5. Activity and outcomes in Bristol

5.1 Mortality outcomes.

5.1.1 Comparison of sources on mortality:

Spiegelhalter (1999, INQ 0015 0021) compares six sources with regard to apparent activity and number of deaths, using both the 13 consensus procedure groups and (except for SWCHR) the open/closed classification. Certain sources do not provide data on all consensus groups, and occasional operations inappropriately classified (using OPCS4 codes) as closed in Surgeons' Logs or Perfusionists' logs are not considered.

In Table 5.1 we consider just Epoch 3, with revised data for Fontans (G9) for the CCR and SL, and using the perfusionists' logs (PL) only for activity. The relative variability between the sources is summarised by the Coefficient of Variation (CV) - values of CV around 20% could be considered as having reasonable agreement, and less than 10% as having good agreement.

For many of the individual procedure groups the agreement is reasonable: for example, Fallot (G1: CV = 15), TAPVD (G4: CV = 22), AVSD (G5: CV = 11) and the sum over procedure groups for which all sources are available (CV = 20). For open operations in general the agreement is remarkably good (CV = 4). There is poor agreement of CSR with other sources for Groups 2 and 3 for reasons discussed in Section 3.2.3 but, if CSR is ignored, agreement is fairly good for G3 (switches). Better agreement may be attributable to procedures that can be fairly unambiguously coded. PAS appears to record more admissions, and the set of clinical records in CCR was partly derived from surgeons' logs, so CCR should include all cases in SL. Disagreement on operation dates between different clinical sources can lead to minor differences between SL, CCR and PAS.

5.1.2 Annual mortality in open operations for under 1s.

Table 5.2 focuses on open operations on under 1s, and the different sources agree reasonably well with regard to both activity and the number of deaths. There is an apparent drop in the mortality rate in 1990, although according to the Surgeons' Logs (SL), only one baby less than 90 days was operated on in 1990 compared to around 7 in other years. However, the numbers each year are too small to draw any firm conclusions on individual years. Linkage of HES with ONS data does not tend to find deaths among those in whom outcome is unknown, suggesting that missing outcomes might reasonably be expected to be survivors.

5.2 Morbidity outcomes.

In response to the findings of Aylin (1999, INQ 0013 0028) of an apparently higher rate of neurological complications in Bristol, Evans (2000, INQ 0029) examined evidence on complication rates in local data sources. The Surgeons' Logs (SL) did not, predictably, contain good information on longer-term outcomes, while both in the coded clinical records (CCR) and PAS the recorded neurological complication rates among survivors of open surgery was very low (1.9% and less than 1% respectively). There was poor agreement between sources and Evans (2000, INQ 0029 0016) concluded that there was under-reporting in all centres, with Bristol possibly being slightly more accurate in its reporting. The Clinical Case Note Review (CCNR) did look in detail at the possibility of disability in those who had not died at 30 days, but with only 40 cases, even though they were preferentially sampled from high risk groups, the number with any disability was very small (4, all "moderate" disability). It is therefore not possible to draw confident conclusions on the true morbidity rate or make comparisons with other centres.

5.3 Conclusions on Bristol activity and outcomes.

There are clear limitations to all sources, and none is subject to defined procedures for data collection, follow-up and validation. It would be fair to say that none is held in high regard as a source of reliable evidence for clinical audit. However, Evans (1999) concludes that where direct comparison is sensible, the pattern is similar and there are no startling discrepancies. Although there is no gold standard for comparison, the Bristol PAS system appears of reasonable quality, and hence this lends confidence to Bristol returns to the national HES database. Our overall comparison suggests that the different sources agree well on the open operations in general and for many specific procedures.

The main findings of interest concern mortality rate for open surgery in under 1s. Overall, sources agree that the mortality rate was around 25 – 30% during the period under scrutiny, although with considerable variability between different procedures. The routine data sources available form an inappropriate basis for any firm conclusions concerning morbidity rates in Bristol.

6 Comparison of Bristol with national performance.

6.1 Analyses carried out.

Murray *et al* (1999, INQ 0014) and Aylin *et al* (1999, INQ 0013) report comparisons between Bristol and elsewhere using CSR and HES data respectively, and including detailed analysis of the relative rank of Bristol in the 12 centres carrying out surgery in England. Spiegelhalter (1999, INQ 0015) synthesised this evidence and discussed each of the procedure groups in detail. Aylin *et al* (2000, INQ 0030) repeated the HES analysis using age categories of under and over 1. Each of these analyses was based on a common method: examine variability among centres other than Bristol, predict the number of deaths expected in a centre with Bristol's activity were it 'typical' of centres elsewhere, and subtract this expected number of deaths from the observed number to estimate 'excess' mortality.

