

Surgeon specific mortality in adult cardiac surgery: comparison between crude and risk stratified data

Ben Bridgewater, Antony D Grayson, Mark Jackson, Nicholas Brooks, Geir J Grotte, Daniel J M Keenan, Russell Millner, Brian M Fabri, Mark Jones on behalf of the North West Quality Improvement Programme in Cardiac Interventions



Abstract

Objective As a result of recent failures in clinical governance the government has made a commitment to bring individual surgeons' mortality data into the public domain. We have analysed a database to compare crude mortality after coronary artery bypass surgery with outcomes that were stratified by risk.

Design Retrospective analysis of prospectively collected data.

Setting All NHS centres in the geographical north west of England that undertake cardiac surgery in adults.

Participants All patients undergoing isolated bypass graft surgery for the first time between April 1999 and March 2002.

Main outcome measures Surgeon specific postoperative mortality and predicted mortality by EuroSCORE.

Results 8572 patients were operated on by 23 surgeons. Overall mortality was 1.7%. Observed mortality between surgeons ranged from 0% to 3.7%; predicted mortality ranged from 2% to 3.7%. Eighty five per cent (7286) of the patients had a EuroSCORE of 5 or less; 49% of the deaths were in this lower risk group. A large proportion of the variability in predicted mortality between surgeons was due to a small but differing number of high risk patients.

Conclusions It is possible to collect risk stratified data on all patients undergoing coronary bypass surgery. For most the predicted mortality is low. The small proportion of high risk patients is responsible for most of the differences in predicted mortality between surgeons. Crude comparisons of death rates can be misleading and may encourage surgeons to practise risk averse behaviour. We recommend a comparison of death rates that is stratified by risk and based on low risk cases as the national benchmark for assessing consultant specific performance.

failures of clinical governance in health care and is tied in with political initiatives about patients' choice. The planned date for publication of surgeon specific data in the United Kingdom is 2004, and although it has been accepted by the secretary of state for health that any such data should be robust, validated, and stratified for case mix to allow meaningful comparisons to be made, this type of dataset does not yet exist for all hospitals and surgeons.¹

The Society of Cardiothoracic Surgeons of the United Kingdom and Ireland had been planning to undertake an analysis on low risk patients, but because of a lack of an appropriate dataset it is now planning to publish individual surgeons' crude mortality data later this year.² Two possible datasets could be analysed in the United Kingdom: hospital episode statistics, which are known to be inaccurate at the level of individual clinicians, and returns of crude mortality that have been made to the Society of Cardiothoracic Surgeons of the United Kingdom and Ireland's annual register on the basis of individual surgeons since 1997. Neither dataset has been subjected to rigorous validation, and the society's returns have no mechanism for risk stratification. Although hospital episode statistics can be partly adjusted for case mix by age, sex, and urgency, these are known to be only a few of the many patient specific factors that contribute to predicted operative mortality.³⁻⁵ The Society of Cardiothoracic Surgeons has collected a more comprehensive dataset since 1996,¹ but this has been voluntary, and not all hospitals and surgeons have contributed. It cannot be used for a comprehensive comparative analysis in the United Kingdom.

We have a long track record of collecting cardiac surgery audit data and validating risk prediction models in Manchester.⁶⁻⁸ Recently we have collected a full dataset on all patients undergoing adult cardiac surgery in NHS institutions in the north west of England since April 1997.⁸⁻⁹ We have analysed this database to explore differences between crude mortality and risk stratified results for surgeon specific publication.

South Manchester University Hospital, Manchester M23 9LT
Ben Bridgewater
consultant cardiac surgeon
Nicholas Brooks
consultant cardiologist
Mark Jones
consultant cardiothoracic surgeon

Cardiothoracic Centre, Liverpool L14 3PE
Anthony D Grayson
regional clinical information analyst
Mark Jackson
head of clinical governance
Brian M Fabri
consultant cardiac surgeon

Manchester Royal Infirmary, Manchester M13 9WL
Geir J Grotte
consultant cardiac surgeon
Daniel J M Keenan
consultant cardiac surgeon

Blackpool Victoria Hospital, Blackpool FY3 HNR
Russell Millner
consultant cardiothoracic surgeon

Correspondence to: Ben Bridgewater ben.bridgewater@smuht.nwst.nhs.uk

BMJ 2003;327:13-7

Introduction

There is an unstoppable momentum towards the publication of surgeon specific mortality as part of the initiative to generate greater accountability and transparency in the NHS. This has been triggered by

