# DATA BASICS

Volume 10
Number 3
2004 Fall

*Promoting Clinical Data Management Excellence*

A PUBLICATION SUPPORTED BY AND FOR THE MEMBERS OF THE **SOCIETY FOR CLINICAL DATA MANAGEMENT, INC.**


## Letter from the Editors

Dear SCDM Members,

As was mentioned in the previous issue of Data Basics, the theme of data quality is so popular, and so much material was received, that we decided to devote two issues to this topic. The previous issue concentrated on the concept of quality, and how it is assessed. This issue provides a number of examples of how quality concepts are applied in various settings. While future issues will have a variety of themes, it is our hope to be able to include an article focused on quality in every issue, preferably on quality as it relates to the theme of the issue. As always, we invite the membership to consider submitting materials for publication.

The next issue of Data Basics will be focused on the Fall Conference. Each speaker has been asked to provide an article based on their presentation, and, given the richness of the program, the results should be very interesting indeed. The conference theme is Clinical Data Management in a Global Environment. Members who are not attending the conference should also consider article submissions, especially those who live outside of North America.

Finally, as you will see elsewhere in this issue, the first cohort of Certified Clinical Data Managers has passed the very rigorous certification exam, and our congratulations go out to them! Data Basics intends to assist data managers who wish to sit for the exam by publishing regular features that address some of the topics that have proved to be the most challenging for examinees to pass.

Again, our best wishes to the new CCDMs, and we hope to see you all in Toronto at the Fall Conference.

Best wishes, Kit Howard & Lynda Hunter ■

## Quality Definitions and Regulatory Requirements

*By: Jonathan Andrus*
*Phoenix Data Systems*

The Society for Clinical Data Management's Spring Forum Session B, Definitions and Regulatory Requirements, was a thought provoking, interactive and informative session. Participants came to discuss best practices in and around data quality. This session identified definitions around data quality, and proposed some recommendations for working well with vendors and suppliers who support the clinical trial enterprise. This session was facilitated by Jonathan Andrus, Vice President of Quality Assurance and Compliance at Phoenix Data Systems.

What are quality data? How can one assure that data are controlled in such a way as to prevent issues at the conclusion of the study? Session participants discussed some requirements for quality data that were taken from the FDA's Guidance for Industry: Computer Systems Used in Clinical Trials. In this guidance document, a five letter acronym was discussed. ALCOA. Attributable. Data must contain attributes that confirm its authenticity and integrity. Legible. Source documents, including case report forms, must be written and captured in such a way as to be able to confirm the content of the record. Contemporaneous refers to a situation where data are current and reflect relevant and up-to-date information. Original records are paramount. The ability to prove and determine original records are critical to quality data. Finally, the Accuracy of the

*Continued on page 4*

## This Issue

# DATA BASICS Editorial Policies

## Editorial Board

(also known as Publications Committee)


**Kit Howard, Co-editor**
Kestrel Consultants, Inc.
Phone: 734-576-3031
E-mail:
kit@kestrelconsultants.com

**Lynda Hunter, Co-editor**
PRA International
Phone: 913-577-2972
E-mail:
hunterlynda@praintl.com

**Cherie Stabell, Editorial Board Chair**
Genentech, Inc.
Phone: 650-225-7672
E-mail: stabell@gene.com

**Felicia Ford-Rice**
Acambis, Inc
Phone: 678-336-2713
E-mail:
Felicia.FordRice@Acambis.com

**Alec Vardy**
CV Therapeutics, Inc.
Phone: 650-384-8519
E-mail: alec.vardy@cvt.com

**Belinda Field**
Pfizer
Phone: 860-732-9880
E-mail:
Belinda_field@groton.pfizer.com

**Chandra Wooten**
Corus Pharma
Phone: (206) 832-2016
E-mail:
cwooten@coruspharma.com

**Tamela Blackstone**
Allos Therapeutics
Phone: (720) 540-5274
E-mail: tblackstone@allos.com


## Submission Deadlines (Articles and Advertising Art Work)

Our quarterly publication schedule for the next three issues requires the following input deadlines:

| | |
|---|---|
| Volume 10, #1 (Spring) | 27 February 2004 |
| Volume 10, #2 (Summer) | 26 April 2004 |
| Volume 10, #3 (Fall) | 27 July 2004 |
| Volume 10, #4 (Winter) | 18 October 2004 |

## Publication Policy

We welcome submission of materials for publication in *Data Basics.* Materials should preferably be submitted in electronic form (MS Word). Acceptance of materials for publication will be at the sole discretion of the Editorial Board. The decision will be based primarily upon professional merit and suitability (i.e. publication may be edited at the discretion of the Editorial Board.)

Neither SCDM nor the Data Basics Editorial Board endorses any commercial vendors or systems mentioned or discussed in any materials published in *Data Basics.*

## Advertising Policy

| AD RATES** | x1 | x2 | x3 | x4 |
|---|---|---|---|---|
| **FULL Page** | $675 each | $640 each($1280) | $610 each ($1830) | $575 each($2300) |
| **HALF Page** | $470 each | $445 each($890) | $425 each($1275) | $400 each($1600) |
| **QTR Page** | $285 each | $270 each($540) | $255 each($765) | $240 each($960) |

**Ads are net, non-commissionable.

Advertisers purchasing multiple ad packages will have the option of placing those ads anytime within the 12-month period following receipt of payment by SCDM.

Quarter Page = (3 5/8 inches x 4 7/8 inches)

Half Page-vertical = (3 5/8 inches x 10 inches)

Half Page-horizontal = (7 ½ inches x 4 7/8 inches)

Full page = (7 ½ inches X 10 inches)

**MECHANICAL REQUIREMENTS:** Do not send logo/photos/images from word processing software, presentation software, or Web sites. Files should be saved in the native application/file format in which they were created. Photos/images should be high resolution and received in the file size you wish to have it printed.

600 dpi or higher for black and white, 300 dpi or higher for color/grayscale

Acceptable file formats include AI, EPS, and high resolution PDF, PSD, JPEG, and/or TIFF

**PAYMENT:** Payment must be received with advertising. Space reservations cannot be made by telephone. There is NO agency discount. All ads must be paid in full.

**CANCELLATIONS:** Cancellations or changes in advertising requests by the advertiser or its agency 5 days or later after the submission deadline will not be accepted.

**GENERAL INFORMATION:** All ads must be pre-paid. Publisher is not liable for advertisement printed from faulty ad materials. Advertiser agrees to hold SCDM harmless from any and all claims or suits arising out of publication on any of his/her advertising. SCDM assumes no liability, including but not limited to compensatory or consequential damages, for any errors or omissions in connection with any ad. The SCDM does not guarantee placement in specific locations or in a given issue. SCDM reserves the right to refuse or pull ads for space or content.