6.2 Mortality outcomes.

6.2.1 Overall summary:

Table 6.1 summarises the results for all open surgery, case-mix stratified open surgery and closed operations, using CSR and HES data for all epochs, under and over 1s, using results contained in Spiegelhalter (1999, INQ 0015) and Aylin *et al* (2000, INQ 0030 0073). Although the CSR data report statistically significant excess mortality for Bristol in over 1s during 1988-1990, the primary finding from both CSR and HES is of excess mortality from 1991-1995 in open operations in under 1s, in which the mortality rate in Bristol was around double that in other centres. This difference is retained after stratifying for operative group, which is the available determinant for case-mix. There is no evidence for excess mortality in closed operations, or for open operations in over 1s from 1991-1995. Reported mortality for open operations in under 1s fell in other centres from 21% in 1984-1987 to 12% in 1991-1995. Bristol appears not to have followed that pattern of improvement. There is no evidence of excess mortality in Bristol during Epoch 4, although activity in Bristol was too small to draw any firm conclusion.

We emphasise that the estimated total excess deaths for HES depends on the age-stratification used: the excess risk is greater in younger children: for all open operations in epoch 3 the total is 30.1 when dividing only into under and over 1s (Table 6.1) and 34.3 when including a < 90 day category (Aylin *et al*, 1999).

6.2.2 Mortality in procedure groups during 1991-1995: under 1s.

Table 6.2 summarises the analyses for Epoch 3 (1991-1995) for under 1s, using results from Spiegelhalter (1999, INQ 0015 0060) and Aylin et al (2000, INQ 0030 0032). For CSR, Groups 2 and 3 have been highlighted as being unreliable for reasons discussed in Section 3.2.3. There are predictable disagreements between the two sources of data. HES identifies excess mortality with 95% confidence for switches (G3), AVSD (G5), ASD (G6), open operations stratified for case-mix, (G1 to G11), and all open operations taken together.

6.2.3. Open surgery – comparison with other centres:

Figure 6.1 shows the mortality rates and 95% confidence intervals for each of the 12 centres carrying out open surgery, based on CSR and HES data for relevant epochs and divided into under and over 1s. The variability between the centres is immediately apparent. In under 1s, Bristol (Centre 1) had the third highest mortality rate reported to the CSR in 1988-1990 and the highest rate in both CSR and HES 1991-1995. The estimated probability that Bristol had the highest true mortality rate in under 1s during 1991-1995 is 88% using CSR data (Spiegelhalter 1999, INQ 0015 0060) and 97% using HES data (Aylin *et al,* 2000 INQ 0030 0073).

Table 6.3 presents the annual mortality rates for open surgery in under 1s for Bristol and elsewhere. The CSR results show that each year between 1988 and 1994 (with the exception of 1990), Bristol had either the highest or near the highest mortality rate for open surgery in under 1s. This is reinforced by the HES data between 1991 and 1994. It is clear that Bristol's activity was consistently below the median in the country, and the possible association of mortality with volume is discussed in Section 8.4.

6.2.4. Divergence of other centres:

Spiegelhalter (1999) and Aylin *et al* (2000) provide estimates of excess mortality for each of the 12 centres, treating each centre in the same manner as Bristol in the main analysis. From this analysis and Figure 6.1 it can be seen that only one other centre, Centre 10, had consistent evidence of divergent performance and this was for open operations in over 1s. It was revealed by the Inquiry in November 1999 that this was Harefield hospital. This finding must be treated with caution. Harefield has been an innovative centre for transplant surgery and these operations are included in the CSR (although not in the HES open category), and it also has a reputation for taking difficult cases from abroad.

Stark (2000b, WIT 0567 0002) observes that the excess mortalities for each centre calculated using HES and CSR data do not always closely coincide. In particular, as observed in Section 3.3 and shown clearly in Figure 6.1, Centre 3 reports far higher mortality to the CSR than may be calculated from HES. Section 6.4.3 considers the sensitivity of the conclusions on Bristol to the removal of this and other centres.

6.3 Non-mortality outcomes.

Aylin *et al* (1999) examined outcomes other than mortality using the HES data for 1991 to 1995, although they emphasise the limitations of this approach. They found that for open operations, Bristol recorded a higher proportion of admissions where central nervous system, cardiac, respiratory and urinary complications occurred, when compared with other centres. However, as reported in Section 5.2, Bristol's reporting of complications may be more complete than in other centres, and in any case the data sources are very unreliable. Aylin *et al* (1999) also report that for both open and closed operations, substantially fewer patients were discharged from Bristol within 7 days compared to elsewhere. This finding must be interpreted with caution, since many factors could influence length of stay.

6.4 Further questions.

6.4.1 Is the excess mortality restricted to switches and AVSDs?

Table 6.2 shows that switches and AVSDs are prominent contributors to the observed overall excess mortality, and Wisheart (2000a, SUB 0009) questions whether there are other contributors. The information in Table 6.4 can be extracted from Tables 7.3.1, 7.3.2 and 7.4.5 of Spiegelhalter (1999).

Excluding switches and AVSDs, the CSR show a significant 83% increase in mortality over other centres. The HES data show a 44% increase in mortality over centres elsewhere, although this is not statistically significant at conventional levels. Because of the known lack of distinction in the CSR between switch (group 3) and inter-atrial repair (group 2), group 2 might also be excluded: Table 6.4 shows that this slightly increases the contrast between Bristol and elsewhere.

