日本臨床試験学会　第 11 回学術集会

シンポジウム 3 ●世界と未来のデータサイエンス

5　Somewhat Wild Speculation on the Clinical Research Data Manager of the Future

University of Texas Health Science Center San Antonio
Meredith Nahm Zozus

Abstract
Clinical Research Data Management, often called CDM has grown into a mature profession. This article reviews that development. A proposed set of clinical research data management goals are articulated. The scope of practice is defined according to established professional competencies in clinical data management and informatics. CDMs of the future will need to command a wide array of tools and methods applicable to data collection and use in clinical studies. Graduate education will likely become an entry point to the profession.

Development of the Clinical Research Data Management Profession

Clinical Research Data Management (CRDM) is maturing as a profession, i. e.,“a disciplined group of individuals who adhere to ethical standards and who hold themselves out as, and are accepted by the public as possessing special knowledge and skills in a widely recognized body of learning derived from research, education and training at a high level, and who are prepared to apply this knowledge and exercise these skills in the interest of others”.[1] The definition of a profession calls out the importance of a defined body of knowledge that is based in research and grows as new discoveries are made. Professions have a code of ethics that is enforced and accepted by the community as governing the activities of members. Such codes require behavior and practice beyond the moral obligations expected of an individual.[1] They define and demand high standards of behavior in respect to the services provided to and interactions with the public and professional colleagues.[1]

The CRDM profession grew from a set of tasks including data collection and management necessary for successful research endeavors. Before the availability of individuals specializing in the collection and management of data, data-related tasks were undertaken by other members of the research team or their support staff and often considered clerical.[2] Various systems were used to store data including paper forms, office software, commercial and custom data systems. Methods were adopted and adapted from survey research, epidemiology, demography, library and other data-intensive scientific fields and evolved in academic and industry settings. Though the scientific community has always acknowledged the importance of quality data to research conclusions, these perceptions and variable practices persisted in the absence of an established body of knowledge and evidence-based practice.

Established in 1994, the Society for Clinical Data Management (SCDM) has worked to establish and advance professional practice in the collection and handling of data for clinical studies. The SCDM established a community of practice and held the first annual meeting in Philadelphia, Pennsylvania in 1997. The Good Clinical Data Management Practices (GCDMP) committee was established in 1998 and published the first version of the GCDMP in the year 2000. The code of ethics was established in 2003, followed by the first Certified Clinical Data Manager (CCDM™) program a year later. Training opportunities outside the annual conference were first offered as webinars and on-line courses in 2005 and 2009 respec-

クリニカルデータマネジャーの未来に関するやや突飛な展望

tively and continue today. In 2016, the evidence-based revision of the GCDMP was initiated. Ongoing maintenance of the professional competencies supporting the CCDM™ exam continues on a regular cycle.[3),4)] Earlier this year, the society announced the transition of its publication Data Basics to an on-line and open access peer-reviewed journal to be published as the Journal of the Society for Clinical Data Management (jscdm. org). The journal and the broad synthesis represented in the evidence-based GCDMP ground the profession in science, serve as a cornerstone of the profession, prompt the development of new knowledge and its translation into practice and, in doing so, support a path for ongoing growth pf the profession.

Establishing professional practice as a profession matures requires transition and redrawing historical role boundaries as the scope of the profession becomes defined. Today information-related professions experience accelerated change as new informatics and data science methods and technologies are developed and translated into practice and as graduate degree programs are established. It is in this context of an established profession and the flourishing of data and information science that the future of data management professional practice will likely blossom.

The Goals of Clinical Research Data Management

Though I haven't seen goals for the profession authoritatively articulated, for my part, they include aims like ensuring that:

①data are available and used for human protection including institutional oversight, human safety, privacy and confidentiality;
②data are fit for use, i. e., ultimately capable of supporting study conclusions and other intended decision-making;
③data are handled in a manner compliant with applicable laws, regulations, and other requirements including research contracts;
④data are documented sufficiently to support reuse, research reproducibility, and replication; and that
⑤data are collected and managed by the most cost effective and efficient means.

