A Response to

"Submissions on behalf of Mr. J D Wisheart: Appendix 2 The Inquiry's Statistical Analysis".

David J Spiegelhalter

Stephen Evans

Paul Aylin

Gordon Murray

May 2000

All views expressed are the responsibility of the authors alone, and do not necessarily represent the views of the Bristol Royal Infirmary Inquiry.

Summary

This Response has been prepared at the request of the Inquiry Panel, and covers those parts of the *Submission on behalf of Mr Wisheart* that concern statistical analyses provided to the Inquiry. We have carefully considered the points made, and can summarise our response as follows:

- Our estimates of excess mortality are not based solely on Switch and AVSD operations - other procedures make significant contributions.
- Excess mortality cannot be explained by identifying additional risk factors for patients with adverse outcomes: the risk profile of the entire series must be considered.
- Discrepancies between Surgeons' Logs and the Inquiry's Data are inevitable, since the former is based on diagnoses and the latter based on procedures. A common coding scheme has been used and so similar discrepancies would be expected in all centres. There is very little disagreement as to whether individual children died or not. This supports the value of overall comparisons of pooled open operations, since these are not so susceptible to coding problems.
- Missing outcomes in HES data has negligible effect on the conclusions.
- When no data source is a gold-standard, corroboration between reasonably independent sources reinforces the conclusions from both.
- Further investigation of the accuracy of the mortality rates derived from the HES data has shown that over 95% of 30-day deaths following open surgery are recorded in HES, and that Bristol's accuracy is typical.
- Risk stratification for surgical risk factors may not be appropriate when evaluating an organisation since it may tend to obscure limitations in pre-operative care.

Conclusions: Although we have had some months to reflect on the issues and carry out further examination of the available data, we see no statistical justification to revise to

any substantial extent the analyses and opinions stated in written and oral evidence to the Inquiry.

Detailed Commentary:

Section 2.2. Findings of Preliminary analysis.

Excess deaths.

The Submission states *'That there are excess deaths in the neonatal switch operations and C-AVSD operations in 1991-1995 is not in dispute',* but goes on to ask *'are the other sub-groups in the under one year of age within an acceptable range either individually or when aggregated'*. Our analysis did not consider results broken down by surgeon, so we can only try to answer this question with regard to overall performance in Bristol.

The following information can be extracted from Tables 7.3.1, 7.3.2 and 7.4.5 of Spiegelhalter (1999). It is not feasible to re-run the full complex analysis for this particular subset of patients, so the 'p-value' is based on a simple comparison between the mortality rate elsewhere and that in Bristol (this p-value is the chance of observing such a difference by chance alone, and is based on a standard 'chi-squared test').

Source	*Mortality elsewhere*	*Mortality in Bristol*	*Estimated excess deaths*	*Simple p-value*
HES	248/2201 = 11 %	21/130 = 16 %	6.4	.12
CSR	279/2257 = 12 %	25/111 = 22 %	11.3	0.003

Table 1. Results for open operations, under one year of age, 1991-1995, excluding switch (group 3) and AVSD (group 5) operations.

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. However, Tables 7.3.1, 7.3.2 show that there can be at least 95% confidence in excess mortality in some subgroups: for example TAPVD in < 90 days, and Closure of ASD in 90 days to 1 year. The data reported to the CSR show significant excess mortality, even excluding switches and AVSDs. (It could be argued, because of the known lack of distinction in the CSR between switch (group 3) and inter-atrial repair (group 2), that group 2 should also be excluded from Table 1. We have repeated the analysis excluding group 2, and it increases the contrast between Bristol and elsewhere).

The individual subgroups contributing to Table 1 are small. We feel a better guide is the Table in the Executive Summary of Spiegelhalter (1999) (INQ 15/0004), that clearly shows that switches and AVSDs are not the only significant contributors to the observed overall excess mortality.

Risk stratification.

We agree that this should be carried out wherever possible. The Submission mentions *'significant additional risk factors in eleven of the fifteen patients'* in a series of C-AVSDs between 1990 and 1994, nine of which died. Our analysis does not identify this particular set of patients, but it is informative to work out what risk these eleven cases would need to have had in order for the mortality rate not to be in excess of that expected.

