DATA QUALITY CONTROL REPORTING IN A MULTISITE CLINICAL TRIAL

Dana LeTendre and Michael S. Shing
Infant Health and Development Program
Stanford University School of Medicine

ABSTRACT

This paper describes the Data Quality Control Reporting System for a multisite clinical trial using paper data collection forms. The system begins with a Data Quality Control Form (DQCF) completed by Data Control Technicians as they identify potential problems with the data collection forms upon receipt. The DQCFs are entered into a microcomputer dBASE III* database before they are sent to the sites for resolution. After a site responds, the resolution is entered into the DQCF database. Using SAS**, the dBASE III database and mainframe forms inventory file are integrated and a variety of reports are produced. The reports are sent to the sites as part of a comprehensive feedback system to monitor and reduce data collection errors. Data collected over a period of two years indicate that the Data Quality Control Reporting System has had a major, positive impact on data collection quality.

BACKGROUND

The Infant Health and Development Program (IHDP***) is a collaborative, multisite study designed to test the efficacy of combining early child development services and family support with pediatric health surveillance in reducing the incidence of health and developmental problems among low birthweight, premature infants. There are eight medical schools serving as data collection sites across the country, with a total of approximately 1000 infants who are being followed for a period of three years. Assessments are conducted at nine different points (birth, 40 weeks gestational age, 4 months, 8 months, 12 months, 18 months, 24 months, 30 months, and 36 months), using a total of over 70 different data collection forms per infant. Each assessment point includes a Family Interview Form, Health Exam Form, Interval Health Form, and a Neurological Exam Form. The content of these forms changes somewhat across the assessment periods in order to accommodate the maturational changes in the infants. In addition, a variety of other psychological, psycho-social, and environmental measures are obtained at the different assessment points. The National Study Office (NSO) for the project is located at Stanford University, and forms are sent to the NSO weekly.

DATA QUALITY CONTROL FORMS

Upon receipt of a batch of data forms, receipt control information is entered into a forms inventory file of the mainframe computer (see Constantine, et al., 1987). This file consists of a record for every form completed on every infant. Each record consists of fields for the infant identification number, form number, data form completed, data form received, and some special status codes.

After the appropriate data have been entered into the forms inventory file, Data Control Technicians review each form according to a form-specific list of visual check specifications. Whenever a problem or error is detected, the Data Control Technician initiates a Data Quality Control Form (DQCF) (see Figure 1).

* dBASE III
** SAS
*** IHDP
The DQCF contains the infant identification number, the form number, the item number, a coded description of the problem, and a variety of other data. The DQCF is a four-part NCR form; the NSO retains one copy of the form, and sends the other three copies to the site. The NSO copy of the DQCF is entered into a dBASE III database maintained on a PC. Once the site has determined the appropriate response to the problem identified on the DQCF, it enters the response on the remaining two of the copies and sending the third copy to the NSO. Upon receipt of the resolved DQCF from the site, Data Control Technicians make the appropriate correction to the data collection form, and the resolution is entered into the dBASE III database.

Prior to the implementation of the DQCF dBASE III system, tracking of unresolved DQCFs was done via a manual filing system. This was time-consuming, inefficient, and provided little monitoring information about the DQCF process. DQCFs remained unresolved for considerable periods of time, and the sites received minimal information regarding the quality of their data as reflected in their DQCFs. Although the sites did receive a weekly report of the total number of DQCFs sent to each site, this information was not particularly informative because each site varied in the number of forms completed each week. Therefore, a simple count of DQCFs provided no information on the actual rate of errors.

DATA QUALITY CONTROL REPORTING SYSTEM

The addition of the DQCF dBASE III system allowed the use of Base SAS software to provide the integration between the mainframe forms inventory file and the PC database of DQCF information. The SAS program uses the forms inventory file to determine the number of forms received from each site in a given time period, and this number represents the denominator of the error rate ratio. The numerator for this ratio is the number of DQCFs sent to each site. The resulting Error Rate Summary Report (see Figure 2) shows the number of forms received from each site, the number of DQCFs sent to each site, and the average number of errors per form received from each site.

In addition to the Error Rate Summary Report, other dBASE III reports are available to provide tracking and monitoring information for managing the DQCF system. The Unresolved DQCF Report facilitates easy identification and followup of unresolved DQCFs, thus preventing needed corrections from "falling through the cracks." The Response Time Report provides information on the number of days that elapsed from the time the DQCF was sent to the time the resolution was made on the DQCF form at the site. The Cumulative Error Reports provide information on DQCFs sent, by site, form and item, or alternatively, across all sites and by form and item.

PROGRAMMING DESCRIPTION AND ISSUES

The Error Rate Summary Report is produced through a combination of SAS and dBASE III programs, operated through the WYLER EXEC... language available at Stanford. An EXEC program is set up so that it prompts the user for the report date and the set of forms to be included in the report. The DQCF dataset is then read into the SAS program to read in the forms inventory file and subset out the particular forms to be included in the report. The DQCF dataset is then read in and merged with the forms inventory dataset. The count of DQCFs and the total number of forms received are obtained for a 10-week period prior to the report date, as well as for the year-to-date period, for each site and for the total sample. The average error rate ratios for the 2 periods are then calculated.