Implicit in these is that data are of sufficient quality. Assessment of data quality is usually operationalized across multiple dimensions.[5)～13)] Dimensions of data quality important to measure and control in clinical research have been variously articulated in the literature.[14),15)] The FDA has articulated through ALCOA (Accurate, Legible, Contemporaneous, Original, Attributable) data quality dimensions important to regulatory decision-making. The Medicines & Healthcare products Regulatory Agency (MHRA) added Complete, Consistent, Enduring and Available.[16)] ICH E6 (R2) added data integrity and traceability.[17)] FDA guidance[18)～20)] also offers higher-level quality aspects of study design, data handling and analysis important to use of Real-World Data (RWD) and through the Real-World Evidence (RWE) program[21)] pursues identification and characterization of RWD dimensions that can be used to judge the suitability of RWD for regulatory decision-making.

A hallmark of professional practice is that products and services are consistently and predictably fit for use. Methods for ensuring data quality across important dimensions are well-established including broad methods for design and maintenance of Quality Management Systems (QMSs),[22)] QMS standards specifically for data quality,[23)] minimum standards and best practices specific to clinical research data management,[24)] and texts on the topic.[25)] Professional Data Managers are expected to expertly apply risk-based quality management practices to design data collection and handling processes and systems capable of producing data of the desired quality. Doing so includes actively and systematically monitoring data quality, identifying and addressing special causes and ensures that data are consistently and predictably fit for use.

The increase in study complexity[26)] and number of data sources[3),4),27)] has consistently been documented. Both are determined by the research question under study and this vary by study. For example, one study may require data obtained from interrogating implanted medical devices while another may necessitate interrogating patients, Electronic Health Records (EHRs) and image interpretations from a clinical reading center; both may collect structured data on a form and data from central labs. Designing and controlling data quality in the context of variability in data sources and technology through which they are collected and processed requires the data manager to command knowledge of general principles and methods for doing so and skill in the application of those principles and methods to new data sources and technology. Quality management systems provide a systematic approach to application of general principles and methods to managing data from multiple sources and in different contexts.

Scope of Practice

As an informatics-focused profession, the competencies for clinical research data management entail a scope of practice spanning the full arc beginning at problem identification and characterization, including solution design, development and evaluation of the solution and culminating in evaluation.[28] As professionals, CRDMs are responsible and accountable for ①designing and bringing to fruition effective data processes and information systems for clinical studies, ②managing the collection and processing of the data, and ③ultimately delivering data that are fit for use. Data Managers do this as a key member of the research operations team along other professionals including statisticians, site personnel, study managers and monitors.

The CCDM™ body of knowledge as defined by the SCDM professional competencies outlines the knowledge, skills and abilities expected of a CDM today. The professional competencies are organized into eight competency domains including design, programming, testing, training, data processing, project coordination and management, personnel management (not tested on the CCDM™ exam), and review. In large organizations, CDMs tend to specialize in one or a few of the domains, for example setting up EDC systems for a studies, integration of external data for studies, or project management and oversight of outsourced data management. In small organizations and academic settings, however, CDM practice seems to include most of the domains. Smaller organizations and academic settings organized by clinical department or individual research groups often do not have the number of studies or people to effectively utilize specialization. Whether one's practice in a particular job is specialized or broad, the competencies outline the knowledge, skills and abilities expected of professional CDMs and achieving certification is the mark of having achieved professional competence across the full scope of practice. As jobs become competitive, certification will become more important as a distinguishing mark.

