

# National quality registries: how to improve the quality of data?

Fieke Hoeijmakers<sup>1,2</sup>, Naomi Beck<sup>1,2</sup>, Michel W. J. M. Wouters<sup>1,3</sup>, Hubert A. Prins<sup>1</sup>, Willem H. Steup<sup>4</sup>

<sup>1</sup>Scientific Bureau, Dutch Institute for Clinical Auditing, Leiden, The Netherlands; <sup>2</sup>Department of Surgery, Leiden University Medical Centre, Leiden, The Netherlands; <sup>3</sup>Department of Surgical oncology, Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands; <sup>4</sup>Department of Surgery, Haga Hospital, The Hague, The Netherlands

*Contributions:* (I) Conception and design: F Hoeijmakers, N Beck, HA Prins, WH Steup; (II) Administrative support: F Hoeijmakers, N Beck; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: F Hoeijmakers; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*Correspondence to:* Willem H. Steup, MD, PhD. Member Data Verification Committee, Dutch Institute for Clinical Auditing, Leiden, The Netherlands; Department of Surgery, Haga Hospital, Els Borst-Eilersplein 275, 2545 AA, The Hague, The Netherlands. Email: w.steup@hagaziekenhuis.nl.

**Background:** Data of quality registries are increasingly used by healthcare providers, patients, health insurance companies, and governments for monitoring quality of care, hospital benchmarking and outcome research. To provide all stakeholders with reliable information and outcomes, reliable data are of the utmost importance.

**Methods:** This article describes methods for quality assurance of data—used by the Dutch Institute for Clinical Auditing (DICA)—regarding: the design of a registry, data collection, data analysis, and external data verification. For the Dutch Lung Cancer Audit for Surgery (DLCA-S) results of data analysis and data verification were assessed with descriptive statistics.

**Results:** Of all registered patients in the DLCA-S in 2016 (n=2,391), 98.2% was analysable and completeness of data for calculations of transparent outcomes was 90.7%. Data verification for the year 2014 showed a case ascertainment of 99.4%. Of 15 selected hospitals, 14 were verified. All these hospitals received the conclusion ‘sufficient quality’ on case ascertainment, mortality (0% under-registration) and complicated course (3.3% wrongly registered complications). One hospital was not able to deliver patients lists, and therefore not verified.

**Conclusions:** Quality of data can be promoted in many different ways. A completeness indicator and data verification are useful tools to improve data quality. Both methods were used to demonstrate the reliability of registered data in the DLCA-S. Opportunities for further improvement are standardised reporting and adequate data extraction.

**Keywords:** Clinical audit; quality assurance; data verification; data accuracy; lung cancer

Submitted Dec 11, 2017. Accepted for publication Apr 18, 2018.