Methods

The northwest quality improvement programme in cardiac interventions is a regional consortium involving all four NHS centres that perform adult cardiac

Table 1 Incidence of risk factors in the study population

Risk factor	No (%) of patients (n=8572)
Mean age at operation in years (range)	62.9 (31.3 to 88)
Female	1809 (21.1)
Angina class IV	1894 (22.1)
Myocardial infarction within 3 months	540 (6.3)
Respiratory disease	1603 (18.7)
Renal dysfunction	197 (2.3)
Peripheral vascular disease	1063 (12.4)
Cerebrovascular disease	703 (8.2)
Ejection fraction <30%	471 (5.5)
Intra-aortic balloon pump support	94 (1.1)
Ventilation support	9 (0.1)
Cardiogenic shock	43 (0.5)
Intravenous nitroglycerin	531 (6.2)
Urgent surgery	1706 (19.9)
Emergency surgery	146 (1.7)

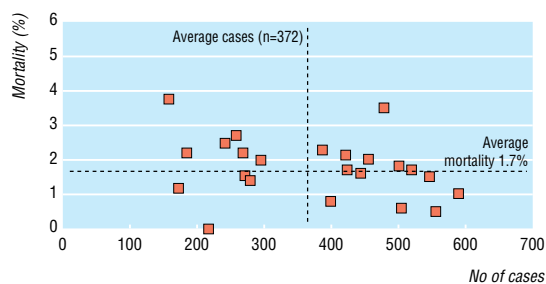
surgery and percutaneous coronary interventions in the north west of England (Blackpool Victoria Hospital, the Cardiothoracic Centre in Liverpool, Manchester Royal Infirmary, and South Manchester University Hospital). The aim of the group is to improve continuously the quality of care for patients receiving cardiac interventions by using a multicentred approach based on studies of institutional systems.⁹

We collected data prospectively on a total of 8572 consecutive patients undergoing isolated coronary artery bypass graft surgery for the first time between 1 April 1999 and 31 March 2002 in the north west of England. Data collection methods and definitions have been described in detail previously.^{8,9} Each patient had a dataset collected, which included data from before and after the operation, to enable a predicted mortality to be calculated. Data were collected in each institution and returned to a central source for analysis. Each centre conducted validation of activity and analysis. Mortality was defined as any in-hospital death. Every patient's record contained an anonymised identifier for each consultant surgeon. Data were analysed for all consultants who were operating in the region on 1 April 2003.

Design

The specific questions we addressed were:

- What was the overall mortality?
- What was the distribution of patients according to predicted operative risk?
- Were there differences in predicted mortality between surgeons?
- Were there differences in observed mortality between surgeons?

**Fig 1** Number of cases and crude mortality by consultant

- Could the dataset be used to stratify according to case mix to allow meaningful comparisons to be made?
- After appropriate analysis did the death rates between surgeons differ significantly?

Statistical analysis

Categorical data are shown as a percentage whereas continuous data are shown as a mean with a range. We determined crude mortality for each surgeon. We calculated predicted mortality for each patient by using the additive EuroSCORE,¹⁰ a scoring system derived from an analysis of 19 000 patients throughout Europe that was reported in 1999. The EuroSCORE ascribes additive points to several risk factors related to the patient and the procedure, to generate a predicted mortality for each patient. It has been shown to be a good overall predictor of mortality for both European and North American surgery.^{11,12} If a patient related factor necessary to calculate the EuroSCORE was missing in the record that factor was assumed to be absent (this occurred in less than 2% of cases). We examined the distribution of patients in each EuroSCORE group and compared the predicted mortality for each surgeon. Owing to the non-normal distribution of predicted mortality between surgeons, we determined variability by the interquartile range. We used the EuroSCORE to determine low (≥ 5) and high (> 5) risk groups¹⁰ and compared observed mortality and variability between the two groups. We calculated the C statistic (equivalent to the area under the receiver operating characteristic curve) to assess the performance of the EuroSCORE. A C statistic of greater than 0.7 indicates a reasonable ability to discriminate between patients who died and those who did not.^{13,14} We calculated the C statistic for the total population, low and high risk groups. We examined surgeon specific mortality in the low risk group by comparing each surgeon's death rates, plotted with 95% confidence intervals, against the mean performance for the region in this group of patients. We analysed the effect of volume of cases on mortality in the low risk group by the χ^2 test for trend after rank ordering surgeons and categorising them as either low, middle, or high volume thirds. We used SAS for Windows version 8.2 to perform all analyses.