Advancing the Profession

Professional practice should also include activities to advance the profession. Many CDMs do this by training new CDMs in their organizations and mentoring early-career CDMs as they develop professionally. Beyond one's organization, there are many opportunities to contribute to the profession regionally. For example, holding local study groups for the CCDM™ exam, holding or participating in local information sharing meetings of CDMs. Opportunities also exist at the national and international levels. In addition to the United States, through regional leadership SCDM has held meetings in Europe, India, China, and a joint meeting with the Drug Information Association (DIA) in Japan. Volunteering for a regional steering committee significantly contributes to advancing the profession locally. Attending, presenting or volunteering to help with a regional or international meeting also advances the profession. Internationally, opportunities abound for advancing the profession. For example, SCDM continually accepts volunteers on standing committees such as the membership, education, innovation and certification committees. The GCDMP and Certification committees are continually looking for Certified CDMs to join GCDMP chapter writing groups, translate GCDMP chapters and draft or review candidate items for revisions of the certification exam. Opportunities abound for professional CDMs to advance the profession through enriching CDM practice and professional development for up-and-coming CDMs within organizations, regionally and internationally.

CDMs can also advance the profession by participating in the generation of new knowledge in clinical research informatics. Similar to the translational cycle in biomedicine, new informatics methods must be tried and evaluated in practice. Implementation in different practice contexts such as different studies, different types of studies, and different organizations further tests new methods. The learning from implementation and demonstration in different contexts furthers improvement of methods and sometimes sparks the innovation of new methods. Conducting careful implementations, evaluating the impact and publishing the results advances the science behind our practice. In addition, professionals encounter challenges in daily practice. These things that are harder to accomplish that they should be, or all together can't be done point out gaps in knowledge, methods or tools. Characterizing and communicating these gaps through the peer-reviewed literature advance the profession by prompting innovation, research and development of methods and tools to address the gaps. These key aspects of applied research ①characterization of challenges and gaps, ②application, and evaluation of new methods in different contexts, and ③adaptation and improvement of methods or tools are significant ways that professional CDMs can directly prompt the advancement of practice as well as the science behind the practice.

SCDM recently announced the transition of Data Basics, the society's publication to an on-line, open access and peer-reviewed scholarly journal, the Journal of the Society for Clinical Data Management (JSCDM).[29] The journal fills a hole in the biomedical by surfacing reports of new informatics methods and tools relevant to and applied in the design, conduct and reporting of clinical studies. The application-focused journal will publish scholarly work with direct relevance to the practice of Clinical Research Informatics and data management. Articles of interest include reports of original research, review articles, design manuscripts, case reports of demonstration projects, education and professional development articles, topic briefs including tutorials and perspectives. The translational cycle of developing new methods and tools, applying and evaluating them in practice is a cycle from research to practice and back. As such, developing new methods and tools and translating them into practice requires both information scientists as well as practicing professionals. Publishing well-characterized challenges and rigorously evaluated innovations is a major way that CDMs can advance the profession.

Summing Up the CDM of the Future

In a newly maturing profession such as clinical research data management, individuals that started their careers two or three decades ago have evolved their competency over-time. Many of them have contributed to making the profession what it is today. They are still contributing and we need those sustained contributions as the profession evolves. Our focus here is on what clinical research data managers of the future will look like.

First and foremost, the CRDM of the future is a professional and a strong member of the clinical research operations team along with statisticians, site personnel, study managers and monitors. The Data Manager is the member of the study operations team responsible for ① designing and bringing to fruition effective data processes and information systems for clinical studies, ②managing the collection and processing of the data, and ③ultimately delivering data that are fit for use. The Data Manager collaborates with clinical leadership and the study statistician on data selection, measurement standardization and data definition. The Data Manager collaborates with site personnel, study managers and monitors on data collection processes and with the Statistician to decide data handling methods, quality acceptance criteria so that the data will be suitable for analysis. Many data managers in clinical research do these things today.

The CRDM of the future commands knowledge of fundamental principles, best evidence, proven methods and new technology to design and build (operationalizes) capable processes for data collection, processing, and use for a study. This includes the design problem of right-sizing the processes and system functionality employed for studies. CRDMs of the future will need to address many different types and sources of data and do so for studies across the spectrum from observational, to quasi-experimental and experimental designs and making use of both prospective and retrospective data. Together these imply formal and likely graduate education in the principles, theories, and methods applicable to data management in clinical research.

