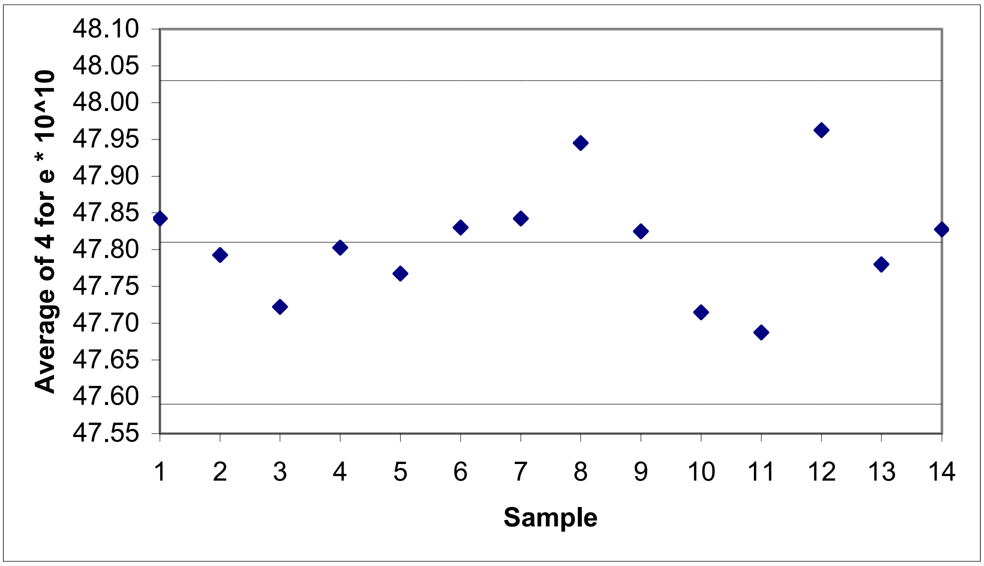

Figure 3: The Bristol Data. Mortality for under 1 year olds following Open Heart Surgery over three epochs. Hospital 1 is Bristol Royal Infirmary.

