A Comprehensive Automated Data Management System Using SAS® Software

Heather F. Eng, Jason A. Lyons, and Theresa M. Sax

Epidemiology Data Center
University of Pittsburgh
Graduate School of Public Health
What is SAS?

Statistical Analysis System
SAS Institute, Cary, NC

Not just a stats package
• relational database
• flexible programming language
• multi-faceted toolkit for data management
Data Management Responsibilities

Ensure quality data collection
• monitor data accrual and study activity
• perform constant quality control
• provide feedback to investigators and coordinators

Maintain timely and complete study database
• create clean and stable datasets
  • interim analysis
  • final analyses
• deidentify data for public sharing
• archive data at study close
Uses for SAS in Data Management

Database loading
  • importing external data files

Quality control

Report creation
  • tables and graphics
  • posting reports to websites and shared servers
  • emailing reports to appropriate personnel

Creating analysis datasets
  • deidentifying data
  • archiving data sets in readable formats
Computer-Assisted Cognitive-Behavior Therapy for Depression Trial (CCBT)

Clinical Sites:
   University of Louisville
   University of Pennsylvania

Data Coordination:
   University of Pittsburgh

Primary Aims:
1) Computer-based cognitive behavior therapy (CCBT) vs. traditional cognitive behavior therapy (CBT)
2) Cost-effectiveness of CCBT and CBT
CCBT Data Sources

- **Patient Assessments**: MS Access + SQL Server
- **Therapy Sessions**: Audio Files
- **Computer-Based Lessons**: Monthly Export

- **VPN Shared Server**
- **SAS Data Library**

3X Daily
CCBT Reporting Requirements

SAS Data Library

Accrual

Quality Control

Analysis Datasets

Safety

Alerts

Study Coordinators

Investigators

Data Managers

IRB / DSMB

Statisticians
Loading the SAS Data Library
DBLOAD program calls a series of programs that load each SQL table into SAS data library

* DBLOAD.SAS: Load CCBT database ;
* jL 5/27/11 ;

%INCLUDE "L:\CCBT\prglib\DBLOAD\EL.SAS";
%INCLUDE "L:\CCBT\prglib\DBLOAD\DEM.SAS";
%INCLUDE "L:\CCBT\prglib\DBLOAD\GAF.SAS";
%INCLUDE "L:\CCBT\prglib\DBLOAD\HAMD.SAS";
...

DBLOAD is run 3X daily, via Windows Task Scheduler
SAS Code Generation: SASGEN

SASGEN writes programs for DBLOAD Excel data dictionary
  • variable names
  • descriptive labels
  • formats (coded values, dates, times etc.)

Generates SAS code to import SQL Server tables
Applies SAS labels and formats

Creates a separate program for each data table
SASGEN executed only once at study start-up
Generated code: Label and Format statements

```
label ID = "Patient ID"
    DATE = "Date completed"
    GENDER = "Gender"
    DOB = "Date of birth"
    MARITAL = "Marital status"
    (etc.);

format Date DOB MMDDYY8.
    Gender SEX.
    Marital MARITAL.
    Ethnic ETHNIC.
    White Black Asian Amind Hawai YESNO.
    Educatn DEGREE.;
```
Quality Control
(Data Editing)
Data Editing

One program for each table (CRF)
- missing data
- range violations
- dependency relationships
- logical checks

Reports posted on shared server
Refreshed 3X daily

Coordinators view and make corrections in Access/SQL Server database
Data Editing

Segment of editing program for one variable

* Initialize ; edit_chk=0;

if DIABTYPE=. then edit_chk=1;
  *(missing)*
else if not (0<=DIABTYPE<=3) and
  (abs(DIABTYPE)>0) then edit_chk=2;
  *(out of range)*
else if ((DIABETES eq 1) and (DIABTYPE=.)
  or (not(DIABETES eq 1) and not(DIABTYPE=.)
  then edit_chk=3;
  *(dependency)*

if edit_chk>0 then output;

This code repeated for each variable in each dataset
SAS Code Generation: EDITGEN

Reads data dictionary from Excel
  • value ranges
  • dependent variables

Generates SAS code for every variable
  • missing
  • range
  • trigger-dependencies

Creates individual table-specific edits programs
EDITGEN executed only once at study start-up
Reporting
Two classes of reports

1. **“Pseudo-Virtual”** – refreshed 3X daily
   - quality control for study coordinators
   - posted on shared server
   - viewed through Access front end

2. **Weekly report packet**
   - includes reports from “virtual” group
   - additional summary (hi-level) overviews
   - emailed weekly to investigators and study staff
Reporting

Two report formats

1. **SAS ODS HTML**
   - drill-down capabilities
   - content-driven traffic lighting
   - refreshed 3X daily

2. **SAS ODS PDF**
   - one copy refreshed 3X daily
   - archive copies stored with date/time stamp
   - HTML also stored in archive as PDF
Forms Status Reports in HTML
Forms Status Reports in HTML

<table>
<thead>
<tr>
<th>SITE</th>
<th>Total Unresolved Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>site 1</td>
<td>154</td>
</tr>
<tr>
<td>site 2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total Unresolved Forms</th>
<th>Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0112ABC</td>
</tr>
<tr>
<td>1</td>
<td>0113HFG</td>
</tr>
<tr>
<td>2</td>
<td>0119POI</td>
</tr>
<tr>
<td>1</td>
<td>0125IUY</td>
</tr>
<tr>
<td>2</td>
<td>0126LKJ</td>
</tr>
<tr>
<td>2</td>
<td>0128IUY</td>
</tr>
<tr>
<td>3</td>
<td>0129UYT</td>
</tr>
</tbody>
</table>
## Forms Status Reports in HTML