Results

Description of the patients

A total of 8572 patients were included in the study. A summary of the incidence of risk factors is given in table 1. Altogether 144 patients died, which is a death rate of 1.7%. The average number of cases per consultant was 372 (range 158 to 598). Not all surgeons were operating throughout the three year period, which accounts for much of this variability. Figure 1 shows observed mortality for the surgeons and ranges from 0% to 3.8%.

Predicted and observed mortality

Figure 2 shows the number of patients in each EuroSCORE group. Predicted mortality in most patients was low. Figures 3 and 4 show the number of deaths and observed percentage mortality in each EuroSCORE group. A large number of patients were in the low EuroSCORE groups, but the percentage mortality was low (less than 2%). In general, observed

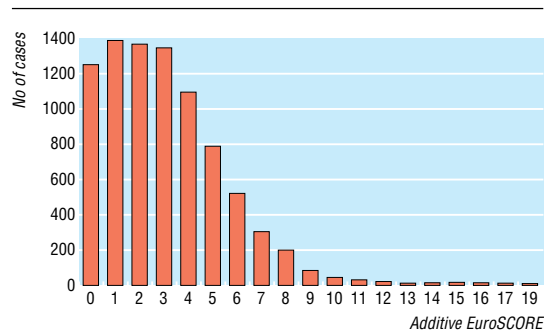


Fig 2 Number of patients in each EuroSCORE group

mortality increased with increasing EuroSCORE. For lower risk patients the EuroSCORE overpredicted observed mortality. In EuroSCORE groupings of 14 and above, the observed mortality was substantially in excess of the EuroSCORE. The EuroSCORE was a good predictor of overall mortality as shown by a C statistic of 0.75. The mean predicted mortality was 3.0% (range 2.0% to 3.7%), indicating a difference of nearly 100% between surgeons at the outer limits of the group. The overall variability between surgeons was high, as shown by an interquartile range of 0.64.

Comparison of low and high risk patients

Eighty five per cent of the total number of patients had a EuroSCORE of 5 or less. Almost half of all observed deaths were in the low risk group (49%). The remaining 51% of deaths were in the 15% of cases in the high risk group (EuroSCORE >5). The proportion of individual surgeons' practices that are high risk ranged from 5.6% to 23.9%. For the low risk group the observed mortality was 1.0% (range 0% to 2.9% between surgeons), with a mean predicted mortality of 2.3% (range 1.7% to 2.7% between surgeons). For the high risk group the observed mortality was 5.7% (range 0% to 13.6% between surgeons), with a mean predicted mortality of 7.4% (range 6.6% to 8.3% between surgeons). The C statistic indicating predictive ability of the EuroSCORE for the low risk and high risk groups was 0.72 and 0.62, respectively, indicating a satisfactory predictive ability for low risk patients but an unsatisfactory ability for those having high risk surgery.^{13 14}

The variability in predicted mortality between surgeons according to the interquartile range in the low and high risk patients was 0.32 and 0.67, respectively, showing that the low risk patients in each

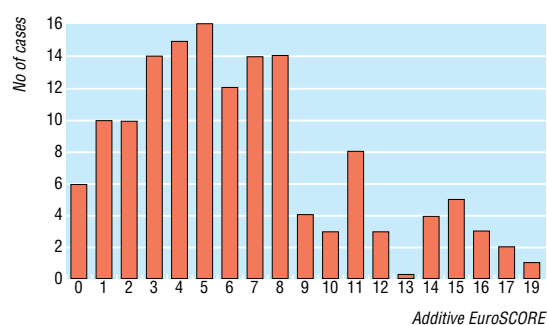


Fig 3 Number of deaths in each EuroSCORE group

surgeon's practice are a relatively homogeneous group, but there is much greater variation between surgeons in the high risk population. Figure 5 shows for the low risk patients that the 95% confidence intervals around mortality for each surgeon operating in the north west overlap the mean mortality for the region, indicating no surgeon is experiencing mortality results that are different from the peer group. We found a strong univariate association between the volume of operations that each surgeon had performed and observed mortality in the low risk group ($P < 0.001$) (table 2).

Discussion

Principal findings

It is possible routinely to collect risk stratified data on all patients undergoing surgery in a defined geographical area in the United Kingdom. Crude mortality analyses may be misleading as variations in the proportion of individual surgeons' practices that are high risk are marked. An accepted risk prediction model is poor at predicting mortality in this high risk population. The low risk group was relatively homogenous between surgeons, and we recommend a comparative analysis based on low risk cases without the need for further risk adjustment.