6.4.2 Is the excess mortality influenced by the missing outcomes in HES?

The majority of the HES analyses have ignored admissions with missing outcome data, Wisheart (2000a, SUB 0009) questions whether this seriously biases the results. We have carried out a simple analysis to examine what the impact of these missing outcomes might be, taking the most optimistic view that in Bristol they all were survivors. The data shown in Table 6.5 were taken from Aylin *et al* (1999, INQ 0013 0055-0057), and only consider pooled open operations. There were 48 cases in Bristol with missing outcomes. If they had been included in the analysis, and had they all survived, then they would have added 0 to the observed number of deaths, and added around 3.6 to the expected number of deaths. Thus the total excess deaths would have been reduced by around 3.6, from 34.3 (the estimated total when using the age-stratification in Table 6.5) to 30.7. Note that this analysis does *not* assume that missing outcomes elsewhere were survivors (although the linkage exercise does not suggest missing outcomes elsewhere are at increased risk of being deaths). Thus, even if we assume that all missing outcomes at Bristol were survivors, there is little effect on the findings. It therefore does not appear that missing outcomes makes the HES analysis unreliable.

6.4.3 Can we base conclusions on imperfect data?

Given that the data sources for these comparisons have such clear limitations, it is reasonable to ask whether in this light any reliable conclusions can be drawn. Statistical significance alone is not a sufficient guide, as it only indicates quantifiable random error and not systematic reporting or coding biases. **The crucial issue is not whether the data are 'true', since they manifestly contain errors, but whether such errors are likely to be great enough to overcome the observed pattern in the data.** To address this issue a number of sensitivity analyses to possible shortcomings in the data have been carried out and the detailed results are shown in the Technical Appendix for the data of primary interest: open operations on under 1s from 1988. The summary conclusions are:

1. *Centres for which doubts exist concerning national data.* Centres 3, 4, 5 were removed from all analyses, since Table 3.3 reveals these as having the highest discrepancy between HES and CSR data, with more than 20% difference in death rates. This has the effect of increasing Bristol's divergence, presumably because the removed centres are smaller and tend to have higher mortality on such patients (see Section 8.1).
2. *Procedure groups.* Procedure Groups 2 and 3 (Interatrial and other repairs of TGA) and 5 (AVSDs) were removed from the analysis, as described in Section 6.4.1. Groups 2 and 3 suffer from known coding overlap in CSR. This increases the divergence in 1988-1990, but reduces it

considerably in 1991-1995. However, as suggested in the simple analysis of Table 6.4, there is still considerable evidence of divergent performance even without these higher-risk groups.

3. *Undercount of mortality in HES.* The mortality rate in the HES data for each centre was increased by the 'undercount' for open operations in under 1s that was detected in the linkage study (Section 3.1.4) and shown in Table 3.2. The undercount ranges from 0 to 12% of deaths, with an average of 5.2%. This has little effect on the conclusions, presumably because Bristol has an average undercount (5%).
4. *An analysis favouring Bristol.* An 'extreme' scenario is one in which we choose, for each centre and procedure group, the results from HES or CSR according to the following rule: for Bristol, we select the results with the *lower* mortality, for each other centre we select the results with the *higher* mortality. This stringent comparison still shows strong evidence of divergent performance for all open operations, but stratifying for case-mix leads to borderline evidence for excess mortality in Bristol, with an estimate of around 50% increase in odds of death.

These sensitivity analyses certainly have an influence on the accuracy with which excess mortality in Bristol can be estimated. However, in our view, the magnitude of the observed divergent performance is such that reasonable variations in assumptions are not sufficient to cast the conclusions into doubt. This is discussed further in Section 9.

7. Comparison with published sources on operative mortality.

7.1 Sources of published data.

Vardulaki *et al* (2000, INQ 0039) have carried out a systematic review of published research on mortality data for five main procedures, corresponding to our consensus groups G3 (switch operations for transposition), G4 (TAPVD), G5 (AVSD), G7 (Truncus) and G9 (Fontan). They acknowledge the difficulty of generalising from published sources, as there is likely to be selective reporting from centres of excellence. There is substantial heterogeneity between sources, and there is a general pattern of improvements over time.

In addition, Hannan *et al* (1998) report a study on 7169 cases in New York between 1992 and 1995 and provide mortality rates for many procedures, including those studied by Vardulaki *et al.* This covers almost exactly the period of Epoch 3.

7.2 Comparison with published sources.

Table 7.1 compares the results derived from Spiegelhalter (1999) with those reported by Vardulaki *et al* (2000) and Hannan *et al* (1998). Direct comparison is difficult, as Vardulaki *et al* (2000) do not stratify for age but do report results at a finer level of detail than our consensus groups. In addition, results for the period 1991-1995 are not directly reported, and so the rates given in Table 7.1 are taken by eye from their Figures and so can only be considered rough estimates. We note that the Hannan results fit closely with those reported by Vardulaki *et al* (although they do not contribute to their analysis since the publication date (1998) lies outside the range adopted by Vardulaki *et al*). The HES and CSR results agree well with the international data.

7.3 Conclusions.

Stark (2000b, WIT 0567) suggests that the mortality rates given for England in Inquiry reports appear low, and reports anecdotal mortality rates from Toronto Children's Hospital of 26% in open surgery between 1991 and 1995 (age-group unknown). However, Table 7.1 suggests that the results from non-Bristol centres derived from both HES and CSR are compatible with published data from elsewhere, and in particular New York State. Bristol appears to have divergent performance from international published sources.