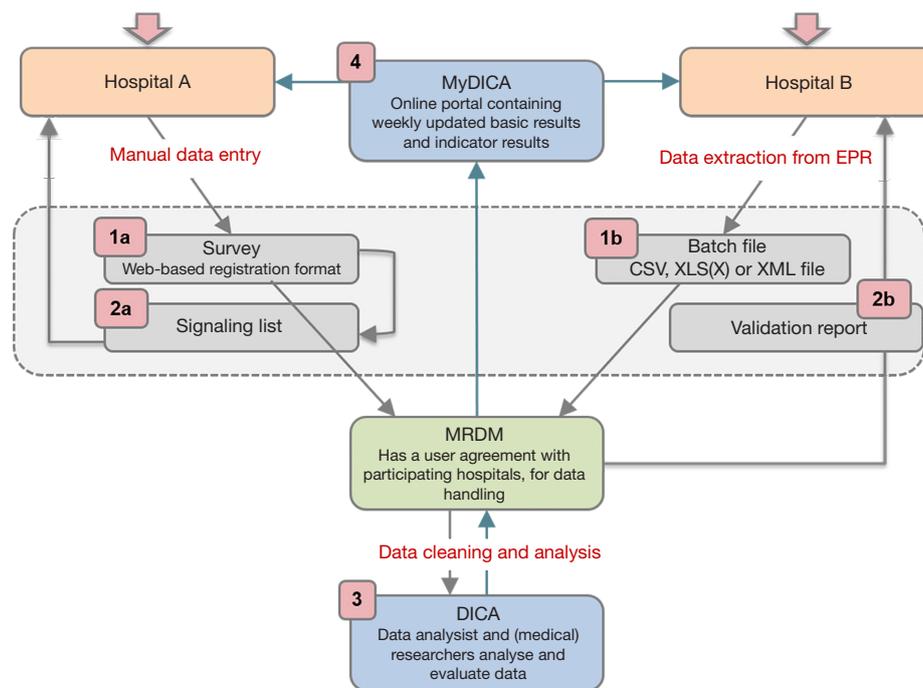
doi: 10.21037/jtd.2018.04.146

View this article at: <http://dx.doi.org/10.21037/jtd.2018.04.146>

## Introduction

The interest in collecting data to monitor quality of healthcare continues to grow. National disease-specific quality registries are operational worldwide (1-6). Healthcare providers use information from these registries, to monitor and benchmark their quality of care and to implement improvement initiatives. The demand for data and transparency on provider-specific results also

increases strongly among other healthcare stakeholders, like patient organisations, healthcare inspectorates, and health insurance companies. Decision support systems for patients are built on these data and in some countries, health insurance companies use comparative data from registries when purchasing healthcare services. In order to maintain affordable healthcare, governments increasingly focus on value-based healthcare and pay-for-performance delivery models. Furthermore, data are used for outcome



**Figure 1** Data collection and feedback processes of DICA registries. 1a, in survey patients are registered one by one by the registrar. Data quality is facilitated directly by required items, conditions, validations and help texts. 1b, by batch processing, hospitals can extract data from EPR and deliver it in ‘bulk’. 2a, a real-time signalling list shows all registered patients for which data are missing or erroneous. 2b, in the validation report hospitals get feedback on their submitted batches, both on patient level as in summary. 3, DICA is responsible for data cleaning, analysis, and feedback of results to hospitals and the scientific committee. 4, in MyDICA, accessible by doctors and hospital boards, weekly updated results are shown in tables and funnel plots. EPR, electronic patient records; MRDM, medical research data management.

research and offer insight into (variation in) everyday clinical practice. These ‘real world’ insights can be of great value in addition to evidence from randomised clinical trials, in which not all subgroups of patients are represented (7). When data are analysed for all of the above-mentioned reasons, reliable data are imperative. Therefore, completeness and accuracy of data in (national) quality registries is an important topic, which should have primary attention.

In the Netherlands, the Dutch Institute for Clinical Auditing (DICA, founded in 2011) is one of the leading organisations in clinical auditing. It facilitates over 20 clinical audits, which number is continuously increasing. DICA is a non-profit organisation, driven by medical professionals, funded by the umbrella organisation of 10 Dutch Health Insurance Companies (ZN) (8). Its main aim is to give doctors and professional associations insight into their quality of care, thereby providing improvement potential (9,10). Over the years, much experience has been gained in improving and verifying data quality. Nowadays,

DICA puts several methods for quality assurance into practice. Data verification is one of these methods and has been described in the literature (2,5,7,11,12). However, about other approaches for quality assurance on the level of national registries, little has been written so far.

This article provides an overview of methods used by DICA to ensure data accuracy and to improve data completeness, with the Dutch Lung Cancer Audit for Surgery (DLCA-S) as a case study. Examples from the DLCA-S, as well as the results of the data verification process of this registry, will be described and discussed.

## Methods

In order to obtain complete and accurate data, various checks and procedures are part of the registration and feedback processes of DICA facilitated registries. *Figure 1* and *Table 1* demonstrate an overview of data collection and feedback processes, including these checks and procedures. In the following sections, these methods will be explained

**Table 1** Steps in data verification and the responsible parties

Steps in the data verification process	(Intended) result	Implementation/responsibility
1. Coordination and support	A coordinated verification process	DICA data verification department
2. Patient and variable selection	Verification items and criteria	Data verification committee
3. Inviting and selecting hospitals	Random sample of hospitals to audit	DICA data verification department
4. Training of data managers	Competent data managers	DICA and scientific committee
5. Providing patient lists and access to (E)PR	List with all patients possibly fulfilling inclusion criteria	Hospitals
6. Providing hospital audit data on a secured laptop	A web-based survey prefilled with registered data by the hospital	MRDM
7. Actual data verification	Results on variable level: 'not-discrepant' or 'discrepant'	Independent data managers
8. Analysis and processing of results	Discrepancy report	DICA data verification department
9. Interpretation and response to discrepancies	Updated discrepancy report	Hospitals
10. Final conclusion	Final report and conclusions	Data verification committee

(E)PR, (electronic) patient records; MRDM, medical research data management.

in more detail and will be supported with examples from the DLCA-S. The design and organisational structure of the DLCA-S—as a part of the Dutch Lung Cancer Audit (DLCA)—are described elsewhere in this issue by Beck *et al.* (13).

