European Journal of Cardio-thoracic Surgery 40 (2011) 91—98

EUROPEAN JOURNAL OF CARDIO-THORACIC SURGERY

www.elsevier.com/locate/ejcts

Task-independent metrics to assess the data quality of medical registries using the European Society of Thoracic Surgeons (ESTS) Database ☆

Michele Salati [a,*], Alessandro Brunelli [a], Marcel Dahan [b], Gaetano Rocco [c], Dirk E.M. Van Raemdonck [d], Gonzalo Varela [e], on behalf of the European Society of Thoracic Surgeons Database Committee

[a] *Division of Thoracic Surgery, Ospedali Riuniti Ancona, Ancona, Italy*
[b] *Thoracic Surgery — CHU Larrey, Toulouse, France*
[c] *National Cancer Institute, Pascale Foundation Naples, Naples, Italy*
[d] *Department of Thoracic Surgery, University Hospitals Leuven, Leuven, Belgium*
[e] *Cirugia Toracica, Hospital Universitario de Salamanca, Salamanca, Spain*

Received 28 May 2010; received in revised form 18 October 2010; accepted 5 November 2010; Available online 17 December 2010

Abstract

Objective: The quality of data collected into a database is of paramount importance in every analysis. No standardized methods are available to quantify the quality of data in medical registries. Expanding the work done in other fields, we aimed at developing a methodological approach to measure the quality of a thoracic surgical database, by using the European Society of Thoracic Surgeons (ESTS) Database. **Methods:** A selection of anonymized data collected in the ESTS Database from 2007 to 2009 was tested using appropriate data quality metrics: completeness, correctness, consistency and believability. Particularly, the believability value is obtained as a result of a min—max operation based on the evaluation of completeness, correctness and consistency. Completeness measures the number of missing values in each checked column of the database, and it is calculated as number of variables registered/number of variables expected. Correctness reflects the number of data units in error referring to a set of clearly defined criteria (number of correct data/number of all data counted) and consistency is calculated by verifying the number of data in conflict in the same recorded patient (number of consistent checks/total number of checks). The threshold selected to indicate good quality was 0.8. **Results:** A total of 49 363 values were reviewed to obtain the quality indicators. The results of the data quality assessment for the analyzed section of the ESTS Database are all above the predefined threshold: completeness is 0.85, correctness 0.99 and consistency 0.98. The believability score of data in the database is 0.85. **Conclusions:** We were able to apply task-independent metrics to measure the quality within the ESTS Database. This study may represent a template to be applied in the medical/surgical field to test the quality of data in clinical registries. Only registries with a high quality of data can be reliably used for scientific, managerial and credentialing purposes.

Keywords: Data; Quality; Registry

1. Introduction

Data quality has gained a growing interest over the last decades. This can be testified by a simple web search that allows opening thousands of Web pages, articles, and books about this topic.

The importance of data quality is well recognized in public and private institutions and organizations. This is particularly true in specific areas (government, business, and companies) where research in data quality definition, measurement and analysis has already progressed to practical tools, methods, and processes to solve critical business problems [1—3].

In this context, the development of management strategies to control and improve data quality has substantial social and economic implications, as confirmed by high-profile public initiatives that underline the relevance of this topic (Office of Management and Budget. Information Quality Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Agencies. http://www.whitehouse.gov/omb/fedreg/reproducible.html).

1.1. What about medical sciences?

The use of medical registries has become mandatory in modern health-care systems for managerial and clinical

☆ Presented at the 18th European Conference on General Thoracic Surgery, Valladolid, Spain, 30 May—2nd June, 2010.

* Corresponding author. Address: Division of Thoracic Surgery, Ospedali Riuniti Ancona, Via Conca 71, Ancona 60020, Italy.
Tel.: +39 0715964439; fax: +39 0715964481.
E-mail address: michelesalati@hotmail.com (M. Salati).

1010-7940/$ — see front matter
doi:10.1016/j.ejcts.2010.11.004

purposes. Looking, for instance, at our specialty, in 2001, the European Association of Cardio-Thoracic Surgery/European Society for Thoracic Surgeons (EACTS/ESTS) Working Group on Structures in Thoracic Surgery recommended the use of a clinical database to allow for quality surveillance [4]. These recommendations were recently endorsed by the ESTS Database Committee, which proposed a quality certification program based on the evaluation of outcomes and processes of care using measures collected in the ESTS Database [5].

Data collection is certainly the most important endeavor of any outcome analysis [6]. In health-care systems, data collection is instrumental to (1) develop outcome analysis, (2) help in clinical decision making, (3) build and optimize pathways of care, (4) define quality of care, (5) quantify costs, and (6) compare results. The information derived from such processing activities is heavily conditioned by the quality of the original data.

