ABSTRACT

SAS*-based systems to monitor receipt of assessment forms and compliance with assessment date windows for the Infant Health and Development Program are described. Upon receipt of an assessment form or missed-assessment report, the coordinating center data control staff log the appropriate information into either the forms-inventory or missed-assessment inventory files. Three types of reports are produced monthly and disseminated to our eight study sites. The Window Monitoring List Report provides listings of form status for individual infants. The Window Monitoring Summary Report provides similar information tabulated by site and form. Finally, the Assessment Monitoring System Graphic Report shows plots of the percent of forms due which have been received, not received, and administered within the window, by week since the beginning of the study. Following the introduction of these systems, the window compliance at all sites has improved steadily and the problem of forms not accounted for at the end of an assessment period has been eliminated. In addition, the study coordinators and data collectors at the sites have responded positively and enthusiastically to the dissemination of this feedback.

BACKGROUND

The Infant Health and Development Program (IHDP)** is a multisite clinical trial involving eight sites. It was designed to test the efficacy of a specific intervention during the first three years of life for low birth weight preterm infants. The coordinating center for the trial is the National Study Office (NSO) located at Stanford University. Among other responsibilities, the NSO supervises data collection and provides data processing and analysis.

The Data Analysis Systems Division (DASD) of the NSO is responsible for producing information from and about the data collected at the eight sites according to the Research Information Systems (RIS) model illustrated in Figure 1. These systems follow the formulation of the research hypotheses and the design of the study, and precede the evaluation of the research hypotheses. Applications of the information produced according to the RIS can be grouped into two functional categories. First, information is routed within the system to provide data quality feedback. Second, information is disseminated outside of the system for the purpose of evaluating research hypotheses.

One application problem faced by the DASD required providing assessment monitoring information extracted during data management for use as feedback to the data collection process at the sites. This particular feedback route is highlighted in bold in Figure 1. This paper describes the problem, the approach we used, and the results we achieved.

DATA COLLECTION

During the three year data collection cycle, 56 different data collection forms are employed across nine separate infant-age assessment time points (e.g., 4 months, 1 year, 2 years, etc.) for approximately 1,000 study infants and their families. The infants were
recruited immediately after birth during a ten month period, therefore, the youngest infant in the sample is approximately ten months younger than the oldest infant. Different forms are used to collect various data on health status, neurological functioning, physical and cognitive development, and psychological and family status. It is critical for many of the variables that the data be collected close to the specified time point because of the rapid developmental progress of the infants.

The assessment time points are specified as exact dates for each infant according to ages calculated from the infant's corrected birthdate, i.e., the estimated date the infant would have been born if it were not premature. For each time point, a window has been established within which data collected will be considered appropriately representative of the infant's development for that time point. The window is defined as either the exact corrected birthday plus or minus one week (for assessments before one year corrected age) or the exact corrected birthday plus or minus two weeks (for assessments beginning with one year corrected age). For example, an infant born three weeks premature on January 4, 1986 would have a 12-month assessment due on its one year exact corrected birthday, January 25, 1987. The allowable window would extend from January 11 to February 8. The narrower windows for the early assessments are necessary because the rate of growth and development of the infant is more rapid during the first year of life; therefore, a two week discrepancy from the targeted assessment time point could make a more significant difference.

The site study coordinators and data collectors receive assessment schedules prior to each assessment time point. These personnel have been rigorously trained in methods to be used to encourage the infant's mother to participate at the appropriately scheduled time and in the importance of collecting the data within the window. After completion, the study coordinators ship the assessment forms to the NSO in weekly batches. At the NSO these shipments are logged in, assessment date and missed-assessment information is entered into the forms-inventory database, and the forms are visually checked and coded by the data control staff. Finally, the forms are keypunched and entered into the research database.

DATA MANAGEMENT

We maintain two primary databases, the research database and the forms-inventory database. The research database consists of five types of files:

1. infant-inventory external files for analysis and non-analysis infants,
2. form-image data entry external files,
3. cumulative SAS datasets containing form-image data for each of the 56 form types,
4. working SAS datasets containing cleaned and reduced data for sets of assessment time points, and
5. analysis SAS datasets containing cleaned and reduced data for individual analysis projects or ancillary studies.