Suppose the risk for each of the remaining four 'standard' patients were 25%, and hence we would expect one death out of the four. This 'explains' one of the nine observed deaths. Then, for the remaining eight deaths not to reflect an excess mortality, the underlying risk for each of the eleven with additional risk factors would have to be at least 8/11 = 73%. This very high figure reflects the fact that risk stratification is not just a matter of identifying additional factors that might explain adverse outcomes of retrospectively identified patients: the risks of *all* patients need to be considered. Thus the excess mortality in open operations in Bristol could only be explained by risk-stratification if a large proportion of the patients had additional risk factors – not just the

ones who had adverse outcomes.

2.3 The Reliability of these Findings.

2.3.2 Comparator data:

The Submission expresses concern about under-reporting and varying definitions of deaths in other centres. There is always the possibility, although it does not seem especially plausible, that Bristol has produced good-quality data, while the bulk of the rest of the country were systematically under-reporting mortality. The current exercise comparing reported mortality with centralised death records could help answer this.

The issue of unknown survival status is dealt with below.

2.3.3 Statistical methods.

2.3.3.1 The effect of coding and grouping.

Coding in paediatric cardiac surgery and cardiology is notoriously difficult. Our coding scheme was developed after extensive consultation and was applied in an unbiased and systematic way to all centres. Specific issues regarding discrepancies with the Surgeons' Log are discussed below. However, it is important to note that 'errors' in coding will tend to make patient groups more homogeneous and hence lead to high-risk groups having lower observed mortality, and low-risk groups having higher mortality. Since there is no dispute about the total number of deaths, it does not seem reasonable only to focus on discrepancies where mortality appears to have been over-stated – if such groups exists, they will be balanced by other groups in which mortality has been under-stated. No formal comparison of the reliability of coding across centres has been carried out.

2.3.3.2 Missing outcomes

We have carried out a simple analysis to examine what the impact of these missing outcomes might be, taking the most optimistic view that they all were survivors. The following data are taken from INQ 13/0055-0057, 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 excess deaths would have been reduced by around 3.6, from 34.3 to 30.7. Note that this analysis does *not* assume that missing outcomes elsewhere were survivors.

Thus, even if we assume that all missing outcomes were survivors, there is little effect on the findings. We therefore reject the conclusion that missing outcomes makes the HES analysis unreliable.

Age group	*Number of missing outcomes in Bristol for open operations*	*Mortality elsewhere for open operations*	*Number of additional deaths expected if Bristol were 'typical'*	*Reduction in excess number of deaths*
< 90 days	7	16%	1.1	1.1
90 days – 1 year	22	7%	1.5	1.5
> 1 year	19	5%	1.0	1.0
	48		**3.6**	**3.6**

Table 2. Impact of including all HES data for Bristol with missing outcomes, and assuming they all were survivors.

2.3.3.3 Aggregation and pooling of data.

The distinction between 'case-mix' (operative procedures) and 'risk-stratification' (clinical risk factors) is very useful. By aggregating over consensus groups we achieve adjustment for case-mix, since excess mortality is only attributed in comparison with mortality elsewhere within the specific stratum defined by operative group, age group and epoch. That is why we present data both for *pooled* open operations, and *aggregated*

over operative group. The summary table on INQ 15/0004 shows this makes little difference in the conclusions.

2.3.3.4 Discrepancies between the Inquiry's data and the Surgeons' Data

It is important to emphasise that the entire analysis of paediatric cardiac surgery at UBHT has been based on operative procedures rather than on diagnosis. This was made very clear in our reports. Two of the major reasons for choosing to use operation were - a) the UKCSR recorded data by numbers of procedures rather than numbers of diagnoses, and b) when comparing different centres, it is likely that agreement about procedures may be greater than agreement on diagnosis. The Submission presents its analyses based on diagnosis rather than on operation, and hence considerable discrepancies must be expected between the analysis of the Inquiry's Data (including that of the Surgeons' Logs) and the analysis in the Submission of the Surgeons' Logs. The Appendix to this report considers the general issues and specific instances in particular.

Further analysis based on linkage of HES records with national death certification records has been carried out by Professor Murray and will be reported to the Inquiry. This shows that in open operations HES identifies around 95% of 30-day deaths (in spite of HES only aiming to capture in-hospital deaths). In conclusion, we do not find statistical evidence to support the statement *'that the estimate of excess deaths based on HES data is substantially wrong'*.

2.4 The Statistical Position in January 2000

2.4.1/2. Unreliability of data. We agree that no source of data can be considered as a gold-standard. However, if two reasonably independent sources of evidence corroborate each other and are largely consistent, then this supports both their conclusions. Furthermore, there is no statistical justification for the claim that using pooled data on open operations is in any way '*unreliable'* – in fact, given the difficulties in obtaining agreed coding categories of diagnoses and operations, such a pooling may be more

reliable than a more sophisticated technique.