However, in the medical field, the development of strategies apt to define, improve and assure the quality of data has not been sufficiently investigated.

Parameters measuring different aspects of data (hereafter 'quality metrics') to quantify the quality of the entire database need to be identified. Borrowing the experience gained mainly in the business and governmental fields, we selected a group of four quality metrics with the aim to test the quality of data of a medical registry.

Data quality definition and measurement is only the first step within the wide process of data quality management, as it is represented in Fig. 1. The intent of this study is to develop a methodological approach to assess data quality in our specialty, providing a template for future analysis and setting the stage for planning future improvement activities.

To this purpose, on behalf of the ESTS Database Committee, we chose to apply this methodology to the ESTS Database, representing an electronic information system ideal for performing such an analysis.

2. Material and methods

After a formal request, the ESTS Database Committee authorized the use of the ESTS Database for the present analysis after a complete anonymization of sensitive information regarding patients, units, surgeons and countries. Only data submitted by units contributing more than 100 pulmonary resections (which constitute more than 90% of the total cases present in the database) during the period July 2007 through October 2009 were used for the present analysis.

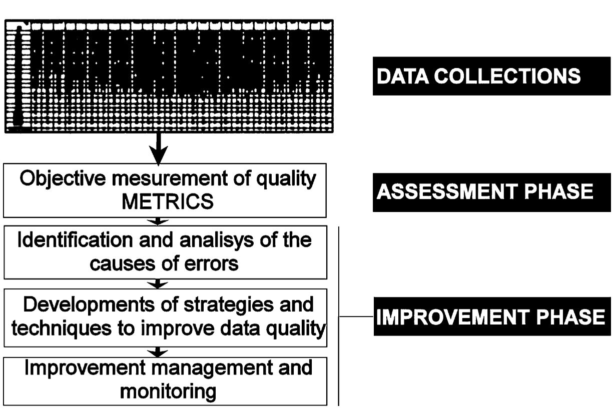

Fig. 1. Phases of the data quality management process.

2.1. Database and analyzed data characteristics

The metrics used in the assessment phase can vary in relation to the type of the analyzed data system. For the present analysis, we chose to use the ESTS Database for the following reasons:

- The ESTS Database is an electronic online registry that presents characteristics typical of both monolithic information systems and Web Information System (WIS) [7]. In fact, this is a single database, where information flows are simple and repetitive as usually occurs in monolithic systems. Considering the low degree of complexity of the data collection and management process for these types of information systems (monolithic), they appear highly suitable for the attempt of developing a data assessment phase model.

  At the same time, the ESTS Database adopts Web technologies to collect and store data. In fact, the data can be submitted by single contributors (clients) to the ESTS Database using a guided Web procedure after a login through the ESTS Web page (https://www.thoracicdata.org/content/index.php). The analysis of the strategies applied for the data quality assessment of WIS involves a different and more complex methodology, which is out of the scope of the present work.
- Fig. 2 describes the types of data usually encountered in a medical registry. The ESTS Database manages structured data, that are 'an aggregation of items (variables) described by elementary attributes defined within a domain. Domain represent the range of values that can be assigned to attributes and usually correspond to elementary data such as numeric values or text string [8].' In most of the cases, medical registries are built as relational tables gathering structured data, which are highly informative and easily usable for statistical purposes.
- The two characteristics described above make the ESTS Database representative of most of the commonly used medical databases.

The items composing the ESTS Database are about 150, describing multiple aspects of different surgical procedures. We filtered the entire data set for specific attributes of specific items (such as <DIAGNOSIS>, <MORPHOLOGY>, <GROUP>, and <SUBGROUP>), to obtain variables and correspondent data only from patients submitted to lung resection for primary lung cancer. Among these, we selected a sample of variables considered critical to allow the analysis of data for future quality initiatives of the Society. Through this process of selection, we were able to obtain a solid, clean, and mono-procedural data set for developing and testing a data quality assessment phase methodology (Fig. 3).

As a consequence of this selection process, the level of quality estimated by our analysis is limited to the pool of selected variables and is not generalizable to the entire database.

