Reasons to participate in European Society of Thoracic Surgeons database

Michele Salati

Unit of Thoracic Surgery, Ospedali Riuniti Ancona, Ancona, Italy

Correspondence to: Michele Salati. Unit of Thoracic Surgery, Ospedali Riuniti Ancona, Via Conca 71-60020 Ancona, Italy. Email: michelesalati@hotmail.com.

Abstract: The process of data collection inevitably involves costs at various levels. Nevertheless, this effort is essential to base our knowledge and the consequent decision making on solid foundations. The European Society of Thoracic Surgeons (ESTS) database collects a large amount of data on general thoracic surgery derived from about 60 units representative of 11 nations. Since its beginning in 2001, the ESTS database has contributed to increase the knowledge and the quality of care in our specialty. The present paper illustrates the ultimate finalities and the obtained results of this data collection, providing a broad overview of the motivations to participate to the ESTS database.

Keywords: Database; data management; thoracic surgery

Introduction

The European Society of Thoracic Surgeons (ESTS) database can be classified as an international multi-institutional medical specialistic collection of data.

In order to understand the importance to participate to the ESTS database, we should proceed at different levels, identifying:

(I) The meaning, role and finality of data in the decision making process (general-philosophical level);

(II) The benefits of a big mono-specialistic clinical database (specific-practical level).

Moving across these two different planes, this paper is aimed at clarifying the science behind the data collection and the following knowledge extraction process. These activities represent the fundament of any decision-making strategy, which is the real ultimate goal and reason to justify a data collection effort.

Data behind actions

The term "DATUM" refers to a single piece, an indivisible quantum of knowledge. It exists a well defined relationship between the datum, the information and the knowledge.

As shown in *Figure 1*, the association of a certain mass of data permits to obtain an information on a subject of the real world. The critique interpretation, using experience and intelligence, of multiple informations generates knowledge about a specific aspect of the real world. This knowledge is able to support and guide the decision making process, that is to say to give direction to our actions (1).

Making an example derived from the medical field. The heart rate represents a single datum related to a person (DATUM). The heart rate may be associated with many other data and reported in a clinical chart or in an electronic support, in order to describe the physical status of that person (INFORMATION). Based on these information, it can be developed a document about the strategy to reduce the risk of a heart attach for that person (KNOWLEDGE).

This means that data are the cornerstone necessary to support the decision making processes. Data are allowed to act using the less arbitrary choice in order to reach our planned results. Moreover, the result should be codified and stored as new data, in order to obtain another substratum to proceed further with the advance of knowledge.

Thus the creation of a baseline data collection, that describes the current state of a certain subject, is one of
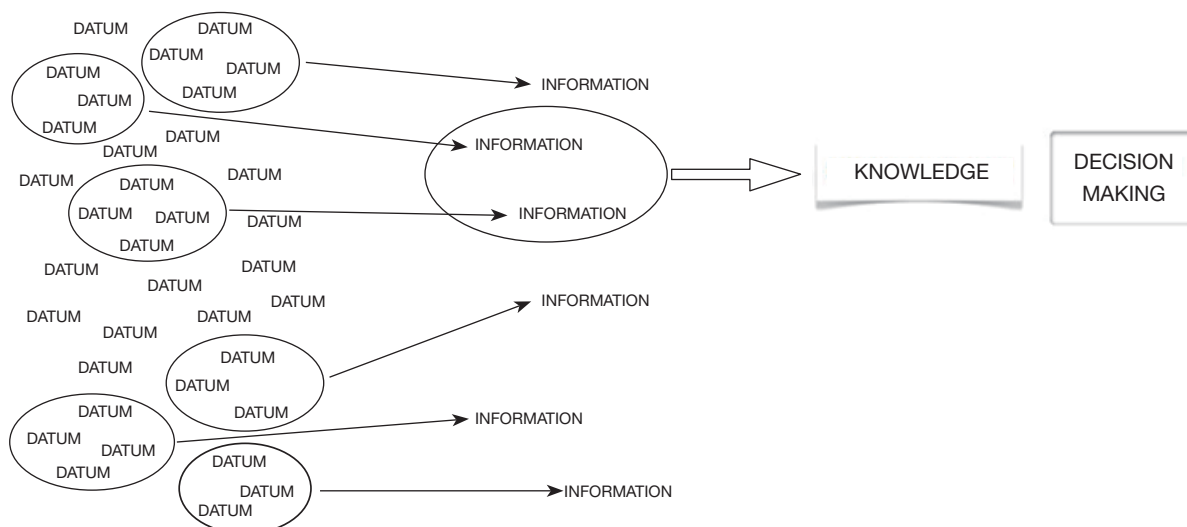
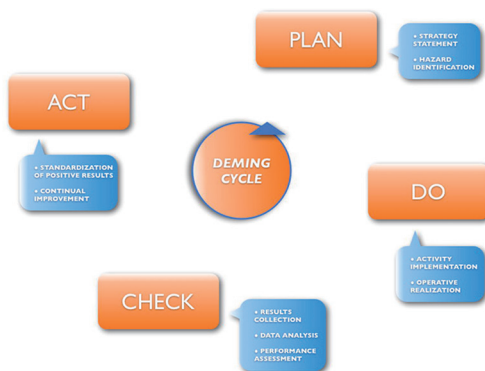**Figure 1** From data to action: the process of decision-making.

