SPECIAL ARTICLE

The National Cardiovascular Data Registry (NCDR) Data Quality Brief

The NCDR Data Quality Program in 2012

John C. Messenger, MD,* Kalon K. L. Ho, MD, MSc,† Christopher H. Young, PhD,‡
Lara E. Slattery, MHS,‡ Jasmine C. Draoui, MS,‡ Jeptha P. Curtis, MD,§ Gregory J. Dehmer, MD,‖
Frederick L. Grover, MD,¶ Michael J. Mirro, MD,# Matthew R. Reynolds, MD, MSc,**
Ivan C. Rokos, MD,†† John A. Spertus, MD, MPH,‡‡ Tracy Y. Wang, MD, MHS, MSc,§§
Stuart A. Winston, DO,‖‖ John S. Rumsfeld, MD, PhD,¶¶ Frederick A. Masoudi, MD, MSPH,*
on behalf of the NCDR Science and Quality Oversight Committee Data Quality Workgroup

*Aurora and Denver, Colorado; Boston, Massachusetts; Washington, DC; New Haven, Connecticut;
Temple, Texas; Fort Wayne, Indiana; Los Angeles, California; Kansas City, Missouri;
Durham, North Carolina; and Ann Arbor, Michigan*

Objectives	The National Cardiovascular Data Registry (NCDR) developed the Data Quality Program to meet the objectives of ensuring the completeness, consistency, and accuracy of data submitted to the observational clinical registries. The Data Quality Program consists of 3 main components: 1) a data quality report; 2) a set of internal quality assurance protocols; and 3) a yearly data audit program.
Background	Since its inception in 1997, the NCDR has been the basis for the development of performance and quality metrics, site-level quality improvement programs, and peer-reviewed health outcomes research.
Methods	Before inclusion in the registry, data are filtered through the registry-specific algorithms that require predetermined levels of completeness and consistency for submitted data fields as part of the data quality report. Internal quality assurance protocols enforce data standards before reporting. Within each registry, 300 to 625 records are audited annually in 25 randomly identified sites (i.e., 12 to 25 records per audited site).
Results	In the 2010 audits, the participant average raw accuracy of data abstraction for the CathPCI Registry, ICD Registry, and ACTION Registry-GWTG were, respectively, 93.1% (range, 89.4% minimum, 97.4% maximum), 91.2% (range, 83.7% minimum, 95.7% maximum), and 89.7.% (range, 85% minimum, 95% maximum).
Conclusions	The 2010 audits provided evidence that many fields in the NCDR accurately represent the data from the medical charts. The American College of Cardiology Foundation is undertaking a series of initiatives aimed at creating a quality assurance rapid learning system, which, when complete, will monitor, evaluate, and improve data quality. (J Am Coll Cardiol 2012;60:1484–8) © 2012 by the American College of Cardiology Foundation

Data Quality Goals of the NCDR

In the past decades, medical outcomes research has generally been conducted with 3 sources of data: randomized clinical trials, administrative claims databases, and data registries. Each of these data sources has a unique set of applications, data quality issues, and requirements. Clinical trials, conducted as part of the U.S. pharmaceutical and device pre-approval process, have highly regulated requirements for source document verification, often with a 100% chart abstraction audit (1). In contrast, quality control of administrative data is primarily limited to fields directly related to claims adjudication. Thus, administrative claims data are significantly limited for the purposes of performing healthcare research (2). Like administrative claims data, registries are nonrandomized, observational datasets that can be generalized to real-world practice, depending on the representativeness of participants and the completeness of enrollment (3). However, as with the data collected in randomized clinical trials, registries include detailed clinical data using standardized data definitions.

In each of these cases, the standards of quality are driven by the purpose for which these data are used. This raises the question of what constitutes sufficient data validation for registries designed primarily to support improvements in healthcare quality and health outcomes research. Because of the large amount of data typically contained in registries, it is not feasible to meet the stringent requirements used in clinical trials (4). However, unlike with administrative claims data, data fields in a registry must be assessed for completeness, consistency, and accuracy to support the central activities of the registry (3).

Since its inception in 1997, the National Cardiovascular Data Registry (NCDR) has been the basis for the development of performance and quality metrics, site-level quality improvement programs, and peer-reviewed cardiovascular health outcomes research. Initially, the assessment of data quality was limited to data completeness checks. Over time, the reach of the NCDR increased with the development of additional registries, and the role of the program expanded. In 2003, the National Quality Forum endorsed the American College of Cardiology Foundation's (ACCF) percutaneous coronary intervention in-hospital risk-adjusted mortality model. This model and others have subsequently been adopted for reporting metrics for the evaluation of hospital performance (Fig. 1).