### Definitions

Critical issues in data collection are clear definitions and the adequate translations of data from electronic records (EPR) or patient records (PR) to these definitions. Influence on data quality therefore already starts before data collection, with specific and consistent data definitions, provided by DICA, and uniform documentation by hospitals. These issues should already be considered in the development phase of a registry, before the data collection—as demonstrated by *Figure 1*—starts.

The scientific committee of each DICA audit, consisting of clinicians mandated by their medical professional associations, is responsible for the content of the audit. This includes clear and unambiguous definitions of the items in the dataset. Detailed descriptions of all registered variables are collected in the data dictionary, of which the most recent version is freely available online (14). Harmonisation of variables that appear in multiple audits, such as comorbidity items, is supported and actively facilitated by DICA.

DICA currently is involved in several initiatives to extract data from existing databases. Automated extraction

of data from the Dutch national pathology registry (PALGA) and linkage to the DICA registries is already functional for several other oncology audits (15). Again, careful matching of definitions and standardisation is key in order to achieve accurate and complete data.

### Data collection

Data collection in DICA facilitated audits is usually performed by doctors, physician assistants, (research) nurses, quality control co-workers or data managers. The responsibility for accuracy and completeness of collected data though, rests with the physician. Roughly, there are two ways of data collection: by using 'Survey' or by 'Batch processing' (visualized as step 1a and b in *Figure 1*). Both methods will be explained in more detail.

#### Data collection—survey

Data can be collected by using a secured web-based survey. This is a pre-defined submission format with several integrated feedback mechanisms, ensuring standardised data. It is designed in a collaboration between the scientific committees, DICA and Medical Research Data Management (MRDM). MRDM is a third trusted party who processes and de-identifies the data before they are received by DICA, so data can never lead to identification of individual patients. MRDM, as a data processor, complies with Dutch and European privacy laws and is NEN 7510:2011 and ISO

**A. Example of a required item**

**!** Which procedure was used to remove the largest quantity of lung parenchyma?

(Pleuro)Pneumonectomy

Bilobectomy

Lobectomy

Anatomic segmental resection

Wedge resection

Exploratory thoracotomy

Other

This is a required field

**B. Example of a condition**

**!** Was a mediastinoscopy performed?

No

Yes

Unknown

**Mediastinoscopy**

Give this value to all fields below, which are not yet filled in

Biopsies taken of lymph node station 2 left

Positive pathology of lymph node station 2 left

**C. Example of a validation**

**!** Predicted FEV1 % (as a percent of predicted)

The value entered must be greater than or equal to 0 and less than or equal to 150

**D. Example of a help text**

**!** Reason for conversion

Reactive conversion

Strategic conversion

There is a distinction between strategic and reactive conversions. One is talking about a strategic conversion when, during the operation, the decision to convert the surgical approach is made due to impracticability via the primary approach. One speaks of a reactive conversion, when it is necessary to convert the surgical approach because of a complication.

**E. Example of the signalling list**

Signalling list DLCA-S						Log out	
Case	Last modified	Validated	Warnings	Errors	Filters		
			Registration year	2017	Number per page	10	<input type="button" value="Apply filter"/>
Patient Xxxx-12-04-1967	31-10-2017 10:25:18	31-10-2017 10:25:21	1	2		<input type="button" value="View"/>	
Patient Yyyy-07-06-1944	07-11-2017 11:02:45	07-11-2017 11:02:46	2	1		<input type="button" value="View"/>	
Patient Zzzz-30-11-1958	30-11-2017 16:33:08	30-11-2017 16:33:09	2	2		<input type="button" value="View"/>	

**Figure 2** Methods for quality assurance during data entry, applied in the web-based Survey of the Dutch Lung Cancer Audit for Surgery (DLCA-S).

27001:2013 certified (16,17).

In the web-based survey (step 1a in *Figure 1*), data quality is already ensured during data entry, by:

- (I) Required items (*Figure 2A*): variables considered important for the data reliability or calculation of quality indicators by the scientific committee or DICA. For these items, error signs appear when data are missing.
- (II) Conditions (*Figure 2B*): built-in restrictions which allow specific questions only to pop-up based on answers to a prior question. This ensures the registrar to take the right steps, makes the registration process logical to pass and saves time.
- (III) Validations (*Figure 2C*): provide direct feedback on

the content of the filled fields. When an impossible or unlikely value is entered, a warning pops up.

- (IV) Help texts (*Figure 2D*): ensure uniform data collection across all hospitals. Help texts with unambiguous definitions are stated by the scientific committee and are added where necessary.

