

FUTURE IMPROVEMENTS IN THE ROUTINE MONITORING OF SURGICAL PERFORMANCE

Jarosláv F. Stark, FRCS, Honorary Senior Lecturer, Department of Paediatric Cardiology and Cardiac Surgery, Institute of Child Health, University College London.



The research on which this report is based was funded by the Bristol Royal Infirmary Inquiry. All views expressed in the report are the responsibility of the author alone, and do not necessarily represent the views of the Bristol Royal Infirmary Inquiry.

Substantial parts of section 3 are taken from Stark JF and Stark J, 2000, Databases for Congenital Heart Disease: A Beginner's Guide, in *Seminars in Thoracic and Cardiovascular Surgery: Pediatric Cardiac Surgery Annual 2000*, ed. Williams WG, Vol. 3, pp 110-116, W.B. Saunders, Philadelphia, PA, USA

EXECUTIVE SUMMARY

INTRODUCTION

This review was commissioned by the BRI Inquiry with the aim of commenting on the key issues involved in the monitoring of surgical performance and on its improvement in the future. My brief was to prepare a report of advice to the Inquiry based on my own personal knowledge and experience as a clinician, rather than on wider consultation or research.

MAIN AIMS OF MONITORING

In this report I define monitoring as the continuous collection of data and its analysis. The main goal of any performance monitoring activity must be benefit to the patient. Monitoring must be perceived as a means of improving clinical practice, and not as a means for assigning blame or punishment.

DATA COLLECTION

Monitoring can be divided into two distinct tasks: the collection of raw data and the subsequent statistical analysis. Data collection needs to be carried out in a complete, accurate and consistent manner. Written instructions and adequate training must be available to all users. Validation of data should be carried out both by users and by outside teams.

SIZE OF THE DATA SET

This needs to be based on the purpose for which the database is to be used and only that data required for this purpose should be collected. For routine monitoring of surgical performance a small data set that requires only a simple database is adequate. A small data set makes it feasible for all the work to be carried out by a single person, preferably a surgeon; this in turn is likely to improve the consistency and accuracy of the data.

Large comprehensive data sets enable more detailed data analysis, risk stratification and research. However, they need significantly larger resources to set up and maintain. Data entry will take longer, which means that a larger number of people need to be involved and data entry is more likely to be carried out by clerical or secretarial staff. This increases the likelihood of errors and makes error detection and validation more difficult.

UNIFIED NOMENCLATURE

In order to pool data from different units, develop national standards and devise methods of monitoring performance, the same nomenclature must be employed in all units. Currently, in paediatric cardiac surgery there are probably as many nomenclatures of diagnoses and operations as there are databases. Recently the American and European cardiothoracic societies proposed a unified nomenclature with the hope that it would be accepted worldwide. This nomenclature is not yet in use in the UK. Mapping data from one nomenclature to another may be feasible, but is undesirable. It is complex, time consuming, expensive, and likely to increase errors.

EXISTING SYSTEMS OF DATA COLLECTION IN PAEDIATRIC CARDIAC SURGERY.

There are many systems used at departmental, national and international level, but none has been formally evaluated. It is therefore difficult to make any recommendations as to which system would be suitable for monitoring performance in the UK.

STATISTICAL ANALYSIS

The range of operations performed by any individual paediatric cardiac surgeon is large and many procedures are performed infrequently. This makes meaningful statistical analysis difficult. In the absence of standards, results of operations can be only compared to results published in the literature. This is often unsatisfactory, as only the best results are usually published. It is important that confidence intervals are always quoted.

In paediatric cardiac surgery neither risk stratification nor standards have yet been developed. I believe that the data routinely collected for purposes of monitoring may not be suitable for these tasks and that risk stratification and the development of standards should be approached as well designed and separately funded research projects.

Statistical methods alone cannot be used as sole arbiter of what is considered acceptable performance. Only if the results of statistical analysis are treated as one element of a broad range of information contributing to the improvement of care will they fulfil a useful role. It is important that the statistical techniques used are comprehensible to clinicians and the public.

OTHER SOURCES FOR PERFORMANCE EVALUATION

A number of other activities can contribute to an overall performance monitoring programme in addition to the collection and analysis of data. An important contribution could be made by departmental audit supplemented by evaluation and advice from the outside teams of experts. Another innovative approach, used by the BRI Inquiry is the evaluation of randomly selected patients notes by a team of experts. This allows the assessment of all aspects of care, not only of surgical performance. Finally, there is the potential to develop for the systematic review of case series into a method for performance evaluation.

CONCLUSIONS

The effective monitoring of surgical performance requires:

1. The collection of accurate and complete data in each department of paediatric cardiac surgery; error checking and validation of this data; provision of training for all users.
2. The introduction of a unified nomenclature and agreed data set in all departments of paediatric cardiac surgery in order to ensure the consistency of collected data.
3. The organisation of national data collection by pooling of departmental data.
4. The establishment of an independent advisory board (surgeon, statistician, manager, IT specialist, patients' representative) which would assist in the development of methods for monitoring surgical performance.
5. The development of risk stratification, standards and performance evaluation methods which take into account the complexities of congenital heart defects.
6. The involvement of parents (patients) in the development of suitable methods of providing information about the outcomes of surgery to parents, clinicians and managers.

None of the above are achievable without the provision of adequate, separate funding for equipment, infrastructure, personnel and training. Close co-operation between clinicians and managers will be essential in ensuring that the monitoring of performance is not an end in itself, but that its results are used for the improvement of patient care.