As the purposes of the NCDR expanded, the responsibility to develop more stringent standards for evaluating data completeness, consistency, and accuracy grew. However, these demands for improved quality had to be balanced with the increasing burden of data collection borne by participating sites. The NCDR developed a Data Quality Program with 3 main components: a data quality report (DQR), internal quality assurance protocols, and a yearly data audit program. The DQR evaluates completeness of data and internal consistency between fields for each facility submitting data. Internal quality assurance protocols ensure data quality before their use in the development of risk-adjusted models and in research. Each export of an analytical file must pass 33 documented quality checks before use for research. The Data Audit Program assesses accuracy of individual fields by comparison of the source documents with the data entered. Together, these processes provide continuous quality monitoring of the data contained within the NCDR (5).
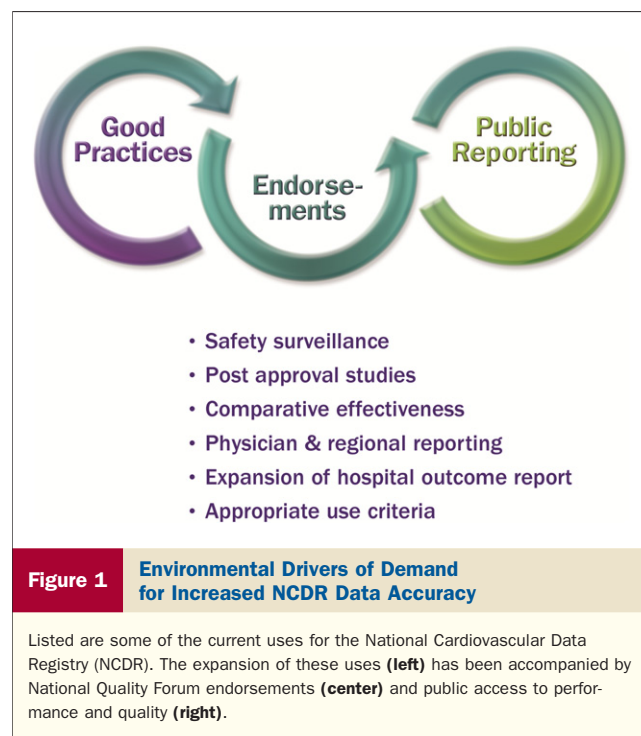
Components of NCDR Data Quality Program

The NCDR Data Quality Program consists of 3 main components: data completeness, consistency, and accuracy. Completeness focuses on the proportion of missing data within fields, whereas consistency determines the extent to which logically related fields contain values consistent with other fields. Accuracy characterizes the agreement between registry data and the contents of original charts from the hospitals submitting data.

Data quality report. The DQR (6–10) consists of registry-specific algorithms that require predetermined levels of completeness and consistency for submitted data fields. Before entering the Enterprise Data Warehouse (EDW), all submissions are scored for file integrity and data completeness, receiving 1 of 3 scores that are transmitted back to facilities using a color coding scheme. A "red light" means

- Safety surveillance
- Post approval studies
- Comparative effectiveness
- Physician & regional reporting
- Expansion of hospital outcome report
- Appropriate use criteria

Figure 1 **Environmental Drivers of Demand for Increased NCDR Data Accuracy**

Listed are some of the current uses for the National Cardiovascular Data Registry (NCDR). The expansion of these uses **(left)** has been accompanied by National Quality Forum endorsements **(center)** and public access to performance and quality **(right)**.

that a submission has failed because of file integrity problems such as excessive missing data and internally inconsistent data. Such data are not processed or loaded into the EDW. A "yellow light" status means that a submission has passed the integrity checks but failed in completeness according to predetermined thresholds. Such data are processed and loaded into the EDW but are not included in any registry aggregate computations until corrected. Facilities are notified about data submission problems and provided an opportunity to resubmit data. Finally, a "green light" means that a submission has passed all integrity and quality checks. Such submissions are loaded to the EDW. After passing the DQR, data are loaded into a common EDW that houses data from all registries and included for all registry aggregate computations. In a secondary transaction process, data are loaded into registry-specific, dimensionally modeled data marts.

Internal quality assurance. The data marts are the source for all data exports and research analytics. Each transactional step is subject to an internal quality assurance process to ensure that data used for reporting and research are error free. The requirements for the data pull are reviewed to ensure that they are complete and logical. On extracting raw data, an independent review of the output is conducted to ensure that data are complete, properly formatted, and without anomalies. On creation of the final analytical file, data are again reviewed to ensure that any created variables are coded correctly. In the final step, analytic output tables are compared with original data to ensure that they correctly reflect the specified requirements. Only after passing these quality steps are data and reports released.