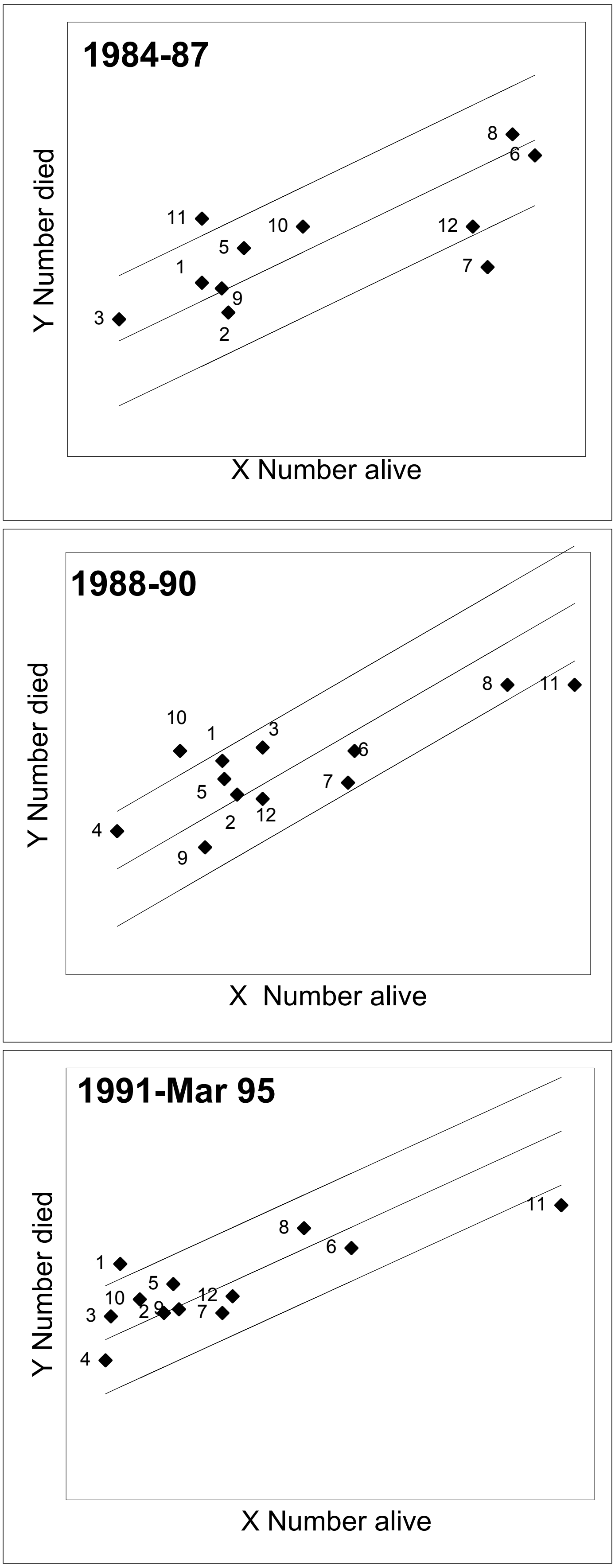

Figure 4: Dr Shipman's data. Comparing Dr Shipman's mortality for women aged 65 years or more in Thameside & Glossop during 1992-98. The three lines indicate the background variation, inclusive of Shipman, in this data set. Dr Shipman's annual mortality rates are imposed on this.

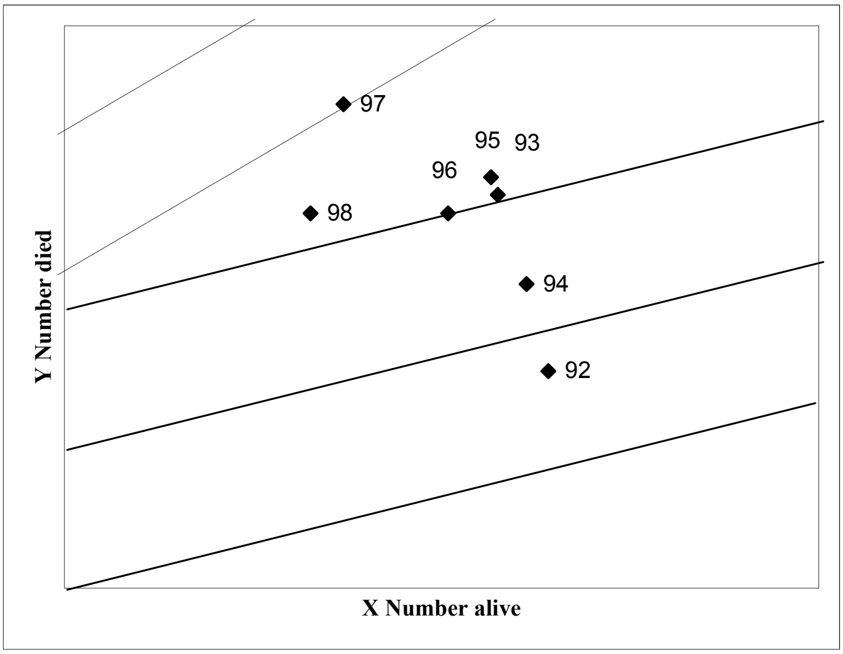

Figure 5: Case-mix adjusted live birth rates from 52 IVF centres in the UK (1996). Clinics above and below the limits indicate special cause variation whilst clinics within the limits indicate common cause variation.