8. Investigation of possible factors associated with divergent performance in Bristol.

Having observed evidence of divergent performance in Bristol, a number of possible explanatory factors have been investigated. For each factor it is preferable to identify two characteristics:
1. The **level** at which it is measured (*i.e.* institutional / patient). Purely institutional factors, such as staffing level, organisation of care, experience of staff, and volume of surgery, can only provide indirect explanation for variability between centres since it is not clear how they directly influence the risk experienced by individuals. The only institutional level factor available for investigation was volume of surgery.

2. The extent to which the factor is **'exogenous/endogenous'** to the system being evaluated, *i.e.* the extent to which the factor is susceptible to influence or change by the system. This is more of a grey-scale than an absolute classification of factors. For example, geographic clustering in births of difficult cases should be an exogenous factor since it is not under any control of the system – the adjustment for broad procedure group attempts to deal with this, although even the choice of procedure is to some extent subject to clinical influence. Similarly, comorbidity factors such as Down's syndrome should be exogenous but the incidence may be influenced by referral practices. Status at admission, comorbidity and timing of surgery have been examined at the individual level – however, for each of these it is not immediately clear to what extent they are directly influenced by the cardiac surgical system under evaluation, and hence they cannot clearly be labelled as either exogenous or endogenous..

8.1 Institutional factors: volume of surgery.

8.1.1 Results of analysis.

Spiegelhalter (2000, INQ 0031) reports an analysis of the association between volume of surgery and mortality outcomes, using data from the CSR and HES. For open operations in under 1s, and for arterial switches and AVSD in particular, there was strong and consistent evidence for an association between mortality rates and volume (not taking into account any data from Bristol), in which higher-volume centres have lower mortality. Stratifying for operation-mix, or including the results from Bristol, strengthened this association. Figure 8.1 summarises the results for open operations in under 1s for 1991-1995, estimating the relative reduction in risk per additional 10 cases per year to be around 3% and 4% in CSR and HES respectively. We note that, according to the HES data, centres carrying out less than 200 cases in four years (one a week) had a mortality rate of 15% (not including Bristol) or 17% (including Bristol), while those carrying out more than one a week had a mortality rate of 10%. The relationship also appears to hold in earlier epochs: the CSR data estimate the relative reduction in risk per additional 10 cases per year to be 9% (95% interval -6% to 22%) in 1984-1987 and 6% (95% interval 2% to 10%) in 1988-1990.

8.1.2 Interpretation.

Spiegelhalter (2000, INQ 0031) estimated that a hospital carrying out 120 open operations a year on patients aged under 1 in 1991-1995 would be expected to have an underlying mortality rate 25% lower than one carrying out only 40 such operations. If the hospitals had exactly the same age- and operations mix, this reduction is increased to 35%. These are percentage changes relative to the underlying risk, and so implications in terms of the difference in numbers of deaths depend on the context. However, considerable caution is needed in interpreting these results, and it does not necessarily follow that concentrating resources in fewer centres would reduce mortality rates, since volume may be associated with lower mortality without being a direct cause. Using the association found in other centres, it was estimated that only around 12% (HES) or 17% (CSR) of the excess mortality observed in Bristol in open operations in under 1s might be explainable by the lower volume of surgery being carried out in Bristol.

8.2 Patient factors: status at admission and comorbidity.

8.2.1 Results of analysis.

Aylin *et al* (1999;2000) explored a number of factors which might account for the high reported mortality following operations at UBHT. Table 8.1 summarises some of these findings for open operations. It suggests that age-mix cannot account for the high mortality at UBHT, not only because age-specific mortality was higher in all age groups compared with elsewhere, but also that UBHT operated on a much smaller proportion of the youngest (higher-risk) babies aged under 90 days (7%) than elsewhere (16%). Mortality in children with Down's syndrome from other centres (excluding UBHT) is not significantly greater (8%) than children operated on without this disorder (7%), so differences in the proportion of children with Down's syndrome treated at UBHT are not likely to account for differences in mortality. Patients transferred from other units to centres (excluding UBHT), have a higher mortality (14%) than patients admitted by other means (5%), but UBHT had a much lower level of transferred patients (6%) than elsewhere (22%), so this again cannot account for higher mortality in UBHT. Emergency admissions have a higher mortality (12%) than non-emergency admissions (7%) in other centres (excluding UBHT), however UBHT admitted a smaller proportion of emergencies (7%) than other units (10%), making this an unlikely explanation for their high mortality. Aylin *et al* (1999) also found that mortality did not vary by levels of socio-economic deprivation of patients and that the distribution of primary diagnoses in UBHT patients was similar to other centres.

8.2.2 Interpretation

The analyses are summarised below.

Factor	What is associated with higher mortality	How Bristol compares to average	Comments
Volume	Low	Lower	Explains small proportion of excess
Age at operation	Low	Higher average age	Marked divergence of practice at Bristol
Proportion of Down's	High	Lower proportion	Does not explain excess
Transfers	High	Lower proportion	Does not explain excess
Emergency Admission	High	Lower proportion	Does not explain excess
Socio-Economic Deprivation	High	No difference	Does not explain excess

HES data is limited in the information it provides about status at admission and comorbidities, and the analyses suggest that these factors cannot explain the high mortality reported at UBHT. The role of age at surgery is now examined in more detail.