CONTENTS

1. Introduction	6
2. Basic Principles of Monitoring	6
2.1. <i>Definition</i>	6
2.2. <i>Aims of Monitoring in Paediatric Cardiac Surgery</i>	6
2.3. <i>Key Characteristics</i>	7
3. Data Collection	7
3.1. <i>Manual v. Computer Based Systems</i>	7
3.2. <i>Errors</i>	8
3.3. <i>Choice of Data Set</i>	8
3.3.1. <i>Small Data Set</i>	8
3.3.2. <i>Large Comprehensive Data Set</i>	9
3.3.3. <i>Different Data Sets for Different Specialists in the Same Team</i>	9
3.4. <i>Nomenclature</i>	10
3.4.1. <i>Nomenclature of Congenital Heart Defects</i>	10
3.4.2. <i>Size of the Nomenclature.</i>	10
3.4.3. <i>Currently Used Nomenclatures</i>	10
3.4.4. <i>Mapping</i>	11
3.5. <i>Written Instructions for Users</i>	11
3.6. <i>Data Entry</i>	11
3.6.1. <i>Completeness, Accuracy and Consistency</i>	11
3.6.2. <i>Complex Defects</i>	12
3.6.3. <i>Coding Deaths</i>	12
3.7. <i>Resources</i>	12
3.8. <i>Existing Systems of Data Collection</i>	12
4. Error Correction and Data Validation	13
4.1. <i>Error Correction</i>	13
4.2. <i>Validation</i>	13
5. Data Analysis	14
5.1. <i>Introduction</i>	14
5.2. <i>Assessment of Surgeons' Performance</i>	14
5.3. <i>Confidence Intervals</i>	15
5.4. <i>Risk Stratification</i>	15
6. Standards	15
7. Other Sources for Evaluation of Performance in PCS	16
8. Conclusions	17
8.1. <i>Recommendations</i>	17
8.2. <i>Controversial Issues</i>	18
8.2.1. <i>Surgeons' Involvement in Data Collection</i>	18
8.2.2. <i>Unified System of Data Collection</i>	18
8.2.3. <i>Central Database</i>	18

9.	APPENDIX: Existing Data Collection Systems	19
9.1.	<i>Departments of PCS in the UK</i>	19
9.1.1.	Data Collected into Departmental Databases	19
9.1.2.	Data Collected by Hospital Administration	19
9.2.	<i>Departments of PCS Abroad</i>	19
9.3.	<i>National Data Collection Systems</i>	19
9.3.1.	UK CSR (original format)	19
9.3.2.	UK CSR (modified format)	19
9.3.3.	Central Cardiac Audit Database (CCAD)	20
9.4.	<i>International</i>	20
9.4.1.	European Cardiac Surgical Registry (ECSUR)	20
9.4.2.	AEPC	20
9.4.3.	Cardiac Consortium [15] - USA	20
9.4.4.	Congenital Cardiac Surgeons Society (CCSS) - USA	20
10.	List of Abbreviations	21
11.	References	22

1. INTRODUCTION

It is now widely accepted that the routine monitoring of results has an essential role to play in the provision of a safe and effective medical service, whether in paediatric cardiac surgery, or in any other discipline. In response to this, the paediatric cardiac community has made a number of attempts to establish a universally acceptable system for the collection and analysis of surgical data. A variety of systems, both in the UK and world-wide, have been developed and implemented to various degrees. Some of these failed and were subsequently abandoned, whilst others have continued to be used, but little is known about their performance. There have been no studies which would systematically evaluate and compare various systems in current use. Most information is anecdotal and cannot be relied on in making recommendations for a system that would satisfy the needs of patients, surgical teams and health service managers. **In my personal opinion, there is currently no system that would fulfil the needs of all these groups of stake holders and that has been tested for use on a national basis.**

In preparing this report, I have been asked to comment on key priorities in improving monitoring systems. In doing so, I believe that it is important to take a step back and establish the fundamental principles that any satisfactory monitoring system will need to satisfy. By far the most important of these is the absolute necessity of setting clear goals, which are agreed by the main stake holders. In particular, paraphrasing [1]

In designing and implementing any data collection system, the first question that needs to be addressed is what information one will want to extract from the system when it is functional, in what form, and to what use this information will be put.

The inclusion of any features that do not directly contribute to the entry, validation, organisation, storage and extraction of the desired data should be viewed with great suspicion. In particular, there is no point expending a great deal of effort entering data which will never be used. This point is unfortunately all too often overlooked, sometimes at great cost.

This closely echoes the final paragraph of section 3.1 of [2]:

Attention was drawn to the danger of collecting data on what is most easily measured, irrespective of its significance. Rather, the focus should be on making the important measurable; ensuring that measures are meaningful and the information generated is disseminated in an accessible form to those who need it and ensuring that it is used appropriately. ...

2. BASIC PRINCIPLES OF MONITORING

2.1. DEFINITION

Although specialised technical definitions of the term exist in the audit literature, for the purposes of this report I will adopt the meaning in common usage, namely the continuous collection of data and its analysis.

2.2. AIMS OF MONITORING IN PAEDIATRIC CARDIAC SURGERY

Ultimately, the overriding focus of any medical activity is the health and welfare of the patient. In the case of paediatric cardiac surgery this primarily translates into maximising the

likelihood of survival, and minimising the risk of serious complications. These should therefore be the main goals of any performance monitoring activity. Within such a broad context, it is possible to identify benefits to the patient at three different levels:

- *Patient* In making decisions about potential procedures, the patients representatives (parents, guardians, GPs etc) need accurate, up-to-date and relevant information about the various risks involved in particular procedures. Ideally, such information should include both short- and long-term results.
- *Clinical Team* In order to continuously strive to improve surgical outcomes, to quickly identify problems, to evaluate the effectiveness of new procedures, to properly advise patients and to determine training needs, surgeons and other clinicians need very similar information to that required by patients.
- *Management* At both the local and national level, information is required to allocate resources and determine health care priorities.

2.3. KEY CHARACTERISTICS

The above goals imply that any clinical monitoring process should exhibit the following features:

1. It must be perceived by patients, clinicians and managers as a means for improvement of clinical practice rather than a mechanism for assigning blame or punishment. The system needs to be structured to give incentives to all those involved to ensure its success.
2. The methods used should be transparent, and perceived as reliable and appropriate.
3. Monitoring must be incorporated into routine care, and carried out by those providing the care, since they are the only ones who can ensure the accuracy of recorded data. This is precisely the point made in the final sentence of section 3.1 of [2].
4. The results of monitoring should be available to all stake holders in a helpful and appropriate format.
5. Monitoring is not an end in itself and must not be considered in isolation. Unless the results of monitoring affect subsequent practice there is little point in undertaking monitoring in the first place. Hence there must be suitable mechanisms for ensuring that the outputs of the monitoring process are used constructively to improve patient care.