Data audit program. Annual audits are conducted to assess data validity and reliability. Each audit is focused on identifying inaccurate data entry and opportunities for improvement through training or further documentation at individual facilities. More than 50 fields are audited in each of the registries each year, with some fields rotating in and out in a 3-year cycle. Within each of the registries, 300 to 625 records are audited annually within 25 randomly identified sites (i.e., 12 to 25 records per audited site). Samples of several hundred observations, even with volume of records as large as seen in the NCDR, are sufficient for the primary task of identifying fields with low accuracy scores (11). For example, if it is deemed that a field should be answered correctly at least 90% of the time (and incorrectly <10% of the time), with a sample size of 300, observing ≤261 correct answers has a predicted probability of 5.493% (based on the binomial distribution with an expected rate of 90%). Using the conventional $p < 0.05$ cutoff, if there are >39 errors, then the variable has been demonstrated to be performing worse than the stipulated 90% accuracy. Based on results such as these, the ACCF has developed additional training programs and materials for fields with lower accuracy scores. These samples are not sufficient either for subgroup analyses or to monitor the quality of data entry at individual hospitals. To address this insufficiency, the ACCF is in the progress of developing a rapid learning

system for quality that will emphasize regional quality assurance efforts as well as increasing the speed at which these improvements are implemented. In addition, the College is currently contemplating additional strategies.

Before the on-site audit, the registry steering committee constructs a list of data elements to be evaluated, which are generally those that are most important in recording and risk-adjusting outcomes. Once all the records have been audited by trained data abstractors, each field is evaluated for raw agreement and for reliability as measured by Krippendorff's alpha or Cohen's kappa statistics, both of which adjust the raw reliability score for the probability of matching by chance (12,13). The results of the audit are presented first to the committees that oversee the registry and then disseminated to the participants via online presentations, new training materials, expanded documentation on the proper coding of fields, and as case studies during yearly NCDR meetings. In the 2010 audits, the participant average raw accuracy of data abstraction for the CathPCI Registry, ICD Registry, and ACTION Registry-GWTG were, respectively, 93.1% (range, 89.4% minimum, 97.4% maximum [Online Table 1]), 91.2% (range, 83.7% minimum, 95.7% maximum [Online Table 2]), and 89.7.% (range, 85% minimum, 95% maximum [Online Table 3]). Further details of the most recent audits are contained in the Online Appendix.

Role of a Rapid Learning System to Support the Data Quality Program

The objective of the next generation of quality assurance is to ensure quality through a rapid learning system that combines mutually supporting components within the NCDR (Fig. 2). A rapid learning system is an arrangement by several organizations to trade data and information to quickly and dynamically produce information (14). To be truly effective, each component should be informed by the activities of other components and in turn should provide timely information and guidance for other components of the quality assurance program (14). The goals of the rapid learning system include to: continue efforts to standardize elements across registries; develop automated algorithms to identify potential data anomalies within the submitted data; enhance the DQR with these algorithms to provide both registry site data mangers and NCDR managers with additional guidance on the quality of the data; increase the efficiency, flexibility, and number of audits by conducting audits remotely using electronic data submission; establish multiple channels of communication between and among registry participants, the ACCF, and other stakeholders.

Figure 2 illustrates a system whereby anomalies are identified by algorithms within the DQR during submission. Shortly thereafter, the anomalies are followed by corrective efforts, including contacting the site registry data manager and, in more extreme cases, targeted audits. These triggers may also suggest further analyses of the registry data. These analyses
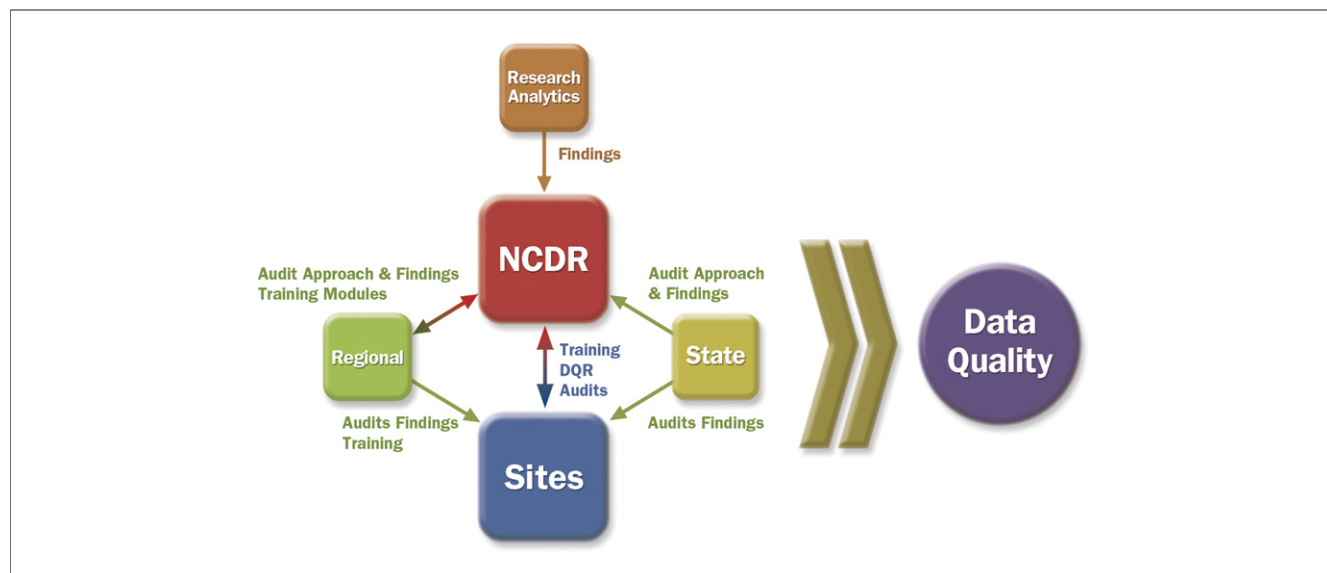
Figure 2   **Rapid Learning System for NCDR Data Quality**

