Data Collection
Database options and Data Management considerations

Paul Donnelly
George Clinical
The George Institute for Global Health
Overview

• Introduction and types of studies
  • Scope of research studies and general considerations in choosing a database

• Database options available
  • Why is a database needed and how is data stored electronically?
  • What options are available

• Data Management support for trials
  • On-line database design, randomisation, data integrations data review and data reporting
GEORGE CLINICAL DATA MANAGEMENT

- The Data Management group provides support to both George Institute (academic Investigator Initiated studies) and also George Clinical commercial sponsors
- Provide database solutions for broad selection of trials
- EDC specialist – all trials conducted using EDC
- Team comprises data entry staff, data managers, clinical programmers and a database administrator
Types of studies

- RCTs – cluster or individual level randomisation
- Cohort studies
- Case-control studies
- Qualitative research
- Survey research
- Data linkage
Why use a database at all?
Why use a database and not a spreadsheet

- More than one person can update a database at a time
- A database stores information more efficiently
- Queries, and the reports based on those queries, are easier to write and run
- In a database, data is stored separately to the application. If the application crashes it will affect all the data in the spreadsheet
- Data recovery is more efficient
- Data is secure and centrally located
Storage and security of data

- All data stored on secured servers
- Password protected files with restricted access to study personnel
- All data is regularly backed up
- Servers are maintained to monitor and ensure efficient performance
- Software is maintained and new versions rolled out effectively and securely
- Disaster recovery databases for all databases
Move from paper to Electronic Data Capture (EDC)

- Need to increase quality data collected
- Speed up data collection
- Have access to real time data
- Resolve basic data collection queries at point of entry
- Cost effective

- Many EDC systems allow for data entry as well as EDC in a ‘hybrid’ system
Choosing an appropriate database

• Choice is dependent on:
  – Type of study
  – How will you collect data
  – Local language requirement
  – Planned use of data
  – Size of study
  – Budget
  – What functionality do you require of the database
Database options

- Incredibly varied array of database solutions available
- Ranging from the ‘free’ Open Source solutions through to the high end “Software as a Service” (SaaS)

- Open-source software (OSS) is computer software with its source code made available for anyone to study, change and distribute the software for any purpose.

- Software as a Service (SaaS), a licensing and delivery model in which software is licensed on a subscription basis and is centrally hosted by the software vendor
More functionality = more cost

- Different systems have very different costing structures primarily due to functionality and validation
- Database administration costs
- Licence fees
- Datapoint fees
- System version upgrade fees
- Infrastructure costs
- Maintenance costs
<table>
<thead>
<tr>
<th>Requirement</th>
<th>Open source</th>
<th>SaaS</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 CFR Part 11 compliant</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Randomisation</td>
<td>?</td>
<td>√</td>
</tr>
<tr>
<td>Multi-language data collection</td>
<td>?</td>
<td>√</td>
</tr>
<tr>
<td>Paper trials / data entry</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Simple data entry</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Basic field checks</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Basic study reports</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Complex multivariate edit checks</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>Query management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>User management tool</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Dictionary coding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard reports</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Ad-hoc reports</td>
<td></td>
<td></td>
</tr>
<tr>
<td>On-demand export</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Cost</td>
<td>Cheap</td>
<td>Expensive</td>
</tr>
<tr>
<td>Functionality</td>
<td>Basic</td>
<td>Extensive</td>
</tr>
</tbody>
</table>
Example of software costs

- “Free” software
  - Server costs
  - DBA costs

- Commercial software
  - Licence $150K pa;
  - Sliding scale of data point fees eg Phase III conducted recently ~$120,000 per quarter

- Some commercial options allow for a lower academic rate; some are free up to a defined threshold
Two database options in use for academic trials within TGI

REDCap

MEDRIO eCLINICAL OVERNIGHT
REDCap (Research Electronic Data Capture)

- REDCap was developed by an informatics core team at Vanderbilt University.
- Designed to address common problems for academic biomedical researchers hoping to use electronic databases.
- REDCap Consortium is a collaborative, international network of more than 250 institutional partners, with more than 20,000 total end-users employing the software with more than 30,000 ongoing research studies and surveys.
REDcap System Features

- REDCap is a secure, web-based application for building and managing online surveys and databases quickly and securely.
- You may create and design projects using:
  - online method from your web browser using the Online Designer.
  - offline method by constructing a 'data dictionary' template file in Microsoft Excel, which can be later uploaded into REDCap.
- Since implementation, functionality has been extended to include randomisation
REDCap @ The George Institute

• REDCap can only be used for non-commercial projects.
• REDCap software is available at no cost
• Technical support is given by our in-house team
• Enables researchers to develop databases
• Requires DBA support at the start of the study, but thereafter researchers are able to cover all tasks
• > 120 projects to date have been set-up within TGI, since 2011
REDCap system

- Easy to use multi-purpose clinical database
- Convenient uniform data entry method
- Reduce data “cleaning” effort during analysis
- Metadata Driven
- Adaptable Features
- Data entry validation
REDCap limitations

• Access – some limitations on restricting access on a role level
• Complex checks across the data are not possible within the system
• Reporting is very limited
  • Filtering; study metrics
Medrio

- Medrio is a commercially available software that provides eClinical SaaS platform
- Medrio provides a secure and regulatory compliant platform (21 CFR Part 11)
- Intuitive point-and-click interface
- Library of forms, variables, and templates
- Dynamic form rules and custom skip logic
- Easy mid-study changes
Collect and manage Data

- Enter data from anywhere via a browser
- Intuitive workflow and task list
- Double-data entry to support Paper or Hybrid studies
- Control data integrity with range checks and queries
- Configurable field level and form level monitoring
- Query management, alerts, and dynamic schedules
Analyse and export

- Capturing and managing data in real-time with a unified database gives you full control
- Ad-hoc reporting and analysis
- On demand export to SAS, Tab-delimited, and others
- Standard and custom reports
- Patient casebooks
Additional Modules as standard

- eClinical Suite offers additional functionality
  - Dictionary coding
  - ePRO
  - Randomisation
  - eCRF file attachments
Data Management is more than a data repository

• Experience – able to advise on all aspects relating to the design, implementation and conduct of a data capture system
• Randomisation
• Database design based on data standards
• Facilitate data integrations (eg lab data; spirometry…)
• Data coding
• Discrepancy management
• Data reporting / metrics
Randomisation

• Originally paper envelopes were sent to site
• There is a trend away from IVRS randomisation (Voice)
• IWRS allows patients to be randomised to a treatment arm on a study via the web (Web)
• Some database packages have an integrated randomisation engine
• Specialised randomisation, eg Minimisation, would need to be set up outside the databases available
  • Require set-up by DM in collaboration with Statistics
Data Integrations

- Integration of data collected on hand held devices
  - Ability top create API’s for integrations
- ePRO – management of electronic diary data
- External data integrations such as lab data, or medical device measurements
Data review and reporting

- Put in place to ensure data quality
- Basic data checks at data entry eg mandatory fields, ranges and simple logic check
- Multivariate checks that generate queries back to site – checks across different data points, forms and complex logic
  - Require review by DM and site
- Generation of study metrics and quality reports
In summary

• The integrity of the data is paramount
• To ensure integrity is maintained, a database will
  • Be held on a secure server
  • Access is restricted by role
  • An audit trail is available
  • All data is regularly backed up and secured

• Ask for help from Data Management wherever possible at the very start of project planning