Broadly speaking monitoring can be divided into two distinct tasks: the collection of raw data and the subsequent statistical (or other) analysis. In line with the principles espoused above, it is important that these two activities are integrated. Thus, the data that is collected should be that which is required to carry out the analysis, and conversely the analytical methods used must take into account the practical limitations of data collection procedures. This issue is particularly important in risk stratification, discussed below. I shall now discuss each of these tasks in more detail.

3. DATA COLLECTION

3.1. MANUAL V. COMPUTER BASED SYSTEMS

Although in principle it is possible to collect the data required for monitoring using a manual system, a well designed computer database for this purpose can have many advantages. In

particular it can facilitate data entry and retrieval, can automate some error checking procedures, can provide sorting and searching operations, can offer a variety of methods for viewing and presenting the data, and can permit non-local access. It must however also be borne in mind that a computer based system also has disadvantages: it requires investment in hardware, infrastructure and support, it necessitates an adequate level of computer literacy amongst all those using the system (including possibly patients), it is more vulnerable to threats to its security (*eg* viruses, attacks from hackers *etc*), and can compromise users critical faculties (it is a common phenomenon that people are more likely to accept computer generated results, even if they are obviously wrong). Care must therefore be taken in designing a database system to minimise such drawbacks. From now, I shall assume that a computer based system will be used.

3.2. ERRORS

Because of the small group sizes in paediatric cardiac surgery even small errors in collected data can have significant effects on any subsequent analysis. As described in [1], errors can enter in a variety of ways. The simplest are pure typographical mistakes such as misspelling a complicated name in a foreign language, or mis-keying a date of birth. Even such innocuous mistakes can have serious repercussions. If a patient has several operations and their name is misspelled only for some of them, then they may appear as two different patients in the database, whilst an incorrect date of birth can turn a neonate into a teenager. Both of these will affect a variety of statistical measures. The second class of errors arises from inconsistencies in entering complex information, either by different individuals, or at different times. The final category arises from structural flaws in a database. A particularly serious example in the context of paediatric cardiac surgery databases occurs with patients who die following a number of procedures during one admission. Such a death is often not due to the last procedure carried out, and hence should not be ascribed solely to that operation, on the other hand recording the death under all the operations in that admission can easily result in a single patient giving rise to several deaths in any statistical summary. In my experience, I have seen cases of a single patient being recorded as dying up to five times in commonly used clinical databases. This is unacceptable as a basis for any meaningful analysis, yet it is not easy to design systems which minimise the occurrence of such errors.

3.3. CHOICE OF DATA SET

The most important factor in choosing a database is the choice of data that is to be collected and stored. This needs to be based on the purposes for which the database will be used.

3.3.1. SMALL DATA SET

If it is decided that the main aim of data collection is routine monitoring of surgical performance a small data set that requires only a simple database will be adequate. This will include basic information about patients, diagnoses, operations and results. In my experience a data set of 20-25 items is sufficient. These can be arranged on a single screen, which makes working with data easy. Even computer beginner can quickly learn how to enter data, how to search for information, how to view and sort data and how to prepare lists of groups of patients or operations. This makes it feasible for all the work to be carried out by a consultant surgeon himself/herself, which in turn makes it easier to ensure the consistency and accuracy of the data. The advantages of a relatively small data set thus include:

1. reduction in the occurrence of errors and increased ease of error checking;
2. minimisation of the resources (particularly time) needed to install, operate and maintain the system;
3. increased transparency and sense of ownership by users: the simpler the system is the more likely it is to be understood, and the more likely it is to be used.

3.3.2. LARGE COMPREHENSIVE DATA SET

On the other hand some surgeons suggest that the data collected should be as comprehensive as possible to facilitate detailed data analysis, risk stratification and research. The advantage of such an approach is that if it is decided at a later stage that a particular data item is important, then there is a higher likelihood that it is already contained in the data set, and hence can be retrospectively analysed from the inception of the database. However, such large comprehensive data sets have a number of drawbacks. They are likely to need significantly larger resources to set up and maintain, and their successful operation is likely to require much greater commitment and expertise from the whole cardiac unit. Data entry will take an order of magnitude longer (*eg* compare entering 25 items for a simple database to say 300 for a comprehensive one). This means that it is more likely to be carried out by clerical or secretarial staff, and a larger number of individuals are likely to be involved. Both factors increase the likelihood of errors and make validation **much more difficult**. Furthermore, even if the error rate per individual item of information remains the same, the increase in the size of the data set will increase the probability of any given operation record containing erroneous data [1].

A comprehensive data set is therefore only appropriate if the team includes at least one enthusiastic and highly computer literate member, who is prepared to accept the overall responsibility for data collection and supervise all the steps. Even then, the rest of the unit must be committed to the system, so that other team members do not feel that they are having the database imposed on them. As far as I am aware, few units in the UK satisfy these conditions, and it is likely that if they were to attempt to collect a large comprehensive data set, the resulting data would be incomplete and inaccurate. **I would thus recommend starting with a small limited set of goals that are achievable, rather than initially attempting to be too ambitious.**

3.3.3. DIFFERENT DATA SETS FOR DIFFERENT SPECIALISTS IN THE SAME TEAM

A paediatric cardiac surgeon usually works as a member of a team which includes cardiologists, perfusionists, anaesthetists, intensivists, nurses *etc*, who may also want to collect data concerning their own work. For a team that has a little or no experience with data collection the adoption of small databases for each subsection of the department may be considered, possibly introduced in stages. This approach may involve some duplication and hence potentially appear to be inefficient, but it is more likely to ensure the accuracy of the data, as each subsection would be responsible for its own data. We can speculate that such independent databases could be eventually integrated into a single system if the databases are designed from the beginning to be compatible and in particular if they adopt the same nomenclature.

