An attempt at data verification in the EACTS Congenital Database

Bohdan Maruszewski, Francois Lacour-Gayet, James L. Monro, Bruce E. Keogh, Zdzislaw Tobota, Andrzej Kansy

The Children’s Memorial Health Institute, Warsaw, Poland
Denver Children’s Hospital, University of Colorado, Denver, CO, USA
Southampton General Hospital, Southampton, UK
The Heart Hospital, Westmoreland Street, London, UK

Women 24 November 2004; received in revised form 15 February 2005; accepted 30 March 2005

Abstract

Objective: The multi-national and multi-institutional collection of data on outcomes in Congenital Heart Surgery (CHS) provides the possibility of analysis of results of treatment and may result in further improvement of the quality of care. The verification of data as far as the completeness and accuracy is necessary to give confidence to all sides—the patients, centers and regulatory authorities. The source data verification (SDV) although difficult, appears possible even in such a large-scale database. Methods: Out of 5,274 patients and 5,612 procedures data of 2003, collected in the database, 1,703 (32.3%) and 1,895 (33.8%), respectively, have been verified at five sites on following fields: IPPV time, date of birth, date of admission, date of surgery, date of discharge/mortality, body weight, case category, CPB time, AoX time, Circulatory arrest time. SDV was performed at five sites by two database officers using the sources of information different to the local copy of the database (patients’ files, operation notes, perfusion charts, OR Books). Verification was performed between June 1st and July 31st 2004. Statistical analysis was performed using R-project software, ver. 2.0.0. and Welch’s t-test for comparison of continuous variables. P-value > 0.05 was used as statistically significant difference between groups. Results: Pre- and post-verification mortalities in all groups showed no significant differences although seven deaths out of 68 (10.27%) were missed. None of the other verified fields showed significant differences after verification. Conclusions: Source Data Verification showed no statistically significant differences between verified and non-verified data on 30 days mortality, LOS, age, body weight, CPB time, AoX and Circulatory arrest time. IPPV time was not available in 58.6% procedures.

Keywords: Congenital heart surgery; Quality of care; Outcome analysis; Database; Data verification; Data validation

1. Introduction

The continuous evaluation of the quality of care became one of the highest priorities in modern surgical practice [1]. Especially, in the field of Congenital Heart Surgery (CHS), where adverse outcomes can be frequent due to the wide spectrum of congenital cardiac conditions and severity of pathologies, as well as high proportion of palliations, it is important to know the exact information on mortalities and morbidities [2,3]. Even more important is the possibility of the comparison of outcomes between surgeons and institutions on large, international, multi-institutional scale [4-8].

In this spirit in early 1990s, a group of congenital; heart surgeons of the European Congenital Heart Surgeons...
Until September 2004, the data on 23,735 CHS procedures have been collected, analyzed and reported online at the EACTS Congenital Database (www.eactscongenitaldb.org) (Table 1).

The reliability of any data collection remains the key issue when the information is used for further outcome analysis and comparison. That is why data completeness and accuracy is essential, although it remains one of the major challenges for database management [4,12,14,15].

The meaning of ‘data validation’ should be understood as testing internal data consistency, e.g. the date of operation should not be earlier than the date of birth.

Data validation is possible at data entry (surgeon, software) and in the database before data analysis. The EACTS Congenital Database developers have provided the internal database validation mechanism that automatically excludes non-accurate data according to agreed criteria (www.eactscongenitaldb.org).

The next key step towards improvement of the data quality is the efficient data verification [17]. According to the Good Clinical Practice guidelines defined and recommended by World Health Organization (WHO), all database contents should be verified by comparison with the original records at the site of their origin—Source Data Verification (SDV) [17,18]. SDV is an evaluation of the conformity of database contents with the source data.

SDV, although very difficult on international, multi-institutional scale due to its costs, time consuming and lingual problems, appears to be possible and remains essential. For all reasons described above the EACTS Congenital Database management decided to make an attempt at data verification and check on the potential discrepancies between verified and non-verified data.

2. Material

Five European CHS Centers volunteered to participate in the project and undergo 2003 data verification at their sites. The verification was conducted between 1st of June and 31st of July 2004 and concerned the following fields:

- 30 days mortality
- LOS
- IPPV

and additionally:

- date of birth
- date of admission

• date of surgery
• date of discharge/mortality
• body weight
• Case category
• CPB time
• AoX time
• Circulatory arrest time

Out of 5,274 patients and 5,612 procedures data of the year 2003 collected in the database 1,703 (32.3%) and 1,895 (33.8%) have been verified, respectively (Table 2).

3. Methodology

SDV has been conducted by EACTS Congenital Database staff (two officers).