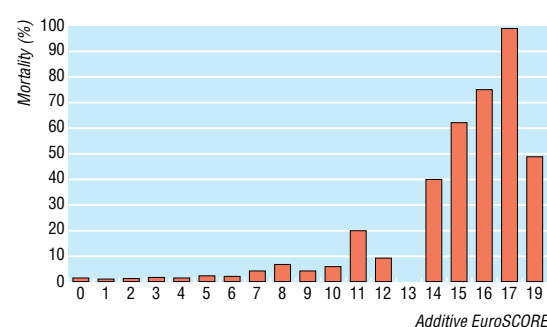


Fig 4 Percentage mortality in each EuroSCORE group

Strengths and weaknesses of the study

Our study has been conducted on a large population of patients undergoing surgery over three years. The average number of cases was 372 per surgeon, which is a reasonable size to allow comparisons to be made. The study has been conducted in the north west of England and includes all patients undergoing surgery in NHS hospitals in a defined geographical area.^{8 9} This is about one eighth of all cardiac surgical activity in the United Kingdom. The data have the confidence of clinicians, which should reassure patients that benchmarking between surgeons is meaningful, helps surgeons believe any differences that emerge, and encourages changes in practice to be made where necessary. This project shows that where there is clinical and management commitment, collecting robust, comprehensive data is possible and useful.

The dataset we have used undergoes local validation in each centre but has not been subjected to external validation, which is a weakness of our study. It has been shown previously that some problems arise with the completeness and reliability of this type of data.¹⁵ We have addressed issues of incomplete data by assuming that any risk factor that has a missing field is

Table 2 Observed mortality by operator volume

	Low volume	Middle volume	High volume	P value (χ^2 test for trend)
Observed mortality (%)	1.6	1.2	0.6	<0.001

negative for that risk factor. The incidence of missing data in our study was less than 2%, but this would lead to a small overall underestimate of predicted risk.

We have used low and high risk groupings to allow meaningful comparisons to be made¹⁰; the low risk group contains most patients and has a low variability in predicted risk between surgeons. Using the low risk group for mortality analysis and benchmarking excludes higher risk patients from comparisons. Clinically, high risk patients are a heterogeneous group, ranging from stable patients with multiple comorbidities to patients who come to surgery as emergencies, often directly from the cardiac catheter laboratory. We have not compared surgeons' death rates in the high risk group as predicted mortality differs between surgeons, the proportion of individual surgeons' patients who are high risk varies, and the EuroSCORE is a poor predictor for this population. However, half of the deaths in the population of patients are seen in this high risk group, and politicians and the public may be wary of excluding this many deaths from comparative analysis. It is also possible that by not analysing the high risk group we may be losing important messages about performance, which may be useful for improving quality.

Strengths and weaknesses compared with other studies

Healthcare outcomes can be benchmarked in several different ways. An approach that has been suggested and used elsewhere is to allocate a predicted mortality for each surgeon on their total practice of coronary artery surgery by using a mortality prediction tool, and comparing predicted with observed mortality to generate an adjusted death rate.¹⁶ In our study patients in the highest EuroSCORE groups (14 and above) have an observed mortality in excess of 50% (fig 5), and a small number of these patients in a surgeon's practice would affect their adjusted mortality adversely. This number varies markedly between surgeons. Because the EuroSCORE is a poor predictor in the high risk group as shown by the C statistic, we think that using adjusted death rates may produce erroneous conclusions.

Although the EuroSCORE is generally regarded to be a good overall predictor of mortality for patients

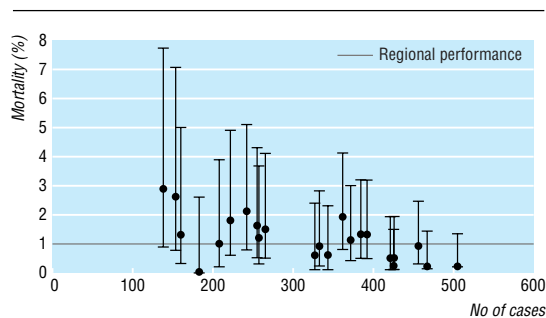


Fig 5 Observed mortality (with 95% confidence intervals) for each surgeon in the low risk group

undergoing heart surgery,¹⁰⁻¹² it has been noticed previously that it underpredicts risk in high risk patients,¹⁷ but the effects of this observation on the publication of surgeon specific mortality has not been described. The EuroSCORE working group have addressed underprediction of the additive score by producing a logistic regression model,¹⁷ the logistic EuroSCORE, which may be a better predictor in high risk patients, but this has not yet been fully validated and is not widely used. We studied the widely used additive model for our investigation, but failure to examine all available predictive scoring systems is a further limitation of this work.