3.4. NOMENCLATURE

3.4.1. NOMENCLATURE OF CONGENITAL HEART DEFECTS

To ensure consistency and accuracy it is essential that a precise set of possible diagnoses and operations is defined and that the database only permits the entry of data items from this set (using pick lists, drop down menus or similar). Paediatric cardiac surgery deals with a relatively large number of heart defects and their variations and consequently with a large number of different types of operations. The nomenclature of diagnoses and operations has not been unified and therefore the same defect may be called by different names. For example, Atrioventricular Septal Defect may be called Atrioventricular Canal or Endocardial Cushion Defect, depending on whether the nomenclature was based on anatomy or embryology.

3.4.2. SIZE OF THE NOMENCLATURE.

The size of the nomenclature is important. If all the variations of individual defects are included in a list of diagnoses, such lists may contain hundreds or even thousand items (*eg* as in the AEPC nomenclature [3]). Data entry from such long lists will be cumbersome and slow and will be more likely to require arbitrary decisions between choices that are only subtly different. The same problem applies to operations.

Lists of diagnoses can be shortened by omitting rare diagnoses, variations or sub-classifications of defects. For example each variant of ventricular septal defect (VSD) such as peri-membraneous VSD, inlet VSD, trabecular VSD, apical VSD, sub-arterial VSD, sub-aortic VSD, sub-pulmonary VSD, doubly-committed VSD *etc.* can be entered under a separate code. Alternatively, only one code (VSD) could be used for all these variants.

Whilst some such simplification is necessary to avoid the nomenclature becoming too unwieldy, too much simplification is also undesirable as it will lead to too many diagnoses and operations will be entered under the category of "other". In my own experience, **a list of diagnoses with 120 - 150 items is adequate.**

3.4.3. CURRENTLY USED NOMENCLATURES

As far as I know there are as many nomenclatures as there are CHD databases. From the limited information that we have about the databases used in the UK, these nomenclatures are probably not compatible. The differences between them make comparison and pooling of data from different units very difficult, if not impossible [4]. I believe that the lack of a unified nomenclature has been the most important single reason why the preparation of standards and guidelines in paediatric cardiac surgery has been delayed and methods for the assessment of the performance of individual surgeons have not been developed. The following are the more widely known nomenclatures available

1. The Society of Thoracic Surgeons (STS) in USA and the European Association for Cardiothoracic Surgery (EACTS) have recognised the importance of a unified nomenclature of diagnoses and operations for the collection of data in paediatric cardiac surgery. A joint committee of these two Societies has developed a "short" and a "comprehensive" list of diagnoses and operations, which were published in April 2000 [5].
2. This US/European initiative of paediatric cardiac surgeons was paralleled by an initiative of European paediatric cardiologists who recommended a different nomenclature of diagnoses and operations. Their lists were published in November 1999 [3].

3. In the UK the two national data collections, UKCSR and CCAD have been using their own nomenclatures.
4. In the UK hospital administrations and the Department of Health have been using ICD-10 codes for diagnoses and CPCS-4 for operations (HES data).
5. Various individual departmental databases have evolved their own local nomenclatures [6-8].

3.4.4. MAPPING

It is sometimes suggested that one can map data from one nomenclature to another. Whilst in principle this may be feasible if moving from a large detailed nomenclature to a more condensed one, the reverse process is typically impossible, since the data in the small nomenclature will not be sufficient to choose amongst a large number of potential terms in the large nomenclature. Even worse is trying to map between two significantly different nomenclatures of similar size. It also has to be stressed that even when possible, mapping is highly undesirable, since there will almost certainly be no way of verifying the resulting data.

Mapping between the two sets of CHD data was used in the analysis of UKCSR data and HES data for BRI Inquiry. The problems associated with this exercise have already been discussed by Murray [9] and Stark [10].

3.5. WRITTEN INSTRUCTIONS FOR USERS

To insure consistence of the data stored in the database written instructions for all users are essential. They should include clear definitions of fields to help to avoid ambiguity, written advice how to approach complex diagnoses, multiple procedures during one operation, or multiple operations during the same admission, operations on separate admissions *etc.* Clear descriptions of complex diagnoses and operations need to be agreed on and all those involved need to be trained in their interpretation. This implies that operating instructions as well as definitions of all terms should be prepared before the database is operational and should be made available to all those involved in data collection, data entry and data evaluation. Adequate training is of course also essential.

Written instructions are particularly important for multi-institutional national or international data collections. Such studies should not be started without written protocols. **I am very concerned that this important step has not been emphasised and has frequently been omitted.**

3.6. DATA ENTRY

3.6.1. COMPLETENESS, ACCURACY AND CONSISTENCY

Data entry needs to be carried out in as a complete, consistent and accurate manner as possible. This implies that all individuals involved require a sufficiently high level of medical training, and ideally should be consultant surgeons or cardiologists. Entrusting data entry to secretarial, clerical or even junior medical staff is likely to lead to much higher error rates, and hence render the database much less useful. For the sake of consistency, it is also preferable if as few different individuals as possible are responsible for data entry: in the case of a simple database, experience has shown that a single individual can easily manage the data entry even for a large unit [6-8].

To ensure consistency of nomenclature, a well designed database should include lists or menus of diagnoses, procedures, clinicians names, outcomes *etc* from which the appropriate choice can be entered with a single click of a mouse or keystroke. This avoids spelling and terminology errors which would arise if such data were typed directly. In modern databases such lists can be presented using intelligible text, rather than meaningless numeric codes.

Many simple typographical errors can be automatically caught by appropriate data entry validation routines built into a database, for instance by not accepting a date of birth later than the date of hospital admission. More sophisticated data entry checks might for example also warn of inconsistent combinations of diagnoses and operations, though I am not aware of any database that currently implements this.

3.6.2. COMPLEX DEFECTS

Clinicians are naturally accustomed to think about complex defects in terms of a single primary diagnosis and associated secondary diagnoses. The same applies to multiple procedures. It is therefore an advantage if the database reflects this and is designed to code only one primary diagnosis and only one primary operation, with as many secondary diagnoses and operations as necessary. This will require some arbitrary decisions on how to enter complex lesions. These decisions must be agreed on and documented to ensure consistency.