Figure 2 Deming Cycle.

the critical phases to understand future roots of action and to measure improvements. This is extensively reported into the Deming Cycle model (2), a quality improvement strategy consisting of a logical sequence of four repetitive steps for continuous improvement and learning, which influenced mostly of the quality management strategies developed during the last 30 years (*Figure 2*).

These concepts should be constantly taken into account in our clinical practice. Being focused on the continuous improvement of our results at any level, we are supposed to apply strategies inspired to the Deming Cycle. Therefore, the data collection should take the upmost consideration, as it is the principal process able to influence the further decision making phase.

Knowledge discovery in medical databases

Applying the previously exposed concepts about data and knowledge to the medical field, we may say that capturing data inherent to our patients is necessary to build the basic evidence and theories to manage our profession.

This is true for mono-institutional databases as well as for large multi-institutional international databases. Obviously, to handle large datasets implies the use of more sophisticated strategies and methods for an efficient extraction of knowledge. For instance the ESTS database has a high data growth rate, due to a constant increase year by year in two ways: (I) number of records collected (vertical size-high increase yearly); (II) number of attributes for each record (horizontal size-low increase yearly). Consequently, specific processes for gathering and handle massive data are needed, as well as methods for assuring the quality of data collected.

Thus different data collection strategies have been developed for the ESTS database, which allowed to import data from multiple sources: (I) direct upload from an ad hoc website (https://ests.dendrite.it/csp/ests/intellect/login.csp); (II) off line upload from single institutional dataset; and (III) off line upload from multi-institutional national datasets (COLLECTION PHASE).

All these preprocessed data have been transformed in an electronic standardized format and cleaned. Moreover activities of quality control have been planned for an ad hoc

verification of the quality of mass of data collected within the ESTS database (TRANSFORMATION PHASE).

So this large amount of data became apt to be managed using appropriate computational techniques. Specific algorithms have been applied in order to discover information from the ESTS database (DATA MINING PHASE).

The proper interpretation of the data mining results finally allowed to extract from the original ESTS data useful pieces of knowledge, not always represented in small institutional or national datasets.

The whole process of data management used for the ESTS database is commonly known as knowledge discovery in databases (KDD) (3). As shown in *Figure 3*, using specific data mining methods (that are dependent to the intended source), we can derive from a data collection, like the ESTS database, different types of information/ discoveries, increasing our:

(I) Descriptive knowledge: the whole process of data management and analysis describes homogeneous characteristics, patterns and behaviors detectable in the observed real-world;

(II) Predictive knowledge: the whole process of data management and analysis creates models with the potential of predict events, results, actions in the observed real-world.