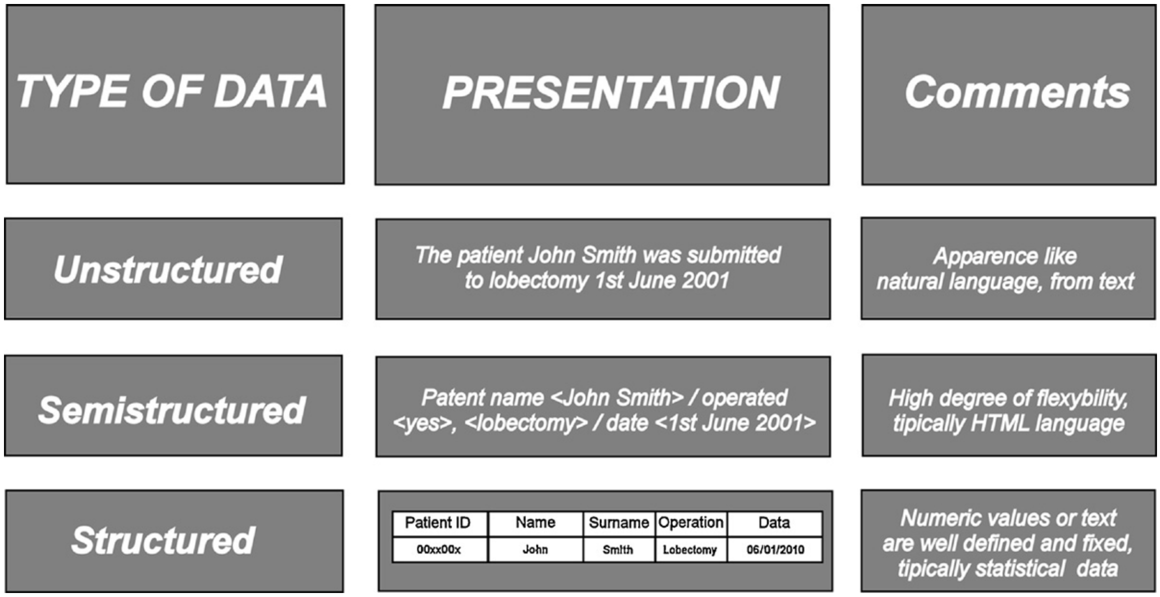

Fig. 2. Classification of data.

2.2. Quality metrics definition

To assess the quality of a database, it is necessary to measure its by using appropriate dimensions and the corresponding metrics [9,10]. Sometimes, as in our case, the metric is unique and the theoretical definition of a dimension coincides with the operational definition of the corresponding quality metrics.

The quality metrics that we used to assess the ESTS Database are among the most important described in the literature [8]. These are defined as follows:

- completeness: The extent to which data is not missing and is of sufficient breadth and depth to describe the corresponding set of real-world objects;
- correctness/accuracy: The extent to which data is correct and reliable; it must be specified that we examined only the so-called 'syntactic accuracy' (closeness of a value to the elements of the corresponding definition domain) rather than the 'semantic accuracy' (closeness of a value to its real-world value);
- consistency: The extent to which data are correspondent and coherent each other (cross-record consistency); and
- believability: The extent to which data is regarded as true and credible.

All these quality metrics could be defined as 'task-independent' insofar as they 'reflects states of the data without the contextual knowledge of the application and can be applied to any dataset, regardless of the tasks at hand [10].'

2.3. Quality metrics calculation

Each of the above-mentioned metrics has been calculated on the basis of a specific formula applied to a group of core variables. The mathematic formulae are reported below as well as the correspondent variables used for the test [10—12].

- completeness: Number of data registered (number of all data expected — number missing data)/number of all data expected;the variables tested were: <DATE OF BIRTH>, <DATE OF OPERATION>, <MORPHOLOGY>, <FEV1>, <DLCO>, <ppoFEV1>, <ppoDLCO>, <COMPLICATION 1>, <OUTCOME AT DISCHARGE>, <OUTCOME AT 30 DAYS>, <Tstage>, <Nstage>, <Mstage>, <NODES>, and <DATE OF DISCHARGE>;
- correctness/accuracy: Number of correct data (number of all data counted — number of wrong data)/number of all data counted;the variables tested were: <FEV1%>, <ppoFEV1%>, <DLCO%>, and <ppoDLCO%>.

Criteria for the definition of 'wrong data': count 'wrong data' if FEV1 or DLCO value is <25 or >150; count 'wrong data' if ppoFEV1 or ppoDLCO value is <15 or >150;

- Consistency: Number of consistent checks (total number of checks — number of not consistent checks)/total number of checks; the variables tested were: <DATE OF BIRTH>, <DATE OF OPERATION>, <DATE OF DISCHARGE>, <FEV1>, <DLCO>, <ppoFEV1>, <ppoDLCO>, <COMPLICATION 1>, and <OUTCOME AT DISCHARGE>;criteria for the definition of 'not consistent checks': count 'not consistent checks' if DATE OF BIRTH value $>$ DATE OF OPERATION value; count 'not consistent checks' if DATE OF OPERATION value $>$ DATE OF DISCHARGE value; count 'not consistent checks' if FEV1 value $<$ ppo-ppoFEV1 value; count 'not consistent checks' if DLCO value $<$ ppoDLCO value; count 'not consistent checks' if OUTCOME AT DISCHARGE value is 'Died in Hospital' and COMPLICATION 1 value is 'None'.
- believability: The value of this metric is calculated using the most conservative of all possible methods used for this computation, which is the minimum operation based on the evaluation of completeness, correctness and consistency; the minimum operation assigns to the dimension the lowest value of the other metrics taken into consideration (completeness, accuracy and consistency).

