Clinical registries: governance, management, analysis and applications

Graeme L. Hickeya, Stuart W. Grantb, Rebecca Cosgriffc, Ioannis Dimarakisb, Domenico Paganoa,
Arie P. Kappeteine and Ben Bridgewaterb,c,*

a Manchester Academic Health Science Centre, The Northwest Institute for BioHealth Informatics, University of Manchester, Manchester, UK
b Department of Cardiothoracic Surgery, Manchester Academic Health Science Centre, University of Manchester, University Hospital of South Manchester, Manchester, UK
c National Institute for Cardiovascular Outcomes Research, The Institute of Cardiovascular Science, University College London, London, UK
d Department of Cardiothoracic Surgery, Queen Elizabeth Hospital, Edgbaston, Birmingham, UK
e Department of Thoracic Surgery, Thoraxcenter, Rotterdam, Netherlands

* Corresponding author. Department of Cardiothoracic Surgery, University Hospital of South Manchester, Southmoor Road, Manchester M23 9LT, UK. Tel: +44-161-2912511; e-mail: ben.bridgewater@uhsm.nhs.uk (B. Bridgewater).

Received 26 October 2012; received in revised form 28 December 2012; accepted 4 January 2013

Summary
Clinical registries will have an increasingly important role to play in health-care, with a number already established in cardiac surgery. This review covers the fundamentals of establishing and managing clinical registries, including legal and ethical frameworks along with intellectual property attribution. Also discussed are important issues relating to the processing of data, data extraction and conducting analyses using registry data.

Keywords: Registries • Databases • Ethics • Management • Research

INTRODUCTION

Clinical registries will have an increasingly important role to play in health-care as information technology infrastructures develop, and the demand for health-care quality improvement and transparency increases [1–8]. Clinical registries can be used to improve the quality of patient care, underpin research, improve cost-effectiveness and provide information for regulatory purposes. Clinical registries are well established in cardiac surgery [2, 9], and will form a vital part of the European Association for Cardiothoracic Surgery (EACTS) Quality Improvement Programme (QUIP) for adult cardiac surgery.

The aim of the QUIP is to provide the framework within which outcomes in cardiac surgery across the entire membership can be measured and improved. The framework will consist of tools for individual health-care teams to measure, predict and improve outcomes. The QUIP will also comprise specific domains aimed at recommending educational standards for health-care providers and reviewing, implementing and designing clinical guidelines. Finally, an Outcome Research Network within this initiative will form the basis for multicentre research in the field. Accurate registries will be paramount to the delivery of the QUIP.

This review covers the fundamentals of establishing and managing clinical registries [10]. Many of the examples used in this review are from the Society for Cardiothoracic Surgery in Great Britain and Ireland (SCTS) National Adult Cardiac Surgery Audit (NACSA) registry, which was established in 1996, and is now managed by the National Institute of Cardiovascular Outcomes Research (NICOR) (www.ucl.ac.uk/nicor). The NACSA represents only one registry that contributed to the most recent EACTS Database report, and it must be acknowledged that there were other major contributors including Germany (http://www.gstcvs.org/); Poland (http://www.krok.org.pl/); Belgium (http://www.bacts.org/); China (http://www.cvs-china.com/) and Italy (http://www.sicch.org/). The issues discussed in this review are generic and largely relevant to any clinical registry project.

THE LEGAL AND POLITICAL FRAMEWORK

Fundamental to any large clinical registry is its ability to function within the parameters laid out by the legal framework governing it [11]. Failure to understand or comply with this framework can be a terminal event for any registry and could potentially have criminal or political repercussions. Issues that fall under this umbrella include data privacy and protection, ethical use of data and intellectual property rights.

Information governance

Information governance is the framework that brings together all the legal rules, guidance and best practice to ensure necessary safeguards for personal information. It is a complex topic and the precise legal framework is often specific to individual countries. An understanding of the appropriate legislation and ethics with respect to clinical data is a prerequisite for establishing any clinical registry.
The general principles of information governance are that patients’ personal data should be accurately collected and stored securely and not shared without appropriate permissions. Personal data also include information that can identify an individual patient in combination with additional data; this is particularly important in an era of record linkage. The publication of non-personal data for small groups of patients with highly specific characteristics could also mean that individual patients become identifiable, and this should therefore be avoided.