All notifications of missing or incorrect data are collected on a patient level and bundled on a signalling list for all patients registered by a hospital (step 2a in *Figure 1*). An example of the signalling list is shown in *Figure 2E*. From this list, a direct link to the specific error in the survey is available, so errors can be checked and rectified easily.

Every year, the functionality and user-friendliness of Survey are reviewed and, when necessary, improved.

### Data collection—batch processing

The second option for hospitals to submit data is by batch processing (step 1b in *Figure 1*). Batch processing allows the user to deliver data ‘in bulk’ (e.g., for a large number of patients at once), instead of manually entering patients one by one in the web-based survey. Users can extract data directly from specific fields in the EPR and collect them in a batch file. The matching of EPR data and audit items is of great importance for the way in which data are entered in a batch file. Because when mismatches occur, data (of all patients) are potentially incorrectly interpreted. Therefore, before a hospital can start with batch processing, a test run is always performed by MRDM. Batches are tested both technically and substantively. When desired by the hospital, MRDM can visit to help matching audit items to the right source in the EPR. A successfully completed test batch procedure is required to start batch processing (step 1b in *Figure 1*).

Once this process has been implemented in the hospital, a batch file can easily be created, with a substantially reduced registration load as a result. Another advantage of batch processing is that data are extracted directly from their source. Misinterpretation of medical record data by a registrar, or mistakes in the translation of these data to the survey questions, is thereby minimised. On the other hand, a registrar can also be of value by noticing missing or ambiguous data in the (E)PR and reporting this to the responsible healthcare provider.

Every uploaded batch is validated and a validation report is generated by MRDM. Similar to Survey and its signalling list, this validation report includes an error overview with more detailed information on missing data, conditional data fields and validations (similar to Survey). Hospitals verify and rectify their submitted data based on this report and provide an adjusted batch file.

For hospitals using batch processing to deliver their data, a maximal lag time between patient inclusion and batch upload of 3 months is required.

### Data cleaning, analysis and feedback

After processing, data arrives at DICA completely anonymised on patient level. When analysing these data, data cleaning is carried out, to remove unreliable records (step 3 in *Figure 1*). Recorded patients are only included for analysis if their data are ‘analysable’, which means that a minimal set of critical items (as defined by the scientific committee) is adequately registered. For the DLCA-S these items currently are: date of birth, date of surgery, type of

surgery and survival status 30 days postoperatively.

Clinicians receive weekly updated feedback on their registered patients for the use of clinical auditing: number of patients, patient characteristics and outcomes on quality indicators. This can help to identify errors early (step 4 in *Figure 1*). To allow hospitals to check their results, DICA is transparent in the methods of quality indicator calculations. These calculations are accessible for all hospitals through the MyDICA portal. In addition, a service-desk is available during working hours for all questions, including those regarding the calculations.

Next to the process and outcome indicators, most DICA clinical audits have a ‘completeness indicator’. This reports the percentage of registered patients in which all items required for calculations of transparent quality indicators are registered. This provides the hospitals with direct feedback on the quality of their registered data. Practice learns that this ‘feedback information’ motivates hospitals to improve quality of the registered data and therefore the reliability of the audit as a whole.

### External data verification

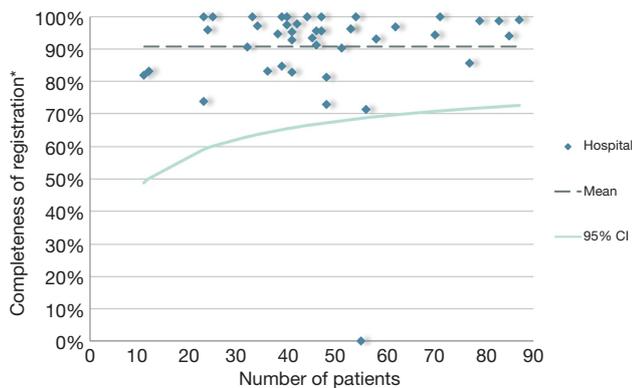
The last step in achieving insight in data quality and to promote accurate data capture is external data verification. Here this process is described in general, supported by a step-by-step overview in *Table 1*. The detailed specifications of the process of data verification of the DLCA-S are explained in the results section.

### Verified items

Patient inclusion criteria and a sample of audit items to be verified are set and selected by representatives of the scientific committee, supported by an independent verification committee existing of two independent clinicians, one expert clinician, a biostatistician, a deputy of the Dutch Healthcare Inspectorate (IGZ), and a deputy from the Dutch patient federation.