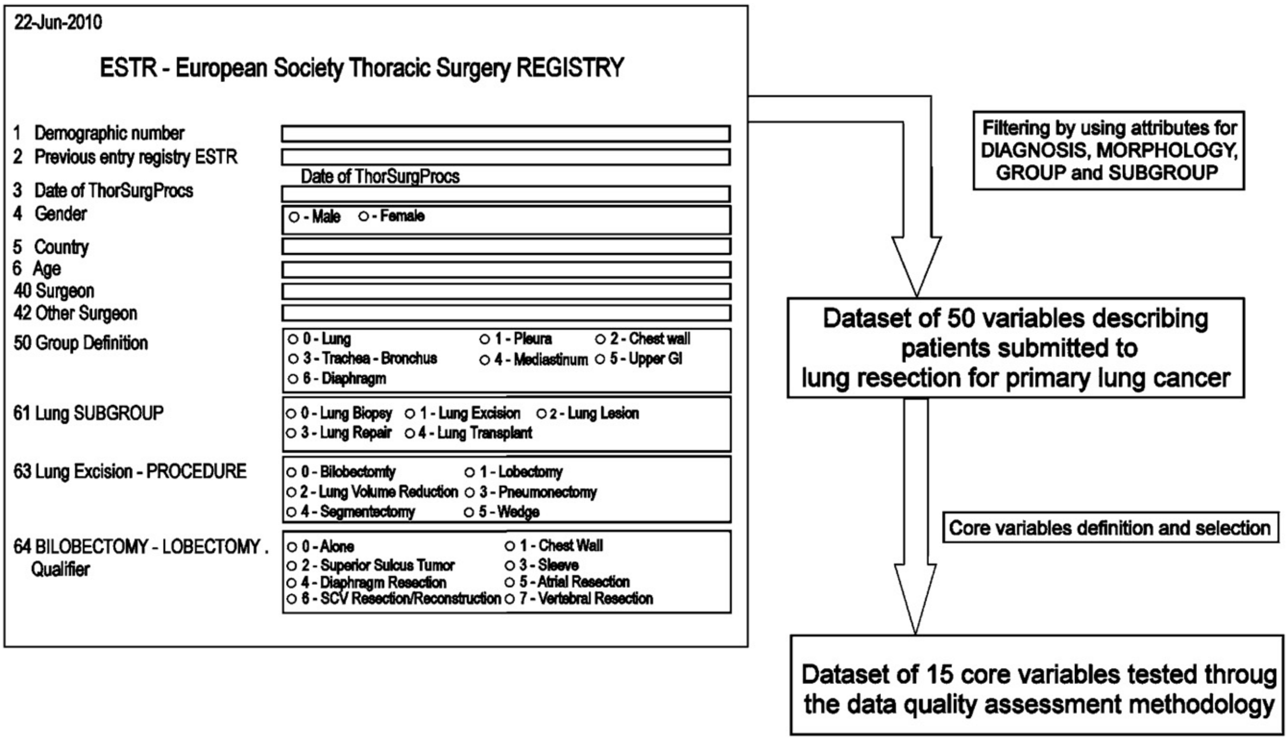

Fig. 3. Process of variable selection to obtain the definitive dataset to analyze.

On the basis of the data quality literature, we assumed that the threshold selected to indicate good quality was 0.8 [10,13,14].

3. Results

A total of 51 variables were available for the analyses after filtering the entire data set for the items <DIAGNOSIS> ('Lung Cancer'), <MORPHOLOGY> ('Primary Neoplastic Malignant'), <GROUP> ('Lung'), and <SUBGROUP> ('Lung Excision').

Table 1 lists the variables tested in the analysis (core variables selected) and their relation to the metrics.

Table 2 illustrates the calculation and the cumulative measurement for the metric completeness. The assessed quality measure for the completeness of the database is 0.85. It should be underlined that <DLCO>, <ppoDLCO>, <OUTCOME AT 30 DAYS>, and <DATE OF DISCHARGE> items exhibit a completeness value below the commonly accepted good quality standards. Nevertheless, given the compensating results of the other items, the overall completeness of the database appears to be acceptable.

Table 1. Core variables selected for the calculation of the correspondent metrics.