Please submit all forms, artwork, and payments to:

Society for Clinical Data Management
555 E. Wells St.
Suite 1100
Milwaukee, WI 53202

Phone: 1 (414) 226-0362
Fax: 1 (414) 276-3349
E-mail: info@scdm.org
Web site: www. scdm.org

# Letter from the Chair

*Marianne Plaunt, Ph.D. STATPROBE, Inc.*

Becoming a Certified Clinical Data Manager has become a reality! This summer, the SCDM Board of Trustees approved the "cut score", i.e., the passing score for the examination and a final version of the exam will be available in the 4th quarter of this year. I would like to thank those who participated in all phases of this project - the creation of the questions, sitting for the Beta version, reviewing the results, and preparation of the final exam. Many of you have generously given your time and talent to make the achievement of this significant milestone possible and we are very appreciative of the contributions of each and every one of you. Special thanks go to Armelde Pitre. Her vision, leadership, enthusiasm and sheer drive have significantly impacted the success of this program. I also want to congratulate the inaugural class of Certified Clinical Data Managers who will be announced and recognized during the upcoming Fall Conference, to be held in Toronto from October 3 - 6, 2004.

Even as we celebrate this important milestone, SCDM volunteers are busy putting plans in place to ensure that appropriate training opportunities and resources are made available to persons interested in becoming certified. Additionally, we continue to expand the certification program to include an advanced level of certification, Certified Senior Clinical Data Manager.

I'm also very excited to share that we are now putting the final touches on the preparations for the Annual Fall Conference, which is titled "Clinical Data Management in a Global Environment". This year, based on your feedback, we have expanded our offerings to include concurrent sessions, thereby providing a broader range of ideas and topics of interest to you. Don't miss out. We hope to see you October 3 - 6, 2004 in Toronto, Canada.

Our other committees have also been hard at work this year. The membership committee conducted two surveys in which many of you participated. Thanks to all of you who took the time out of your busy schedules to respond. It is only through your input that we can continue to ensure that the activities and services of SCDM are truly addressing the needs of our members. Look for the results of these surveys to be shared during the Fall Conference and in future SCDM publications. The External Liaison committee continues to network and collaborate with various organizations with similar interests – the FDA, DIA, PHARMA, INCDMA , CDISC , and the DQRI, to name a few. The GCDMP committee is in the process of reviewing the current version of the document and preparing for additional updates. This group is committed to ensuring the GCDMP is a living and useful document that remains up-to-date and reflective of current trends and best practices in the industry. I'm also very pleased to share that we've recently launched a new web-based repository that enables our hundreds of committee volunteers to collaborate and share documents much more efficiently.

We have made wonderful progress so far this year, and are on track for completion of our main objectives for the year. This would not be possible without the volunteer membership of SCDM. Thank you for your continued support of this organization. ■

*Please Note:* SCDM does not sell its membership list and does not condone the use of the on-line membership database for electronic broadcast marketing activities.

## *SCDM Committees*

The following are currently active Committees within the Society for Clinical Data Management.

**Certification Committee**
*Chair* : Armelde Pitre
Phone: (203) 226-6456
E-mail: pitrea@snet.net

**GCDMP Committee**
*Chair* : Christine Little
Phone: (919) 408-8000
E-mail: clittle@rhoworld.com

**Membership Committee**
*Chair* : Brenda Hoeper
Phone: (513) 984-0450
E-mail:
Brenda.Hoeper@quintiles.com

**Publication Committee**
*Chair:* Cherie Stabell
Phone: (650) 225-7672
E-mail: stabell@gene.com

**Web Site Committee**
*Chair:* David Borbas
E-mail: dave@borbas.net

## Web Sites to Check Out

**ACDM** - www.acdm.org.uk
**CDISC** - www.cdisc.org
**FDA** - www.fda.gov
**ICH** - www.ich.org

There are more links to be found on our web site!
**SCDM** - www.scdm.org

Please let the Web site Committee know about any other "hot" web sites that you feel would be of interest to the SCDM membership.

## Quality Definitions and Regulatory Requirements

*Continued from cover*

data can only be assured if the other aspects of the acronym have been satisfied. ICH also speaks to the requirements for quality data. Within ICH, sponsors are required to implement and maintain SOPs that cover topics related to quality assurance and quality control. Quality control should be applied to each stage of data handling.

Often, organizations struggle over agreement on the definition of terminology. What is IQ/OQ? What does a term mean to the organization internally, and what does it mean to your vendors? What are clean data? What are self evident corrections? The development of a quality agreement was discussed as an option to help organizations identify quality requirements, identify roles and responsibilities, and denote any and all terms that are germane to developing and ensuring a well-defined relationship. The biggest advantage of the quality agreement is to define the controls, communications and quality assurance measures that need to be taken between the sponsor and its vendors. Finally, the sponsor's identification and solidification of their requirements and definitions is important to establish before trying to forge a relationship. Develop internal SOPs, seek agreement on definitions and process, then train, enforce, and review regularly.

Topics discussed included the elements of a data quality plan and the data listing audit. Other sessions during the spring forum covered planning for data quality and ensuring quality data at the end of a study through database audits, but this session focused on the identification of data clarification/edit checks, clinical coding, and risk management. Topics related to the data quality plan included:

- How well defined is your edit check specification?
- What type of dictionary is to be used?
- How will you maintain the dictionary over the course of the study?
- How frequently will you conduct coding activities?
- When a protocol deviation occurs, what actions will you take to address this within the data management scheme?

The database audit: what should be included in the plan? Session participants indicated primary variables, secondary variables, cutoff dates, data transfers, and the definition of an acceptable error rate? Heavy emphasis was placed on who should receive the audit report. Often, a database audit will be conducted, but the report is not sent to those who can affect change and ensure that the issues identified in the report are not repeated on subsequent studies. A corrective and preventative action program was discussed in the context of preventing database issues. Participants indicated that they have some form of this program in place, but it is not as structured as they would like it to be.

In conclusion, this session proved to be an eye opener for some, a refresher for others, and a session for new ideas to tackle old data management problems. From identifying the requirements for quality data to clearly defining internal and external terminology, the use of a quality agreement was found to be embraced by nearly all. Identifying the controls, communication and quality assurance measures upfront, prevents finger pointing at later points in the relationship. ■

# A Risk Approach to Software Validation

*By: Anthony Costello, Nextrials, Inc.*

If you're reading this article, chances are you spend a significant portion of your professional life worrying about data quality. As data managers in the clinical research industry, SCDM members deal regularly with data ranging from incomprehensible scribble on 3-ply paper to electronic data transferred in the latest XML file formats. In many cases, you may see a wide range of delivery formats between your various projects - even within a project. Despite this variation, a couple of undeniable facts persist:

1. The quality of these data is squarely your responsibility and
2. You have no hope of managing this information without some kind of validated, reliable database system

Enter the Data Quality Research Institute (DQRI). Founded by industry expert and CDM advocate Kaye Fendt, the DQRI treats quality as the cornerstone of its mission and has begun to move aggressively toward standards in this area. The DQRI is already attracting significant support from industry, standards groups and regulatory bodies and why not? Quality of data is perhaps the most central obligation of the research community and it is certainly complex enough a topic to evade state of the art technologies and even the most brilliant among us. Quality deserves an Institute.

