How to create a surgical database?

Dania Nachira¹, Luca Bertolaccini², Mahmoud Ismail¹, Marco Chiappetta¹, Elisa Meacci¹, Stefano Margaritora¹

¹Department of General Thoracic Surgery, Fondazione Policlinico Universitario “Agostino Gemelli” IRCCS, Università Cattolica del Sacro Cuore, Rome, Italy; ²Department of Thoracic Surgery, Maggiore Teaching Hospital, Bologna, Italy; ³Competition Center of Thoracic Surgery, Department of Surgery, Charité-Universitätsmedizin Berlin, Berlin 10117, Germany

Correspondence to: Dania Nachira, MD. Department of General Thoracic Surgery, Fondazione Policlinico Universitario “Agostino Gemelli” IRCCS, Largo A. Gemelli, 8, 00168, Rome, Italy. Email: danynac@libero.it.

doi: 10.21037/jtd.2018.10.58
View this article at: http://dx.doi.org/10.21037/jtd.2018.10.58

Introduction

In an increasingly fast-changing medical world, above all in surgery, it has become essential for physicians not only offering patients the best, updated and safe treatment but also contributing in clinical research by sharing their results.

Indeed, above all in academic settings, surgeons are frequently asked to provide postoperative results about new treatments, fast-tracking protocols, outcomes in old patients with several comorbidities, etc. and also hospitals are very interested in evaluating quality of care in a more and more cost-conscience setting.

There is, therefore, an increasing need of goof data in all clinical fields (1).

Prospective databases can provide powerful information about patients’ clinical data, surgical results, postoperative outcomes, success and failure of treatments. Sharing the data with other centers, like it is in the aim of online international databases (2,3), allow surgeons and researchers to obtain faster and easier larger series of information for answering clinical queries in more powerful way. This allow, for instance, to evaluate the most common types of complications of a surgical treatment and identify the high-risk class of patients, in order to create risk scores and develop personalized solutions. Only recording, collecting and sharing clinical successes and failures, it is possible to learn from them and growing. And growing in surgery means better and safer treatments for patients.

Obviously, there is not only one perfect database but the ideal one must be able to collect the most important and necessary variables that can answer the main queries on a certain topic.

Far away from claiming to be a “holy text” of methodology, the aim of this paper is to provide the basic principles to young surgeons and researchers for creating their own solid and useful database for surgical research.

Online large-volume databases

The advances in all fields of technology and consequent changes in healthcare have stimulated birth of national and international large-volume databases in surgical research for stratification of risk factors, cost evaluations and investigations of postoperative complications and other topics (3).

The commonly used national databases are administrative or clinical databases.

The administrative databases are usually created for having billing details, while clinical ones for recording medical information and evaluating specific surgical specialty arguments (3).

One of the first example of surgical national database is the field of cardiothoracic surgery, for instance, is the Society of Thoracic Surgeons (STS) national database, established in 1989.

Originally, it recorded only cardiac operations, then the program expanded to include general thoracic surgical procedures in 2002 and now STS database includes more than 345,000 general thoracic operations, involving more than 800 surgeons (4). Since 1 January 2018, the STS National database was further implemented, having now 4 components on different area of cardiothoracic surgery (adult cardiac surgery, general thoracic surgery, congenital heart surgery and Intermacs Database—on mechanical circulation support).

On the other side of the ocean, for the same surgical field,
The European Society of Thoracic Surgeons (ESTS) Database Committee founded the ESTS database, in 2001 (5). It was developed to assess the performance of thoracic surgery units across Europe by risk-adjusted instruments (5), with the final aim of improving patient care. The second version of ESTS database was launched online on July 2017 and now it counts the participation of about 190 thoracic surgical units.

The accurate methodology in expanding and including more procedure-specific-information represents the power of these databases and allows more and more the reproducibility of results and database-driven surgical research (3).

**Surgical databases: methodology and road map**

A well-structured database should have the capability to answer a wide array of questions in surgical research (3), drawing epidemiological, technical, oncological and life-quality conclusions.

Furthermore, the main goals of a good database are: easy data input, personal and medical information stored and protected, readily accessible by staff physicians from any institutional workstation thanks to a password (1), availability of encoded electronical data for promoting the development of multicenter studies. Indeed, a good institution database can easily be transferred into nation- and internationwide databases (1,6).

Data collection form should include all crucial variables of baseline characteristics, clinical, surgical information and outcomes (7).

Therefore, the first step is the correct selection of data to be collected that should include all items necessary to carry your project of research out, effectively.

The better the data, the better the results will be.

**Some tips on how select variables**

A good data form is the basis of everything and it can be easily transformed into an electronic well-structured database.

The main questions to be answered in building a data collection form are:

- Is it easy to be completed properly?
- Are all necessary variables inside (avoiding secondary and unnecessary variables...)?
- Are all data about surgical technique and operation reported?
- Are all main risk factors and possible predictors evaluated?
- And what about main primary and secondary outcomes (including life-quality results)?
- Are main complications and morbidity inside?
- And what about follow-up data?
- And for research about oncological pathology: are all necessary oncological, staging and pathological information recorded?
- Is this database suitable for future researches apart from the planned and upcoming one?
- Is this database suitable for sharing data with other centers?