As the stewards of study data and information systems CRDMs of the future will be called upon for significantly more real-time and intense exchange and use of data during a study. For example, use of data to constrain and automate workflow, to produce alerts, support decision-making at sites, direct study operations and to provide study oversite. Designing these requires a firm grasp of cognitive engineering, human computer interaction, data visualization, and software development lifecycle principles. Bringing these to fruition requires the ability to work closely with information technology and research information systems professionals.

The expected scope of practice will traverse the full arc from problem identification and characterization, including solution design, development and implementation, and culminating in evaluation and continuous improvement. CRDMs of the future will apply these skills across the breadth of the eight professional competency domains, the full spectrum of clinical study designs, and all types and sources of data. Because most research is regulated in some way, professional CDMs of the future will regularly interpret and apply regulations relevant to clinical studies.

Educating the Next Generation

Today's CRDMs learned on the job and through training in their organizations or through clinical research professional societies. However, the scope of the practice is expanding over an increasing number of data sources, types of data and research contexts. CDMs may encounter new data sources on each study. New types of data are often not manageable by the same processes as previously encountered data. In this case, learning by doing will not suffice because nuances of the new data that should be checked or that could be exploited for efficiencies can easily be overlooked.

Professional practice today where new data, data sources and studies are frequently encountered requires a full methodological tool box and systematic approaches to deciding the optimal methods and tools. Such comprehensive methodological knowledge and skill in applying it implies formal education in the underlying principles, theories and methods. Many of these come from informatics and data science. For example, acquisition, representation, management, and use of rules-based knowledge in data cleaning and decision support. Other examples include methods for processing structured and unstructured data, data mining and machine learning. Theories and principles from other disciplines such as systems analysis and software development lifecycle approaches, human centered design and usability widely used in clinical research are also taught in informatics programs. Educating the next generation of clinical research informatics and CRDM professionals may benefit from basic biomedical informatics and data sciences courses. These programs likely offers a future entry point to the profession.

CDM is also informed by data collection and processing methods that have evolved in other data-intensive scientific disciplines such as survey research, health services research and epidemiology. CRI and CRDM education programs should place these approaches within the framework of general principles and arsenal of informatics methods applicable to the design, conduct and reporting of clinical studies. Likewise, professional competency involves expert use of common information system functionality in clinical studies as well as adapting or designing new functionality to accomplish automation, decision support, interoperability and knowledge extraction. As such, skills development in the use and configuration of common research information systems such as CTMS, EDC, LIMS, ePRO, and eConsent should be undertaken.

Curriculum for educating clinical research informatics and CRDM professionals will also address foundational knowledge[4] needed in the performance of professional competencies. Many of these are covered in basic graduate courses in biomedical informatics and data science. Topics such as research design, therapeutic development and regulation in clinical research not commonly covered in biomedical informatics or data science curricula will require additional courses. To prepare graduates for professional practice degree programs will include practicum and internship experiences. Example curricular topics and competencies are provided in **table 1**.

While courses that teach informatics methods in the context of specific application areas has long been considered a best practice in biomedical informatics education,[30] cost pressures in academia favor offering general courses that are used by multiple programs to decrease teaching costs. Relying on general courses sacrifices applicability to specific professional contexts and job-readiness of graduates. For example, general basic biomedical informatics and data science courses do not usually contain exercises dealing with clinical studies. While these general informatics courses are likely to contain coursework with exercises using EHRs, they will not likely contain the same using common clinical research information systems. Lacking integration of general principles and methods with specific problems in clinical research means that new graduates will enter the workforce lacking exposure to clinical research vocabulary, concepts and requirements.