### Verification process

DICA’s data verification department coordinates the process of external data verification, though the process itself, for privacy reasons, is carried out by MRDM. All registering hospitals in the audit receive an invitation to participate, free of charge and on a voluntary basis. In selected hospitals, case ascertainment is evaluated by comparing registered patients to a patient list derived from the hospital’s administrative system. To access accuracy, data registered



**Figure 3** Funnel plot with the hospital results on the ‘completeness indicator’ of all hospitals registering in the Dutch Lung Cancer Audit for Surgery (DLCA-S) in 2016. \*, completeness of registration: ‘the percentage of registered patients in which all items required for calculations of transparent quality indicators are registered’.

in the audit are manually compared on a patient level with the information in the (E)PR, using a web-based survey prefilled with registry data. This on-site verification process is carried out by independent data managers provided by another third party. They are trained by DICA, though perform the verification within the privacy domain of MRDM. The data managers report their findings to DICA. Data managers of the data verification department analyse these data and feedback results to the visited hospitals in a ‘discrepancy report’. Hospitals can check the detected discrepancies and respond accordingly within 3 weeks. Based on the appeal of hospitals, necessary alterations in the discrepancy report can be made, or a statement can be added.

#### Assessment of verification results

The independent verification committee ultimately draws a final conclusion by judging the discrepancy report and the possible appeal of the hospital. This results in a final conclusion, for each hospital, of ‘sufficient quality’ or ‘insufficient quality’. Hospitals receive a final summary report with the detailed results, to learn from the discrepancies and to help them optimise their registration procedure.

## Results

Results of data analysis and external data verification of the

DLCA-S are discussed as a case study here. As one of the longest existing registries, the DLCA-S is a mature audit and representative of DICA’s approach.

#### Analysability and completeness of DLCA-S data

The DLCA-S was started in 2012 and from 2015 all Dutch hospitals that perform lung cancer surgery, register their performed procedures for malignant and benign lung disease. In 2016, 2,391 patients with proven or suspected lung cancer were registered, of which 98.2% (n=2,349) met the criteria of analysability. Most patients underwent parenchymal resection for non-small cell lung cancer (NSCLC; n=1,839). (13) The DLCA-S has transparent quality indicators mostly on non-small cell lung cancer (NSCLC). Therefore, the completeness indicator is described as follows: “The percentage of patients undergoing surgery for NSCLC and registered in de DLCA-S, in who all items required for calculations of the transparent indicators are complete.” The specific items required for this calculation can vary over the years because new quality indicators are developed and others are dropped. These developments are based on new techniques and treatments, growing insights in clinical practice and decreasing inter-hospital variation over the years. Over 2016, the completeness of items used for quality indicator calculations was 90.7%. Between hospital variation in this completeness is shown in *Figure 3*. In 2012, 2013 and 2014 the percentages of completeness were respectively 97.9%, 99.4% and 99.3%. In 2015 it dropped to 88.3%, which can be explained by a stricter definition of the indicator.

#### Specification and outcomes of 2014 external data verification of the DLCA-S

In 2016, external data verification for the DLCA-S took place for the first time. Data from the year 2014 were checked and compared to the (E)PR data.

#### Verified items

The scientific committee and verification committee agreed on verification of case ascertainment and all variables required for the outcome indicators ‘mortality’ and ‘complicated course’, for patients undergoing surgery for NSCLC in 2014. Criteria for ‘sufficient quality’ were determined upfront and defined as follows:

- ❖ Case ascertainment: quality is sufficient when at least 98% of patients who fulfilled inclusion criteria, are

registered in the DLCA-S.

- ❖ Mortality (defined as ‘the percentage of patients that died within 30 days after surgery or during primary hospital admission’): quality is sufficient when no patients who, according to the inclusion criteria of the registration, should have been registered and who also are deceased, are not registered or registered as deceased.
- ❖ Complicated course (defined as ‘mortality within 30 days after surgery or during hospital stay, a performed re-intervention, or a length of hospital stay of >14 days in combination with one or more complications’): quality is sufficient when less than 5% of the possible number of complications, of all patients who should have been registered according to the inclusion criteria, were not recorded.

### Verification process

From the 43 participating hospitals in the DLCA-S, 29 hospitals (67%) voluntarily signed up. This is in line with other external data verification projects performed by DICA, where participation ranges from 60–88% (12). By random selection 15 of 29 hospitals were chosen to participate. These hospitals were asked to prepare and deliver the patient list of all patients operated for NSCLC in the year 2014. One of the hospitals could not fulfil this criterion and therefore received a final conclusion of ‘insufficient quality’.