An example of a data collection form is reported in Table 1. Demographic, biological, surgical and outcome variables can be divided into quantitative or qualitative (8).

Among quantitative variables there are categorical and continuous ones. A particular type of categorical variables is the Boolean type that assumes always a binary value (yes or not, 0 or 1, male or female etc.). In reporting continuous variables, for not losing information (that can be precious for future aims of multicentric researches) it is always useful to record the value as it is, without creating a cut-off value, if not in case of final analysis.

Qualitative variables, instead, need to be categorized and encoded.

Obviously, it is quite impossible to create a unique database that can be suitable for any type of research or necessity, also because a large database, full of variables and information becomes difficult to be maintained and updated.

If a surgical department is already involved in a multicentric or international research group, it is useful considering in the main structure of its own database the variables required for the multicentric project, that the involved centers can be asked to reverse on multicentric or online database.

And a good database should be enough complete, well encoded for speeding up its filling and as less as possible time-consuming, both in updating phase and in analysis phase, if questioned by appropriate queries (1).

An electronic database can be easily set up using several types of programs.

The most known and widespread programs for creating data fields are Microsoft Excel, Microsoft Access or SPSS (IBM SPSS, Armonk, NY, USA) or, if preferred, web-based application can be also suitable (1).

The choice is related to personal preferences, experience or institutional budget available and consequently to the figures that will manage the database, if medical doctors or
### Table 1 Example of data collection form

#### Patient demographics
- Name
- ID code
- Date of birth
- Age
- Gender
- Height (cm)
- Weight (Kg)
- BMI
- ...

#### Comorbidities
- Current smoker
- Former smoker
- Pack-year
- Diabetes
- Hypertension
- Cardiovascular diseases
- ECOG performance status
- ASA score
- Symptoms (type; duration…)
- ...

#### Other syndromes or paraneoplastic syndromes (type; appropriate classification if any…)
- Previous tumors (yes/no; type…)
- Treatment of previous tumor (CT, RT, surgery…)

#### Oncological data (in case of oncological disease)
- CT (yes/no; date…)
- PET-CT (yes/no; date…)
- MRI (yes/no; date…)
- Other radiological and instrumental examinations (type; date; results…)
- Preoperative biopsies (type and date)
- Preoperative diagnosis
- Clinical stage

### Table 1 (continued)

| Lung functionality (in case of pulmonary surgery) | FEV1, DLCO, pO2, pCO2… |
| Pulmonary scintigraphy | … |
| Preoperative cardiological evaluation | Echocardiography |
| Stress test | … |

#### Preoperative treatment
- Induction therapy (yes/no; type; number of cycles, in case of neoplastic pathology)
- Antibiotic therapy (type and duration)
- Clinical stage (ycTNM, in case of tumor)
- ...

#### Surgical info
- Date of operation
- Type of surgery
- Type of surgical access
- Technique (other surgical info…)
- Duration
- Operator
- Assistant
- Use of advanced materials, sealants, energy devices, etc…
- Number of drainages
- Intraoperative complications (minors, majors, type…)
- ...

#### Postoperative data
- ICU admission (yes/no; number of days…)
- Postoperative complications
- Re-operation
- Hospital stay
- Drainage length
- Postoperative pain and VAS scale
- Outcomes (level of satisfaction, quality of life, effectiveness, cosmetic results etc.)

...
Table 1 (continued)

<table>
<thead>
<tr>
<th>Oncological and pathological data (in case of oncological disease)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Histology</td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td></td>
</tr>
<tr>
<td>Tumor dimension</td>
<td></td>
</tr>
<tr>
<td>Number of lymph nodes removed (number of positive lymph nodes)</td>
<td></td>
</tr>
<tr>
<td>R-status and site</td>
<td></td>
</tr>
<tr>
<td>Receptor mutations</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Follow-up data

| Date of follow-up |  |
| Recurrence of disease |  |
| Overall survival |  |
| Long term results (level of satisfaction, quality of life, effectiveness, cosmetic results...) |  |
| Any other type of treatment or surgical re-treatment |  |
| Disease-free survival (for oncological diseases) |  |
| Oncological adjuvant treatment (yes/no; type; duration...) |  |
| ... |  |

BMI, body mass index; ECOG, Eastern Cooperative Oncology Group; ASA, American Society of Anesthesiologists; CT, computed tomography; RT, radiotherapy; PET, positron emission tomography; MRI, magnetic resonance imaging; VAS, Visual Analogue Scale.

Conclusions

The creation of surgical databases, following the main suggestions given above, can help surgeons in providing patients the best, updated and safe treatments, through the constant self-evaluation of their clinical practice, but also in contributing in international clinical research.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References
