

Institute of Primary Care
University of Sheffield
Sheffield S5 7AU
8th October 1999

Mrs Liz Baldock
Bristol Royal Infirmary Enquiry
2-10 Temple Way
BRISTOL BS2 0BY

Dear Mrs Baldock

Commentary on statistical analyses re letter from Dr Chadwick 4/10/99

Thank you for the Annexes A,B,C and D which compare neonatal deaths following open-heart surgery at Bristol Royal Infirmary with the rest of the UK and with data for selected European Centres.

Annex A

This contains analysis of paediatric cardiac mortality data from UBHT 1990-1992 and compares it to the national average year of 1991. It omits the neonatal arterial switch operation. It shows Bristol to be statistically significantly worse than the rest of the UK on a number of operations.

I have checked the chi-squared tests and they seem correct.

Comments

- 1) The data were analysed by chi-squared test without Yates' correction. Some of the expected numbers were small, rendering the chi-squared test less reliable. Some statisticians prefer the use of Yates' correction, which gives larger p-values and so the results are less likely to be significant, or Fisher's exact test for tables with small expected values. I carried out these tests, and the only change I got was for the A-V canal (age <1 year) where the mid-P two sided value from the Exact Test came out as $p=0.051$ (just above the critical level of 0.05).
- 2) It is better practice to provide estimates of the difference in proportions between Bristol and the rest of the UK and confidence intervals (say 99% because there are several comparisons). Also one should give exact p-values rather than (say) $p<0.005$.
- 3) Some allowance in the interpretation of the p-values should be given for the fact that 12 comparisons are being made. A very conservative approach would be to multiply the p-values by 12, (the Bonferroni correction) but even then most would remain significant (i.e. <0.05).
- 4) The assumption must be that the patients are comparable in prognostic factors for Bristol and the rest of the UK.

Annex B.

This compares Bristol with data from the European Congenital Heart Defects Database (ECHDD) , for years 1992-1994.

The important analyses are Figures 4-10, showing meta-analyses comparing the ECHDD with Bristol for 1993, and Figures 11-18 demonstrating the CUSUM analysis.

There are no conclusions from these analyses in the text. In particular the CUSUM charts are there largely for ease of visual inspection of the data. The analyses would appear to be sensible and valid.

Annex C

This compares Bristol 1992-1995 with the UK Cardiac Surgical Register 1992-1994. There are no data to comment upon.

Annex D contains a letter from Dr Bull to Dr Bolson (spelling?) discussing Cusums

Advice requested in your letter

i) *Are the statistical methodologies used (Chi-squared, Cusum etc) appropriate for the purpose of monitoring differences in mortality rates?*

The chi-squared test is a simple method for assessing whether differences in proportions could have arisen by chance. It is valid but has the usual drawbacks of a significance test:

- 1) the p-value is dependent on sample size,
- 2) a statistically significant difference is not the same as a clinically important one,
- 3) a lack of statistical significance does not mean that there is not a clinically important difference present.

The CUSUM chart is a well established technique for monitoring quality, especially in the manufacturing industry. It may be worth pointing out that the paper '*Quality control: an application of the cusum*' Williams SM, Parry BR and Schlup MMT . BMJ 1992, 302 1359-1361 had appeared during the period in question 1984-1995. In general it is a decision aid, and most importantly it requires a reference or target failure rate. This is easier to obtain in industry than in medicine where one can set an 'acceptable' failure rate on, say, economic grounds.

ii) *Are the analyses carried out feasible and appropriate and are the results soundly based? (some analyses are carried out on non-comparable years)*

The analysis in A compares Bristol 1990-1992 with UK in 1991 The authors argue this probably favours Bristol since the national average improved over the time period (i.e. if they had chosen 1992 for the UK Bristol would have looked worse).

The analysis in B compared Europe and Bristol only for 1993.

Thus I don't think the analysis has been compromised by the use of different time periods.

iii) *With specific reference to the period 1984-1995, were there widely available and accessible statistical methodologies that would have been more appropriate and robust for use in monitoring differences in mortality outcomes?*

I know of no other simple methodologies that might have been used and understood locally. Of course the CUSUM methodology was available before the reference period and might have been used for monitoring, but it is not a technique that was appreciated by the medical community at the time.

I hope this is of use.

Michael J. Campbell
Professor of Medical Statistics