A representation of the actors within the National Cardiovascular Data Registry (NCDR) rapid learning system for data quality. Different actors cooperate to continuously improve the quality of the data. DQR = data quality report.

would inform educational efforts and potentially further enhance the DQR to meet new challenges.

Conclusions

As the uses of registry data expand, the need for data validation increases. Enhanced data validation is necessary to meet stakeholder requirements such as those implemented by insurance payers for pay-for-performance, consumer coalitions for direct-to-consumer reporting, and federal and state agencies such as the U.S. Food and Drug Administration for post-approval surveillance using registry data. Such initiatives often carry implicit or explicit consequences for patients, providers, and manufacturers, further highlighting the need for accuracy. Concomitant with these external stakeholder demands, the ACCF continues to expand the uses of the NCDR into appropriate use criteria and risk-adjusted models that also require more precise data to draw valid conclusions.

The NCDR, in designing and constructing the DQR, has initiated a dynamic quality assurance process that quickly provides feedback to the sites entering the data. Our goal for the future is to expand this effort by increasing the sophistication of the DQR and building explicit links between the DQR and communications to participants as well as targeted virtual audits. When complete, this will result in a rapid learning system of data assurance that is constantly monitoring, evaluating, and improving data quality in these growing clinical registries.

Reprint requests and correspondence: Dr. Christopher H. Young, ACC, Heart House, 2400 N. Street, NW, Washington, DC 20037. E-mail: cyoung@acc.org.

REFERENCES

1. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems, implications for outcomes research. Ann Intern Med 1993;118:844–50.
2. Shahian DM, Silverstein T, Lovett AF, Wolf RE, Normand SL. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. Circulation 2007; 115:1518–27.
3. Gliklich R, Dreyer N. Registries for Evaluating Patient Outcomes: A User's Guide. 2. Agency for Healthcare Research and Quality Publication No. 10-EHC049. Rockville, MD: Agency for Healthcare Research and Quality, 2010.
4. Wennberg JE, Roos N, Sola L, Schori A, Jaffe R. Use of claims data systems to evaluate health care outcomes. JAMA 1987;257:933–6.
5. McNamara RL. Cardiovascular registry research comes of age. Heart 2010;96:908–10.
6. CathPCI_v40DQRCompanionGuide Feb2012 (available on request from NCDR@acc.org). Accessed August 2012.
7. IMPACT_v1_DQRCompanionGuide Updated 112011 km Feb2012 (available on request from NCDR@acc.org). Accessed August 2012.
8. ICD_v2_DQRCompanionGuide Feb2012 (available on request from NCDR@acc.org). Accessed August 2012.
9. ACTION_v2_DQR_CompanionGuide Feb2012 (available on request from NCDR@acc.org). Accessed August 2012.

10. CAREv1_DQRCompanionGuide Feb2012 (available on request from NCDR@acc.org). Accessed August 2012.
11. Blalock H. Social Statistics. Revised 2nd edition. New York, NY: McGraw-Hill Book Company, 1972:208–15.
12. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Measurement 1960;20:37–46.
13. Krippendorff K. Content Analysis: An Introduction to Its Methodology. Thousand Oaks, CA: Sage, 2004.
14. Abernethy AP, Etheredge LM, Ganz PA, et al. Rapid-learning system for cancer care. J Clin Oncol 2010;28:4268–74.

Key Words: audit ▪ data quality ▪ National Cardiovascular Data Registry.

▶ **APPENDIX**

For supplemental material and tables, please see the online version of this article.