The 1995 European Union (EU) Data Protection Directive [12] regulates the processing of personal data within EU Member States. Its guidelines relate to transparency, explicit and legitimate purpose and proportionality of data processing. Although it authorizes the transfer of personal data from a Member State to a third country that has an adequate level of data protection, international registry projects need to abide with each country’s specific national data protection legislations, even if they are stored and analysed in a single country [13]. The interpretation of the Directive by the UK, and existing legislation are described in Table 1. There have been conflicting legal views on the interpretation of these laws for practice in health-care research [11], and this has subsequently disrupted a number of registry projects due to lack of legal clarification [14].

Intellectual property and project governance

Intellectual property (IP) with respect to registries is a complex issue [15]. There are a number of different aspects, including database rights, copyright, confidence and contract rights, all of which are governed by National and EU legislation. The exact framework for defining IP with regards to registries has not yet been developed, and because different groups often manage various components, registries are potentially exposed to the risk of contention.

Owing to the large number of organizations involved and the importance of the registry, it is not difficult to conceive of a situation whereby conflict regarding IP and data control might arise. It is important therefore that any registry should give consideration to these issues as it is being established. Project governance—the management framework within which project decisions are made—is key to this. A list of all the stakeholders involved in managing and participating in the NACSA is shown in Table 2.

As an example, in the UK, for pragmatic purposes the actual ‘owner’ of the data submitted from NHS organizations is the Secretary of State for Health. The dataset structure and definitions are the IP of the professional society. As sponsors of the NACSA, the Healthcare Quality Improvement Partnership (http://hqip.org.uk) exerts control over the utilization of the pooled national dataset and is responsible for determining the purpose and manner in which any personal data are, or will be, processed [16]. The hospitals that submit data to the UK audit do so under ‘goodwill’, and we aim to ensure that hospitals provide consent for publication of outcomes data in which they are named.

We would regard a similar process of ‘consent’ as part of any upload of data to the EACTS Database as an essential part of future developments, prior to any subsequent publications. By virtue of its investment in collating and harmonizing multiple smaller registries, one might infer that EACTS will become the ‘owner’ of the data by proxy when it is used on the international platform, governed by its own elected body. In principle, the ownership of the raw national-level data should remain unchanged. However, we would emphasize that this interpretation is equivocal and lacks necessary legal scrutiny. The full legal framework regarding IP with respect to national and international registries is highly complex and beyond the scope of this review and will, in our view, require further consideration as part of the EACTS Database project [17].

Table 1: The primary legislation and guidelines that govern the use of health-care registry data in the UK

<table>
<thead>
<tr>
<th>Legislation</th>
<th>References</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Common Law Duty of Confidentiality</td>
<td>NA</td>
<td>A paradigm of law based on precedents set by verdicts made in court cases, which applies independently of the UK Data Protection Act 1998. It ensures that patient data are collected with the understanding that it is confidential and will not be shared with other bodies without the express consent of the patient.</td>
</tr>
<tr>
<td>The UK Data Protection Act 1998</td>
<td>[76]</td>
<td>Defines how ‘personal data’ must be collected and stored. It also places duties on those responsible for holding and using such information to ensure that it remains secure and is not susceptible to abuse. Personal data routinely collected in health care is covered by the Act, which states that it must be stored securely within that organization and must not be shared without appropriate legal permissions.</td>
</tr>
<tr>
<td>The Caldicott Principles</td>
<td>[77]</td>
<td>Recommendations and principles developed following a Department of Health review of the transfer of patient-identifiable information between NHS units and non-NHS bodies for all purposes. Within each NHS organization, a designated ‘Caldicott Guardian’ is required to monitor and ensure that the six principles of the report are enforced. Patient identifiable information should only be used with a justifiable purpose, the minimum necessary data should be used and access should be granted on a strict need-to-know basis. Responsibility is placed on all individuals involved in the process in addition to the organization as a whole to understand and comply with the law.</td>
</tr>
<tr>
<td>Section 251 of the NHS Act 2006</td>
<td>[78]</td>
<td>Allows the common law duty of confidentiality to be set aside in specific circumstances for improving patient care or in the wider public interest, where anonymized information is not sufficient and patient consent is not practical. For example retrospective consent for using registry data in a separate study (e.g. a record linkage project) would not be feasible, and patient identifiable information is required to link records and track long-term mortality.</td>
</tr>
</tbody>
</table>