But DQRI is not just another pretty acronym hanging out a shingle. The first working group to be formed by the institute has been hard at work for almost 10 months now and has presented results at several industry conferences including the SCDM Spring Forum in San Antonio, TX this March. This group, known as the Software Validation Committee is focused on the somewhat vague regulation regarding the validation of clinical trials software. Membership on this committee includes industry, vendors, CROs, academia, and regulatory expertise - an intentional variety of backgrounds designed to bring a wide breadth of experiences to a discussion about revolutionizing the way validation is performed.

The committee began with the premise that existing guidance documents provide insufficient clarity on the topic of data systems validation. Perhaps the most well known requirement stated in the now infamous 21CFR Part 11 guidance is "validation of systems to ensure accuracy, reliability, consistent intended performance…" While this concept is largely self-evident, the specific guidance wording is hardly a prescription for performing acceptable validation activities. By developing a risk analysis model, the DQRI working committee hopes to provide a tool for industry to use when evaluating variations in risk of different software and variations in validation activities based on the risk analysis. Simply put, software with more risk should be more validated. Seems pretty clear but that's where the simplicity ends. Those of us who currently bear responsibility for software development and/or validation know that passing audits or any 3rd party evaluation of QA and validation activities is often a crap shoot - each reviewer having a different take on how much planning, process, QA, documentation is enough. The DQRI committee hopes that the risk model will ameliorate much of this confusion in the following ways:

1. By using a DQRI risk model tool, companies will be prepared to evaluate software tools, programs, features and applications with a consistent methodology.
2. After determining a risk score using the model (briefly described below) companies can define and proceed with a validation plan knowing that the model provides a framework for them to defend their plan of action.
3. Regulatory bodies aware of the risk model will be prepared to evaluate the process a company followed to determine risk and validation approach.

This model may not give companies an exact prescription for validation. Nor will it likely become a blueprint for auditors to follow in their evaluations. More realistically, it will become a common language spoken by these and other stake-holder entities trying to understand how clinical trials software has been designed and validated to perform its function. Debate may still rage about gradations of 'risk' and iterations of validation activity but now, within some kind of commonly understood framework.

Early versions of this model continue to flux and the working committee is seeking widespread industry input to both validate the current direction and generate ideas to continue evolving the model. Feedback from SCDM members attending the Spring Forum was very positive. The model was presented to 60 manager and director level attendees from numerous Bio-Pharmaceutical companies. Clearly there is agreement that variations in risk call for variations in validation practice. Data managers at the conference also endorsed the notion that a multi-tier model tool would be beneficial when struggling with the somewhat subjective notion of risk assignment. Specific feedback from SCDM as well as other conferences is already being fed back into the committee's discussion sessions. Plans are also being made to present results to the FDA who is aware of the working committee and interested in reviewing progress as it happens.

The work of this first DQRI committee is earning praise so far but fine-tuning a model that can be widely applied will require considerable more work and industry input. More broadly, DQRI is actively seeking corporate membership investments to spawn several new working committees on issues related to clinical data quality. SCDM members interested in participating are encouraged to contact Kaye Fendt directly at KFendt@dqri.org. Specific questions about the risk model can be addressed to me at costello@nextrials.com. ■

# How to Win Friends and Influence Data: Quality Considerations in CRF Design

*By: Lauren Shinaberry, PRA International*

Case report form (CRF) design is one of the most complex and important endeavors in clinical trials. Second to the protocol, the CRF is the most critical piece of the process. While the protocol describes the purpose of the trial, it is the CRF that is the "real life" embodiment of the protocol. It is the gatekeeper to obtaining sufficient quality data supporting the assertion of the protocol.

A quality CRF is one that not only collects the data specified in the protocol, but does it in a way that the data collected are consistent, reliable and analyzable. Survey design research and engineering concepts for design of human machine interfaces (HMIs) can provide valuable and interesting concepts which can be applied to CRF design to improve the quality of the data collected.

Taking a page from human-machine interface design, and engineering in general, the most important concept is that of usability. In other words, a good CRF is one that is easy to use. Given the myriad of users of a CRF, this can be a difficult balancing act. The CRF should provide data management and statisticians with the data needed to support the research objectives, but it also must meet the needs of the site personnel completing the form and the study monitors comparing the CRF to source documents.

According to the International Engineering Consortium, "a well designed HMI must fit the user's mental map of the task he or she wishes to carry out." By involving site personnel and monitors in the review cycle during CRF design, a better understanding of what the mental map is for each user can be gained. It is important to keep in mind that users develop their own conceptual model of their work and this conceptual model is *never* the same as the designer's model. [1] The ISO 9241 standard defines three components of quality of use that may be applied to CRF design:

- Effectiveness: Does the product do what the users require? Does it do the right thing?
- Efficiency: Can the users carry out their tasks with minimum expended effort, including a minimum of errors?
- Satisfaction: Do users express satisfaction with the product? Does the new product reduce stress? Do the end users now have a more satisfying job?

It may be easy to think, "What difference does it make if the site is dissatisfied and stressed as long as we are collecting the data we need?" A poorly designed CRF can severely damage the relationship between the study sponsor and the investigators. Frustration from the site while they are completing the CRFs as well as frustration while responding to queries resulting from confusing forms can produce poor quality data, delays in query responses and angry investigators. Monitors also suffer when CRF design does not take into consideration their use when comparing with source documents, causing missed timelines and extra work. Statistical power may be diminished if there is not sufficient data fit to be included in the analysis. By involving all users of the CRF in the design process, the usability of the forms will be improved and, therefore, the quality of the data.

Involving the site in the CRF design process can help avoid confusing directions and page layouts. Investigators can also provide expertise on wording and terminology that may be unique to a particular therapeutic area. You may also wish to consider whether the site personnel are accustomed to using a 24 hour clock or a 12 hour clock. If the site is not familiar with the 24 hour clock convention, the likelihood of errors in the data as they record times is increased. If the site is more comfortable with the 12 hour clock, it is possible to convert the times to 24 hour convention programmatically rather than on the CRF itself. Similarly, if a site has thermometers that only measure in Fahrenheit, allowing the site to record in Fahrenheit should be considered and performing a conversion to centigrade programmatically after the data has been collected.