8.3 Patient factors: timing of surgery.

8.3.1 Results of analysis.

The age at which surgery takes place may be influenced by the system of care, and hence may be an explanatory factor in divergent performance. Only HES data provide comparative data with precise dates of operation, and this is only available for Epochs 3 and 4 (April 1991 to December 1995). Aylin *et al* (2000) derived Figure 8.2, which shows the number of open operations taking place at each month of age up to 18 months, in Bristol and in the other centres combined in Epoch 3. It is clear that Bristol has a peak of activity at 11 months, in contrast with a steady decline in activity with increasing age seen in other centres. Aylin *et al* (2000) show that this peak in activity apparently only occurs before March 1994, and an 11th -month peak in Bristol between 1990 and 1993 is confirmed by examination of local data sources, specifically PAS, CCR, Surgeons' Logs and perfusionists' logs: Figure 8.3 shows the age distribution recorded in the PAS system. Figure 8.4 shows the age-specific activity in all 12 centres between April 1991 and March 1994, which shows the 11th -month peak was unique to Bristol. The peak is apparent in the larger groups (G2 Inter-atrial repair, G5 AVSD, G7 VSD) but other groups are too small to judge.

Table 8.2 considers age at operation in 3 month intervals up to two years old, and annually and five-yearly thereafter. Bristol had only carried out 21% of its surgery before the age of 9 months, compared to 39% elsewhere. However, in the following three months up to their first birthday, 14% of all surgery (60 operations) was carried out compared to 4% elsewhere. The mortality rate elsewhere is reasonably constant after the first three months, and can be used to estimate the expected number of deaths expected in Bristol. The estimated total of excess deaths is 34.1 (slightly different from previous estimates due to the finer age-stratification): those operated on in the first three months of life contribute 16.8 and those in the final three months before their first birthday contribute 7.8, approximately 25% of the total. This excess of 7.8 is due both to the number of operations taking place, mainly in the 11th month, and the fact that the mortality rate of 18% (19/60) is significantly higher than the mortality elsewhere (5%, 49/381, $P< 0.001$).

8.3.2 Interpretation.

According to the HES data, around a quarter of the age-stratified excess mortality (7.8 out of 34.1) in open surgery in 1991-1995 is associated with operations performed within three months of the first birthday. Aylin *et al* (2000) identify AVSD operations as a primary contributor to this: 41% (14/34) of AVSD surgery was in this period with a 50% mortality rate (7/14).

The pattern for timing of surgery shown in Table 8.2 suggests that the operations carried out just prior to the first birthday may have been delayed from earlier rather than brought forward, since Bristol had carried out only 21% of its open operations before the age of 9 months, compared to 39% elsewhere. In particular, 40% of AVSD operations performed elsewhere during 1991-1995 were carried out in the first six months of life, compared to 9% in Bristol.

8.4 Patient factors: the process of care.

A Clinical Case Note Review (CCNR) (Hamilton and Silove1999, INQ 0016) was carried out on a stratified sample of UBHT medical records in order to provide peer judgement as to the adequacy of care received. Full interpretation of the results of the CCNR is given elsewhere (Hamilton and Silove, 2000). Their Executive Summary concludes that the care received by 70% of the children was adequate, leaving 30% whose care was less than adequate to different degrees. For just over 5% of children, it was considered that different management would reasonably be expected to have made a difference to outcome. The reviewers provided criticisms of a range of aspects of process of care throughout the surgical system, including delays in treatment, shortcomings in cardiological contribution and organisation of intensive care, the split site and general organisational failings. The conduct of surgery was one of the criticised factors but was not particularly highlighted. In the stratified sample, over half the deaths (21/40) were considered to have received less than adequate care in which different management might have made, or would reasonably be expected to have made, a difference in outcome. We do not, however, have a comparative group in order to see whether other centres have similar systemic difficulties, and so cannot know whether similar criticisms could be levelled at other centres over this period.

9. Overall Conclusions.

9.1 The available data sources.

The two national sources, HES and the CSR, are admittedly imperfect. Both suffer considerably from lack of agreed operating procedures for ensuring completeness and accuracy of activity, coding and outcome results. Both the OPCS4 coding scheme and the use of non-clinical coders lead HES to be viewed with suspicion by clinicians. There are also strong concerns about variability between centres in the CSR's coding procedures and recording of mortality. Even if they were meticulously completed, agreement between the two sources could not be expected due to their different criteria. However, HES was found to be surprisingly accurate in its recording of in-hospital mortality and, with certain clear exceptions, the sources described the same broad picture.

The local sources were found to provide good agreement on activity and overall mortality, although comparison at a finer level was sensitive to the coding conventions used. Nevertheless, the six sources on Bristol's activity and outcome agree well for open operations in general and, to a lesser but still reasonable extent, for finer consensus procedure groups of interest. Where there is disagreement, then there are clear reasons, usually resulting in transfer of operations between two groups.