The forms-inventory database consists of two files: the forms-received file and the missed-assessment file. These external files which are updated weekly using keypunched transaction files and SAS programs employing the UPDATE command. The forms-received file contains one record for each form received for each infant, consisting of the infant identification number, form number, assessment date, date received, and special status codes. The missed-assessment file contains one record for each missed-assessment report received, consisting of the infant identification number and data identifying one or more missed assessments.

ASSESSMENT MONITORING FEEDBACK SYSTEMS

Three types of information are needed to facilitate full receipt control and assessment monitoring: listings of form status for individual infants, summary tables showing site and across site performance for specific forms within a reporting period, and summary graphs showing performance trends over time. The systems we developed to provide this information are described below.

Window Monitoring List Report

This report shows compliance information for the set of assessments associated with a specified time point for each included infant with an assessment due during a specified reporting period. A sample page from a Window Monitoring List Report is reproduced in Figure 2. To produce this report, the parameters passed to the SAS program by the menu-driven exec program specified a report for site 2, reporting period August, 1986, the 18-month assessment time point (forms 26 through 31, and 33), and the data collection sample group (as opposed to the analysis sample group). The
assessments per infant, and then using PROC PRINT to create the list report. With the latter approach, the program and the output both became more easily readable.

Window Monitoring Summary Report

This report provides site and form tabulations of the information presented in the Window Monitoring List Report. A sample tabulation is shown in Figure 3.

The Window Monitoring Summary Report is produced from the same program and temporary dataset as the list report, and is created using PROC TABULATE. After many programming hours and much trial and error, the proper specifications necessary to get PROC TABULATE to correctly produce this report were discovered.

Assessment Monitoring System Graphic Report

These reports show plots by week of the percent of assessments completed within the allowable window, together with the percent of assessments for which forms have been received and not received. Since multiple forms are grouped together into assessment time points (e.g., the 8-month assessment consists of a family interview, health exam, health interview, and neurological exam), one form type must be specified to represent each assessment point before running the graphs. Typically, the family interview is chosen. The graphs illustrate, for each week plotted, the composite compliance picture based on the most recently due assessment for each infant.

To produce this report, the Window Monitoring System program creates a temporary dataset with multiple observations per infant. Each selected infant observation from the infant inventory file is duplicated into the number of forms for the specified assessment time point. These observations are then merged with the relevant section of the forms-inventory file to extract the assessment dates for each infant/form combination.

In an earlier version of the program, the list report was created using PUT statements while cycling through multiple observations per infant in the temporary dataset. An improved approach was developed using PROC TRANSPOSE to recombine the multiple infant

Figure 2. Window Monitoring List Report

report shows for forms included in the specified assessment, and for each infant, whether the assessment form was received and administered within the allowable assessment window, if administered outside the window how many days outside, and if not received whether a missed-assessment form has been received or the assessment has not been accounted for.

To produce this report, the Window Monitoring System program creates a temporary dataset with multiple observations per infant. Each selected infant observation from the infant inventory file is duplicated into the number of forms for the specified assessment time point. These observations are then merged with the relevant section of the forms-inventory file to extract the assessment dates for each infant/form combination.

In an earlier version of the program, the list report was created using PUT statements while cycling through multiple observations per infant in the temporary dataset. An improved approach was developed using PROC TRANSPOSE to recombine the multiple infant
Depending on the infant's corrected age during a particular week, different assessments are counted for different infants on the same week, and for the same infant on different weeks.

A sample graph is reproduced in black and white in Figure 4. Note that the forms received and not received categories are mirror images of each other and sum to 100 percent, i.e., for any week an infant's assessment forms are counted as either received or not received. In contrast, the within-window forms are a subset of the forms received.

Before interpreting trends in the plots over time, it is important to remember that the size of the window increases from two weeks to four weeks for each infant at the one year assessment point. This adds to any positive trend in window compliance. On the other hand, the denominator used to calculate the within window percent is based on forms due rather than forms received. This means that missed assessments (frequently due to subject attrition) are counted as being outside of the window, thereby attenuating any positive trend to the extent that attrition increases over time.