3.6.3. CODING DEATHS

Definitions have to be specified. The currently used 30 day mortality is probably not ideal. On one hand, very few departments have developed mechanisms how to track patients after discharge from the hospital. Thus some deaths occurring after discharge of the patient but within 30 days of operation may be not included. On the other hand, some patients after complex repairs, particularly small infants or premature babies may die in the hospital well outside the 30 day limit. Yet, re-coding such deaths as "late deaths" would be clearly inappropriate.

Careful thought needs to go into how deaths are recorded, so that they are assigned to the most appropriate of a sequence of operations on the same admission, and no patient is recorded as having died more than once.

3.7. RESOURCES

Most paediatric cardiac surgeons carry heavy clinical commitments and work long hours. Many of them have little training or experience in the use of computers. To successfully implement and use a database requires knowledge, and a significant investment of time and resources. Until recently, no extra (ear marked) resources were made available for the establishment, running and validation of data collection systems in the departments of PCS in the UK.

3.8. EXISTING SYSTEMS OF DATA COLLECTION

A variety of systems are in current use at the departmental, national and international level. I have not attempted a comprehensive survey of available systems, since this is outside the brief of this report. In the Appendix, I comment on the systems of which I have some knowledge. The range of the systems currently in use reflects different purposes for which they have been set up. None of these systems has been designed specifically for monitoring sur-

geons' performance and as far as I know none is ideally suited for this task without further modifications. It has to be stressed that before any data collection system is recommended for national adoption for monitoring surgeons' performance it must be tested and evaluated.

4. ERROR CORRECTION AND DATA VALIDATION

No matter how well designed the database, and how well trained the users, it has to be accepted that some errors will be introduced when the data is entered. It is therefore important that there are procedures for detecting and correcting as many errors as possible, and for validating the database. By the latter we mean that at an appropriate stage a decision is made whether or not the data for a particular period of time is sufficiently accurate to be used for its intended purposes.

Both error checking and validation are important to the success of any monitoring system. It has to be emphasized that both are greatly facilitated if the data is held on a local departmental database. If instead, data is collected and analysed centrally it is difficult to see how effective error checking and validation procedures can be implemented.

4.1. ERROR CORRECTION

Once data is entered into a database, it is essential that it should be checked for completeness, accuracy and consistency. In the same way that, as indicated above, some errors can be automatically caught by data entry validation routines built into a database, it is desirable that a database incorporates features that aid error checking. This includes flexible searching facilities (including for duplicate or blank fields) and the ability to see different summaries of the same data (to allow cross-checking). The day-to-day use of the database is another important mechanism for uncovering errors [6]. For instance, if a list of all operations in a given category is printed (say in order to give advice to a patient), inconsistent, aberrant or incorrect data may often easily be spotted. It may be helpful to prepare a protocol so that such "by chance" checking is developed into a systematic system. Procedures also need to be in place for correcting any errors found, and for notifying corrected data to a national pool if appropriate.

4.2. VALIDATION

Even with the most painstaking error checking, it is very unlikely that all errors will be ever detected. At some stage, however, it has to be decided that the data is of sufficient quality to be used, and analysed further. Ideally this should be done by a formal validation process, carried out on a regular basis. As far as I am aware there is little experience with methods for such validation and further research to develop appropriate procedures is required. A possible approach could be based on random checks of selected patients' records from the database against corresponding patients' notes. The data in the database could also be compared with the data obtained from other sources (perfusionists records, register of deaths *etc.*). This procedure could be carried out by users, but it would be advantageous if the validation could be also performed by an independent team. This team would ideally include a paediatric cardiac surgeon, an IT specialist and a statistician. The Society of Cardiothoracic Surgeons or the Royal College of Surgeons could establish such "validation teams", which would visit departments of paediatric cardiac surgery at regular intervals. Resources for the validation process would have to be made available.

Ideally validation procedures should yield an estimate of the error rates in the data and hence indicate the level of confidence one can place on the conclusions of subsequent analysis. As far as I am aware, there is currently little information of what are reasonable error rates to expect, and even less on the effect of different rates on subsequent conclusions. As indicated above, given the small group sizes encountered in paediatric cardiac surgery, it is likely that even low levels of errors can have very adverse effects.

5. DATA ANALYSIS

5.1. INTRODUCTION

The collection of data at the departmental level fulfils a number of the basic aims of monitoring, namely in providing patients and clinicians with information about the risks of particular procedures within that department. However, sooner or later patients, clinicians and managers will all want to compare the results from one unit with those from others, both in the UK and internationally. This will require a smaller or greater degree of statistical analysis. Such analysis may also be useful in indicating whether variations within one unit (with time, or from one surgeon or another) are significant, or can be ascribed to chance.

In exactly the same way that there is no universally agreed system for data collection, the statistical techniques for comparing results in paediatric cardiac surgery are immature. Some sophisticated approaches have been attempted [11], but the interpretation of these is difficult if not impossible for those without statistical training, *ie* both patients and clinicians. Much work therefore needs to be done to determine the best ways in which to analyse and present data to satisfy the requirements of patients and clinicians.

5.2. ASSESSMENT OF SURGEONS' PERFORMANCE

The range of operations undertaken by any individual paediatric cardiac surgeons is large and many procedures are performed infrequently. Very few departments of paediatric cardiac surgery in the UK have enough experience with regular monitoring of surgical performance. To evaluate mortality rates for rarely performed operations over a short period of time is difficult. In most departments it is left to individual clinicians how to approach this task. In only a few departments are they helped by statisticians.