The industry trend toward pre-competitive collaboration and the move in education towards on-line and distance education may relieve some of all of the pressure within universities toward use general courses. There is significant opportunity for distance-ready programs to partner with the clinical research industry to meet the workforce need through development of programs specific to professional practice in clinical research informatics and data management. On-line and distance programs designed to accommodate working professionals will addi-

Table 1　Example Curricular Topics and Competencies Relevant to Clinical Research Data Management

Topic	Competencies
Therapeutic development	Design and implement processes and systems for data collection and management appropriate for the phase of a study.
Regulation in clinical research and therapeutic development	Identify and interpret relevant regulations to design compliant processes and systems and to ultimately provide data suitable for use in regulatory decision-making.
Informatics and data science methods applicable to clinical research	Select and apply appropriate methods in the design, development, implementation and evaluation of processes and systems for the collection and management of data for clinical studies.
Quality management systems	Understand and successfully function within organizational quality management systems. Design effective quality assurance and control for the collection and management of data for clinical studies.
Biostatistics	Understand how study data will be analyzed. Work with study statisticians to co-plan data collection, cleaning and processing. Recognize opportunities for bias and data errors impactful to the planned statistics analysis.
Quasi-experimental and experimental research design	Understand the design of studies for which data are managed. Identify research questions outcomes and endpoints in study protocols and data elements important to them. Identify threats to internal validity and generalizability and suggest data elements or data collection approaches that may minimize them.
Information systems in healthcare and clinical research	Identify reasonable sources of data for clinical studies. Measure and control the quality of clinical data. Leverage information system functionality to automate and quality assure clinical studies.
Personnel, project and change management	Manage direct reports and outsourced data management. Manage development, validation, implementation and change control of information systems. Manage the collection and processing of data for multicenter studies.

tionally attract students already working in CDM and wanting to advance their careers. Where such workforce development is needed, regional needs assessments with companies, universities and regulatory agencies employing significant numbers of CDMs will provide estimates of the number of working professionals interested in furthering their education and the number of openings for new graduates. Needs assessment will also help determine the sustainability of professional degree program within a region. Where targeted workforce development programs are not feasible, current and future CDMs could seek training through general biomedical informatics or data science programs. In this case, new entrants to the clinical research workforce will need organizational or professional society training for the clinical research-specific knowledge and competency development.

Summary and Conclusions

Based on the stability of the professional competencies, the scope of practice for professional clinical research data managers seems established. CDMs of the future will need to command a wide array of tools and methods applicable to data collection and use in clinical studies. New data sources and types of data should always be expected. The past two competency surveys indicate an increase in the breadth of studies for which data are managed. In addition, comparative effectiveness research, pragmatic trials conducted within health systems using routine care data and regulatory interest in real world data indicate openness to quasi-experimental designs for generation of some types of evidence and desire for health record, claims, and device data. Informatics, data science and other scientific disciplines relevant to the practice apply equally to these newer sources of data for clinical studies and are consistently delivering new methods and tools. These will need to be applied and evaluated in and possible adapted to clinical research settings. Graduate education in informatics or data science broadly applies across data sources and types of data and will be needed to prepare CDMs of the future.

REFERENCES

1）AustralianCouncilofProfessions. *What is a profession?*,(2003). https://www.professions.org.au/what-is-a-professional/

2）Keralis, S. D., Halbert, M. & Moen, W. *Research Data Management in Policy and Practice: The DataRes Project*(Catalogue No. CLIR Publication No. 160, Council on library and information resources（CLIR), Washington DC, 2013).

3）Zozus, M. N. *et al*. Analysis of professional competencies for the clinical research data management profession: implications for training and professional certification. *J Am Med Inform Assoc* 24, 737–745, doi: 10.1093/jamia/ocw179（2017).

4）Aboulelenein, S., Williams, T., Baldner, J. & Zozus, M. N. Analysis of professional competencies for the clinical research data management profession. *Data Basics* 26, 6–17（2020).

5）Wand, Y. & Wang, R. Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM* 39, 10（1996).

6）Wang, R. & Strong, D. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems* 12, 30（1996).

7）Chan, K. S., Fowles, J. B. & Weiner, J. P. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. *Med Care Res Rev* 67, 503–527, doi: 1077558709359007［pii］10.1177/1077558709359007（2010).