NHS: National Health Service.
further branched Indication of a procedure within one of these groups unlocks categorized into four major groups: coronary artery bypass graft medical history, preoperative measurements, intraoperative Fields are classi comprehensive communications with all contributors and external Consequently revised in 2003 and 2010. Each revision required comp... may not be good case ascertainment and data completeness; however, they participants and small datasets facilitate high participation rates, good case ascertainment and data completeness; however, they may not be fit for the purpose. Datasets with too many data fields may be limited by significant missing data. A registry that can easily evolve to capture new data sources or fields is likely to be expensive and complicated, but one that is inflexible can become outdated as clinical practice develops. The first agreed dataset for the SCTS registry was finalized in June 1996 and subsequently revised in 2003 and 2010. Each revision required comprehensive communications with all contributors and external software developers.

For the NACSA dataset, approximately half of the 168 data fields are applicable to every patient, with the other half of the fields 'branched', meaning that they are only relevant for specific procedures or if certain patient characteristics are identified. Fields are classified into patient identifiers, patient characteristics, medical history, preoperative measurements, intraoperative fields and postoperative fields. Cardiac surgical procedures are categorized into four major groups: coronary artery bypass graft (CABG), valve, major aortic and other cardiothoracic procedures. Indication of a procedure within one of these groups unlocks further branched fields. For example, indicating a patient had a CABG procedure would unlock fields to allow completion of the number of grafts, graft sites and conduit type. If more than one graft is selected, then the data-collection software is programmed to expect multiple inputs for each unlocked field.

### DEVELOPING THE REGISTRY

#### Dataset design

Dataset design is a critical factor in establishing a registry. Fewer participants and small datasets facilitate high participation rates, good case ascertainment and data completeness; however, they may not be fit for the purpose. Datasets with too many data fields may be limited by significant missing data. A registry that can easily evolve to capture new data sources or fields is likely to be expensive and complicated, but one that is inflexible can become outdated as clinical practice develops. The first agreed dataset for the SCTS registry was finalized in June 1996 and subsequently revised in 2003 and 2010. Each revision required comprehensive communications with all contributors and external software developers.

For the NACSA dataset, approximately half of the 168 data fields are applicable to every patient, with the other half of the fields 'branched', meaning that they are only relevant for specific procedures or if certain patient characteristics are identified. Fields are classified into patient identifiers, patient characteristics, medical history, preoperative measurements, intraoperative fields and postoperative fields. Cardiac surgical procedures are categorized into four major groups: coronary artery bypass graft (CABG), valve, major aortic and other cardiothoracic procedures. Indication of a procedure within one of these groups unlocks further branched fields. For example, indicating a patient had a CABG procedure would unlock fields to allow completion of the number of grafts, graft sites and conduit type. If more than one graft is selected, then the data-collection software is programmed to expect multiple inputs for each unlocked field.

#### Collecting the data

For the NACSA all NHS, some Irish and a number of private centres performing adult cardiac surgery in the UK contribute data. The database now contains over 450,000 records. Data are collected through a variety of local specialized database systems [18]. Many of these systems are commercially developed; however, a number of centres use software developed ‘in-house’. The data remain in the individual centres for internal validation and local auditing. Following this process, the data are uploaded to central servers housed at NICOR. This upload is done using a sophisticated registry-import software tool that first executes a preliminary data screen to validate data quality. Data issues are flagged and invalid records rejected. Data are subsequently merged into a single homogenized file structure and encrypted.

Data validation is highly important to ensure that the data are accurate for reporting and research purposes [9]. Most centres in the UK employ a local database manager who has responsibility for working with the surgeons to ensure that data collection is complete and robust. Typically, they will cross-check the local database against operative logs or independent administrative data to check that case ascertainment is complete. In addition, they can utilize local independent sources of mortality tracking to ensure that mortality returns are accurate; monitor data completeness rates for the various database fields and may undergo formal data validation exercises on subsets of records. Effective local data management protocols are a vital aspect of any large clinical registry.

Following local data validation, central data cleaning and external validation are undertaken. Prior to any information being
released into the public domain, summaries of centre- and surgeon-specific data are returned to individual centres for validation. When centres report unexpected statistics, they are resolved by identifying and amending misleading records or improving the data-cleaning algorithms. In Fig. 1, we outline the flow of data between the various organizations involved in the production of the SCTS Sixth National Adult Cardiac Surgical Database Report [18].