<table>
<thead>
<tr>
<th>Subject</th>
<th>0129UYT, Randomized 03/07/13</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0=Delinquent .. 1=Complete .. N=Incomplete .. V=To be verified .. X=Missed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EL</th>
<th>DEM</th>
<th>HAM</th>
<th>D</th>
<th>SR</th>
<th>S</th>
<th>CID</th>
<th>D</th>
<th>M</th>
<th>PH</th>
<th>L</th>
<th>I</th>
<th>F</th>
<th>G</th>
<th>AF</th>
<th>BD</th>
<th>CE</th>
<th>T</th>
<th>PA</th>
<th>CT</th>
<th>AS</th>
<th>WA</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>1</td>
<td>1</td>
<td>V</td>
<td>V</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>V</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>X</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>X</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

| Initial EL eval | V | 1 | 1 | V | V | 1 | - | - | - | 1 | - | - | - | - |
| Second EL eval  | V | - | 1 | - | - | - | 1 | 1 | - | 1 | - | - | X | 1 |
| Week 4          | - | - | 1 | - | - | - | - | 1 | 1 | - | - | X | 1 |
| Week 8          | - | - | 1 | - | - | - | - | 1 | 1 | - | 1 | X | 1 |
| Week 16         | - | - | 1 | - | - | - | 1 | 1 | 1 | - | 1 | 1 | X |
| Month 3         | - | - | 1 | - | - | - | 1 | 0 | 1 | - | 1 | - | - | - |
# Forms Status Reports in PDF

<table>
<thead>
<tr>
<th>Form</th>
<th>Complete</th>
<th>Incomplete</th>
<th>To be Verified</th>
<th>Missed</th>
<th>Delinquent</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL</td>
<td>3729 (89%)</td>
<td>32 (0.8%)</td>
<td>26 (0.6%)</td>
<td>327 (7.8%)</td>
<td>98 (2.3%)</td>
</tr>
<tr>
<td>EL</td>
<td>167 (100%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>DEM</td>
<td>62 (98%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>1 (1.6%)</td>
</tr>
<tr>
<td>HAMD</td>
<td>328 (92%)</td>
<td>0 (0.0%)</td>
<td>1 (0.3%)</td>
<td>26 (7.3%)</td>
<td>3 (0.8%)</td>
</tr>
<tr>
<td>CSSRSB</td>
<td>44 (70%)</td>
<td>0 (0.0%)</td>
<td>19 (30%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>SCID</td>
<td>61 (97%)</td>
<td>0 (0.0%)</td>
<td>2 (3.2%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>SCIDII</td>
<td>63 (100%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>MPHx</td>
<td>62 (98%)</td>
<td>0 (0.0%)</td>
<td>1 (1.6%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
</tr>
</tbody>
</table>

(continued)
Event Monitoring
Event Monitoring: ALERTS

Data manager alerts

• serious adverse events
• protocol violations
• early terminations (patient withdrawal)
• may require DSMB, IRB reporting
• review in weekly staff meetings

ALERTS program

• reviews SAE, PV, ET datasets for new activity
• emails report to data management staff
Event Monitoring: LOG_SCAN

Automated programs run in background
• errors or warnings missed?

SAS execution logs written to text file

LOG_SCAN program
• scans each log for error and warning notes
• emails report to data managers
• usually null
Automation
Sending emails in SAS: EMAILER

*Set up mail server ;
options emailsys =smtp
  emailhost =mail.edc.pitt.edu
  emailauthprotocol=LOGIN
  emailpw =Cxyz16mail;

FILENAME outmail EMAIL
  Subject="CCBT Weekly Report"
  From  ="CCBTMessenger@edc.pitt.edu"
  To    =("X@some.edu""Y@other.edu")
  CC    =("Z@another.edu")
  ATTACH ="U:\Weekly Report\CCBTRprt.pdf";

DATA _NULL_;  FILE outmail;
  put / 'Attached is the CCBT Weekly Report. ' '
       // 'Please review at your convenience.'
       / '[etc.]';
run;
Scheduled Task: **AUTO_REPORT**

One SAS driver program runs everything 3X daily:

- Initiates the execution log file
  - calls DBLOAD  (calls individual CRF programs)
  - calls EDIT_RPT  (calls individual CRF programs)
  - calls HTML_RPT  (calls individual report programs)
  - calls WEEKLY_RPT  (Mondays only)
  - calls ALERTS
  - calls EMAILER

Closes execution log file

Calls LOG_SCAN to report any log issues
Summary of Components

SAS code generation

SAS data libraries (relational DB)

SAS programming features
  • data manipulation
  • string handling
  • array processing

SAS driver programs calling smaller programs

SAS Output Delivery System (ODS)

SAS Emailer

(and Windows Task Scheduler)
CCBT Investigators

University of Louisville
Jesse H. Wright, MD, PhD
Tracy Eels, PhD
Katherine deVoogd, MA

University of Pennsylvania
Michael Thase, MD
Marna Barrett, PhD
Gregory Brown, PhD

University of Pittsburgh
Stephen Wisniewski, PhD