Be careful when including extra data in the CRF either because it might be needed "just in case" or because it has always been collected before. While it might seem harmless or even beneficial to include extra data, the additional efforts to collect, monitor, enter and clean this extraneous information can jeopardize the quality of the truly crucial data. You should clearly be able to explain why each data point is included in the CRF. Legacy data that may have been collected on previous studies may no longer be relevant. It is rare that data collected "just in case" is truly useful. If the purpose of this data is not clearly defined in the protocol, it is unlikely that the data will be sufficient to support any claims without an additional trial focusing on that facet anyway.

The field of survey design research can offer some interesting insights into the power of wording and response choices on the quality of your data. Balance is the key - you want to make sure that you are asking the questions you really want answered without being leading or confusing. Consider the following:

- "How are you feeling?"
- "Has anything unusual happened since your last visit?"

Anyone who has worked with patient diary data knows that you can get unexpected answers to these questions. In response to the first question, you may get a vague answer and miss some potential adverse events, but the second question may elicit a lengthy discussion of the fact that the subject's cat has suddenly learned to do calculus! While that is unusual, it certainly has no relevance to an adverse event (except maybe concern that the drug causes hallucinations).

Survey research indicates that when asked to qualify something, people will be more consistent if the rating scale is worded in the positive instead of the negative. For example, given a hypothetical group of people that all have had symptoms worsen to the same

*Continued on page 7*

# How to Win Friends and Influence Data: Quality Considerations in CRF Design

*continued from page 6*

degree since their last visit, you will get a wider variety of responses to the question "Are you feeling worse today?" than when you ask "Are you feeling better today?". Individuals will differ greatly in their assessment of "how bad does it have to be to be considered worse?" but tend to agree more when assessing "how good does it have to be to be considered better?" There are also words that can produce more than a 25% difference in opinion when asked. These high variability words include "most", "many" and "several". In place of these words, "lots", "almost all", "hardly any", "a couple" and "a few" will give more consistent responses from person to person.

Because wording can have such a powerful effect on the consistency of responses, it is crucial when using validated rating scales or other instruments that no changes are made in the order of questions or the wording without first confirming with the developer of the instrument that these changes do not impact the validity. Even if wording is not changed when you feel only certain questions in the rating or questionnaire are relevant, you may inadvertently be affecting the validity of the questionnaire by excluding questions .

Survey design research also shows that the choices available for a response can influence your data in ways you may not have realized. The FDA is recommending that the relation of an adverse event to the study drug be expressed as either related or unrelated. Traditionally, this question has had responses that reflect the varying degrees to which the investigator is comfortable attributing the event to the study drug. Is the response of "unlikely" the same as "not related"? How certain does the relationship need to be in order to be considered related? Based on survey research, limiting the choices to two options should provide the most conservative results. For example, in a survey about the death penalty [2], when asked simply if they were in favor of it or not, 70% responded that they supported the death penalty. When asked to select between life in prison and the death penalty, support for the death penalty dropped to 50%. Adding in a third choice of paying restitution to the victims, the support for the death penalty dropped to 30%. If you apply this same rationale to the relation of an event to the study drug, you can see that by providing many gradations (unlikely, possibly, probably, definitely, etc.) you may be under-representing the number of events that could be considered related to the study drug, when compared to forcing a yes/no response.



Also at play with the event causality is the issue of having an even number of choices or an odd number. Odd numbers of choices allow for a "middle of the road" response, while even numbers of choices force the respondent to pick either the negative side of the scale or the positive. People can get frustrated if they must pick from an even numbered list, especially when a neutral answer may be the most accurate.

Responses from trained medical personnel based on measurements are not as open to influence as patient diary responses or verbal questions posed by the site personnel. It is important to consider the population that the subjects come from. Education level, income level and culture can all impact the quality of the data you collect from subjects. In third world countries, respondents have a tendency to exaggerate answers or respond in a way to please the researchers. There can be the perception that the site personnel are in a position of power with the ability to punish or reward. Literacy must be considered when writing patient diaries in a manner that is easy to understand.

Rating scales are also affected by education level. Research has been done that indicates a scale from 1 to 10 is most affective for subjects with a college degree. It is easier for people with less than a high school education to select from a scale of 1 to 5. Populations with little or no education may give the most consistent response when faced with only 'good', 'OK' or 'bad'. Additionally, it is important to realize that a '2' on a scale of 1 to 5 is not the same as a '4' on a scale from 1 to 10. Further complicating matters is the fact that while people in the US view the '1' on the scale as the worst or lowest intensity and '10' as the best or highest intensity, other cultures are accustomed to '1' being the best or highest intensity and '10' being the worst or lowest intensity.

The order of questions can influence the quality of the responses. If you have a long list of questions, the respondent may give up

*Continued on page 8*

# How to Win Friends and Influence Data: Quality Considerations in CRF Design

*continued from page 7*

halfway through and later questions will either be left unanswered or the responses will not be thought through to the degree that the earlier ones were. While rearranging the order of questions on a patient diary can cause difficulties in the data management side, if a long list of questions is critical to the outcome of the trial, you may want to consider having questions in a different order from visit to visit. Long lists of responses are also subject to bias, based on order. People will tend to pick the choices nearest to the top of the list when they read through it themselves. Survey designers will put the most important questions first, but in clinical research all the questions are likely important, so this solution becomes more difficult to implement.

Eye-tracking studies have shown that it is easier and more efficient to have all the responses lined up along the right margin of the page or screen. When completing a long list of questions, this layout can improve user satisfaction with your form and therefore improve the chances of collecting accurate responses on not just the first few questions, but those at the end of the list as well.

Determining exactly what data you will collect on your case report form is only the first step in data quality. Equally important is determining exactly how you will collect the data. By overlooking the influence wording and response selection has on consistent and accurate collection of the data, you may jeopardize the outcome of the study. Fields of study outside of clinical trials and data management can provide valuable insight into how to create well designed CRF through the principles of human-machine interface design and survey research. ■

**References:**


1. The Human Machine Interface Tutorial, *International Engineering Consortium. Online: http://www.iec.org/online/tutorials/hmi*
2. Survey Design, *Creative Research Systems. Online: http://www.surveysystem.com/sdesign.htm*
3. Survey Design Considerations, *WebSurveyor. Online: http://www.websurveyor.com/pdf/designtips.pdf*
4. *Spilker B, Schoenfelder J.* Data Collection Forms in Clinical Trials. *New York, NY: Raven Press; 1991.*
5. *Society for Clinical Data Management.* Good Clinical Data Management Practices Version 3. *September 2003. Online: http://www.scdm.org/gcdmp/*

# Standardization Supports Quality in the Interchange of Laboratory Data

*Susan Bassion, Clinical Data Interchange Standards Consortium (CDISC) and Phillip Pochon, Covance Central Clinical Laboratories*

CDISC is an open multidisciplinary non-profit organization committed to the development of industry standards to support the electronic acquisition, exchange, submission and archiving of clinical trial data and metadata for the medical and biopharmaceutical product development. The mission of CDISC is to lead the development of global, vendor-neutral platform-independent standards to improve data quality and accelerate product development in our industry.