**DATA PREPROCESSING**

Real-world clinical registries are inevitably ‘messy’ or ‘dirty’; cardiac registries are no exception [19]. Generally, as the number of records, data fields and complexity of the registry increases, the quality of the data decreases. This ‘messiness’ is defined by an accumulation of transcription errors, logical inconsistencies, missing information, duplicate records and measurement errors. The individual components are caused by multiple data entry points from heterogeneous data sources, and are rarely mutually exclusive. Prospectively, data quality can often be improved via validation checks at the point of entry, provision of comprehensive ‘user guides’ for contributors, technical and clinical help-desks, training, feedback mechanisms and communication plans. In most situations, however, retrospective identification and resolution tools are required.

**Tracing data problems**

Problems with data can arise at any stage of the data collection or processing pathway and may have serious implications. Understanding the causes of data problems is vital if the data are to be correctly utilized. Problems most commonly occur at the data input stage. Data inputted using handwritten data-forms are more likely to contain inaccurate information than software systems that capture the required dataset. Options available for data input and the logic built into each system will vary according to the software developer, and therefore, particular errors are often prone to single centres or clusters of centres. Even with the best-designed data capture systems, human and clerical errors can still occur.

Problems with data accuracy can also arise at other stages in the data pathway from surgeon to researcher. At the central repository, dataset revision followed by centres attempting to retrospectively edit or submit data in a prior format can lead to disastrous results. Human error can also lead to data extractions for researchers being unknowingly corrupted. For example variables that list multiple options separated by a marker might be arbitrarily truncated, meaning that not all data are transmitted. Specific examples of erroneous data from the NACSA include patients who have their heights recorded as negative values (e.g. −160cm), procedures on five valves, deceased patients being discharged home and aortic root replacements being performed on the abdominal aorta.

**CLEANING DATA**

Fundamental to the success of any project reliant on registry data is the ability to clean the data [20]. Simply removing records or fields that do not fully meet standards of accuracy and coherency will lead to an increase in bias and variance for any analysis conducted using the registry [21]. Data cleaning is the process of
detecting and resolving data problems to improve data quality. Appropriate resources must be allocated to this process, which will usually require the attention of experienced clinicians and database managers working in close collaboration.

Erroneous and conflicting data

A primary step is to resolve homonyms and synonyms into a homogenous and research-ready classification system [20]. A common error is to simply input option numbers as opposed to the alphanumeric string in the registries definitions file, e.g. ‘1’ inputted instead of ‘1. Male’ in the gender field. Other simple irregularities included case sensitivity, spelling errors and symbol choice (e.g. using a ‘-‘ instead of an ‘&’ symbol). The most time-consuming element to resolve is where free-text has been used in place of predefined options because in most cases, this will require review by a clinician.

The next step is to resolve numerical conflicts. For example (1) five heart valves repaired in a single procedure; (2) patients heights entered in millimetres; or (3) a patient with 15 grafts for a CABG procedure. In Example 1 there is an error, and in Example 3 an unlikely value. The correct values cannot be ascertained, meaning that the data are effectively missing. For Example 2, the data can be adjusted as the error is traceable.

Some records will contain fields that are in conflict with other fields within the record. For example a patient listed as having off-pump surgery but with a cardiopulmonary bypass time recorded. Conflicts can be resolved using clinical judgment, knowledge of data-inputting processes and other available data in the record. However, it is unlikely that every possible conflict can be identified a priori for large registries due to the sheer number of combinations. Conflicts may often only be identified when a focused project is undertaken on a specific area of the registry.

Duplicate records

It is important that any cleaning process also accurately identifies and removes any duplicate records in the registry [22]. Duplicate records can be the result of multiple data inputs for the same record by different individuals (often with some variables entered slightly differently), software problems or data-upload errors. Identification of duplicate records can be difficult and computationally expensive. If the duplicate identification process is too simple, it may result in the incorrect identification of duplicate records. If the identification process is too complicated, it becomes impractical for routine use by researchers. In surgical registries, records that correspond to a within-admission re-operation for the same patient are at the greatest risk of being incorrectly identified as duplicate records since they will share the same database ‘key’.

Dealing with missing data

Missing data is a common and recurring problem experienced by all clinical database analysts [23]. In addition to true missing data (i.e. no value recorded), variables where an ‘illegal’ or non-discernible input is recorded should also be treated as missing data. For all intents and purposes, the different missingness mechanisms are treated identically. There are a number of approaches for handling missing data including multiple imputation and basal risk imputation. Multiple imputation uses iterated sequential conditional modelling to generate multiple complete datasets. The analyses are combined using a sophisticated methodological framework. Basal risk imputation defaults missing data to the baseline level. While the latter imputation method appears crude and has been demonstrated to be poorly performing in some contexts [24], it is the approach currently used by the NACSA for audit purposes based on expert understanding of the data flow and validation exercises.