9.2 Evidence for divergent performance of Bristol.

There is no evidence of excess mortality in closed operations carried out in Bristol, and limited evidence in open operations on children aged over 1 year. However, there is strong and consistent evidence of excess mortality in open operations in children less than 1 year old at operation. It is estimated from HES data that in the period 1991-1995, 24.1 (95% confidence interval 12 to 34) of 41 recorded deaths are in excess of that expected were Bristol a 'typical' centre: finer age-stratification increases the estimated excess mortality. CSR data suggest the excess mortality dates back at least to 1988. Open procedures on children aged less than 1 that can be identified with reasonable consistency as having excess mortality include 'switches', operations for TAPVD, AVSD and, although rare in this age group, ASD. It is to be expected that excess mortality is easier to detect in higher risk groups.

The excess mortality was not just restricted to AVSDs and switch operations, and the conclusions are robust to admissions with missing outcomes. National mortality rates were comparable to those in the international literature. One other centre had a consistent pattern of excess mortality in open operations in children over 1 year, but there were no other centres with consistently divergent raised mortality in the younger age group.

9.3 Explanation for divergent outcomes.

At an institutional level, Bristol is a low-volume centre and other low-volume centres have been associated with higher mortality rates. Regardless of the policy implications of this finding, it is apparent that only a limited proportion of Bristol's excess mortality can be 'explained' by this indirect risk factor.

Bristol differed from the national pattern in some aspects of status and comorbidity (Section 8.2), but these characteristics do not apparently explain divergent performance. There is also no evidence that Bristol had systematically higher-risk case-mix. The most striking factor is the high incidence of surgery in the period immediately preceding the first birthday between 1990 and March 1994. Around 25% of the excess mortality in open surgery is associated with a peak of operations in the three months before their first birthday, mainly in the 11^{th} month. The evidence in Table 8.2 suggests that these cases may have been delayed from earlier surgery, rather than being operations that might normally have been carried out after their first birthday.

9.4 What might have been known?

It is possible to consider what simple analyses might have been performed using the data and the statistical tools that would have been readily available to the surgeons at the time. The participating centres in the CSR were supplied with detailed annual reports giving mortality rates split by age and procedure, aggregated over all participating centres. This would have allowed a centre to compare its mortality rates with corresponding national figures.

Open surgery performed in children aged under one year is an appropriate subgroup to monitor, since these children are at high risk and include the majority of deaths. In this group the ratio of the overall mortality rates at Bristol to the rates for other centres in England for 1985 through to 1995 were 1.18, 1.21, 1.24, 2.04, 1.93, 0.79, 2.05, 1.19, 3.18, 1.67 and 0.50 respectively. A chi-squared test performed each year would have given a crude indication of whether the local mortality rate differed from the national rate by more than could be explained by chance. Using such a test, the data for 1988, 1989, 1991 and 1993 are statistically significant at the 5% level. If years had been pooled in pairs or triplets to give larger numbers, then the results for 85/86 and 86/87 are non-significant, as are the results for 85/86/87, but the results for 87/88 and 86/87/88 are statistically significant. Thus with any of these approaches, it is not until the data for 1988 were included that the divergence from the national rates became statistically significant, and this was reinforced by the data for 1989. Given that there was a delay of the order of 18 months before the CSR data were fed back to centres, it would have been 1990 before the data from the CSR might have given any reason for concern, and the independent reinforcement for the 1989 data, which would become available during 1991, would have heightened this concern. However, the data for 1990 then came back into line with national figures (see Table 6.3), which might have been taken as reassurance that any problems which might have existed previously had been resolved.

This final point illustrates the difficulty of interpreting crude data based on small numbers of patients each year. Taking running totals from three year periods the data are statistically significant for 86/87/88, 87/88/89, 88/89/90, 89/90/91, (borderline non-significant for 90/91/92), 91/92/93, 92/93/94 and 93/94/95. Clearly there is a consistent and on-going pattern of poor outcomes, but it is difficult to know what weight should have been put on these data at the time, with there being questions over the data quality and with inadequate statistical tools to adjust for case mix and to analyse accumulating data from many different centres. A related difficult question is the extent to which the responsibility lay with individual centres to interpret their own data, versus the role of the Society of Cardiothoracic Surgeons which with access to the full data for each centre was in a better position to analyse and interpret the data. Of course, statistical analysis is only one aspect of monitoring clinical performance.

9.5 Conclusions.

We again emphasise that statistically significant findings, taken on their own, are insufficient grounds for confidently identifying divergent performance when there are grave and well-founded doubts about the quality of the data sources. It is also important to emphasise that there are many areas in which there was no evidence of poor performance in Bristol. Nevertheless, although no data source can be considered as exactly representing the true state of affairs, their consistency, and the fact that they are derived in very different manners, suggests that their findings reinforce each other.

The single most compelling aspect of the data is the magnitude of the discrepancy between the outcomes observed at Bristol and those observed elsewhere. For children aged under one year undergoing open surgery between 1988 and 1994, the observed mortality rate at Bristol was roughly double that observed elsewhere in 5 out of 7 years. While the national trend over this period was for mortality rates to fall substantially, no such trend was seen in the Bristol results. In spite of the many flaws in the data sources, we do not believe that statistical variation or any systematic bias in data collection can explain a divergence of this magnitude. We therefore conclude that there is strong evidence of divergent performance at Bristol in the areas identified above, and we believe that the imperfections of the data do not cast serious doubt on these conclusions.