In fact, as more extensively reported in the following paragraph, analytic models applied to the ESTS database have been finalized at enhance the descriptive knowledge of our profession at an European level. Examples in this sense could be considered most of the data elaborations and analysis performed in order to obtain the reports presented within the Silver Book (http://www.ests.org/_userfiles/pages/files/ESTS%20Report%202014_final_on-line.pdf). This document offers information about the thoracic surgery activity in Europe recognizing more or less evident groups or classes. At the same time it describes emerging habits or strategies of our profession and trends over the time.

On the other side, the data collected within the ESTS database have been used to apply analytic models aimed at verifying the correlation between patient characteristics, preoperative, intraoperative or postoperative activities and outcomes. These are examples of predictive knowledge that, based on a European data knowledge discovery process, strongly contributed to enhance the level of consciousness and perception of our activity as surgeon and to help the consequent decision making process.
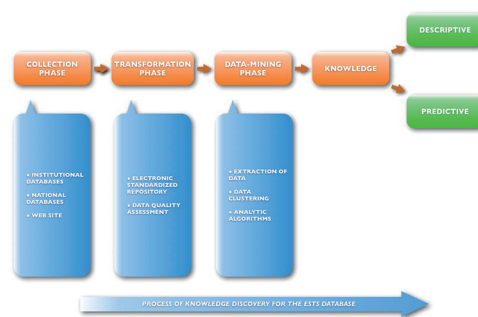

Figure 3 The process of KDD. KDD, knowledge discovery in databases.

Knowledge discovery (and more) from the ESTS database

The ESTS database is a data collection with the following characteristics:

(I) Mono-specialistic: data are inherent to different procedures of the thoracic surgery specialty;

(II) Multi-dimentional: multiple attributes describing the preoperative-baseline characteristics, the operative procedures, the postoperative course and outcome are reported in the data collection;

(III) Multi-institutional: data are uploaded by multiple contributors using a final common electronic repository (as stated before, the data can be uploaded both on-line using an ad hoc platform or off-line through specific data import procedures);

(IV) International: collecting most of the data from units of thoracic surgery across Europe.

The activity of data collection started in 2001 using a simple Filemaker Pro support and today the ESTS database has a web platform, where data can be directly uploaded on line. The participation to the ESTS database is voluntary and reserved to all the ESTS members. Each contributor, after obtaining a personal login account requested through a specific application form to the ESTS, can upload his data at any time. The information sent to the database are completely anonymized, but each unit has the privilege to download his own data for analysis at a local level. As previously reported, the possibility exist to import off line blocks of data from national databases as well as from databases maintained in single centers.

This large amount of data is gathered with the support of a professional society (Dendrite Clinical System Italia

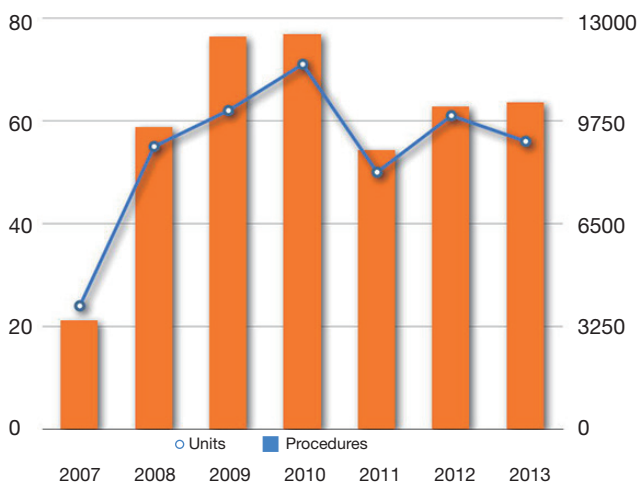|	2007	2008	2009	2010	2011	2012	2013
Units	24	55	62	71	50	61	58
Procedures	3431	9583	12421	12498	8833	10200	10326

Figure 4 Trends of contributors and procedures collected over the years.