USING REGISTRIES

National and international registries can be used for audit purposes, risk-prediction model development, epidemiological research, health services research and hypothesis-driven scientific research. In addition, registries have motivated developments in statistical methodology. As the number of contributors and number of records increase, the value of a registry to research and audit increases. Achieving and maintaining high participation and case ascertainment rates rely heavily on the perceived value of the outputs generated, combined with effective clinical leadership and project management.

Identifying cohorts for analysis

For the purposes of analysis, an entire registry is usually unnecessary as often the focus of an investigation or clinical hypothesis is limited to a specific patient cohort or procedure type. An example of a particular group often used for cardiac surgery audit purposes is patients who have had isolated aortic valve replacement. To identify this cohort, a series of individual filters must be generated and combined using logic rules; they are (i) aortic valve patients; (ii) patients having a single procedure and (iii) valve replacement. Application of these filters may be complicated by conflicting or missing data.

An analysis often requires multiple iterations at this initial filtering and data exclusion stage. The individual components that form an analysis, for example imputation, propensity matching and model building must all be repeated if there is an error in the data inclusion stage. This can be computationally demanding and often wastes considerable time. For this reason, a good knowledge of the registry structure and data-cleaning process combined with careful planning of any analyses is vital for project success.

Risk models

Risk models are commonly developed and validated using registry data and play a vital role in the analysis of registry data for governance purposes. Risk prediction models are well established in cardiac surgery and are most commonly used to estimate a patient’s risk of short-term mortality following surgery [25]. They can be used to inform clinicians and patients about the risk of surgery, to facilitate clinical decision making in the presence of competing treatment options and to risk-adjust clinical governance analyses.
The EuroSCORE is probably the most utilized cardiac surgery risk model across Europe and was developed on data from patients undergoing cardiac surgery in eight European countries in 1995 [26]. EuroSCORE now significantly over-predicts the risk of contemporary cardiac surgery [27] and has recently been superseded by EuroSCORE II [28]. This over-prediction of risk or ‘miscalibration’ is caused by a fall in the mortality associated with cardiac surgery despite a concurrent increase in the EuroSCORE-predicted risk over time.

Calibration drift is important because it can result in patients being quoted inappropriately high risks of surgery that could potentially influence treatment decisions. If calibration drift is not addressed in clinical governance analyses, it will result in falsely reassuring results. Regular validation and, if necessary, updating of risk models using registry data are therefore important [29]. To address this issue, the SCTS has used NACSA data to recalibrate the EuroSCORE models for contemporary surgery [18, 30].

The EuroSCORE is a general model for cardiac surgery developed using data specifically collected for risk model development. A different approach has been taken by the Society of Thoracic Surgeons (STS), who have used their existing registry data to develop-risk prediction models specific to CABG surgery, isolated valve surgery and combined CABG and valve surgery [31]. Procedure-specific models are potentially more useful due to the heterogeneous nature of the general cardiac surgery population. The STS models have also been developed to predict important postoperative morbidity such as sternal wound infection, renal failure and stroke.

Models developed on individual registries often experience problems in gaining widespread external application due to differences in datasets. For example the STS valve model includes a number of variables not recorded in the NACSA. Transitioning from the logistic EuroSCORE to EuroSCORE II will present similar problems for the NACSA [32].

Clinical governance analyses

Clinical governance analysis for cardiac surgery is the means by which centre- and surgeon-specific performances are assessed and areas for performance improvement identified. Published cardiac surgery clinical governance analyses based on registry data are now well established. In the UK, data at the centre level have been published for over a decade and at the individual surgeon level since 2005 [2]. This has been well received by patients, politicians and the media, and a similar process is undertaken by the STS [33] and the New York State Department of Health [34]. Two obvious questions about this process are: (i) Has it improved the quality of care for patients? (ii) Have high-risk patients been denied surgery as a result?