Selected variable [a]	Related metric
Date of birth	Completeness, consistency
Date of operation	Completeness, consistency
Morphology	Completeness
FEV1	Completeness, correctness/accuracy, consistency
DLCO	Completeness, correctness/accuracy, consistency
ppoFEV1 [used for the analysis only in patients submitted to lobectomy, pneumonectomy, sleeve resections]	Completeness, correctness/accuracy, consistency
ppoDLCO [used for the analysis only in patients submitted to lobectomy, pneumonectomy, sleeve resections]	Completeness, correctness/accuracy, consistency
Complication 1	Completeness, consistency
Outcome at discharge	Completeness, consistency
Outcome at 30-days	Completeness
T [used for the analysis only in patients submitted to lung excision for lung cancer]	Completeness
N [used for the analysis only in patients submitted to lung excision for lung cancer]	Completeness
M [used for the analysis only in patients submitted to lung excision for lung cancer]	Completeness
Nodes [used for the analysis only in patients submitted to lung excision for lung cancer]	Completeness
Date of discharge	Completeness, consistency

[a] The reported variables are named as well as within the ESTS Database.

Table 2. Completeness measurement.

Selected variable[a]	Formula: number of data registered (number of all data expected – number missing data)/number of all data expected	Correctness value
Date of birth	(2265 – 9)/2265	0.99
Date of operation	(2265 – 0)/2265	1
Morphology	(3432 – 78)/3432	0.98
FEV1	(2265 – 162)/2265	0.93
DLCO	(2265 – 1430)/2265	0.37
ppoFEV1 [used for the analysis only in patients submitted to lobectomy, pneumonectomy, sleeve resections]	(2113 – 318)/2113	0.85
ppoDLCO [used for the analysis only in patients submitted to lobectomy, pneumonectomy, sleeve resections]	(2113 – 1164)/2113	0.45
Complication 1	(2265 – 106)/2265	0.95
Outcome at discharge	(2265 – 114)/2265	0.95
Outcome at 30 days	(2265 – 597)/2265	0.73
T [used for the analysis only in patients submitted to lung excision for lung cancer]	(2109 – 143)/2109	0.93
N [used for the analysis only in patients submitted to lung excision for lung cancer]	(2109 – 153)/2109	0.93
M [used for the analysis only in patients submitted to lung excision for lung cancer]	(2109 – 152)/2109	0.93
Nodes [used for the analysis only in patients submitted to lung excision for lung cancer]	(2109 – 52)/2109	0.97
Date of discharge	(2265 – 645)/2265	0.71
Total	(34 214 – 5123)/34 214	0.85 (95%CI 0.004)

[a] The reported variables are named as well as within the ESTS Database.

Table 3 illustrates the calculation and the cumulative measurement for the metric correctness/accuracy. All variables tested had a value of 0.99, reflecting an optimal correctness. As a consequence, the assessed quality measure for the correctness of the database is also 0.99.

Table 4 illustrates the calculation and the cumulative measurement for the metric consistency. Of particular importance is the very high consistency level of the two items describing the clinical outcomes (<COMPLICATION 1> and <OUTCOME AT DISCHARGE>). The assessed quality measure for the consistency of the database is 0.98.

After performing a conservative calculation based on the previous metrics results, the assessed quality measure for the believability of the database is 0.85.

All the cumulative metrics assessed using data derived from the ESTS Database were above the predefined good quality indicator value 0.8.

4. Discussion

Managed health-care systems demand systematic data collection to plan clinical and management activities [4,6]. Collected data are critical information that can be used to evaluate the performance of surgeons, allocate institutional resources, inform third parties about the hospital or institutional clinical activity, implement a pay-for-performance policy, and support political and medico-legal actions, among the others [15]. Given their pivotal role, the quality of medical registries needs to be carefully scrutinized by reliable instruments.

The definition of data quality reported by International Standard Organization is based on 'the features and characteristics of a data set that bears on its ability to satisfy the needs that result from the intended use of data [11].'

Table 3. Correctness/accuracy measurement.

Selected variable[a]	Formula: number of correct data (number of all data counted – number of wrong data[b])/number of all data counted	Correctness value
FEV1	(2103 – 1)/2103	0.99
DLCO	(835 – 7)/835	0.99
ppoFEV1 [used for the analysis only in patients submitted to lobectomy, pneumonectomy, sleeve resections]	(1795 – 1)/1795	0.99
ppoDLCO [used for the analysis only in patients submitted to lobectomy, pneumonectomy, sleeve resections]	(949 – 7)/949	0.99
Total	(5682 – 16)/5682	0.99 (95%CI 0.003)

[a] The reported variables are named as well as within the ESTS Database.

[b] Definition of 'wrong data': count 'wrong data' if FEV1 or DLCO value is <25 or >150; count 'wrong data' if ppoFEV1 or ppoDLCO value is <15 or >150.

Table 4. Consistency measurement.