## *CDISC Clinical Laboratory Standard*

The CDISC LAB Team has defined a model for the standardized representation of laboratory data generated during the conduct of clinical trials. The LAB model is a superset of data fields that follows a recognizable and logical hierarchy of clinical trial laboratory data. The base LAB model can be used to represent all routine "one test, one result" testing, required in clinical chemistry, hematology and urinalysis. A fully tested production version of the model is available on the CDISC website, www.cdisc.org/standards.

The reasons for the development of data standards for laboratory data interchange are well known to those involved in day to day handling of this data, accounting for as much as 60-80% of data generated during the conduct of a trial. Central laboratories may support over 1200 different data formats and one pharmaceutical company may have multiple standards. Each format requires development, maintenance and support, and each receiver or sender in the data chain must work to validate their systems and verify exchanges. Standardization of laboratory data interchange formats reduces cost, complexity and resource requirements, while improving quality.

The LAB model, developed in response to these inefficiencies, has not been tied to a single specific computing environment or file type This approach was taken in recognition of the varying formats and systems being used in the industry. In order to avoid rejection due to technical incompatibilities, text file as well as SAS and XML schema implementations have been developed.

In addition to the base model, extensions of the model are being defined for use with more complicated testing algorithms such as those seen with microbiology testing. A draft version of the microbiology extension is available for public comment on the CDISC website. Future extensions of the model are planned for handling of genomics data, specimen handling data and ECG interpretations, often handled by laboratory data managers.

The LAB model can be used for data interchange between sender and receiver, but is not a plug-in application that addresses all data issues. A transmission agreement between sender and receiver must be developed that covers implementation type (e.g., SAS), fields used (as not all fields are required) and use of standardized text fields and codes, in addition to other issues. (See LAB 1-0-1 Specification pdf posted at www.cdisc.org/standards.)

## *Code Lists and LOINC*

Where appropriate and possible, the effort at standardization is extended to recommendations for use of existing code lists. Fortunately, the lab industry has existing standards developed by LOINC. Logical Observation Identifier Names and Codes (LOINC) was developed and is copyrighted by the Regenstrief Institute, Inc. and the LOINC Committee. The use of LOINC® codes is recommended in the LAB model as a supplement to test codes developed by each laboratory for internal use. Some pharmaceutical companies have already moved from proprietary internal test codes to LOINC.

Use of standardized codes improves reliability and accuracy of data interchange and should permit higher-level analyses of databases. Such higher-level analyses may be useful in drug or legacy databases, and the FDA has expressed an interest in use of LOINC codes for such analyses in the future.

(NOTE: The Consolidated Health Informatics (CHI), an eGOV initiative, has recommended the use of LOINC as the standard for federal agencies. In 2002, President Bush outlined a management agenda for making government more focused on citizens and results, including expanding Electronic Government or eGOV. eGOV uses improved Internet-based technology to streamline citizen- and business-to-government communications.)

The LOINC database provides a set of universal names and codes for identifying laboratory and clinical tests. The purpose is to facilitate the exchange and pooling of laboratory test and observation results. Each LOINC code corresponds to a single test definition. The LOINC test code definition includes fields for specifying:

1) Component (analyte) — e.g., potassium, hemoglobin, hepatitis C antigen.
2) Property measured — e.g., a mass concentration, enzyme activity (catalytic rate).
3) Timing - i.e., whether the measurement is an observation at a moment of time, or an observation integrated over an extended duration of time — e.g., 24-hour urine.
4) The type of sample — e.g., urine; blood.
5) The type of scale — e.g., whether the measurement is quantitative (a true measurement,) ordinal (a ranked set of options), nominal (e.g., E. coli; Staphylococcus aureus), or narrative (e.g., dictation results from x-rays).
6) Method, where relevant, i.e., the method used to produce the result or other observation.

The LOINC codes are not intended to transmit all possible information about a test or observation. They are only intended to *identify*

*Continued on page 10*

# Standardization Supports Quality in the Interchange of Laboratory Data

*continued from page 9*

the test result or clinical observation. Other fields in the message can transmit the identity of the source laboratory and very detailed information about the sample or test method.

When matching local codes to LOINC codes, one local code may have several possible LOINC codes. For most laboratory tests the LOINC database contains a methodless code and one or more method specific codes. If a clear match by method is possible, the method specific code is preferred. Conversely, variations in a method may produce several local codes that all match a single LOINC code. In such a case the data transmitter should match all such local codes to the appropriate LOINC code, and then use a separate field in the data transmission to distinguish variants of the base method. If the variations are felt to be significant, a request for new LOINC codes may be submitted to the LOINC committee.

The Regenstrief Institute (Indiana University Medical Center) maintains the LOINC database, supporting documentation and the RELMA mapping program for matching local codes to LOINC. RELMA (Regenstrief LOINC Mapping Assistant) is a Windows-based mapping utility available for free use. The CDISC LAB Team is working with the Regenstrief Institute to develop that a subset of test codes commonly used in drug development clinical trials. When posted with the October 2004 LOINC update, RELMA will contain a sub-setting tool that can used to search the listing "Commonly Ordered Tests in Clinical Trials".

Use of LOINC codes should increase efficiency of lab data transfers as common terminology decreases the need to learn, load, utilize and maintain proprietary codelists for each organization, department or program. And use of LOINC should improve data quality as standardized terminology provides an enhanced ability to combine and analyze data.

The World Wide Web URL: loinc@regenstrief.org provides downloads of all LOINC and RELMA files and documentation. ■

# What is the Data Quality Research Institute?

*By: Kit Howard, DQRI, and Kestrel Consultants*

## *Introduction*

As has been discussed elsewhere in this issue, the environment and requirements for managing data have changed dramatically in the 30 years since the inception of the field of clinical data management. Not only have the data capture tools and the storage media evolved, but the sheer volume of data being managed has exploded. While techniques have been developed for addressing some of the resulting challenges, little rigorous and published quality analysis has occurred, and what has been done is scattered. With the advent of standardized approaches to data storage and transmission such as the ICH E2B and CDISC, an opportunity exists to define the quality of the data themselves. This is the role that the Data Quality Research Institute (DQRI) hopes to fill.

The DQRI is a non-profit 501(c)3 organization established by Kaye Fendt (co-founder of CDISC and initiator and champion of the GCDMP for SCDM). It is physically located on the University of North Carolina campus in Chapel Hill.