The simplest approach probably consists of estimating overall mortality rates for all cardiac operations and mortality rates for open and closed procedures in two or three age groups over a period of 6-12 months. Such estimates can easily be automatically calculated by the database used to collect the data, and hence can be carried out without statistical support. However, great care needs to be taken in the interpretation of such simplistic measures. The outcome of surgery is influenced by many factors. Some differences may be due to natural variations and others due to various risk factors (*eg* associated non cardiac defects, prematurity, preoperative ventilation *etc.*). Furthermore different surgeons' case mixes will affect their overall mortality rates. Ideally, therefore, any results should be presented with appropriate statistical confidence intervals, and should take into account risk and case mix factors. Currently, as far as I am aware, **appropriate methods for this do not exist, and their development will require a significant research effort.**

Until suitable techniques are developed, the reliable assessment of surgeons' performance will be limited, particularly over a short timespans. This difficulty should not discourage the

collection of crude data, as the accumulation of such data over a longer period of time will facilitate more sophisticated approach in the future. In the meantime it has to be accepted that in the absence of methods of risk stratification and because of small data groups it is not always possible to provide parents with adequate information about the risk of a proposed operation. Similarly, it is **difficult to determine whether the performance of a surgeon or of the whole team can be considered adequate by statistical means alone.**

In the absence of standards, the results of individual operations can be only compared to results published in peer review journals. This is clearly unsatisfactory for monitoring purposes, as only the best results are usually published, periods of reporting are often different and not all consecutive patients are necessarily reported.

5.3. CONFIDENCE INTERVALS

To avoid misleading conclusions the data must be properly analysed. In particular, significance should not be ascribed to differences which may be explained by natural chance variations. It is therefore essential that confidence intervals are always quoted. In medicine 95% confidence intervals are often used. If we monitor 20 or more variables (*eg* mortality rates for 20 different age/operation combinations) per unit or per surgeon, then purely due to chance, we would expect to see one data point outside the confidence interval in each unit. We could use more stringent confidence levels, say 99%. Even then with over 10 units in the UK and 20 variables per unit, we would expect two points outside the interval due to chance. At the same time, since the data groups are small, even 95% confidence intervals can be very broad. We thus run the risk of on one hand spuriously identifying performance as below standard, and on the other of having such broad standards that they are useless for warning of potential problems. Given the small data groups, **it is not clear that this contradiction can be meaningfully resolved.** Even if sophisticated statistical tools can overcome this difficulty, it is important that they do so in a way that is meaningful to patients and clinicians. Statistical analysis is therefore a useful tool, but cannot completely be a replacement for careful ongoing scrutiny of all aspects of care.

5.4. RISK STRATIFICATION

For adult cardiac surgery there are well developed techniques for taking into account important factors such as case mix [12-14]. In PCS risk stratification has been used by Moller [15] and Hannah and colleagues [16], but their methods have not been accepted universally. Most paediatric cardiac surgeons believe that risk stratification is essential for a reliable assessment of their work. At the same time they are concerned, that risk stratification in CHD surgery may be too complicated and for routine monitoring may prove impractical. To develop methods of risk stratification will require a large collection of reliable data, which will need to include all those variables considered important for the outcomes of CHD surgery. I believe that data collected for routine monitoring is not suitable for this task and that **risk stratification should be approached as a separate, well designed and funded research project.**

6. STANDARDS

The dictionary definition of "standards" appropriate to task of performance monitoring is the "degree of excellence required for a particular purpose". In the present context is likely to be measured in terms of the maximum mortality and rate of complications that is considered to

be within acceptable limits. Currently the development of such standards in PCS in the UK has not started, at least partly due to the practical problems of collecting data and hence the current lack of reliable data on which standards could be based. I believe that development of standards should be carried out as a well designed research project and could perhaps proceed in stages:

1. A system for data collection which would be used for setting standards should be organised and then supervised. Standards both for all operations and for specific operations must be based on reliable data which are trusted by patients and parents, as well as the profession. I would stress that the use of results for comparison purposes should be based on careful analysis and results should be interpreted cautiously in order to avoid misleading conclusions [17].
2. It is important to agree how the standards should be set and then applied before they are set-up. There is need for careful statistical analysis and further research in this area.
3. Standards should be sufficiently broad to minimise the possibility of them being broken due to chance, while they should also provide early warning about the potential problems. This is not an easy task!
4. We must also bear in mind that even if results are satisfactory, or even excellent, serious mistakes might still be occurring in the treatment of individual patients. This can be illustrated by a theoretical example. A department may record a mortality rate for say arterial switch operation of 5%, which is low. When this result is analysed in more detail, it is possible that a number, say 10% of those surviving the switch operation, required postoperative ECMO. This should be considered a near miss [18]; it is likely, that such patients would have prolonged ITU and hospital stay. Results may also be apparently good due to chance variation, even if some aspect of care is sub-optimal.
5. The above examples illustrates that statistical methods alone cannot be used as a sole arbiter of what is considered acceptable performance. There are many other practical and ethical issues which need to be considered that go beyond the scope of the present paper. Only if standards are treated as one element of a broad range of information contributing to the improvement of care, then they may fulfil a useful role.

7. OTHER SOURCES FOR EVALUATION OF PERFORMANCE IN PCS

There are other possible approaches to the evaluation of performance which could be considered.

The results of operations are reviewed monthly by the whole team. Special attention is paid to patients with difficult postoperative course and to those who died. All aspects of patients management are discussed in order to elucidate reasons for the unsuccessful outcome. The reasons are often not found. Several options could be than considered.

1. If the death for a particular operation was an isolated death, careful monitoring of that particular operation should continue.
2. If the department's mortality for an operation is higher than that expected by the team (based on the information from other departments or literature review) careful evaluation of all patients who undergone that operation should be undertaken. This can be initially done by a member of the team and the result presented to the whole team.

3. It should be possible to invite a colleague from another department to help evaluating the management of those patients.
4. If the problem appears to be a serious one, invitation of a "Rapid Response Team" (RRT) could be considered. RRT's were suggested by the Royal College of Surgeons. These teams were originally envisaged to visit in response to a complaint against the team or against an individual member of the team. There is clearly no reason why the RRT could not visit at the invitation of the department of PCS which is seeking help. The RRT could help evaluating their results and prepare the plan how to rectify the situation.
5. Evaluation of randomly selected clinical notes by a team of experts, as used by BRI Inquiry in Bristol. Such a review can concentrate on wider issues of care as provided by all members of the team and not to be restricted to evaluation of mortality rates only.
6. The systematic review of literature (meta analysis). Although reviews of published papers on selected topics have been periodically published in specialist journals, the first systematic review was prepared recently for BRI Inquiry by Vardulaki *et al.* [19]. Preparation of such an analysis is clearly beyond the possibilities of an individual clinician. Only institutions with adequate resources and access to various specialists could undertake such task. The systematic review of literature may contribute to the assessment of performance in paediatric cardiac surgery.