2.4.3 Unreliability of conclusions.

The statistical evidence does not support the claim that *'the uncertainties in the preliminary data render them unreliable as the basis for any judgement'* – the strength and consistency of the 'signal' dominates the indisputable 'noise' that exists.

2.4.4 Team activity.

The acknowledgement of the importance of the team activity serves to downgrade the need for an analysis stratifying for factors present at surgery. Care prior to surgery may affect the presence or knowledge of such factors, and hence 'adjusting' for these could tend to obscure important differences between centres in pre-operative care.

References

Spiegelhalter D J (1999) An initial synthesis of statistical sources concerning the nature and outcomes of paediatric cardiac surgical services at Bristol relative to other specialist centres from 1984 to 1995. *Bristol Royal Infirmary Inquiry, INQ15.*

Appendix: Notes on apparent discrepancies between the Inquiry's Data and Surgeons' Logs.

Diagnostic categories

The submission (SUB 0009 0025, section 2.3.3.4) refers to Atrial Septal Defects ("ASD") and suggests that Mr Wisheart and Mr Dhasmana carried out operations on 102 children in the period 1991-1995. It is not known exactly how Mr Wisheart has obtained these numbers. His own computerised records of his log has 102 operations with a diagnosis of "ASD" over the whole period of the Inquiry (1984-1995), and a total of 39 in the period 1991-1995 (6,9,11,8,5 in the individual years). No children in this group are recorded on his log as dying in that period.

The Inquiry's version of the Surgeons' logs (SL) has been compared to Mr Wisheart's log (WL). For the 97 operations described as "ASD" in WL that are able to be unequivocally linked with operations in SL, 95 of them are classified with ICD9 code 745.5, as one of the diagnostic codes used. The text describing this code is "Congenital Atrial Septal Defect". The concordance appears to be considerable, but there are a number of operations where the code 745.5 is used, but where the description by Mr Wisheart is not a simple ASD. The list of other diagnoses in WL for these operations from the SL is given in table 1 below.

Diagnosis	Freq.
C-AVSD	3
Fallot	4
Fontan-SV	4
Fontan-TA	3
MISC	9
P-AVSD	1
PA+IVS	10
PA+VSD	1
PS	5
TAPVD	4
TGA	7
TGA+VSD	1
VSD	7
VSD+PS	6
Total	**65**

Table 1 Diagnoses in Mr Wisheart's log for those cases that have an ICD9 code of 745.5 in the coded surgeons' log, but are not "ASD" in Mr Wisheart's log.

In all cases, the code 745.5 is accompanied by other codes. There are 70 of the ASD operations where there is code 745.5 on its own as a primary diagnosis; Mr Wisheart classes all of these as "ASD". There are "ASD"s where there is more than one ICD code recorded as a primary diagnosis as well as 745.5.

Operation codes

Examining the operation codes (from SL) for those operations in WL described as "ASD"; 81 of them are classed as K10 using OPCS-4 operation codes; there are a further 20 that are classed as K10 that are not ASD. For all the statistical analyses it is clear that the operations were grouped by operation code. {Details at INQ 0013 0054} The group corresponding to operations for ASDs is group 6. This group was defined by those

operations that were K10 (any code beginning K10) or K09.4 or K20. Of the patients with a diagnosis of ASD, in addition to the 81 with K10, there are 6 with K20 and 10 with other operations not in group 6.

The text for these operations from the OPCS 4 code is:

```
K10     Closure of defect of interatrial septum
Excludes: When associated with correction of tetralogy of fallot (K04)

K10.1   Closure of defect of interatrial septum using prosthetic patch
K10.2   Closure of defect of interatrial septum using pericardial patch
K10.3   Closure of defect of interatrial septum using tissue graft nec
K10.4   Primary closure of defect of interatrial septum nec
K10.5   Revision of closure of defect of interatrial septum
K10.8   Other specified
K10.9   Unspecified

K09.4   Closure of persistent ostium primum

K20     Refashioning of atrium
K20.1   Correction of persistent sinus venosus
K20.2   Correction of partial anomalous pulmonary venous drainage
K20.8   Other specified
K20.9   Unspecified
```

It can be seen that there are a number of operations that involve closure of a defect, and it is likely that they will not all have had the same diagnosis. Hence the tables in the reports and those data provided by Mr Wisheart are not referring to the same children.