Since the NACSA registry was introduced, risk-adjusted in-hospital mortality in the UK has fallen by >50% despite more elderly and high-risk patients having surgery each year [18]. It is increasingly accepted that the collection and feedback of data to health-care teams, and publishing them openly, is an effective way of driving quality improvement [35]. Despite concerns that publishing mortality results by named centre or surgeon might encourage risk-averse rather than risk-proportionate clinical decision-making [36], evidence is inconclusive. Despite some anecdotal evidence that it does, the magnitude of any effect has not been measureable in the UK to date [37].

Health services research

Health services research relies on clinical data to support national-level clinical guidelines. National registries are ideally placed to provide these data. For example the STS registry was used to illustrate the effectiveness of mitral valve repair for infective endocarditis [38]. Registries can also be longitudinally augmented through record linkage to address highly relevant health service research questions, such as ‘what are the reasons and predictors of readmission following cardiac surgery?’ [39]. This can be used to develop strategies to reduce readmissions and associated health-care costs.

Clinical hypothesis testing

In the era of evidence-based intervention and clinical guidelines, registries provide an invaluable opportunity to test the impact of these on outcomes in the ‘real world’. When there are no clinical trials to recommend treatment methods, which may be a result of the associated expense or due to an incomparable population subset, clinical registries can inform guidelines based on observational studies of outcomes. For less frequently occurring clinical conditions, registries can be used to inform consensus guidelines.

At the most basic level, the format of conducting a clinical hypothesis test using registry data generally entails generating the hypothesis based on expert knowledge; applying filters to the registry to exclude irrelevant records; using statistical methodology to analyse the data subset and test the hypothesis. Research that has stemmed from this type of application of registry data includes assessing the impact of endoscopic vein harvesting on clinical outcomes [40], assessing whether multiple arterial grafts improve late survival of patients undergoing CABG [41], comparing whether drug-eluting stents are safer than bare-metal stents [42]; comparing whether mechanical or biological valve prosthesis choice provides better long-term event-free survival [43].

Surgical epidemiology

Epidemiological techniques can be used to extract patterns and characteristics of surgical indexes within a population from national or regional clinical registries. Established databases can be used to show long-term trends in patient characteristics, demographics and risk factors [27]. Socio-epidemiological analyses have also used linked regional registries collected as part of quality outcomes programmes to assess whether social deprivation is predictive of intermediate-term mortality [44].

Registry data have been used to demonstrate that access to optimal treatment varies considerably with geographic area, for example in the case of mitral valve repair vs replacement in for those suffering from degenerative mitral valve disease [18]. The identification of such health-care inequalities can be used to inform both local and national health-care policy. Another potential application of registry data is to facilitate research into the epidemiology of postoperative infection that could be used to develop optimal prevention strategies [45].
Developments in statistical methodology

The demand for information and increased scrutiny of health-care performance has motivated a number of developments in statistical methodology [46, 47]. Methods include funnel plots for health-care provider performance [48]; hierarchical modelling [49]; predictive modelling [50]; time-series modelling [51]; statistical process control [52] and novel methodology for detection of performance outliers [49]. Methodology for inferring causal effects such as propensity score matching [53, 54] and handling missing data [23] have also been the developed. Understanding the differences between the methodological approaches is important for those using registry data, especially in health-care companion studies, as they can significantly impact on the results and interpretations.

STRENGTHS AND LIMITATIONS: LOOKING TOWARDS THE FUTURE

Administrative data

Clinical registries are considered the gold standard of observational data [55]. However, many medical and surgical specialties have yet to establish adequate national registries. In the absence of clinical registries, administrative data such as Hospital Episode Statistics [56], which are primarily collected for economic purposes, can potentially be used for research or governance purposes. However, such administrative data frequently present an array of problems for such analyses, especially with respect to governance and accounting where the stakes are higher [57]. Reported concerns include the lack of clinically appropriate variables for adequate risk adjustment [58]; the fact that the data may not be representative of the overall population [59]; increased inaccuracies due to coding by administrative rather than clinical staff, particularly with respect to comorbidities and postoperative complications [60]; lack of agreement between analyses [61]. Administrative data can complement clinical data because many socio-demographic variables can be captured administratively and augment clinical registries through record linkage. Linkage has the potential to greatly increase the value of clinical registries [62] and allows the relative strengths of each data source to be assessed [63].

Record linkage

Record linkage will play an important and central role in future registry programmes [64]. Registry data can be linked to other data sources such as administrative databases, official census records (e.g. Office for National Statistics), national medical council databases (e.g. General Medical Council) and other associated registries to enhance the scientific value of the registry. Linkage can be either probabilistic or deterministic, and this is determined by the available data in each database [65].