The data verification process was organized in the following steps:

- 100% of patients data of 2003 have been supplied to the Database before site visit.
- The data underwent standard internal validation procedure.
- Person locally responsible for the verification indicated the sources of information at the site, different to the local copy of the database or related to the database which were as follows:
  - patients’ files
  - hospital computerized system
  - operation notes
  - perfusion charts
  - OR books

- The database staff visited the site and checked on the accuracy of 100% of 2003 records for the fields undergoing verification. The central database printouts have been compared with the local sources.

### Table 1

| Standard EACTS Congenital Database online report of 18th of September 2004 | | | | | |
|---|---|---|---|---|
| No. of cases | % of all | Min | Mean | SD | Max |
| IPPV (h) | 13.146 | 55.39 | 1 | 55.21 | 165.63 | 300 days |
| Total CPB time (min) | 16.725 | 100.00 | 2.00 | 103.89 | 71.61 | 999.00 |
| Total AorticX time (min) | 15.121 | 90.41 | 1.00 | 56.55 | 40.43 | 998.00 |
| Circulatory arrest (min) | 1.844 | 11.03 | 1.00 | 30.61 | 22.75 | 153.00 |
| Weight (kg) | 22.583 | 95.15 | 0.50 | 14.36 | 17.23 | 128 |
| Age at operation (months) | 23.684 | 99.79 | 0.00 | 52.08 | 102.50 | 932.06 |
| LOS (days) | 23.339 | 98.33 | 0.00 | 18.95 | 26.11 | 895.00 |

IPPV, intermittent positive pressure ventilation (measured as intubation time), refers to all procedures; CPB, cardiopulmonary bypass; AorticX time, aortic crossclamp time, refers to CPB procedures; LOS, length of stay, surgery-discharge time; Circulatory arrest refers only to CPB procedures.

### Table 2

| Material | | | |
|---|---|---|
| Center | Patients | Procedures |
| A | 281 | 305 |
| B | 444 | 480 |
| C | 199 | 227 |
| D | 362 | 385 |
| E | 417 | 498 |
| All | 1703 | 1895 |

No. of patients and procedures by centers.
The statistical analysis was performed on both sets of data (verified, non-verified) for comparison (Table 3).

4. Statistical methodology

Results were expressed as mean values. All analyses were performed with appropriate software (R-project, ver. 2.0.0). Comparisons of continuous variables were performed by Welch’s $t$-test. $P$-value greater than 0.05 means that difference between groups is not statistically significant.

5. Results

Two patients, out of 1.705 after verification, have been missed.

There have been no statistically significant differences between pre- and post-verification mortalities in all verified groups of patients, nevertheless seven deaths were missed out of 68 deaths found after verification (10.27%) (Tables 4-7).

There have been no statistically significant differences found between verified and non-verified data in three groups of patients: all, infants and neonates, as far as age, AoX, Circulatory arrest, CPB time, LOS, body weight.

Though only two patients were not reported out of 1.705 (after verification), there have been bidirectional shifts between age groups due to the correction in dates of birth and/or of operation (Table 8).

IPPV time was not available in three Centers.

6. Conclusions

Source Data Verification showed no statistically significant differences between verified and non-verified data on 30 days mortality, LOS, age, body weight, CPB time, AoX and Circulatory arrest time IPPV time was not available in 58.6% procedures.
7. Discussion

We believe that this study is the first one to present the verification process on the big, multi-institutional and international data registry of the outcomes of the CHS.

It is obvious that 30 days mortality measure carries important limitation as far as the assessment of even early survival [8]. Undoubtedly, modern intensive care technology extends critical patients survival time what does not necessarily mean increase in long-term survival and improvement of the quality of life. Nevertheless this has been the only objective measure available 6 months after study time frames (2003). In future works, 30 days mortality measure should be avoided and probably survival analysis should concentrate more on mid-term outcome (6 months or 1 year). This is hard to believe that such a data can be obtained completely and accurately, especially in international big scale registry.

Another limitation of the study is due to arbitrary selection of verified Centers. Relatively optimistic results of this first attempt at data verification, performed on the Centers being ‘good data submitters’, may not reflect the overall quality of data in the whole data collection. The only solution to decrease the negative impact of arbitrary choice of Centers will be using ‘by chance’ mechanism for the next 2005 attempt or include the Centers with relatively short experience in database works.

The methodology employed in this project was very demanding, since 100% of 2003 records for nine fields have been checked, which finally verified seven investigated parameters.

The verification process according Good Clinical Practice for the clinical trials, especially for the pharmaceutical products should cover 100% of data. This rule is very rarely applied to any verification on the big data collection, although some institutions have used it to verify their clinical data (e.g. CHOP, Pennsylvania). There are important differences between various verification protocols used for medical data sets [15]. The verified samples vary from as small as 10 up to 100%. It shows that an International Committee of experts is needed to define common Data Verification methodology and to apply it in future works on outcome analysis in CHS. Such a group has been recently established between the EACTS and STS. This is the work in progress and some feedback from this Committee is expected before the end of 2005.