Next to verification of case ascertainment in the registry (for all patients), in the 14 remaining hospitals at least 25 patients were verified on accuracy of data. When over 25 patients were registered per hospital, a random sample of patients was taken. The size of the sample was calculated by the formula:  $25 + (n-25)/10$ . On average 26 patients per hospital were verified, which makes the total number of verified patients 388.

### Assessment of verification results

All 14 hospitals got the judgment ‘sufficient quality’ on all three subcategories: case ascertainment, mortality and complicated course. Results on a patient level are shown in *Table 2*. Five patients (0.6%), distributed over 14 hospitals, fulfilled inclusion criteria but were not registered. Overall case ascertainment for the verified hospitals was therefore 99.4%. For none of the hospitals, this was less than 98% of the includable patient. Of the 5 patients wrongly not registered, 3 had a complicated course and none was deceased. All registered mortality statuses were filled in

correctly. Of 7 patients (1.8%) the mortality status was not registered at all, none of these patients were deceased. Of all recordable complications, 3.3% were not or were wrongly registered.

### Discussion

When reliable, the opportunities for the use of quality registry data are immense. Data from national quality registries are used for evaluating and improving quality of care, outcome research, evaluation of (new) treatments and to gain insight into the most cost-effective treatment approaches based on the value-based healthcare principles (7,18). Quality assurance and verification of data is therefore of utmost importance.

As demonstrated in this article, DICA promotes data quality in several ways. Regarding the DLCA-S, analysability of data is, with 98.2%, high. Despite the stricter criteria for data completeness since 2015, completeness of DLCA-S data in 2016 was over 90% and again increasing, which endorses the value of this indicator for improving data quality. Because of the specific definitions used, these percentages are difficult to compare to other registries—both other DICA registries and registries on lung cancer in other countries (19,20). However, other DICA registries show increasing completeness of data up to 97% (21). This suggests there is improvement potential left for the DLCA-S.

External data verification showed that the quality of data in participating hospitals was high for the verified patients and items. From the 829 patients in 14 hospitals which fulfilled inclusion criteria, only 0.6% was missing in the registry. And although these results might not be (fully) representative for hospitals that did not sign up for external data verification, this implies high case ascertainment in the DLCA-S. This high level of case ascertainment, also in comparison to other countries, is an important strength of the DLCA-S (2,19,20). It emphasizes the reliability of data and the generalizability of outcomes for daily medical practice.

Data verification provides reliable information on completeness and accuracy of data used for quality indicator calculations, such as for mortality and complicated course. It can prove the robustness and reliability of indicator outcomes, and it is only then that transparent outcome information can be valued properly. Information which is highly important for all stakeholders, because hospital policies may be changed and (management) decisions may

**Table 2** Results of external data verification over 2014 of 15 hospitals registering in the Dutch Lung Cancer Audit for Surgery (DLCA-S)

Verified items	N (%)
<b>Case ascertainment</b>	
Number of patients verified on presence in the DLCA-S	829 (100.0)
Patients verified and registered	824 (99.4)
Patients verified and wrongly not registered	5 (0.6)
Wrongly not registered and deceased	0 (0)
Wrongly not registered and with a complicated course	3 (0.4)
Total number of patients registered by selected hospitals	824
<b>Mortality</b>	
Number of patients in sample test	388 (100.0)
Mortality status according to EPR	381 (98.2)
Mortality status not according to EPR	7 (1.8)
According to EPR 'deceased' but registered as 'alive'	0 (0)
According to EPR 'alive' but registered as 'deceased'	0 (0)
According to EPR 'alive' but no value registered	7 (1.8)
<b>Complicated course</b>	
Number of possible complications <sup>^</sup>	6,596 (100.0)
Number of complications according to EPR	6,380 (96.7)
Number of complications not according to EPR	216 (3.3)
Complications not registered ('no' or no value)	179 (2.7)
Complications wrongly registered (wrong value)	37 (0.6)

Case ascertainment: agreement between patients on a list extracted from the hospital's administrative system and patients registered in the DLCA-S. Mortality: mortality within 30 days after surgery or during hospital stay. Complicated course: mortality within 30 days after surgery or during hospital stay, a performed re-intervention, or a length of hospital stay of >14 days in combination with one or more complications. <sup>^</sup>, possible complications: number of patients in sample test (n=388) × the number of registrable complications in each patient (n=17). EPR, electronic patient records.

be made based on these outcomes. Previous research showed that transparency of outcome data contributes significantly to the effectiveness of auditing and feedback (22).