In WL, there are several different operations used for those with “ASD” as their diagnosis. These are shown in Table 2 below.

Operation	Freq.
Clos-patch;ligPDA;SutVSDs	1
Clos;patch;corrPAPVD	1
Closure	60
Closure-Patch	1
Closure-patch	25
Closure-patch;cor	1
Closure-patch;enlgt SVC	1
Closure-patch;pul valvot	1
Closure; corr PAPVD	1
Closure; inspect PV & PA	1
Closure; lig PDA	1
Closure; pul valvot	2
Closure;Pul Valvot	1
Closure;Pul valvot	1
Closure;pul valvot	1
Closure;reconstrLPA	1
ClosureA	1
Repair	1
Total	102

Table 2 Operation descriptions from Mr Wisheart's log for those with a diagnosis of Atrial Septal defect

Outcome

In the coding of the SL there are 11 children in group 6 having operations between 1991 and 1995 who are recorded as having died. We shall label these as A to K. All but *D* were under Mr Dhasmana. *D* had a diagnosis of "MISC" in Mr Wisheart's grouping of diagnoses, and the operation was recorded in his version of the log as "RA Thrombect; cl PFO;ExplPAs". This child is recorded as dying in Mr Wisheart's log.

Of the 10 cases in Mr D's log, three had a 745.5 diagnosis recorded; four had 745.1 (congenital anomalies of great vessels); two had 745.6 (congenital ostium atrioventriculare commune); and one had 746.5 (congenital mitral stenosis). Most had more than one diagnosis recorded.

All these children also appear in the CCR as having died. Eight of them have codes in group 6 from the coding of the CCR records. It should also be noted that Group 6 is ranked 11, so that other open operations in the grouped procedures will take precedence if they also occur with an operation coded as 6.

There is strong agreement between the SL, CCR and PAS in the diagnostic categories assigned to the 11 children recorded as dying. In most cases they have more than one diagnostic code assigned to them, but not all of them have a 745.5 code assigned. For the one case in Mr W's log (*D*) the PAS has a code of 745.51 assigned. In the PAS eight of them are classed with an operation within group 6. In each of the sources they are all recorded as having died.

Operations for valve surgery

Similar problems occur for groups 10 and 11, which Mr Wisheart summarises by diagnostic group. Mr Wisheart has a total of 245 different terms for coding his operations, though some of these are a result of inconsistent spelling *etc*. It is very difficult to be certain which operations are for the different valve operations from his own log (WL). It should also be noted that he has 88 children in the "MISC" group, some of

whom have been classified in other groups by SL, CCR or PAS. The mortality rate in this group is very high (33 % early and 20% late mortality).

The point at issue is the classification of operations. There is no evidence that deaths have been recorded when they have not occurred in more than a very few instances overall. The problem is that the classification of operations is difficult. With random misclassification of type of operation, but accurate determination of death, then will be a tendency for mortality rates in the different groups to be more similar to one another than would be the case if no misclassification occurred. In particular groups there may be a higher rate, but in other groups there will be a lower rate than there should be. Focussing only on the groups with a higher rate is biased. It is for this reason that examination of all open operations was also done in the statistical analysis. The other issue is that coders in different centres, who are each familiar with the OPCS4 system, will tend to code operations in a way that reflects that coding system, rather than clinicians' views. The key comparisons are made between centres, and no doubt, individual clinicians in those other centres are also likely to have different ways of classifying their operations. Random misclassification is likely to make the different groups more similar across centres also.

Summary

There is very little disagreement between the sources of data in regard to individual children as to whether they died or not. There is disagreement between Mr Wisheart's grouping by diagnosis, and the other sources that are grouped by operative procedure. While it is possible that some groups seem to show a higher rate in the statistical reports provided to the Inquiry than in Mr Wisheart's grouping of the data, there will be other groups where Mr Wisheart's data would seem to have a higher mortality rate than the statistical reports. He has not drawn attention to these, since his own comments apply only to selected groups. It is not clear from Mr Wisheart's submission how he obtained the numbers of operations or deaths for those operations carried out by Mr Dhasmana. If it was by hand searching the original logs, it seems possible that errors were made in

the numbers of deaths. It is clear from the Inquiry's coded version of Mr Dhasmana's log that there were some deaths in the period 1991 to 1995 for operations that were coded solely as ASDs, and not zero deaths as stated by Mr Wisheart.