Although this study is only an attempt at data verification within big, international, multi-centric database, it has proven the scientific value of such a data collection. The verification process will be continued and further works will focus on improvement of the quality of data. This should result in better understanding of the specific risk factors and areas of weakness in the quality of care in CHS.

Acknowledgements

We would like to acknowledge doctors Duccio Di Carlo, Mark Hazekamp, Victor Hraska, Giovanni Stellin, and all staff members of five verified hospitals for the support and understanding during demanding process of data verification.

References


Appendix A. Conference discussion

Dr J. Monro (Southampton, United Kingdom): I do congratulate you on all this work and the remarkable database that you’ve collected for all these years. The results of the verification in these five units are amazing. With the British experience in mind, it is incredible that they only missed two patients and those seven deaths. I wonder if this is because you chose the five best, most organized centers, who therefore did have these very accurate data. When you’ve looked into the others, I think you may find that it’s not quite so good. But as you’ve said before, everything has got to be 100%, both centers and the data they’re collecting.

Dr Maruszewski: And of course, you are right. We have taken the centers that offered us cooperation. And I have to use the opportunity to thank my colleagues from the five centers for their very hard work, because this is a lot of very hard work. In the future, we want to make the decision about the centers that we choose by chance, so machine will choose the centers and then we will see.

Mr V. Tsang (London, United Kingdom): The work you have done is very admirable, but can I talk about how to tackle a big problem in general terms? For example, your database is huge. And in psychological terms, when you have a huge problem people feel overwhelmed as to what to do next. I think that the key to solve that sort of problem is to break the big problem into smaller blocks. The progress made in each of the smaller blocks, when added together, would become major progress.

Dr Maruszewski: Those of you who saw the outcome analysis that we were showing on various occasions, including yesterday at the business meeting, probably understand that is exactly what we are trying, we are attempting to do. So take part of the data and clear them and make sure they are very fine, and then to conclude on one question that you want to answer. That’s what we are doing with neonates, Norwoods, influencing the overall outcomes.

Dr B. Williams (Toronto, Ontario, Canada): I have tremendous respect for what you’ve done with this database and the enormous amount of work that you put into it.

I’d like to comment on your findings that the time to extubation was much more difficult to find. That doesn’t surprise me because it’s data that’s out of our domain. It’s kept by other people in other areas. And my bias would be that we should not try to capture that information in a surgical database. Surely the intensivists have their own database and would it not be far more logical to create a link between your surgical registry and the database in the other areas of the hospital, specifically the intensive care unit.

If you carry this analogy one step further and look at the genetics, I think our data about genetics is going to be extremely poor and we will need to link to the genetics department to find out which patients have genetic disorders.

Dr Maruszewski: As you know, we are meeting beginning of October, CHSS and our association, in Montreal. One of the tasks is to make sure that both associations realize that IPPV time is a weak measure. And I think we should not carry on promoting this as the quality of care measure, because, first of all, we don’t have this measure, we don’t have the data, I agree.

Dr G. Ziemer (Tuebingen, Germany): Regarding ICU-data we have to keep in mind that ICU care is organized so differently across Europe and even within certain countries, that uniform data collection is not practical at the moment, although the idea I think is a great one.

Dr W. Mrowczynski (Poznan, Poland): Although the database is voluntary, the data verification may require agreement of the authorities governing, for example, cardiac surgery department. Don’t you think it can be an obstacle in data verification?

Dr Maruszewski: It hasn’t been for those who keep sending data for years; and this problem has been sorted out at the European union level and the same in States. The parents of the patient who is admitted to the hospital, on the admission accept that the data are used for the medical record and for the medical database. It may be different in various states and, for example, I’m not aware of that in details, but so far we are able, in the legal manner, to collect this data.

Dr F. Lacour-Gayet (Denver, Colorado, USA): I would like to congratulate you and your group in Warsaw for the incredible amount of work that produced for the last 10 years in finally building what is probably the best congenital heart surgery database in the world.

Let me give you some information from the STS database. We have also decided to verify data at the STS. It’s a bit more complicated and there are some authorizations that are needed, but we are proceeding. We have decided on a protocol that is based essentially on comparing the data of the hospital registry to those of the department and check that there is no death missing. Ultimately, there is a need for a common protocol to verify data.

Dr Maruszewski: Before I finish, I would like to say just a few words. I would like to give my best thanks and a lot of respect to our Association for the support that has always been given to the database even in a very premature state when the outcome was not sure.

Then I would like to thank Bill Williams for his stimulation and motivation and criticism that we received some time ago and that motivated us to work.

I would like to thank all the colleagues in the European countries for putting the data in and the effort made on verification.

At the end, I would like to thank Dr Tobota from Warsaw who is running, in fact, this database. He is doing 90% of the work and I am just presenting. So thank you, Dr Tobota, very much.