10. Proposals for the future.

In the light of our combined experience in working on the Inquiry data sources, we would now like to make a range of proposals regarding future monitoring systems in paediatric cardiac surgery, which may also have more general relevance to other settings.

10.1 What kind of comparative studies?

Given the limitations of key data sources, the Inquiry's statistical evidence is necessarily focussed on short-term mortality outcomes of those who received surgery. However, our investigation suggests that a much broader perspective is appropriate if comparing systems of care, perhaps more in line with public-health investigations. Such a population-based approach examines all cases of interest, whether or not they come to surgery, and is ill served by current systems and initiatives. Such a perspective becomes particularly important if, for example, surgeons started to avoid operating on high-risk patients in order to improve apparent mortality rates – this may be a consequence of a "blame culture" that emphasises penalties for apparent poor performance.

Overall, there needs to be clarity as to the precise objectives of any comparative exercise. A crucial distinction is whether the objective is to identify grossly discrepant performance, or whether the aim is more educational, with individual surgeons or units following their performance year by year, looking for minor problems, or seeking to identify the benefits of minor changes in practice ('closing the audit loop'). Many articles on clinical audit see the latter as being the aim of audit. However, the statistical work commissioned for the Inquiry shows that, even given perfect data sources and even if there were no difference in case-mix, statistical variability would mean that data would need to be accumulated over many years to detect modest but important differences in mortality rates. Given the many flaws that have been identified in existing data sources, it is clear that only gross divergence could have been identified with any degree of confidence. If, for example, the mortality rate for open operations in under 1s observed at Bristol had been 50% higher than elsewhere rather than 100% higher, it would have been very difficult to exclude the possibility that the difference had arisen through a combination of differences in case-mix, in the coding of operative procedures, and in the thoroughness of achieving follow-up data.

Existing data sources can and should be improved, for example by introducing routine linkage of HES records to national mortality records in order to confirm mortality data. Equally, data collection procedures require much greater standardisation, with adequate training of the staff involved, and regular feedback of data so that quality can be maintained. The objectives of any audit exercise need to be reviewed carefully in the light of the sample size that is likely to be available for any comparisons, and the magnitude of the likely biases.

10.2 What source of comparative data?

There are currently two broad approaches to sources of data for comparative exercise: an administrative model and a clinical model. A simplistic comparison of their advantages and disadvantages might comprise:

Administrative model	**Clinical model**
Example: PAS/HES	Example: Clinical databases
Pro:	**Pro:**
Established system for pooling data	Contribute data to refined CSR
Trained coders	Data controlled by clinicians
Facility for linkage for population studies	Clinical data
Accurate mortality records	Individualised risk-assessment
All centres contribute	
Anti:	**Anti:**
Non-medical coders	Lack of standardisation between centres
OPCS4 not ideal	No agreed coding scheme
No adjustment for clinical risk factors	Lack of linkage for mortality *etc*
Only mortality outcomes	Voluntary involvement

This is clearly a simplification and many compromises are possible between these archetypes. We believe that each approach has a role, but that development in isolation to each other is wasteful and inefficient. Our experience in this exercise has been that neither approach has been satisfactory.

We therefore strongly recommend the development of linkage schemes between ONS national statistics and administrative systems, and between administrative and clinical systems.

A separate but related issue is the question of how to raise the credibility of routine data, especially for clinicians. It is clear that, for whatever reason, many clinicians have no confidence in the HES data. Any future developments of routine data systems needs to address the issue of how best to ensure data are clinically valid and meaningful, possibly based on the promotion of a sense of 'ownership' of the data by clinicians.

10.3 What kind of coding scheme and groupings?

This investigation has revealed the difficulty in developing an agreed coding scheme for complex cases in paediatric cardiology, that allows both accurate description of individuals and a facility for pooling cases in a clinically acceptable way. The fact that such a scheme was not in general use in this country forced the Inquiry to use data sources and coding schemes that were criticised by clinicians. Furthermore it is unclear how coding is this context will develop, in the light of the recent publication of two independent schemes under the auspices of the Society of Cardiothoracic Surgeons (Mavroudis and Jacobs, 2000) and the Association of European Paediatric Cardiology (Franklin *et al*, 1999) respectively.

While it is desirable to make comparisons between precisely-defined homogenous groups of patients, we feel this has been over-emphasised and that for monitoring purposes it is better to develop broad groups into which activity can be allocated with reasonable accuracy. Finer distinctions can always be made for more focused clinical purposes.

We recommend the adoption of a scheme in which each procedure is placed in one of a small number of risk categories. Whatever detailed clinical coding scheme is adopted, it is important that it can be mapped both onto such a simplified system for monitoring, and into the codes used by administrative systems.

10.4 How can statistical methods help in analysing performance?

Comparative data may be useful in many ways, and a variety of statistical tools are available to help exercise due caution.