8）Kahn, M. G., Raebel, M. A., Glanz, J. M., Riedlinger, K. & Steiner, J. F. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Med Care* 50 Suppl, S21–29, doi: 10.1097/MLR.0b013e318257dd67（2012).

9）Zozus, M. *et al*. Assessing Data Quality for Healthcare Systems Data Used in Clinical Research（Version 1.0): An NIH Health Care Systems Research Collaboratory Phenotypes, Data Standards, and Data Quality Core White Paper.,(Healthcare Systems Research Collaboratory, 2014).

10）Weiskopf, N. G., Hripcsak, G., Swaminathan, S. & Weng, C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform* 46, 830–836, doi: 10.1016/j.jbi.2013.06.010（2013).

11）Feder, S. L. Data Quality in Electronic Health Records Research: Quality Domains and Assessment Methods. *West J Nurs Res* 40, 753–766, doi: 10.1177/0193945916689084（2018).

12）Johnson, S. G., Speedie, S., Simon, G., Kumar, V. & Westra, B. L. Application of An Ontology for Characterizing Data Quality For a Secondary Use of EHR Data. *Appl Clin Inform* 7, 69–88, doi: 10.4338/ACI-2015-08-RA-0107（2016).

13）Liaw, S. T. *et al*. Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *Int J Med Inform* 82, 10–24, doi: 10.1016/j.ijmedinf.2012.10.001（2013).

14）Weiskopf, N. G. & Weng, C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 20, 144–151, doi: 10.1136/amiajnl-2011-000681（2013).

15）Zozus, M. N., Kahn, M. & Wieskopf, N. in *Clinical research informatics*（eds Rachel L. Richesson & James E. Andrews）Ch. 10,(Springer, 2019).

16）MHRA.(Medicines & Healthcare products Regulatory Agency, 2018).

17）ICH. E6（R2）Good Clinical Practice: Integrated Addendum to ICH E6（R2）Guidance for Industry, March 2018.(U. S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research（CDER）Center for Biologics Evaluation and Research（CBER), 2018).

18）FDA. Use of real-world evidence to support regulatory decision-making for medical devices: guidance for industry and food and drug administration staff.(FDA, 2017).

19）FDA. Guidance for industry and FDA staff best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data.(2013).

20）Food and Drug Administration, U. S. D. o. H. a. H. S.(ed United States Department of Health and Human Services)（Food and Drug Administration, 2018).

21）FDA. Framework for FDA's Real-World Evidence Program.(U. S. Department of Health and Human Services, 2018).

22）ANSI/ASQ/ISO. Vol. ASQ/ANSI/ISO 9000: 2015（2015).

23）ISO. Vol. ISO/TS 8000-1: 2011（2011).

24）*Good clinical data management practices*（*GCDMP*), ＜https://scdm.org/gcdmp/＞(2020).

25）Zozus, M. N. *The data book: collection and management of research data*.(CRC Press Taylor & Francis Group, 2017).

26）Getz, K. A. & Campo, R. A. New Benchmarks Characterizing Growth in Protocol Design Complexity. *Ther Innov Regul Sci* 52, 22–28, doi: 10.1177/2168479017713039（2018).

27）Wilkinson, M., Young, R., Harper, B., Machion, B. & Getz, K. Baseline Assessment of the Evolving 2017 eClinical Landscape. *Ther Innov Regul Sci* 53, 71–80, doi: 10.1177/2168479018769292（2019).

28）Friedman, C. P. What informatics is and isn't. *J Am Med Inform Assoc* 20, 224–226, doi: 10.1136/amiajnl-2012-001206（2013).

29）Zozus, M. N. Letter from the Editor. *Data Basics* 26（2020).

30）Friedman, C. P. *et al*. Training the next generation of informaticians: the impact of "BISTI" and bioinformatics--a report from the American College of Medical Informatics. *J Am Med Inform Assoc* 11, 167–172, doi: 10.1197/jamia.M1520（2004).