Several studies have looked at outcomes of individual surgeons or institutions and their relation to volume of surgery.¹⁸⁻²⁰ Some of these have been on crude mortality data and others corrected for case mix. Although we have observed a strong association between volume and outcome in our data, we did not design our investigation to look for this as a primary end point. We believe that this observation should be treated with caution as there are numerous possible effects, including time and learning curve effects, which were not controlled for by our study design.

What is already known in this topic

The release of surgeon specific mortality data for coronary artery bypass surgery in the United Kingdom is planned for 2004

Outcomes after surgery are known to depend on several patient related factors

Currently no dataset is available to allow an appropriately risk stratified comparison of all surgeons in the United Kingdom

Proposed analyses would be undertaken on crude mortality data

What this study adds

It is possible to collect risk stratified mortality data on all patients undergoing coronary artery bypass surgery in a defined geographical area in a multicentre study

Most patients have a low predicted mortality

Predicted mortality differs between surgeons, which is largely due to differing proportions of high risk patients

An accepted risk tool is not good at predicting mortality in the high risk group

Crude mortality comparisons can be misleading and may encourage surgeons to practise risk averse behaviour

Risk stratified analyses should be encouraged as the basis for assessing consultant specific performance

Meaning of the study

We believe that publishing surgeon specific, crude mortality data,² as is planned in the United Kingdom, is not in the best interests of patients, and our study shows that surgeons cannot be compared fairly in this way. Cardiac surgeons already work in a stressful environment, and the perception that a “bad run” might jeopardise their career or result in suspension and investigation may lead to a tendency to turn down high risk cases. The easiest way to obtain low mortality is to do only straightforward operations—so called risk averse behaviour. This has already been identified as a potential problem after a survey of all cardiac surgeons in the United Kingdom in 2000, where 94% of responders agreed that high risk patients were being turned down for surgery.¹ Death rates in these patients often approach 100% if the patients are denied surgery and patients at heightened risk from surgery are, in general, those who have the most to gain from a successful operation.²¹ Our recommendation of benchmarking only low risk patients seems scientifically justified and pragmatic and should help to prevent risk averse behaviour.

Unanswered questions and future research

Some evidence from North America sheds light on the effects of publication of surgeon specific data on patients, cardiologists, and surgeons,^{1 22 23} but we do not know to what extent initiatives to publish crude mortality data for individual surgeons will actually deny operations to high risk patients, and what implications this will have on patients' survival, quality of life, and use of healthcare resources. This is an important area for future studies. Further investigations are also needed on high risk patients, to improve the quality of risk prediction in this group, and to understand variability in outcomes following high risk surgery for quality improvement purposes.

This study has been conducted on behalf of the North West Quality Improvement Programme in Cardiac Interventions, and the participating consultant surgeons are listed as follows: John Au, Ben Bridgewater, Colin Campbell, John Carey, John Chalmers, Walid Dhimis, Abdul Deiraniya, Andrew Duncan, Brian Fabri, Elaine Griffiths, Geir Grotte, Ragheb Hasan, Tim Hooper, Mark Jones, Daniel Keenan, Neeraj Mediratta, Russell Millner, Nick Odom, Brian Prendergast, Mark Pullan, Abbas Rashid, Paul Waterworth, Nizar Yonan. We would like to acknowledge the assistance of the audit officers working in each centre for their hard work in collecting and validating the data.

Contributors: BB had the idea for the study and with ADG and MJ was responsible for the study design. Data analysis was performed by ADG and MJ. The manuscript was prepared by BB and ADG. All authors contributed to writing the paper, which was written on behalf of the North West Quality Improvement Programme in Cardiac Interventions. BB will act as guarantor.

Funding: All primary care trusts in the north west of England.

Competing interests: None declared.

Ethical approval: The project was conducted on routinely collected prospective data. All patient identifiers were anonymised. The study therefore did not need ethical approval.

1 Keogh B, Kinsman R. *National adult cardiac surgical database report 2000-2001*. London: Society of Cardiothoracic Surgeon of Great Britain and Ireland, 2002.

2 Dyer O. Heart surgeons are to be rated according to bypass surgery success. *BMJ* 2003;326:1053.

3 Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired heart disease. *Circulation* 1979;13:112.