Selected variable[a]	Formula: number of consistent checks (total number of checks − number of not consistent checks[b])/ total number of checks	Correctness value
Date of birth versus date of operation	(2265 − 22)/2265	0.99
Date of operation versus date of discharge [used for the analysis only in patients with date of discharge available]	(1620 − 50)/1620	0.97
FEV1 versus ppoFEV1 [used for the analysis only in patients with FEV1 and ppoFEV1 > 0]	(1897 − 58)/1897	0.97
DLCO versus ppoDLCO [used for the analysis only in patients with DLCO and ppoDLCO > 0]	(781 − 23)/781	0.97
Outcome at discharge versus complication 1	(2265 − 0)/2265	1
Total	(9473 − 798)/9473	0.98 (95%CI 0.003)

[a] The reported variables are named as well as within the ESTS Database.

[b] Definition of 'not consistent check': count 'not consistent checks' if DATE OF BIRTH value > DATE OF OPERATION value; count 'not consistent checks' if DATE OF OPERATION value > DATE OF DISCHARGE value; count 'not consistent checks' if FEV1 value < ppoFEV1 value; count 'not consistent checks' if DLCO value < ppoDLCO value; count 'not consistent checks' if OUTCOME AT DISCHARGE value is 'Died in Hospital' and COMPLICATION 1 value is 'None'.

To quantify these characteristics, it is necessary to define, implement and apply a specific data quality methodology. The main processes making up a data quality methodology are represented by: (1) state reconstruction, which collects information about organizational processes, data collection and management procedures, quality issues, and cost of a given database; (2) assessment/measurement, which measures quality of data collections through the development of relevant quality dimensions; and (3) improvement, which defines steps, strategies, and techniques for reaching new data quality targets [8,16].

The present study represents a methodological approach for performing the assessment/measurement phase of data quality management in our specialty.

To this purpose, we developed and applied a data quality analysis to the ESTS Database. The use of the ESTS Database was justified for two main reasons: first, this is a database characterized by the collection of structural data. These are the most usual type of data encountered in medical registries and, at the same time, highly suitable for statistical analysis. Second, the ESTS Database is a well-known and well-accepted registry model in our specialty, including a minimum set of standardized core variables and end points, hence representing a benchmark for data collection within our field. This offers the advantage of performing an analysis of data quality, considering data previously defined, and skipping the phase of state reconstruction.

Quality metrics were defined and applied to quantify the quality of data in the registry. This methodology has been developed in the business and managerial sectors. We tried to verify its applicability in a medical registry. The main intent of this work was not to judge the quality of the ESTS Database but to test a method that can be generalized to any medical database, through specific and reproducible equations.

By using these metrics and commonly accepted standards, we were able to measure the quality of selected variables in the ESTS Database, showing their overall good level of quality.

Very few articles have been published with the aim to verify data quality in cardiothoracic registries, and none of them was specifically focused on general thoracic surgery [17—20]. Moreover, most of these previous works were focused in defining the so-called semantic accuracy, the process aimed at assessing the correspondence between the values reported within the database and the ones of the real world [8].

At variance with those previous investigations, the present study described a strictly defined procedure to test data quality by the use of metrics assessing multiple dimensions of data in relation to predefined criteria and domains. In this respect, this data quality approach is completely independent from the original real world that data should describe.

This leads to the development of a data quality assessment procedure that is reproducible and usable for further analysis:

- compare data quality between different registries;
- data quality evolution in different periods;
- assess the data quality related to different data collection tools; and
- plan data quality improvement programs.

The study has potential limitations:

- evaluating the structural quality of a database is different from checking the conformity of the collected data with the source data; this latter process should be performed independently by a source data verification (audit) at the site of their origin,[20] a process not yet implemented in the ESTS database;
- certain variables have an intrinsic low value of completeness (i.e., DLCO); this is due to a low compliance in measuring this parameter across multiple European centers; the strict application of a quality metrics may not fully account for local clinical policies restricting the collection of a variable for legitimate reasons; future studies are needed to investigate how to most reliably measure specific variables subject to large clinical variability and their influence on the overall believability score; and
- the method described in this study regards only one phase of the data quality methodology (assessment/measurement); further investigations are needed to evaluate the state reconstruction and improvement phases.

In conclusion, we were able to apply a quality metrics system to verify the quality of data in a sample of variables of the ESTS Database. Our analysis may provide a methodological template that can be reproduced for the same purpose in other clinical registries and ideally to evaluate the data quality of the entire ESTS Database. It would be desirable for clinical registries to incorporate an automated system using these measures to periodically check the accuracy, completeness, correctness, consistency and believability of the most relevant variables.

References

[1] Laudon KC. Data quality and due process in large inter-organizational record systems. Commun ACM 1986;29:4—11.

[2] Strong DM, Lee YW, Wang RY. Data quality in context. Commun ACM 1997;40:103—10.