Besides giving insights, data verification also improves data quality in several ways. First, feedback from the discrepancy report learns the hospitals to understand the pitfalls in their registration process, resulting in improvement actions. Second, the final report often reveals ambiguities and inaccuracies in the registered variables. For example, after the DLCA-S data verification, the variable 'intensive care unit (ICU) admission after surgery' turned out to be multi-interpretable. Some hospitals registered only 'ICU-days because of complications', while other registered also 'planned ICU-days for observation'. The

scientific committee then specified the definition. Third, the performance of external data verification—even when not mandatory—increases the awareness of hospitals about the importance of accurate data registration. This might result in more thorough data collection in hospitals because registrars are aware of the possibility of verification. For all these reasons, data verification is an important tool that will improve the quality of the registrations and collected data in the future.

Regarding the DLCA-S was concluded that completeness, mortality and complicated course can be verified effectively and efficiently. Ideally, in future verification projects, more quality indicators and casemix factors will be verified. However, external data verification is a labour-intensive,

and therefore costly, project. To expand in the future, more efficient ways of data verification are needed.

Despite all described implemented procedures, there are still steps to be taken to improve data quality. Registration burden is a significant challenge. Hospitals are responsible for data registration themselves, which is a time-consuming process for doctors, physician assistants or (administrative) nurses. Lack of time can lead to registration mistakes and under-registration.

Ideally, data used for quality registries are extracted directly from the EPR, requiring healthcare professionals to register a patient and its treatment characteristics only once. This would also avoid misinterpretation of source data by a registrar and might even make extensive data verification processes unnecessary in the future. However, in practice, this is challenging, because of multiple different (E)PR systems and lack of uniform registration of data in (E)PR.

Even between the different DICA registries, there is no uniformity yet. Over the years DICA has gained experience in data collection and by having the clinician in the lead, clinically relevant and clearly defined variables were selected and evaluated for each registry. The reverse side of this is, that without adequate coordination between different scientific committees, definitions discrepancies between registries occurred. With an increasing number of registries over the last years, uniformity is needed. DICA recently started to harmonize definitions between all registries where possible.

### *Future perspectives*

Important challenges for the (near) future are (I) uniform definitions for data recording at the point of care, and (II) efficient extraction of these data from EPR or other structured sources.

To address these issues and to improve future data quality, DICA already works on several projects, often in collaboration with hospitals, Health Insurance Companies and other stakeholders as the Dutch Federation of University Hospitals (NFU) and the national centre of expertise on EHealth NICTIZ. For standardised terminology, in these projects 'building blocks' (e.g., agreements on registration of a (medical) concept, such as a diagnosis or a medical action) and standardised multidisciplinary meeting and surgery reports are developed (23).

Clearly, this is not only a problem DICA is facing. Worldwide these issues are addressed and plenty organisations are making big efforts to facilitate efficient and reliable data

capturing. Initiatives and organisations as ICHOM, CDISC, and ISPOR respectively define standard sets of outcomes per medical condition, develop data standards to allow data 'to speak the same language' and promote health economics and outcomes research excellence by education and publishing research tools (24-26). These are only a few examples, which underline the need for standardised data capturing.

With more integration and uniform definitions, collaborations with other initiatives, such as the ESTS database, will become easier. The thereby created opportunities for international comparisons of outcomes can lead to critical analysis of differences, which will contribute to higher quality of care in thoracic surgery.

### **Conclusions**

For DICA, quality of data is of major importance. This article describes how this is currently addressed in: the design of a registry, data collection, analysis of data and by external data verification. By the various procedures to achieve complete and accurate data DICA aims for high-quality data for hospital benchmarking and outcome research.

### **Acknowledgements**

The authors thank all clinicians, physician assistants, nurses and data managers that register patients in the DICA databases, as well as the DLCA-S scientific committee and the data verification committee.

### **Footnote**

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

*Ethical Statement:* For analysis of the results of external data verification, DICA only received on patient level anonymised data from MRDM. In accordance with the regulations of the Dutch Central Committee on Research involving Human Subjects (CCMO), this type of study does not require approval from an ethics committee in the Netherlands.