Institutional comparisons: Curnow (1999, WIT 0361 0002) emphasised that statistical techniques may be used to indicate when an institution may have passed either a 'warning' threshold, which might trigger further investigation, or an 'alarm' threshold which might indicate immediate action. The setting of such thresholds requires a combination of statistical and clinical judgement, and allowance for random error and inevitable between-institution variability. Outcomes should be risk-adjusted where feasible, although this might be only into broad groups (see Section 10.3) as to much disaggregation reduces precision. Statistical methods can also prevent undue attention to spurious ranking into 'league tables' (Marshall and Spiegelhalter, 1998).

Clinical comparisons within institutions: Availability of good data sources would allow, for example, the cumulative monitoring of risk-adjusted excess mortality (or another performance indicator) for individual clinicians, as is being increasingly adopted in adult cardiac surgery. Care is required if formal thresholds are used for monitoring.

Patient information: There is likely to be increased demand for patients to be given numerical risk assessments when asked for consent for surgery. This is not a straightforward matter: does one give the data for the individual surgeon, institution, or nationally, and for what period? How much should data-based statistics be adjusted for subjective opinions concerning the individual patient? There are statistical methods that can help with individualised risk-assessment, discounting historical data, pooling local with national data, and critiquing past numerical risk assessments.

We recommend the informed introduction of formal statistical procedures for institutional comparisons, monitoring individual clinical performance, and providing for informed consent of patients.

Acknowledgements

The help of the Secretariat to the Inquiry, particularly Ruth Chadwick, is gratefully acknowledged. Mr Leslie Hamilton and Dr Kate Bull provided invaluable help on coding, and members of the statistical Expert Panel to the Inquiry provided valuable suggestions. The authors, however, take full responsibility for the opinions expressed in this report.

References

Aylin P, Alves B, Cook A, Bennett J, Bottle A, Best N, Catena B and Elliott P. (1999). Analysis of Hospital Episode Statistics for the Bristol Royal Infirmary Inquiry. *Bristol Royal Infirmary Inquiry*, INQ 0013, November 1999.

Aylin P, Alves B, Cook A, Best N (2000). Supplementary analysis of Hospital Episode Statistics for the Bristol Royal Infirmary Inquiry. *Bristol Royal Infirmary Inquiry*, INQ 0030.

BRI Inquiry (1999a) Issues list for Part I of the Inquiry. *Bristol Royal Infirmary Inquiry*, February 1999.

BRI Inquiry (1999b) The Inquiry's approach to making use of relevant data sources. *Bristol Royal Infirmary Inquiry*, March 1999.

BRI Inquiry (1999c) Preliminary overview of existing data sources relevant to the Inquiry's remit. *Bristol Royal Infirmary Inquiry*, July 1999.

BRI Inquiry (1999d) Synthesis of statistical sources: a note on expert consultation on key analytical issues. *Bristol Royal Infirmary Inquiry*, November 1999.

Curnow R N (1999). Potential contribution of statistical evidence to the Inquiry. *Bristol Royal Infirmary Inquiry*, WIT 0361, November 1999.

Dhasmana J D (2000) Response to supplementary statistical evidence to the Inquiry. . *Bristol Royal Infirmary Inquiry.*

Evans SJW (1999). A report on local data relating to children who received cardiac surgery under the terms of reference of the Bristol Royal Infirmary Inquiry. *Bristol Royal Infirmary Inquiry*, INQ 0012, November 1999.

Evans S J W (2000). Further Reports based on Local Sources of Data for the Bristol Royal Infirmary Inquiry. *Bristol Royal Infirmary Inquiry*, INQ 0029.

Gallivan S (2000). Issues related to papers by Professor John Yates presented to the BRI Inquiry. *Bristol Royal Infirmary Inquiry*, INQ 0036.

Franklin R C G, Anderson, RH, Daniels, O, Elliott, M, Gewillig, MHML, Ghisla, R, Krogmann, ON, Ulmer, HE, Stocker, FP (1999) Report of the Coding Committee of the Association for European Paediatric Cardiology. *Cardiology in the Young*, **9**, 633-658.

Hamilton, L and Silove, E (1999): Preliminary Report on the Clinical Case Note Review. *Bristol Royal Infirmary Inquiry*, November 1999.

Hamilton, L and Silove, E (2000) Report on the Clinical Case Note Review. *Bristol Royal Infirmary Inquiry.*

Hannan E L, Racz M, Kavey R E, Quaegebeur J M, Williams, R (1998) Pediatric cardiac surgery: The effect of hospital and surgeon volume on in-hospital mortality. *Pediatrics,* **101**, 963-969

Lawrence A E and Murray G D (2000). The UK Cardiac Surgical Register: assessment of data quality issues for the Bristol Royal Infirmary Inquiry. *Bristol Royal Infirmary Inquiry*, INQ 0033.

Marshall EC and Spiegelhalter DJ (1998). League tables of *in vitro* fertilisation clinics: how confident can we be about the rankings? *British Medical Journal*, **316**, 1701-4.

Mavroudis C and Jacobs J P (2000) Congenital heart surgery nomenclature and database project: overview and minimum dataset. *Ann Thoracic Surgery*, **69**, 2-17.

Murray G D, Lawrence AE and Pollock J (1999). UK Cardiac Surgical Register and South West Congenital Heart Register: a statistical analysis and review of key data sources relevant to the inquiry's remit. *Bristol Royal Infirmary Inquiry*, INQ 0014, November 1999.