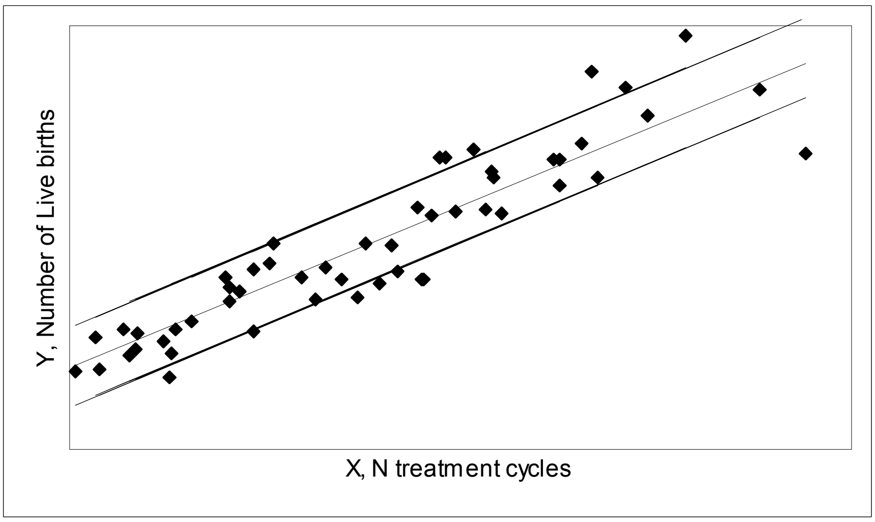

Figure 6: Mortality following surgery for fractured hips in 8 hospitals in the East Anglian Audit.

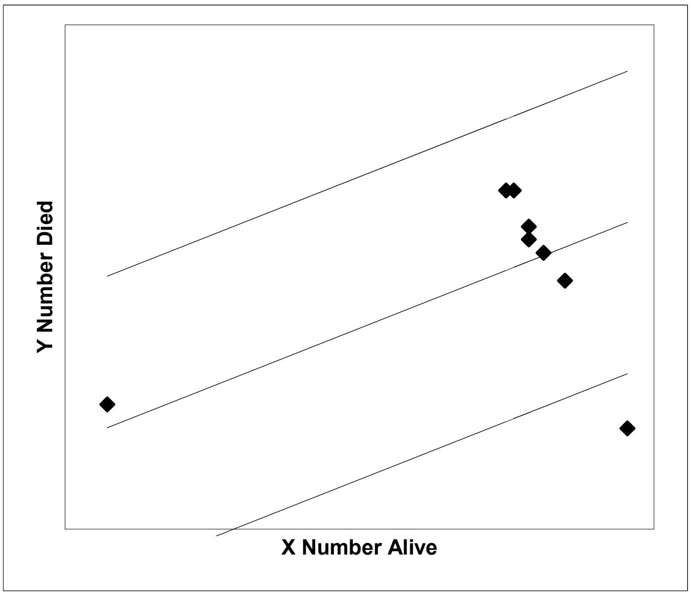

Figure 7: Mortality in nine neonatal units,

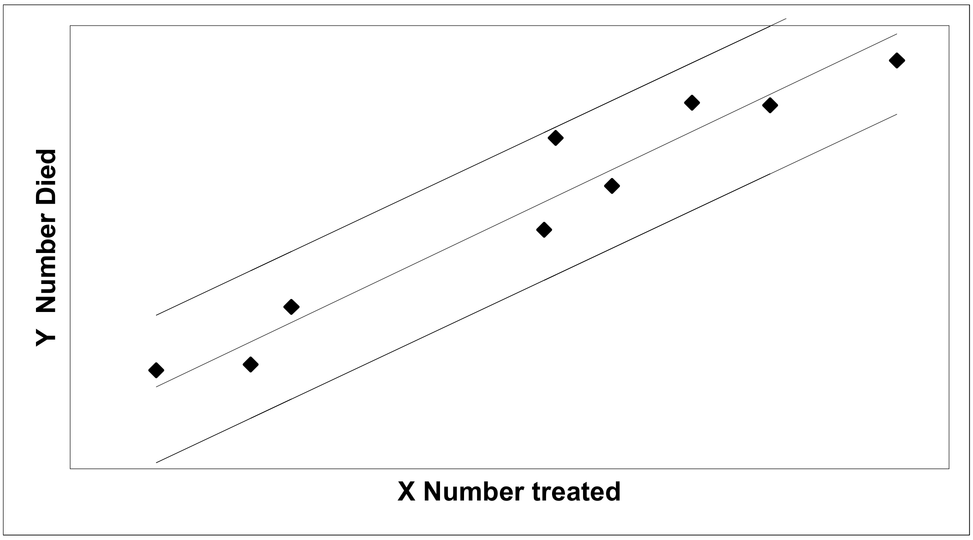

Figure 8: Prevalence of CHD in 16 general practices in one primary care group in Birmingham.