Srl), which assures a proper storage, transformation and security of the data, using dedicated software and hardware instruments.

Nowadays, the ESTS database collects data from about 60 different thoracic surgery units (if we consider only those contributors uploading at least 50 procedures), ranging from a minimum of 24 units, at the moment of dataset institution in 2007, to a maximum of 71 units in 2010 (*Figure 4*). This offered the possibility of analyzing data about our specialty from 11 nations.

At the end of 2014 were uploaded within the ESTS database about 74 thousands procedures. The large majority of themes are grouped as lung procedures (78%), followed by procedures on mediastinum and thymus (8.2%), pleura (7.7%), chest wall (2.7%) and so on. Considering that each procedure is described by 60-70 different attributes this means that the ESTS database handles about 4-5 millions of data.

Based on the ESTS database, specific analysis have been performed to build models able to stratify the risk in lung resection candidates and to compare the outcomes of different European thoracic surgery units, once adjusted for the complexity of patients treated. These results were described in two different studies published in 2005 and 2008 (4,5), that explained and applied the so called European Society Objective Score (ESOS). This score, based on simple baseline characteristics of the patient such as the age and the forced expiratory volume at first second predicted postoperative (ppoFEV1), is nowadays available for each thoracic surgery unit to graduate the risk and to monitor and compare the quality of care even at a local level.

Other studies led to the creation of a multi-dimensional score (Composite Performance Score—CPS), that evaluates entirely the performance of a Unit taking into account measures of the preoperative, intraoperative and postoperative processes of care (6,7). The solidity of the model used for deriving the CPS is consequent to the analysis of a large number of cases reflecting the real activity of multiple international Units. It would have been hazardous and unreliable to derive such a score, which measures the quality of clinical practice in our specialty, analyzing data from a single institution dataset.

At the same time, the ESTS database was used to develop and apply analytic models of data quality evaluation. Specific data quality metrics for clinical datasets were tested and the ESTS database become one of the few international clinical datasets periodically explored for advancing data quality management strategies (8).

Moreover, since 2009, the ESTS database committee published an annual report known as Silver Book, which describes the thoracic surgery activity across Europe and outlines benchmarks for standardize the quality of care in our specialty. Over the years the reported analyses have described with increasing accuracy multiple aspects of the thoracic surgery practice, starting from the epidemiological and physiologic characteristics of our patients, passing through our operative strategies and postoperative management approach, ending to the obtained outcomes. Interestingly, the last ESTS database annual report reserved a dedicated section to show the comparison of surgical outcomes between the years 2007-2010 and 2011-2013 (http://www.ests.org/_userfiles/pages/files/ESTS%20 Report%202014_final_on-line.pdf).

Apart the evidences published in scientific literature, the ESTS database plays a central role in the process of accreditation of the European thoracic surgery units (http:// www.ests.org/collaboration/ests_quality_certification_ programme.aspx). In fact the participation to the ESTS database is a mandatory activity for those units interested in participating to the European Institutional Accreditation Program. This program is aimed at "standardize and improve practice of European thoracic surgery units by peer-driven, voluntary and specific instruments of clinical

Table 1 Advantages in contributing to the ESTS database

Benefits for the clinical practice, specialty and patients

Development of European benchmarks of performance through the analysis of outcomes and processes of care indicators

Performance assessment by risk-adjusted outcome and/or process indicators, which will allow you to compare your own institutional performance against European benchmarks

Analysis and development of new potential outcomes and processes of care indicators that may complement/substitute current quality of care measures

Implement a provider-led quality monitoring and improvement program with the aim to improve your practice

Feedback to document quality efforts and areas for improvement in quality of care

Data for research projects, which can be used to assess new technologies/pathways of care that can ultimately lead to improved patient care and outcomes

Maintain your own data if data is requested or mandated by third parties

Use for local hospital administration resource allocation

Use for individual negotiations, public relations and expert witness

Opportunity to participate in a European quality improvement effort for general thoracic surgery that has a positive impact at the local, national and international levels