This Assessment Monitoring System program is similar to the Window Monitoring System program in its creation of a temporary dataset containing multiple observations per infant. The process is complicated by the need to cumulatively track assessments that vary independently across infants and weeks. Although we did manage to get an earlier macro-free version to run adequately for more than a year, the use of macros in the current version has tremendously simplified the program and facilitated maintenance and longitudinal expansions.

**COSTS**

The primary costs involved in developing these systems include systems design time, programmer time, and CPU development time. For the Window Monitoring System, design time required approximately 10 hours, programmer time involved approximately 35 hours, and computer time was approximately 15 CPU minutes (IBM-3090). For the Assessment Monitoring System, these costs were approximately 12 design hours, 50 programmer hours, and 18 CPU minutes.

The production cost is approximately 40 CPU seconds for one full set of monthly Window Monitoring List and Summary Reports for all eight sites. For the Assessment Monitoring System graphs, approximately 267 CPU seconds and 6 hours of HP 7550A plotter time are required to produce a full monthly set of 27 graphs (by site and treatment group and totals).

**RESULTS**

The results of the implementation of these systems have been encouraging:

1. The problem of forms not accounted for at the end of an assessment period has been fully eliminated.

2. Window compliance has improved steadily at all sites and remains at approximately 86 percent for all sites and groups combined. For some site/treatment group combinations, the within-window line has merged with the forms-received line, indicating that all forms received have been administered within the window.

3. The site study coordinators and data collectors have responded positively and enthusiastically to the dissemination of this feedback. They feel that it provides a useful and powerful tool to enable them to better accomplish their jobs. It also serves as an exceptionally timely means of providing to the data collectors interesting information about the data they collect, since most of the research analyses will not be disseminated until after the completion of all data collection.

4. The NSO and the Research steering Committee have begun to rely on these reports to monitor data collection and sample attrition. Site and treatment group differences in compliance and attrition receive especially careful monitoring through the use of these reports.

---

![Assessment Monitoring System Graphic Report](image-url)

Figure 4. Assessment Monitoring System Graphic Report

672
ACKNOWLEDGMENTS

We thank Helena Kraemer and Wendy Kreitman for contributing to the design of these systems, Dana LeTendre for lobbying for systems improvements, and the IHDP data control group for their diligent efforts in taking the data from data collection to data management.

The Infant Health and Development Program is supported by grants to the Department of Pediatrics, Stanford University, the Frank Porter Graham Child Development Center, 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 Youth Development.

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

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

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

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. Kreitman, Director of Field Operations (The Stanford University Center for the Study of Youth Development).

Program Development Office:
Craig T. Ramey, Ph.D., Director; Donna Bryant, Ph.D., Associate Director; Joseph Sparling, Ph.D. and Barbara Wemk, Ph.D., Co-Directors of Curriculum Development; Isabelle Lewis and Claudia Lyons, Curriculum Development Specialists; Kaye Yendt, 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:
Pamela E. Casey, M.D., Arkansas; Cecelia M. McCarron, M.D., Einstein; Michael W. Yogman, M.D., Harvard; Charles R. Bauer, M.D. and Keith S. Scott, Ph.D., Miami; Judith Barnbaum, M.D., Pennsylvania; Jon E. Tyson, M.D. and Mark Swanson, M.D., Texas; Clifford J. Belsky, M.D. and Fong, M.D., Washington; David T. Scott, Ph.D., Yale.

Education Directors:
Dorothy Kizer, M.Ed., Arkansas; Patricia Wagner, M.D., Einstein; Kerrie Martley, M.S., Harvard; Dana Graham, Ph.D., Miami; Joanne Croome, M.Ed., Pennsylvania; Beverly A. Mulvihill, M.Ed., Texas; Rebecca A. Pauwels, Ph.D., Washington; Sandie R. Malmquist, M.A., Yale.


Research Steering Committee:
Helena C. Kraemer, Ph.D., Chair (Stanford University); Charles S. Bauer, M.D.; J. Brooke-Guth, Ph.D. (Educational Testing Service); Marie C. McCormick, M.D. (University of Pennsylvania); Craig T. Ramey, Ph.D.; David Scott, Ph.D.; Sam Shapiro (The Johns Hopkins University School of Hygiene and Public Health); Donna Spiker, Ph.D.