## *Purpose*

Its primary purpose is to provide an international scientific forum for academia, healthcare providers, industry, government, and other stakeholders to develop the science of quality as it pertains to clinical research data. It intends, among other things, to:

- Develop methodologies to assess and assure data quality
- Apply principles of risk assessment and risk management to clinical research data (future direction of FDA)
- Develop new standards and metrics for research data quality in context of what's already available
- Provide methods for independent assessment of software and data quality
- Develop metrics for describing and communicating data quality
- Develop educational materials and programs related to research data quality

In other words, to sponsor the research and communicate the results to all interested parties

## *Projects*

The institute will sponsor research projects designed to address such topics as:

- The science of errors
- Risk factors in clinical trials software development
- Risk factors in instrument design, data collection and analysis
- Statistical models for assuring quality in clinical research data
- Areas of vulnerability in drug development and risks of action and inaction
- Electronic Data Capture risk assessment and quality analysis

The initial project involves developing risk approach to validation of software systems that will allow those who need to validate software to determine how much validation is appropriate and required based on the type and uses of the software. This project is well underway, and preliminary results are expected to be presented at the Society for Clinical Trials meeting in May.

Some additional projects are expected to be initiated soon, and include:

- developing statistical methodologies for eliminating unnecessary data querying
- define the metrics, quality parameters, that assess data quality at each step of the data lifecycle. E.g., how do you measure and report quality in study design? How do you measure and report quality in study execution? And so forth.
- what is database auditing, what does it tell us, what should it be, how do we know if we have quality, and what are the appropriate statistical approaches for ensuring that quality

## *Institute Structure*

The institute is headed by Kaye Fendt, and is run through an Operating Board that is drawn from industry, academia, and support services. Currently it includes Dave Christiansen, Bill Sollicito, Kaye Obenshain, Sue Carroll and Imogene Grimes. DQRI sponsors research on its own, but also partners with other organizations to accomplish its goals. The National Institute for Statistical Sciences (NISS) is a partner, and similar relationships are being developed with other groups. Perhaps most importantly, an FDA liaison, Steve Wilson, has been assigned to DQRI, and the FDA is fully aware of the work that the institute does.

## *Further Information*

DQRI has presented at the International Conference on Software Quality and led a workshop session at the 2004 SCDM Spring Forum. Additional presentations will be made at the Society for Clinical Trials meeting in May, and the American Society for Quality (ASQ) in May. Informational sessions are being organized for other relevant meetings as well. Copies of descriptive materials and slides are available from the Institute at www.dqri.org or by emailing info@dqri.org. ■

# Congratulations to the Inaugural Class of Certified Clinical Data Managers (CCDM)

The SCDM would like to recognize the following people for their outstanding achievement in passing the SCDM Certified Exam and becoming professsionally Certified Clinical Data Managers. In order to be eligible to sit for this rigorous exam, they had to meet these requirements:

- Bachelor's degree or higher plus 2 or more years full time Clinical Data Manager experience;
- Associate's degree (2 years) plus 3 or more years full time Clinical Data Manager experience;
- No degree plus 4 or more years full time Clinical Data Manager experience or part time work experience that can be translated into full-time equivalents.

In addition, they provided the SCDM Certification Committee with valuable feedback regarding the beta test process. The Certification Committee is grateful for their contributions and is in the process of addressing their comments and updating the final exam. Once again, our congratulations!

CERTIFIED CLINICAL
CCDM
DATA MANAGER

*Paul Clarkson*
*Lynda Clark*
*Deborah Cole*
*Ellen Coull*
*Colleen Cox*
*Lara Dague*
*Stuart Donaldson*
*Jennifer Duggan*
*Arleen Eppinger*
*Susan Hale*
*Debra Harris*
*Christopher Helker*
*Lynda Hunter*
*Shannon Labout*
*Michele Langton*
*Mark Laskowski*
*Amanda Lieberum*
*Kathryn Long*
*Brian McCourt*
*Christel McMillan*
*Laurie Macomber*
*Ivey Moebus*
*Meredith Nahm*
*Becky Page*
*Amy Pease*
*Renee Pridgen*
*Sarbaree Raha*
*Friedrich Riess*
*Chris Santoro*
*Patricia Seidl*
*Dawn Shewchuk*
*Lauren Shinaberry*
*Anne Smith*
*Judith Stein*
*Janet Welsh*
*Janice Wiggan*

# SCDM launches first certification exam for Clinical Data Managers

MILWAUKEE – The Society for Clinical Data Management (SCDM) announces the launch of the first and only validated and industry-recognized certification exam for clinical data managers. Available this fall for industry use, the exam was established in order to create an accepted standard of knowledge, education and experience by which clinical data managers would be professionally recognized by the pharmaceutical, biological development and medical device therapies community.

"Certification has been and continues to be an exciting opportunity to help shape the future of the profession and to recognize the important contribution that clinical data managers make to clinical trials," said Armelde Pitre, Chair of SCDM's Certification Committee. "The rollout of SCDM's certification exam is a cornerstone of SCDM's commitment and strategic approach to enhancing the profession."

SCDM is the only organization in North America that dedicates itself exclusively to the advancement of clinical data management as a profession. SCDM has spent the last 5 years developing the exam. It will provide clinical data managers with the credentials they need to demonstrate that they successfully met eligibility criteria and passed a rigorous exam.

The exam features approximately 200 multiple choice questions, focusing on the broad spectrum of clinical data management tasks, including protocol review, coding, lab management, processing of local and central lab data, and database design. For more information on how to apply, the cost and where the certification exam is offered, visit the SCDM Web site, www.scdm.org.

The Society for Clinical Data Management, Inc. is a rapidly growing national organization founded in 1994 for the advancement of the discipline of clinical data management. Its membership is comprised of over 1,300 individuals who work in the pharmaceutical and related health care industries, and related third party support service providers.

# Sponsor Strategies for Improving the Quality of Outsourced Data Management

*By: Debbie Stein-Kurz, DSK Consulting, Inc.*

Outsourcing clinical data management to a contract research organization (CRO) is a common practice, yet sponsors are not always fully satisfied with the results. This article will address quality strategies to improve the success of outsourced data management. Sponsors play a critical role in the success of such projects and there are several things sponsors can and should do to better enable CROs to meet their customer's expectations.

## *Requirements*

Written requirements are the foundation to quality. A fundamental principle of validation is that it is not possible to validate a computer system or software without written requirements. In the same way, it is not possible for the CRO to verify that deliverables meet customer expectations without written requirements in the form of specifications. Written specifications must be provided for the database, edits, datasets or output, and data transfers. The sponsor should formally approve such specifications and ensure that they are accurate. In this auditor's experience, sponsors frequently fail to provide adequate and/or accurate specifications. Problems can be communicated by frustrated sponsors in situations where the deliverable was consistent with the specification but the specification was not correct. In order to better manage sponsor expectations, the CRO should have a change control process in place with charges assessed to the sponsor when changes are made to specifications that the sponsor previously approved.