Benefits for the unit

Participation to the ESTS database is a pre-requisite to participate in the European Institutional Quality certification program

Your own data, collected in a standardized ESTS-endorsed dataset, can be downloaded at local level and used for your internal quality analyses or institutional research purposes

As a future project, participants will receive a periodic confidential feedback on the quality of their data and their performance against International benchmarks

Participants can propose their own research projects based on the total data present in the database

ESTS, European Society of Thoracic Surgeons.

audit based on outcome and process of care evaluation". Each unit contributing for 2 years to the ESTS database, uploading an adequate amount of high quality data, can be evaluated using the Composite Performance Score. In case of good CPS the unit is eligible for an external audit and a final evaluation by the ESTS database committee, which assigns the accreditation.

Special benefits of the ESTS database contributors

Considering the reasons why to participate to the ESTS database with a different perspective, using the more focused view represented by the benefits accessible for a single contributor, the ESTS database committee have clearly stated these benefits in several documents.

Table 1 reports extensively the utilities and advantages of the units contributing to the ESTS database.

It is evident that contributing to the ESTS database permits to increase the critique analysis and to derive evidence about our specialty at an international level. These

results should lead an improvement of the quality of care with consequent advantages available to each thoracic surgery patient. At the same time to the contributors is offered the opportunity to do specific research on their own data, once downloaded, as well as on the entire amount of data collected within the ESTS database. Finally, participating to the ESTS database it is possible for a single unit to evaluate the quality of care offered in comparison to many others units in Europe and to certificate it accessing to the European Institutional Accreditation Program.

Conclusions

The ESTS database is one of the larger multi-institutional collections of data in thoracic surgery. It has a constant growth rate with a large potential to increase the yearly data upload, considering that only 15% of the European thoracic surgery units, among those that have at least one staff surgeon as full member of the ESTS, are contributing to the ESTS database.

Participating to the ESTS database means to facilitate the increase of the scientific evidence in our specialty, with a positive impact on the quality of care offered to our patients.

At the same time the ESTS database is an instrument able to rigorously lead the processes of monitoring and standardization of the thoracic surgery activity across Europe.

Each unit contributing to the ESTS database has the opportunity of obtaining his own cleaned data for analysis at local level. Moreover each contributor may enter the ESTS Quality Certification Program.

Tacking into account these reasons to participate to the ESTS database, all the present contributing units as well as the future ones should be commended for the valuable effort aimed at offering benefits to the patients, the colleagues and the entire scientific community.

Acknowledgements

References

1. Pazzani MJ. Knowledge Discovery from Data? IEEE Intelligent Systems 2000;15:10-12.
2. Deming WE. eds. Out of the Crisis. Cambridge, Mass.: Massachusetts Institute of Technology, Center for Advanced Engineering Study, 1986.
3. Fayyad U, Piatetsky-Shapiro G, Smyth P. The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM 1996;39:27-34.
4. Berrisford R, Brunelli A, Rocco G, et al. The European Thoracic Surgery Database project: modelling the risk of in-hospital death following lung resection. Eur J Cardiothorac Surg 2005;28:306-11.
5. Brunelli A, Varela G, Van Schil P, et al. Multicentric analysis of performance after major lung resections by using the European Society Objective Score (ESOS). Eur J Cardiothorac Surg 2008;33:284-8.
6. Brunelli A, Berrisford RG, Rocco G, et al. The European Thoracic Database project: composite performance score to measure quality of care after major lung resection. Eur J Cardiothorac Surg 2009;35:769-74.
7. Brunelli A, Rocco G, Van Raemdonck D, et al. Lessons learned from the European thoracic surgery database: the Composite Performance Score. Eur J Surg Oncol 2010;36 Suppl 1:S93-9.
8. Salati M, Brunelli A, Dahan M, et al. Task-independent metrics to assess the data quality of medical registries using the European Society of Thoracic Surgeons (ESTS) Database. Eur J Cardiothorac Surg 2011;40:91-8.