8. CONCLUSIONS

8.1. RECOMMENDATIONS

The collection of reliable data is a prerequisite for monitoring of performance in PCS. In order to establish the system for collection of basic CHD data in all departments of PCS in the UK as quickly as possible, the following measures should be considered.

1. Provide financial support for establishing data collection systems in all departments.
2. Make available the same database to all departments of PCS. The most important feature of such database should be a short data set and easily manageable lists of diagnoses and operations.
3. Prepare guidelines on collecting data from original sources, on entering it into the database and on validating it.
4. Select one consultant in each department who would assume responsibility for data collection. He/she should receive appropriate training. Ideally he/she should enter all data to ensure the completeness and consistency of the data.
5. Ensure that monitoring is not seen as an end to itself. Consider a suitable mechanism for ensuring that the outputs of the monitoring process are used constructively to improve patients care.
6. Approach risk stratification as a separate, properly funded research project.
7. Consider pooling data from individual departments of PCS into a "national database".
8. Consider developing other measures, apart from statistical methods, for evaluating surgical performance.

9. Establish an independent advisory board (surgeon, statistician, IT specialist, patients' representative) which would assist in the establishment of data collection and monitoring in PCS in UK.

8.2. CONTROVERSIAL ISSUES

These recommendations raise several potentially controversial issues, which will need to be discussed and resolved.

8.2.1. SURGEONS' INVOLVEMENT IN DATA COLLECTION

Clinical governance requires that each surgeon should be responsible for monitoring his/her own performance. This implies that each surgeon should acquire adequate skills in working with database, performing basic statistical analyses, estimating risks etc. Yet most paediatric cardiac surgeons work excessive hours and are often on call every other day and every other weekend. Without adequate training, resources and support they may not be able to fulfil this requirement.

8.2.2. UNIFIED SYSTEM OF DATA COLLECTION

Unless the same data set and the same nomenclature is used in all departments of PCS in the country, the data cannot be pooled easily into a national database. If it is recommended that all departments use the database with the same data set and nomenclature, the departments which have been already collecting data into their own database may resent the change. If they change the system and do not manage to establish the continuity between their original database and the proposed new system, they would rightly feel that a lot of their previous work could be wasted.

8.2.3. CENTRAL DATABASE

If the decision is made to collect the data in a central database instead of at the departmental level, different problems will arise. It is likely that clinicians and patients would not have direct access to the data which would defeat one of the main purposes of setting up a data collection system. It would be very difficult if not impossible to validate the data. Such a system would not be transparent. According to Moller's experience [15], development of a centralised system that suits all users takes a considerable time.

9. APPENDIX: EXISTING DATA COLLECTION SYSTEMS

9.1. DEPARTMENTS OF PCS IN THE UK

9.1.1. DATA COLLECTED INTO DEPARTMENTAL DATABASES

Some departments or individual surgeons in the UK have been collecting their data into electronic databases for some time. In 1997 five departments with established electronic data collection systems agreed to participate in a study with the aim of evaluating their results of operations for CHD over a two year period. Results from the first year (1997-98) were recently published [4]. To carry out the analysis the results from each department were collated centrally. It was not possible to pool the data because each department was using a different database with a different nomenclature. From a statistical point of view this was considered less than optimal.

Not all departments of PCS in the UK have installed an electronic database system and instead use manual data collection.

9.1.2. DATA COLLECTED BY HOSPITAL ADMINISTRATION

In general data collected by hospital administration has not been used by clinicians for the evaluation of their clinical work. Such data is collected by trained non-medical coders who follow written instructions. ICD-10 codes are used for diagnoses and OPCS-4 codes for procedures. **I believe these codes are unsuitable for the evaluation of surgeons' performance.**

To my knowledge, with the exception of data concerning the treatment of children with CHD in Bristol, there has been no attempt to validate the CHD data collected by hospital administrations. Recent comparisons of hospital data with data collected into departmental databases in Southampton and Great Ormond Street showed considerable discrepancies [10].

9.2. DEPARTMENTS OF PCS ABROAD

There are many departmental databases in use in other countries. I can mention the systems at the Hospital for Sick Children in Toronto [6] and at the Children's Hospital in Helsinki [7], as both these systems have been successfully used for over 20 years. Data from these systems is available to the members of the department at any time; they are both operated by a single person.

9.3. NATIONAL DATA COLLECTION SYSTEMS

9.3.1. UK CSR (ORIGINAL FORMAT)

UK Paediatric cardiac surgeons and paediatric cardiologists decided at a meeting in January 2000 to continue collecting data into the original format UKCSR. The register's format has been repeatedly criticised. The structure of the register was described in detail by Murray *et al.* [9].

9.3.2. UK CSR (MODIFIED FORMAT)

At the same meeting it was also decided to collect data into the modified format of UK CSR. The modified format is simpler; it collects data on open and closed procedures and data on seven selected operations.

9.3.3. CENTRAL CARDIAC AUDIT DATABASE (CCAD)

Since 1.4.2000 all departments of PCS in the UK also send data about each individual patient to the CCAD. This is a newly created data collection system supported by the Department of Health. The system uses its own data set and nomenclature. As the system has not been used before and information about its structure and function is sparse, it is difficult to assess the role which this database might play in the monitoring of surgeons' performance.

9.4. INTERNATIONAL

9.4.1. EUROPEAN CARDIAC SURGICAL REGISTRY (ECSUR)

From 1.1.2000 the PCS database of the Warsaw Medical School has started to collect data from European departments of PCS. This database has incorporated the data set and the nomenclature recommended by STS/EACTS [5]. Data is available from the Centre, not at the departmental level. The project is supported by EACTS and the European Congenital Heart Surgeons Foundation.

9.4.2. AEPC

The Association of European Paediatric Cardiologists has recently proposed a new nomenclature [3]. To my knowledge the software incorporating this nomenclature has not been developed.