As each patient in the UK has a unique NHS number, it is reasonably straightforward for the NACSA to link patient records across registries using a deterministic matching procedure [66] to access long-term mortality-tracking data. Such linkage may not be so straightforward in other health-care systems. If record linkage requires the use of information such as name, address and date of birth, then in principle, deterministic matching should be straightforward. However, matching is made more complicated due to phonetic and typographical differences as well as missing data. For example ‘John Smith 01-04-1954’ would not deterministically match to ‘J. Smith xx-04-1954’. Matching on surname and year of birth (i.e. ‘Smith’ and ‘1954’) would probably bring up multiple matching records in this case. A series of deterministic matching rules can be created; however, some researchers are exploiting probabilistic record linkage techniques that can potentially improve the handling of slight discrepancies [67].

Although record linkage opens up significant research possibilities, it also brings new challenges and legal obstacles to those entrusted to protect and uphold the privacy and ethical uses of the data [68]. For example the identification of records is difficult in a single registry due to the limited data collection and the pseudo-anonymization of personal data. However, when linked to one or more other databases, the increase in data leads to an inevitable increase in the chance a record could be matched to the wrong individual or other marker, or patients could be identified.

Registry data vs randomized clinical trials

Although considered the gold standard of observational data, registry data are often considered second best to randomized clinical trial (RCT) data in health-care assessment. However, each data source has different strengths and weaknesses [69].

The main advantage of RCT data is that confounding variables that might lead to biased treatment effects have been formally accounted for by randomization. On the other hand, as registries contain observational data, inferring causal effects is difficult due to inherent biases stemming from confounding variables. Although there have been developments in statistical methodology, a registry is not a replacement for a well-designed RCT. The limitations in the collection of registry fields listed mean that there might be unmeasured confounders and potential selection bias [70].

Two primary limitations of RCTs are (i) the expense and resource requirements to run them and (ii) the often narrowly focused patient cohorts included in RCTs may not reflect the general target population [1]. In contrast, registries feature observational data that reflect common practice and the general target population [5]. Registries can be used to either validate or augment RCTs [71]. Registries can also be superior to RCTs for monitoring long-term outcomes that exceed the study window of a trial.

Device monitoring

Over recent years, there has been an increase in the number of devices implanted into patients. The now defunct UK Heart Valve Registry was initially set up after problems occurred in identifying patients implanted with faulty heart valves [72]. The transcatheter aortic valve implantation registry has recorded all procedures in the UK since the introduction of the technique in the UK in 2007 [73].

A recent high-profile case that demonstrates the importance of device-monitoring registries is that of Poly Implant Prothèse breast implants, which were shown to have a significantly higher rupture rate than other implants, detectable after ~5 years [74]. Registries for devices such as these would lead to earlier detection of unacceptable failure rates by health-care authorities.
Commercial value

Registry projects, particularly those on a national and international scale, require considerable resources, infrastructure and sustained funding if they are to deliver long-term benefits. Funding for registries can come from a variety of sources including government budgets, professional societies and local health-care commissioners. It makes sense that the value of the data housed in these registries be exploited as a source of revenue. The STS have identified two revenue sources for their national database: (i) non-funded major or minor data requests and (ii) regional activities [8]. The first source allows for researchers to access information from the database. The second would allow for regional governments to access high-quality reports in order to steer health-care policy.

SUMMARY

The success of a clinical registry project can be measured on the database completeness, accessibility of information and proven usefulness [11]. Benefits include improvements in informed patient decision making, improvements in treatment and advances in health-care research and governance. In this review, we have discussed some of the associated fundamentals that reflect our experience in setting up and managing a cardiac surgery clinical registry.

Cardiac surgery databases have been established worldwide and are growing in number and size, and will have increasingly important uses locally and nationally for the various purposes described above. Well-configured clinical registries give some clear advantages over administrative databases and maximum information will come from the linkage of clinical databases with other sources of information. The EACTS Database report published in 2010 included data from over 1 million patients from 29 countries [75]. As registries increase in size, become more commonplace and are used for increasingly important purposes, it is vital that they are established and maintained appropriately. It is imperative that any EACTS registry is compliant with all appropriate National, European and International legislation (for the membership outside the EU). This will require the central registry administration to be clear on how the laws affect them, and it will also require each national group to be aware of the relevant aspects that govern them locally and to give appropriate consent for holding and using the data prior to uploading data. For registries to be effective, dedicated clinical input alongside high-level analytical and data management expertise is required. EACTS will be supporting national registries in achieving this through the forthcoming QUIP programme.