[3] Wang RY. A product perspective on total data quality management. Commun ACM 1998;41:58—65.

[4] Klepetko W, Aberg THJ, Lerut AEMR, Grodzki T, Velly JF, Walker WS, Ahren C, Arsovski A, Bellenis I, Bequiri S, Berrisford RG, Branscheid D, Bibicic J, Csekeö A, Drescik I, Dussek J, Erzen J, Furmanik F, Godinho MT, Goldstraw P, Gotti G, Halezeroglu S, Hamzik J, Harustiak S, Hasse J, Hartl P, Hostrup P, Horvat T, Ilic N, Jeyasingham K, Kappetein AP, Kecskes L, Laisaar T, Lampl L, Levasseur Ph, Maggi G, Magnanelli G, Massard G, Moghissi K, Molnar T, Noirhomme PhH, Orlowski T, Pafko P, Petricevic A, Pujol JL, van Raemdonck EM, Ramos Seisdedos G, Ris HB, Salo J, van Schil P, Schmid RA, Thorpe A, Toomes H, Varela A, Varela G, Venuta F, sa Vieira VM, Weder W, Wihlm JM, Zannini P, Marta Gabriel M. Structure of general thoracic surgery in Europe: by the EACTS/ESTS Working Group on structures in thoracic surgery. Eur J Cardiothorac Surg 2001;20:663—8.

[5] Brunelli A, Rocco G, Van Raemdonck D, Varela G. Lessons learned from the European Thoracic Surgery Database: the composite performance score. Eur J Surg Oncol 2010;36:S93—9.

[6] Ferraris VA, Ferraris SP. Risk stratification and comorbidity. Card Surg Adult 2003;2:187—224.

[7] Iasakowitz T, Bieber M, Vitali F. Web information systems — introduction. Commun ACM 1998;41:78—80.

[8] Batini C, Cappiello C, Francalanci C, Maurino A. Methodologies for data quality assessment and improvement. ACM Comput Surv 2009;41:1—52.

[9] Wang R, Strong D. Beyond accuracy: what data quality means to data consumers. J Manage Inf Syst 1996;12:4—19.

[10] Pipino LL, Yang WL, Yang R. Data quality assessment. Commun ACM 2002;45:211.

[11] Arts DG, de Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc 2002;9:600—11.

[12] Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc 1997;4:342—55.

[13] Barrie JL, Marsh DR. Quality of data in the Manchester orthopaedic database. BMJ 1992;304:159.

[14] Wagner MM, Hogan WR. The accuracy of medication data in an outpatient electronic medical record. J Am Med Inform Assoc 1996;3:234—44.

[15] Varela G. Pay for performance in thoracic surgery. Thorac Surg Clin 2007;17:431—5.

[16] Cappiello C, Francalanci C, Pernici B. A rule-based methodology to support information quality assessment and improvement. Stud Commun Sci 2004;2:137—54.

[17] Herbert MA, Prince SL, Williams JL, Magee MJ, Mack MJ. Are unaudited records from an outcomes registry database accurate? Ann Thorac Surg 2004;77:1960—4.

[18] Grover FL, Shroyer LW, Edwards FH, Pae Jr WE, Ferguson Jr TB, Gay Jr WA, Clark RE. Data quality review program: the Society of Thoracic Surgeons adult cardiac national database. Ann Thorac Surg 1996;62:1229—31.

[19] Shroyer AL, Edwards FH, Grover FL. Updates to the data quality review program: the Society of Thoracic Surgeons Adult Cardiac National Database. Ann Thorac Surg 1998;65:1494—7.

[20] Maruszewski B, Lacour-Gayet F, Monro JL, Keogh BE, Tobota Z, Kansy A. An attempt at data verification in the EACTS Congenital Database. Eur J Cardiothorac Surg 2005;28:400—4.

Appendix A. Conference discussion

Dr B. Kozower (*Charlottesville, VA*): This is a very forward thinking approach to assessing the quality of a clinical database in surgery. The audience knows the importance of a well-constructed and well-administered database. As the demand for improving and documenting quality continues to increase, an approach like the one you have used to understand and to document quality will be essential. It will also be extremely valuable in the future to have a way to compare the quality of different databases. So I have three questions for you.

First, why did you exclude centers submitting fewer than 100 cases, and how many centers did you exclude?

Dr Salati: Yes, you are right, we decided excluding those units contributing less than 100 consecutive patients to the ESTS database, and, applying this criteria, we lost 18 centers, but the remaining 20 centers granted us the opportunity of performing the analysis on a large amount of data as well, accounting for 90% of the data collected within the database. So we think that this sample of the database could be considered sufficiently representative of the entire database as well.

Dr Kozower: Secondly, you chose a quality threshold of 0.8. How exactly did you choose that threshold and is it high enough for a voluntary clinical database?