### **References**

1. Jakobsen E, Palshof T, Osterlind K, et al. Data from a national lung cancer registry contributes to improve outcome and quality of surgery: Danish results. *Eur J*

- Cardiothorac Surg 2009;35:348-52; discussion 352.
2. Linder G, Lindblad M, Djerf P, et al. Validation of data quality in the Swedish National Register for Oesophageal and Gastric Cancer. *Br J Surg* 2016;103:1326-35.
  3. European Society of Thoracic Surgeons. ESTS international database. Available online: <http://www.ests.org>. Accessed 29-11-2017.
  4. Society of Thoracic Surgeons. STS national database. Available online: <http://www.sts.org/national-database>. Accessed 29-11-2017.
  5. Khakwani A, Jack RH, Vernon S, et al. Apples and pears? A comparison of two sources of national lung cancer audit data in England. *ERJ Open Res* 2017;3:00003-2017.
  6. Stirling RG, Evans SM, McLaughlin P, et al. The Victorian Lung Cancer Registry pilot: improving the quality of lung cancer care through the use of a disease quality registry. *Lung* 2014;192:749-58.
  7. Garrison LP Jr, Neumann PJ, Erickson P, et al. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health* 2007;10:326-35.
  8. Dutch Health Insurance Companies (Zn). Available online: <https://www.zn.nl/1483931648/English>. Accessed 29-11-2017.
  9. DICA. Dutch Institute for Clinical Auditing. Available online: <http://www.dica.nl/>. Accessed 24-11-2017.
  10. Van Leersum NJ, Snijders HS, Henneman D, et al. The Dutch surgical colorectal audit. *Eur J Surg Oncol* 2013;39:1063-70.
  11. Cundall-Curry DJ, Lawrence JE, Fountain DM, et al. Data errors in the National Hip Fracture Database: a local validation study. *Bone Joint J* 2016;98-b:1406-9.
  12. Data verification reports of DICA audits. Available online: <http://dica.nl/dica/dataverificatie>. Accessed 29-11-2017.
  13. Beck N, Hoeijmakers F, Wiegman EM, et al. Lessons learned from the Dutch Institute for Clinical Auditing: the Dutch model for quality assurance in lung cancer treatment. *J Thorac Dis* 2018;10 Suppl 29:S3472-85.
  14. MRDM. Data dictionaries of (DICA) registries. Available online: <https://www.mrdm.nl/showcase/downloaden>. Accessed 29-11-2017.
  15. The nationwide network and registry of histo- and cytopathology in the Netherlands (PALGA Foundation). Available online: <http://www.palga.nl/en/about-stichting-palga/stichting-palga.html>. Accessed 29-11-2017.
  16. MRDM. Medical Research Data Management. Available online: <https://www.mrdm.nl/page/about-mrdm>. Accessed 29-11-2017.
  17. Personal Data Protection Act (WBP). Available online: <https://www.government.nl/privacy>. Accessed 29-11-2017.
  18. Porter ME. What is value in health care? *N Engl J Med* 2010;363:2477-81.
  19. Jakobsen E, Rasmussen TR. The Danish Lung Cancer Registry. *Clin Epidemiol* 2016;8:537-41.
  20. Royal College of Physicians. National Lung Cancer Audit annual report 2016 (for the audit period 2015). London: Royal College of Physicians, 2017.
  21. DICA annual report 2016. Indicator results of the Dutch Upper GI Cancer Audit (DUCA). Available online: <https://www.dica.nl/jaarrapportage-2016/DUCA>. Accessed 29-11-2017.
  22. Larsson S, Lawyer P, Garellick G, et al. Use of 13 disease registries in 5 countries demonstrates the potential to use outcome data to improve health care's value. *Health Aff (Millwood)* 2012;31:220-7.
  23. Facilitating clinical documentation at the point of care. Sharing and optimising healthcare information. Available online: <https://www.registratieaandebron.nl/middelen/communicatie/film-registratie-aan-de-bron/>. Accessed 29-11-2017.
  24. Mak KS, van Bommel AC, Stowell C, et al. Defining a standard set of patient-centred outcomes for lung cancer. *Eur Respir J* 2016;48:852-60.
  25. ISPOR. International Society for Pharmacoeconomics and Outcomes Research. Available online: [https://www.ispor.org/workpaper/real\\_world\\_data.asp](https://www.ispor.org/workpaper/real_world_data.asp)
  26. CDISC. Clinical Data Interchange Standards Consortium. Available online: <https://www.cdisc.org/resources/cdisc-brochure>. Accessed 29-11-2017.

**Cite this article as:** Hoeijmakers F, Beck N, Wouters MW, Prins HA, Steup WH. National quality registries: how to improve the quality of data? *J Thorac Dis* 2018;10(Suppl 29):S3490-S3499. doi: 10.21037/jtd.2018.04.146