## *Quality Control*

The CRO commonly provides test data transfers during the course of the study. If the sponsor does not provide feedback the CRO may assume that there are no issues with the test transfer. Unfortunately, this may only mean that the sponsor did not look at the data in any detail. In order to detect and correct problems as early as possible, the sponsor should carefully review data in test transfers. However, the sponsor should not be conducting quality control activities that the CRO should do prior to providing the test transfer. If the sponsor concludes that the CRO has not completed an adequate verification of the deliverable against specifications, the sponsor should promptly return the deliverable for rework prior to conducting a thorough review.

## *Process*

Standard processes are another foundation of quality. Processes at the CRO should be evaluated by the sponsor to ensure they are adequate. This is often done as part of a pre-contract assessment by quality assurance. Once CRO processes are determined to be adequate, the sponsor should strictly limit requiring changes to the CRO's processes for the sponsor's study. Forcing the CRO out of its standard processes can have a negative impact on quality.

## *Contract*

Contracts often focus on legal and financial language and do not contain adequate detail regarding deliverables and quality. The contract should contain specific expectations that can also serve as audit standards in evaluating the quality of the CRO's work. For example, the contract should indicate if the CRO is required to conduct a database quality review prior to delivery of final datasets with results of the database quality review provided to the sponsor. Acceptable error rates can also be defined in the contract.

## *Communication*

Frequent communication is also necessary to ensure the CRO understands and can meet customer expectations. The sponsor should ensure that all decisions and assignments from phone calls and meetings are documented and provided to both parties to ensure consistent understanding and adequate implementation.

## *Quality Assurance Audits*

While a comprehensive discussion of sponsor quality assurance pre-contract assessments and post-contract audits of CROs is beyond the scope of this article, there are several areas that should be emphasized in order to maximize the effectiveness of these audits. Quality assurance audits should focus on ensuring appropriate validation or quality control activities are routinely in place for all deliverables. For example, how does the CRO verify that the CRF collects all data as defined in the protocol? How does the CRO verify that all edits are programmed and firing correctly? How does the CRO verify that output, e.g., SAS datasets, are a complete and accurate reflection of the database and consistent with sponsor-approved specifications? How does the CRO measure database quality, e.g., error rates, of final datasets?

Sponsors who adopt these strategies will be providing the tools to the CRO that will enable the CRO to meet the sponsor's expectations. ■

# SCDM Spring Forum 2004

## *Session C: Data Quality during the Life of a Project*

*Presented by: Abdelhak Oualim, PharmD, Aventis Pharmaceuticals*
*Report by: Alec Vardy, CV Therapeutics, Inc.*

In this session, Abdelhak asked participants to focus on a Clinical Data Management (CDM) data quality strategy targeting a clean database as its ultimate deliverable. He included in this strategy two related elements that, in this clinical data manager's opinion, are too often overlooked among all the other stresses and challenges involved in planning a successful clinical trial: (1) proactivity, and (2) the need for data managers to build quality into **every** aspect of the trial, even (especially?) those "owned" by other functions.

In this report, I have chosen to highlight CDM input into those tasks owned by other functions. While much of what I heard and read at this session about our potential to contribute on an equal footing to overall trial and data quality is not new, it was very impressive and almost awe-inspiring to see it all displayed in a single location. My "take-away" thought was: CDM has so much to offer, if only we could routinely be granted the opportunity and empowerment to do so !!

My choice should in no way be seen as a negative comment on the basic CDM quality standards expressed at this session. Participant input to and discussion of the Data Management Plan, the CRF Study Book, Database Structure, and other CDM basics are well worth reading, and many of us will find at least one or two new ideas to discuss with our colleagues. But the potential impact of CDM on the Study Team came through so strongly as to be simply irresistible.

It has not been my general experience that CDM is one of the signators of protocol approval (although I know this is the case in many companies), but seeing the list of protocol components that have a major impact on CDM activities and resources encourages me to lobby for this sign-off. Is the definition of the endpoints clear, including ancillary conditions such as study drug compliance that may be used to "qualify" endpoint data for certain analyses? Are the event schedule and visit windows well defined? Are the event schedule and the textual description of events internally consistent? Where interim analyses are called for, can the reader immediately understand which patients and patient data will be included in such analyses? How are protocol violations to be handled? Does the protocol include an efficient and comprehensive accountability of study drug? And so on.

Many of the above considerations carry over to the Statistical Analysis Plan (SAP). If we assume that the SAP is a more detailed presentation of the statistical section of the protocol, and may includes mock-ups of tables, listings and figures, CDM can apply its protocol review process at a greater level of detail. Additionally, the SAP gives CDM information which might be very valuable. Can we identify data which we might routinely collect, but which will not figure heavily in the study report (such as many variables included on the concomitant medication form)? If so, what should we do – remove them from the CRF, reduce the intensity of cleaning, or nothing? The SAP will also help us identify critical variables for a future database audit.

CDM's input to the Site Monitoring Plan was also seen as very important, to the extent that the proposal was made that Monitoring and CDM should share this document as an expression of joint accountability. Clearly, these two functions bear most of the burden for routine cleaning of trial data, and many ideas generated during the session were aimed at increasing the understanding of one another's strengths and constraints. Examples included: informing site monitors more extensively of the checks implemented in CDM, providing guidance for reporting of adverse events, a more consolidated approach to CRF harvesting, establishing clear expectations of cycle times of both functions, jointly defining metrics to monitor data quality, and many more.

One of the more obvious collaborative suggestions expanded during the session was the Data Surveillance Plan (aka Blinded Data Review Plan, Interim Data Review Plan or Data Monitoring Plan). This plan sets out the details of ongoing monitoring of data by various functions (CDM, Biostatistics, Monitoring, etc.), including (but not limited to) the times at which the data should be closely reviewed, and the purpose of the review. Proposals for timing of the review were of the type "at 25%, 50% and 75% enrollment". Items listed under the purpose of the review included consideration of recruitment, fraud, protocol deviations, critical endpoints, informed consent issues, study schedule issues, systematic missing values, outliers, query rates and types, etc. Critical to the success of the implementation of this plan is that any corrective measures decided as a result of the review need to be fully documented. I fully support the implementation of such a plan, not only because it promotes a more extensive examination of the data than routine data review methods allow, but also since it emphasizes further a multi-functional responsibility for data quality. I would even go further and suggest that many of the items included in the purpose of this review can productively be examined even earlier than the 25% enrollment point.

I have been fortunate to work in a few companies where Clinical Data Management was involved in most of the above activities (but never all of them – until now !!). However, I recognize that many of my data management colleagues are not so lucky. I hope that they can use the content of this session, and the list of obvious benefits of CDM involvement in these activities to the clinical trial process as a whole, in negotiations toward a greater, and more fulfilling role for clinical data managers.