Dr Salati: Yes, we chose the threshold of 0.8, but, to be clear, this threshold is completely arbitrary. We indicated this threshold after a careful review of the data quality literature, obviously, and particularly we took into consideration a paper published at the end of the '90s by Hogan, which showed the correctness and completeness of values reported by several studies assessing data quality of different medical registries. So taking into account these results, we indicated this threshold of 0.8.

I would like to point out another detail that, yes, we indicated this threshold of 0.8, but we also think that this threshold could be changed depending on the metric that we are assessing and also on the variable that we are testing. Because, for instance, a threshold of 0.8 could be considered as sufficiently high for a variable such as the 'ASA score', for instance, but not for another variable such as 'operative mortality', where we would like to reach, for instance, a threshold of 0.95 of completeness. So basically we think that this threshold could be tailored depending on our scenarios and also on the goals that we want to reach.

Dr Kozower: My third question has to do with the data quality aspect of correctness/accuracy. I understand that you are trying to look at a task-independent way of assessing this quality metric. Traditionally when we look at accuracy or correctness we have performed audits of the data. However, you define accuracy as syntactic accuracy, and I just want to make sure everyone understands what it is you are measuring. It is not actually comparing what is entered with the real or actual value. You are comparing what is entered in the database with a category domain or range of plausible values.

So, what is the audit process for the ESTS database to see if indeed there is accurate data? And second, should we really consider this syntactic accuracy as accuracy?

Dr Salati: I think I should clarify this concept. In general, we can distinguish two types of accuracy, the syntactic accuracy and the semantic accuracy. The syntactic accuracy is the only type of accuracy that data quality assessment methodologies are interested in assessing, and the syntactic accuracy measures the closeness of a particular value to the possible elementary attributes of the definition domain. In other words, in syntactic accuracy we are only interested in verifying that a specific value is correspondent to one of all the possible values of the domain. So, for instance, if we consider a variable such as name of the patient, a value such as Gene is syntactically accurate even if the real world value is John. This is completely not true for the semantic accuracy, because in the semantic accuracy, we want to assess if a particular value is exactly correspondent to the original real world value. So coming back to the example, if we consider the variable name of the patient, in this case a value as Gene is semantically not accurate if the real world value is John. The only value that we expected is Gene again. This kind of accuracy, the semantic accuracy, could also be assessed even if it is not the main metric that we should consider in data quality assessment methodologies. And all the processes that want to assess this kind of accuracy are desirable and very important, but this kind of activity should be planned very carefully because of the impact on cost and time-consuming activities. And this is particularly true for large multi-national, multi-institutional databases, as, for instance, in our case, in the ESTS database, because we have to consider lingual problems, too.

Finally, we didn't assess the semantic accuracy. At present, the ESTS doesn't have a formal audit process of data based on the data source verification.

Dr E. Lim (*London, UK*): When you introduce the term 'data completeness', how can you guard against people who are selecting the data fields which are more complete to use them as outcome measures themselves?

Dr Salati: To force them?

Dr Lim: No. The degree of completion depends on the data variables you choose. Some presumably will be very, very badly completed. How can we guard against manipulation of the original question so that you don't actually choose the fields which are already very well filled in as markers for good data completion?

Dr Salati: If I understand your question, we selected at the beginning of the process 15 core variables, because we think that these kind of variables should be considered the most important for the ESTS database because they have the most important impact on outcomes at different levels, and within this selection, we made the analysis of the completeness of each of these variables.

Dr N. Cohen (*Baltimore, MD*): To expound on that question, is the ESTS database a voluntary database or an involuntary database?

Dr Salati: Voluntary.

Dr Cohen: Voluntary, exactly. So there is a vast world of difference in the data quality that is collected from people that want to participate, that volunteer to participate, and that you select to be very high volume. It is very different than, say, a transplant database. In the United States, if you want to do transplants you must participate in UNOS, and the quality of the data is different. So you have set the bar arbitrarily at 8, but in fact I would argue that the bar should be at 1.0, because these people want to do this and they should do it 100% accurately. Do you have a claim style database which is involuntary to test for true quality of the whole world of your data?

Dr Salati: The problem of a voluntary or involuntary registry could influence our position in assessing data quality, obviously, but this is just a methodology. We would like just to present a methodological approach to assess the quality. We didn't try to define if the threshold assessed is enough or not enough.

Dr Cohen: That is fine. The question becomes the standard to which you hold the data and that is what is going to reflect the rest of the data analysis when you go to a larger registry. In the United States, the Society of Thoracic Surgeons database is only 111 centers, but there are thousands of centers that do thoracic surgery that don't participate. So the quality of the data is the actual best that we can see. That should be held to a much higher standard than some arbitrary math number.