ACKNOWLEDGEMENTS

We acknowledge the European Association for Cardio-Thoracic Surgery (EACTS) Quality Improvement Programme (QUIP) for their role in the European registry project. This review has been configured to support the EACTS QUIP. We are grateful to the two anonymous reviewers and the Associate Editor for their comments, which helped us improve this manuscript.

Funding

This research was partly funded by Heart Research UK (Grant number RG2583). The National Adult Cardiac Surgery Audit is commissioned and funded by the Healthcare Quality Improvement Partnership (HQIP).

Conflict of interest: none declared.

REFERENCES


G.L. Hickey et al. / European Journal of Cardio-Thoracic Surgery
Clinical registries: yes, but then appropriately!

Hendrik Tevaearai and Thierry Carrel*

Clinic for Cardiovascular Surgery Inselspital, Berne University Hospital and University of Berne, Berne, Switzerland

* Corresponding author. Clinic for Cardiovascular Surgery, Berne University Hospital, 3010 Berne, Switzerland. Tel: +41-31-6322111; e-mail: thierry.carrel@insel.ch (T. Carrel).

Keywords: Clinical registries • Randomized clinical trials • National database • Results of cardiac surgery procedures

In the last two decades, several cardiac surgical procedures have become standard, and the results can be considered good or even excellent. Yet, unless a procedure is considered 100% safe and efficient, cardiothoracic surgeons should not be satisfied with a ‘good enough’ mentality, but must perpetually strive for further improvement. One critical limitation in this endeavour is the assessment of the results, because a significant improvement in morbidity and mortality rates may be difficult to detect if these end-points already occur at very low rates. Randomized controlled trials (RCTs) remain the gold standard for reaching Level 1 of evidence. By carefully selecting the end-points as well as the inclusion and exclusion criteria, RCTs aim to rapidly provide solid information regarding a new therapeutic approach (percutaneous intervention, surgical procedure, device or drug performance) that may eventually be further extrapolated to a larger (or more global) population. However, results of RCTs may be biased because of the selection of ‘best-performer’ institutions and, therefore, that of such powerful studies may not adequately represent the average level of care. This is where registries may give a better picture of the real world: but, if poorly conducted, they may also lead to a distorted understanding of the reality. In the current issue of the European Journal of Cardio-Thoracic Surgery, Hickey et al. review various critical aspects of administering a registry [1]. The authors highlight the growing interest in registries and the potential they offer for future clinical, but also socio-economical, developments.

In its ideal form, a registry should provide enough evidence to support the development of individuals or collectivities and to reflect the changes in performances. Indeed, one critical asset of registries is the principle of reciprocity, requiring each participant to comprehensively and unreservedly include all of his/her data. In exchange, all participants should be granted access to a larger source of information, by means of which they would be encouraged to learn from each other and to improve the quality of their own institution. The unrestricted use of these large databases will, however, remain sub-optimal as long as not all relevant data of all consecutive patients constituting the specific registry (transcatheter aortic valve implementation, coronary artery bypass grafting and others) are included. Geographically, registries should also be as broad as possible or at least be conceived so that different entities are compatible (same criteria and definitions) and can be technically merged into one larger database. Finally, registries should allow for longitudinal analysis in order to provide results over the long-term [2, 3].

To achieve this ambitious but inevitable objective, critical rules must be discussed, accepted and sooner or later implemented. Some already exist and are obligatory, such as the legal and ethical structures (which may vary from country to country), and others are suggested by guidelines or recommendations, such as the way data should be managed, but several are yet to be clarified. Among them, the quality of the ‘raw data’ is probably the most critical issue. Currently, and with only few exceptions, submission of data is facultative, and data are not subjected to obligatory third-party checks (independent audits). Inconsistencies, missing information, duplication or transcription errors, among others, can be checked and corrected after they have been submitted to the registry; however, the truth behind submitted data remains, for most centres, a matter of trust. One major exception—as an example—is the registry of the European Congenital Heart Surgery Association (ECHSA), running for more than 10 years, and that includes independent audits to allow institutional certification.

Typically, the surgeons themselves or locally dedicated persons (database managers) with varying experience and skills manage their own institutional database and, most of the time, there is no local audit or control process defined. One can reasonably suspect...