IMPACT: TRENDS OVER TIME

At the time the SAS-generated Error Rate Summary Report was initiated, the average number of errors per form across all sites was .107. Six months later, this rate had dropped to .097, and the sites had become

1676
productively competitive in making appropriate efforts to keep their error rates low. When the new system was implemented, the average response time, i.e., the amount of time it took the sites to respond to a DQCF, stood at 30.6 days. Six months later, the average response time had dropped to 7.6 days. The Cumulative Error Reports provided both NGO staff as well as site staff with detailed information on which staff were making what types of errors on which items on which forms. This information was used to focus staff training, clarification of policy and procedural matters, and redesign of specific items on subsequent forms.

ACKNOWLEDGMENTS

Thanks to Kevin Su for his early work on the DQCF reporting system, Norm Constantine for his strong support and encouragement, the IHDP data control staff for their constant efforts at making the system a success, and Nancy Greenwood for making everything look nice.

The Infant Health and Development Program is supported by grants to the Department of Pediatrics, Stanford University, the Frank Porter Graham Child Development Center, the University of North Carolina, and the eight participating universities by the Robert Wood Johnson Foundation, with additional support to the Department of Pediatrics, Stanford University, from the Division of Maternal and Child Health of the U.S.P.H.S., the National Institute of Child Health and Human Development, the Pew Trusts, and Stanford University Center for the Study of Families, Children and Youth.

For more information or copies of the documented SAS code contact the authors at:
Infant Health and Development Program Building 460, Stanford University Stanford, California 94305

NOTES

dBASE III is a registered trademark of Ashton-Tate Inc., Culver City, CA, USA.

**SAS is a registered trademark of SAS Institute Inc., Cary, NC, USA.

***The Infant Health and Development is a collaborative effort. The senior staff involved in the study are listed below.

****WYLBUR EXEC is a registered trademark of Stanford University, Stanford, CA, USA.

National Study Office:
Ruth T. Gross, M.D., Director; Donna Spiker, Ph.D., Deputy Director; Norman A. Constantine, Ph.D., Director of Data Analysis; Wendy L. Keititzman, Director of Field Operations (The Stanford University Center for the Study of Families, Children and Youth.)

Program Development Office:
Craig T. Ramsey, Ph.D., Director; Donna Bryant, Ph.D., Associate Director; Joseph Sparkling, Ph.D. and Barbara M. Wasik, Ph.D., Co-Directors of Curriculum Development; Isabelle Lewis and Claudia Lyons, Curriculum Development Specialists; Kaye Fendt, M.S.P.H., Director of Data Management and Statistical Computing (The Frank Porter Graham Child Development Center, University of North Carolina at Chapel Hill).

Participating Universities:
University of Arkansas for Medical Sciences (Arkansas); Albert Einstein College of Medicine (Einstein); Harvard Medical School (Harvard); University of Miami School of Medicine (Miami); University of Pennsylvania School of Medicine (Pennsylvania); University of Texas Health Science Center at Dallas (Texas); University of Washington School of Medicine (Washington); Yale University School of Medicine (Yale).

Site Directors:
Patrick H. Casey, M.D., Arkansas; Cecilia M. McCarton, M.D., Einstein; Michael W. Yogman, M.D., Harvard; Charles R. Bauer, M.D. and Keith G. Scott, Ph.D., Miami; Judith Barnbaum, M.D., Pennsylvania; Jon E. Tyson, M.D. and Mark Swanson, M.D., Texas; Clifford J. Sells, M.D. and Forrest C. Bennett, M.D., Washington; David T. Scott, Ph.D., Yale.

Education Directors:
Joan Forex, M.Ed., Arkansas; Katy Lutzius, Einstein; Marcia Hartley, M.S., Harvard; Mimi Graham, Ph.D., Miami; Joanne Crooms, M.Ed., Pennsylvania; Beverly A. Mulvihill, M.Ed., Texas; Rebecca R. Fellow, Ph.D., Washington; Sandra E. Mainquist, M.A., Yale; Bettye Coldwell, Ph.D., Educational Consultant, Arkansas; Paul Turner, Ed.D., Dallas Independent School District Liaison, Texas.

Research Steering Committee:
Helena C. Kramer, Ph.D., Chair (Stanford University); Charles R. Bauer, M.D.; J. Brooks-Gunn, Ph.D. (Educational Testing Service); Marie C. McCormick, M.D. (Brigham and Women's Hospital); Craig T. Ramsey, Ph.D.; David T. Scott, Ph.D.; Sam Shapiro (The Johns Hopkins University School of Hygiene and Public Health); Donna Spiker, Ph.D.

REFERENCES