SCDM members can view the session content in full at: https://www.scdm.org/members/RecentEvents/Spring2004/session_c.doc ■

# Fun Stuff

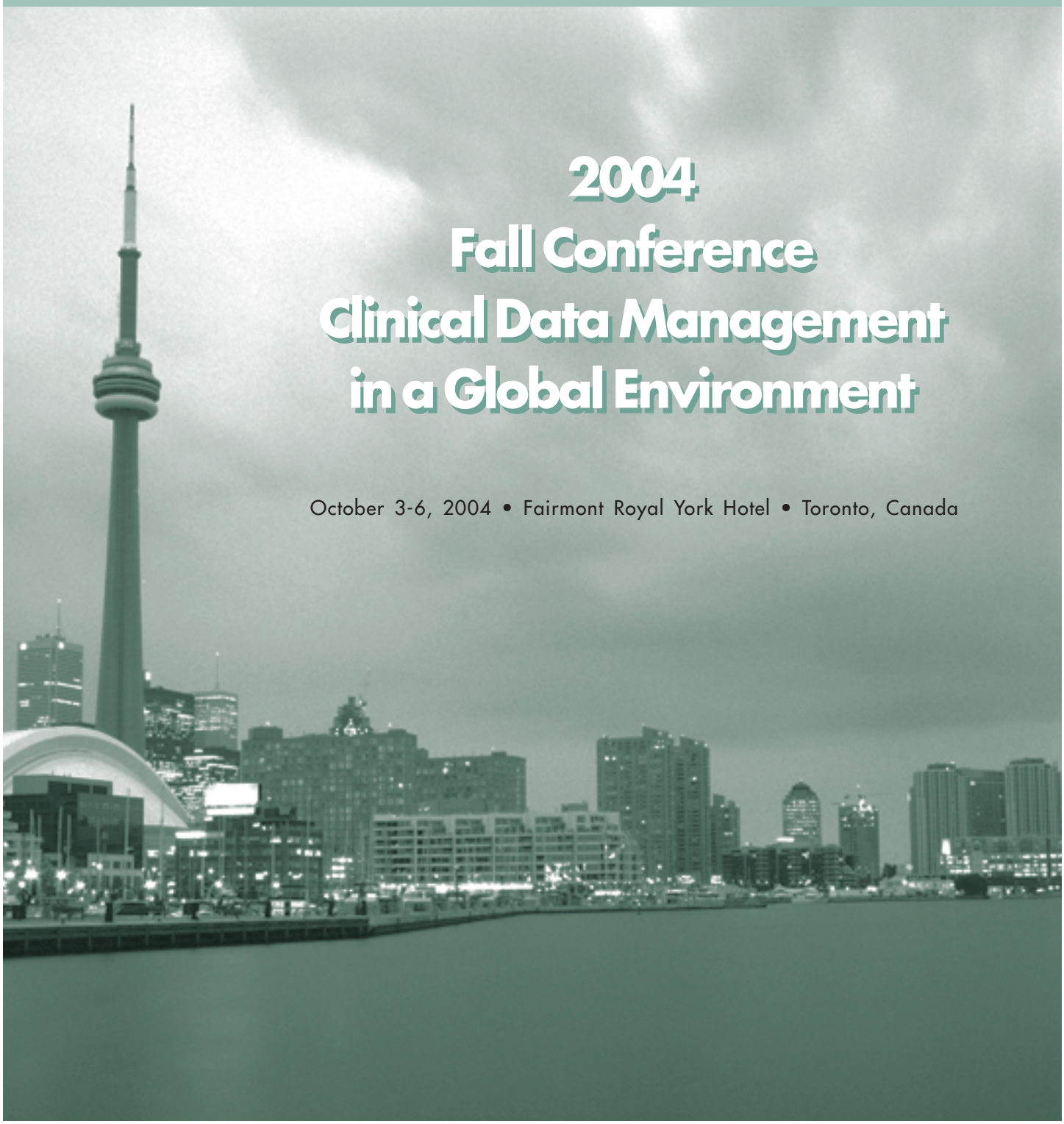
This issue's Fun Stuff is a word search with a twist. All the words have been drawn from this issue, but instead of listing them, clues have been provided. The numbers in parentheses following each clue indicate the number of letters in the word or words. Spellings may be European or North American, depending upon the context. Words may appear up, down, left, right or diagonally. There is a secret phrase contained in the letters that remain after all the words have been circled, and the phrase contains 27 letters, not counting the spaces between words. The solution will be printed in the next issue of Data Basics. Have fun!

C N M B U I L D E T A R U C C A
I O I O S I S Y L A N A K S I R
N I N U D A T A B A S E G A S G
D T V T A E G N A H C R E T N I
T A E S E L L A T Q U D N I O Y
S S S O T M A A U L I E O S I G
D I T U A C P N B T M S M F T O
R N I R R U A O I E Y I I A A L
A O G C R V D R R G E G C C D O
D M A I O V E I T A I N S T I T
N R T N R E U Y T N N R R I L A
A A O G R Q Y D A O O E O O A M
T H R R E S E A R C H C O N V E
S L O R T N O C Y T I L A U Q H
A S P E C I F I C A T I O N S Y
R A T I N G S C A L E G I B L E

1. Correctly reflective of the source (8)
2. Confirming data authenticity and integrity (12)
3. Review or inspection (5)
4. Happening at the same time (15)
5. Legal document (8)
6. Company that performs work for others (3)
7. Electronic file containing information (8)
8. Create or invent (6)
9. Measure of correctness (5,4)
10. Study of DNA (8)
11. Process of synchronizing (13)
12. The science of blood (10)
13. Exchange (11)
14. One who researches (12)
15. Readable (7)
16. Prototype or experiment (5)
17. Not a copy (8)
18. Contracting work to another party (11)
19. Degree of excellence or value (7)
20. Processes for ensuring data validity (7,7)
21. Multiple choice assessment (6,5)
22. List of needs (12)
23. Examine or study (8)
24. Assessment of threat level (4,8)
25. Degree of fulfillment or contentment (12)
26. Characteristics of system or job to be done
27. Accepted best ways of doing something (9)
28. Questionnaire (6)
29. Process for ensuring software correctness (10)

SOCIETY FOR CLINICAL DATA MANAGEMENT

2004 FALL CONFERENCE



# 2004 Fall Conference Clinical Data Management in a Global Environment

October 3-6, 2004 • Fairmont Royal York Hotel • Toronto, Canada

Please contact info@scdm.org if you have questions about registration for upcoming meetings, advertising, renewal of membership, or if you need to provide updated mailing/contact information.

**Society for Clinical Data Management, Inc.**
555 E. Wells Street
Suite 1100
Milwaukee, WI 53202-3823
Phone: 414-226-0362
Fax: 414-276-3349
E-mail: info@scdm.org

Visit us at
**www.scdm.org**

92-0804-005

Society for Clinical Data Management, Inc.
555 E. Wells Street
Suite 1100
Milwaukee, WI 53202-3823

FIRST CLASS PRSRT
U.S. POSTAGE
PAID
PERMIT NO. 2066
EAU CLARIE, WI 54701