- 4 Tu JV, Jaglal SB, Naylor D. Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital length of stay after cardiac surgery. *Circulation* 1995;91:677-84.
- 5 Roques F, Nashef SA, Michel P, Guadalucho E, de Vincentiis C, Baudet E, et al. Risk factors and outcome in European cardiac surgery; analysis of the EuroSCORE multinational database of 19,030 patients. *Eur J Cardiothorac Surg* 1999;15:816-23.
- 6 Nashef SA, Carey F, Silcock MM, Oomen PK, Levy RD, Jones MT. Risk stratification for open heart surgery: trial of the Parsonnet system in a British hospital. *BMJ* 1992;305:1066-1067.
- 7 Bridgewater B, Neve H, Moat N, Hooper T, Jones M. Predicting operative risk for coronary artery surgery in the United Kingdom: a comparison of various risk prediction algorithms. *Heart* 1998;79:350-5.
- 8 Wynn-Jones K, Jackson M, Grotte G, Bridgewater B, and the Northwest Region Cardiac Surgery Audit steering group. Limitations in the Parsonnet score for measuring risk stratified mortality in the north west of England. *Heart* 2000;84:71-8.
- 9 The Northwest regional quality improvement program in adult cardiac surgery. www.nwheartaudit.nhs.uk (accessed 10 Jun 2003).
- 10 Nashef SAM, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R, the EuroSCORE study group. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999;16:9-13.
- 11 Roques F, Nashef SAM, Michel P, Pinna Pintor P, David M, Baudet E. Does the EuroSCORE work in individual European countries? *Eur J Cardiothorac Surg* 2000;18:27-30.
- 12 Nashef SAM, Roques F, Hammill BD, Peterson ED, Michel P, Grover FL, et al. Validation of the European system for cardiac operative risk evaluation (EuroSCORE) in North American cardiac surgery. *Eur J Cardiothorac Surg* 2002;22:101-5.
- 13 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
- 14 Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285-93.
- 15 Fine LG, Keogh BE, Cretin S, Orlando M, Gould MM. How to evaluate and improve the quality and credibility of an outcomes database: validation and feedback study on the UK Cardiac Surgery Experience. *BMJ* 2003;326:25-8.
- 16 Coronary artery bypass surgery in New York State 1997-1999. Albany, NY: New York State Department of Health. 2002. www.health.state.ny.us (accessed 10 Jun 2003).
- 17 Roques F, Michel P, Goldstone AR, Nashef SAM. The logistic EuroSCORE. *Eur Heart J* 2003;24:1-2.
- 18 Spiegelhalter DJ. Mortality and volume of cases in paediatric cardiac surgery: retrospective study based on routinely collected data. *BMJ* 2002;324:261.
- 19 Hannan EL. The relation between volume and outcome in health care. *N Engl J Med* 1999;340:1677-1679.
- 20 Hannan EL, Kilburn H, Bernard H, O'Donnell JE, Lukacik G, Shields EP. Coronary artery bypass surgery; the relationship between in-hospital mortality rate and surgical volume after controlling for clinical risk factors. *Med Care* 1991;29:1094.
- 21 Westaby S. League tables, risk assessment and an opportunity to improve standards. *Br J Cardiol (Acute Interv Cardiol)* 2002;9:AIC5-AIC10.
- 22 Schneider EC, Epstein AM. Use of public performance reports: a survey of patients undergoing cardiac surgery. *JAMA* 1998;279:1638-42.
- 23 Hannan EL, Stone CC, Biddle TL, DeBuono BA. Public release of cardiac surgery health outcomes data in New York: what do New York State cardiologists think of it. *Am Heart J* 1997;134:55-61.

(Accepted 17 June 2003)

Corrections and clarifications

ABC of diabetes: cardiovascular disease, hypertension, and lipids

Owing to an electronic problem, we had to retype many of the symbols in this article by Peter J Watkins, and inevitably we slipped up on one (19 April, pp 874-6). The fourth paragraph in the section on blood pressure management should read: “Blood pressure > 140/80 mm Hg [not < 140/80 mm Hg] should be treated if there is evidence of organ damage ... The target pressure is < 140/80 mm Hg.”

Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review

An error crept into the diagram of the Ottawa ankle rules in this article by Lucas M Bachmann and colleagues (22 February, pp 417-9). The medial view of the ankle should have been labelled: “Posterior edge or tip of the medial [not lateral] malleolus—6 cm.”