9.4.3. CARDIAC CONSORTIUM [15] - USA

Valuable information about collecting data into the multi institutional central database was published by Moller [15]. The project started in 1982 by collecting data from a few North American departments treating children with CHD. Gradually more departments joined the program; by 1996 the number of participating departments had increased to 39. The data is collected and analysed centrally, but the validation is carried out locally and the results are then made available to individual departments. The results are available with a 1-2 years delay; the database provides excellent information about trends in mortality rates but the system may not be suitable for the assessment of surgeons' performance.

9.4.4. CONGENITAL CARDIAC SURGEONS SOCIETY (CCSS) - USA

This Society started collecting data from several leading departments of PCS about specific defects and operations in late 1980's. Several important papers about operations for transposition of the great arteries, coarctation of the aorta and pulmonary atresia were published from this data collection.

10. LIST OF ABBREVIATIONS

PCS: *Paediatric Cardiac Surgery.* Surgery on the heart and great vessels performed on children. The age is defined differently by different people, but in general patients under the age of 16 years are included

CHD: *Congenital Heart Defect.* Malformation of the heart and great vessels with which the child is borne (as opposed to acquired heart defects, which may be the result of infection, trauma *etc.*).

AVSD: *Atrioventricular Septal Defect.* Complex CHD involving the walls between both atria (collecting chambers of the heart) and the ventricles (pumping chambers, as well as valves between those chambers. Repair of this defect is complex.

VSD: *Ventricular Septal Defect.* A “hole in the heart” defect between the two pumping chambers.

ECMO: *Extra Corporeal Membrane Oxygenation.*

ITU: *Intensive Therapy (Care) Unit.*

UKCSR: *United Kingdom Cardiac Surgical Register.* System used by the Cardiothoracic Society of Great Britain and Ireland to collect data on outcomes in adult cardiac, congenital cardiac and thoracic surgery.

CCAD: *Central Cardiac Audit Database.*

HES: *Hospital Episodes Statistics.* Data collection by hospitals, finally collated by the Department of Health.

STS: *Society of Thoracic Surgeons.* Based in US but with a large international membership, it is one of the two largest cardiothoracic societies.

EACTS: *European Society for Cardiothoracic Surgery.* This is the leading European professional body of cardiothoracic surgeons.

AEPC: *Association of European Paediatric Cardiologists.*

11. REFERENCES

1. Stark JF and Stark J, 2000, Databases for Congenital Heart Disease: A Beginner's Guide, in *Seminars in Thoracic and Cardiovascular Surgery: Pediatric Cardiac Surgery Annual 2000*, ed. Williams WG, Vol. 3, pp 110-116, W.B. Saunders, Philadelphia, PA, USA.
2. Bristol Royal Infirmary Inquiry, 2000, summary report of public seminar on *Systems: Safety and Risk Management, Quality and Information*, 16 March 2000.
3. The European Paediatric Cardiac Code, 1999, *Cardiology in the Young*, **10**, Suppl. 1, 1-146
4. Stark J, Gallivan S, Lovegrove J, Hamilton JR, Monro JL, Pollock JC, Watterson KG, 2000, Assessing Mortality Rates after Operations for Congenital Heart Defects and Surgeons' Performance From Five UK Centres, *Lancet*, **355**, 1004-1007
5. Mavroudis C, Jacob JP, 2000, guest editors: The Society of Thoracic Surgeons Congenital Heart Surgery Nomenclature and Database Project, *Ann. Thorac. Surg.*, **69**, S1 - S372.
6. Williams WW, 1999, *Database: How I Went About It*, presented at International Symposium, 10.9.1999, Institute of Child Health, UCL, London, UK.
7. Sairanen H, 1999, Helsinki Children's Hospital database, personal communication.
8. Stark J, 1998, *SimpleCARE (Simple CARDiac REgistry)*, presented at the 1st International Conference on Nomenclature of Congenital Heart Defects/Operations, 19.9.1998, Chicago, USA.
9. Murray GD, Lawrence AE, Pollock J, 1999, *A Report on the UK Cardiac Surgical Register and the South Wales Congenital Heart Register for the Bristol Royal Infirmary Inquiry. A Statistical Analysis and Review of the Key Data Sources Relevant to the Inquiry's Remit*, report to the Bristol Royal Infirmary Inquiry.
10. Stark J, 2000, *Comments on Statistical Analysis and Review of Outcomes of Paediatric Cardiac Surgical Services at Bristol and Other Specialist Centres*, submitted to the Bristol Royal Infirmary Inquiry.
11. Spiegelhalter DJ, 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*, report to the Bristol Royal Infirmary Inquiry.
12. Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S, 1997, Monitoring the Results of Cardiac Surgery by Variable Life-Adjusted Display, *Lancet*, **350**, 1128-30.
13. Parsonnet V, Dean D, Bernstein AD, 1989, A Method of Uniform Stratification of Risk for Evaluating the Results of Surgery in Acquired Adult Heart Disease, *Circulation*, **779**, Suppl. 1, 1-12.
14. Spiegelhalter DJ, 1992, Risk Stratification for Open Heart Surgery, *BMJ*, **305**, 1500.
15. Moller J, 1998, *Paediatric Cardiac Care Consortium 1984 - 1995*, Futura Publishing Company Inc., Armonk, NY, USA.

16. Hannah EL, Racz M, Kavey RE, Quaegebeur JM, Williams R, 1998, Paediatric Cardiac Surgery: The Effect of Hospital and Surgeon Volume on In-Hospital Mortality, *Pediatrics* **101**, 963-969.
17. Rothwell PM, Warlow CP, 1999, Interpretation of Operative Risks of Individual Surgeons. *Lancet*, **353**, 1325.
18. de Leval MR, Carthey J, Wright DJ, Farevell VT, Reason JT, 2000, Human Factors and Cardiac Surgery: A Multi-center Study, to appear in *J. Thorac. Cardiovasc. Surg.*
19. Vardulaki KA, Bennet-Lloyd BD, O'Riordan PA, Reeves B C, Tsang VTC, Black NA, 2000, *A Systematic Review of the Outcomes of Open-Heart Surgery*, submitted to the Bristol Royal Infirmary Inquiry in April 2000.