



DATA MANAGEMENT IN A TRI-NATIONAL AIDS TRIAL

Abdel Babiker, Joel Singer, Tassos C. Kyriakides, Mirek Piasieczny, John Russo, Ruth Goodall;
For the OPTIMA Team



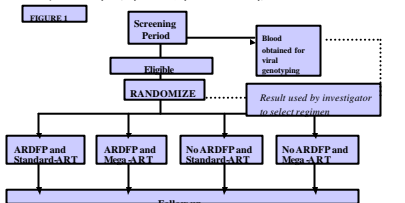
The on-going OPTIMA Management with Antiretrovirals (OPTIMA) Trial, a collaboration between three governmental agencies in the USA, UK and Canada is a large-scale, multicenter, randomized controlled trial designed to compare the relative efficacy of different therapeutic strategies in HIV disease.

The collaboration of three data coordinating centers introduces unique data management issues including: (i) use of different data systems for managing "country" trial data; (ii) two-way data transfer between the coordinating centers and the center where OPTIMA data is merged and analysis files are generated; (iii) translation of certain data forms (mainly patient completed questionnaires) into French and Spanish.

The involvement of three data centers has provided a challenge in planning, designing and executing data management procedures in OPTIMA. Processes have been implemented to ensure the trial's successful execution at the tri-national level, without disregarding country-specific requirements and regulations. The OPTIMA model provides insights on a systematic approach for conduct of such multi-national trials and requires a significant time and resource investment on behalf of all agencies involved.

DESIGN

- 2X2 open-label randomized study design
- Trial Objective:** To assess the effect of antiretroviral drug free period (ARDFP) and mega-ART on clinical outcomes, virologic and immunologic responses, and other healthcare outcomes in the context of clinical management of patients with advanced HIV disease, in whom currently available anti-retroviral therapy (ART) has failed
- Study Duration:** 5.5 years (4.5 years accrual, 1 year minimum follow-up)

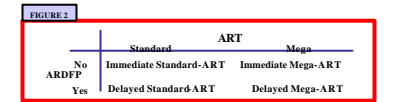


STUDY POPULATION

- Patients in the United States, Canada and the United Kingdom with advanced HIV disease whose anti-HIV regimens that included drugs from all three classes have failed

RANDOMIZATION

- Central randomization list prepared for the whole trial
- Variable block size for large sites or groups of smaller sites within each country
- Screening CD4 cell count stratification (CD4 > 100 or CD4 < 100)
- Each country is responsible for its own randomization process (telephone or fax)
- Patients randomized to either mega-ART (5 or more drugs) or standard-ART (up to 4 drugs) start treatment either immediately or following an ARDFP (Figure 2). For patients randomized to ARDFP, assignment to standard or mega-ART is not communicated until the end of the ARDFP (i.e. the time of drug re-initiation) since immediate knowledge of their eventual drug allocation could influence the initial phase of the study.



STUDY CONDUCT AND MONITORING

- Trial management shared between various committees
- Trial Management Committee (TMC):** (a) day-to-day trial management; (b) resolution of questions regarding trial eligibility, enrollment, randomization, regimen determination, length of drug free periods, grading and disposition of toxicities and adverse events; (c) determination of endpoint questions that need resolution by the Endpoints Review Committee (ERC)
- Trial Steering Committee (TSC):** (a) general oversight of the trial conduct, accrual, retention, quality and data systems, and management in accordance with the protocol and in light of reports or recommendations from any of the trial committees; (b) review and consideration of relevant information or developments in the field of HIV therapy and in legal, statutory, or regulatory requirements
- Data Safety Monitoring Board (DSMB):** (a) monitoring of patient intake, adherence to the allocated treatment strategy and the protocol, baseline assessment of study participants, completeness of data retrieval, and adverse events; (b) safety and treatment strategy efficacy
- Three coordinating centers: (i) the Canadian HIV Trials Network; (ii) the MRC UK Clinical Trials Unit; (iii) the VA Cooperative Studies Program Coordinating Center (CSPCC) (New Haven) oversee the day-to-day operation within each country; randomization, orienting and monitoring clinical sites, quality and timeliness of data and site-specific issues
- Review/monitoring team in each country ensures the trial conduct according to Good Clinical Practice (GCP) as well as country and trial-specific guidelines.

DATA MANAGEMENT

- Data management at country level (Figures 3-4) and tri-national level (Figure 5).
- Each coordinating center is responsible for data processes in its country. Once collected, processed and verified, data are sent to the Canadian coordinating center.
- Data from three countries is merged into the trial database and analysis files are generated

OPTICAL CHARACTER RECOGNITION SYSTEM/CASE REPORT FORMS

- Case Report Forms (CRF) developed using an optical character recognition (OCR) software package (Teleform® Elite v 8.0, by Cardiff Software, Inc.) and shared between the three coordinating centers. All CRF data fields and their formats are uniform across the three countries. Some country-specific CRF changes were made to accommodate for other local requirements and/or parameter units (e.g. different metric systems).
- Guidelines for CRF completion in accordance with GCP and the OCR/software requirements are provided to the participating sites.

ORIGINAL DATA FORMS COLLECTED AT THE PARTICIPATING SITES ARE PROMPTLY SENT TO THE COORDINATING CENTER AND A COPY KEPT AT THE PARTICIPATING INVESTIGATOR'S OFFICE.

SCANNING OF CRF

- Each form is scanned into the OCR application and an image file is generated.
- Data fields not interpretable by the OCR program are presented to the reviewer for verification. Critical fields such as patient ID number are presented for re-view and verification regardless of the OCR program's interpretation. After a form has been verified, the data is automatically exported into a predefined data file.
- Scanning errors are corrected using standard computerized data checking programs. The computerized data check is reviewed and any errors caused by the scanning process are corrected prior to forwarding data queries to the sites.
- OCR software reports are produced periodically in order to review the efficiency of the system and the operators. This report includes a review of manual verifications, per data field and the time needed for processing each CRF.
- Any changes to data entry fields during the course of a trial (revision of forms etc.) are documented and validated to ensure that all previously recorded data is correctly entered.

FILED OF CASE REPORT FORMS

- Paper copies of the CRF are filed in secure file cabinets by site number and patient ID.
- CRF images generated by scanning are archived in a secure database at the coordinating center. The database must bear a one-to-one correspondence with the CRF maintained at the site.

DATA SOURCES

- Two separate data sources: (1) files created from the CRF scanned at each coordinating center; (2) analysis file generated from the merged database. The analysis files incorporate all information collected from each coordinating center from the start of the study and are used for the generation of interim reports and final analyses.
- Oracle 8.0 (UK and Canada) and SAS 8.2 (US) are used for data management. UK and Canada data are mapped in Oracle tables developed at the Canadian coordinating center. US data are sent as SAS transport files and then mapped into Oracle tables.
- Analysis files are extracted from the Oracle tables, formatted into space-delimited files and then posted on a secure FTP site.
- Data flow process was successfully tested prior to the study start-up.

PROCESSING OF INTERNAL DATA

- Internal data processing at each coordinating center (Figure 3) includes data checking and editing according to that center's standards and methods.
- On a monthly basis, each coordinating center submits via a secure FTP protocol data into our database to the Canadian coordinating center. Prior to submitting its database, each coordinating center performs data checks and resolves a majority generated queries as possible. The next set of files sent completely replaces the previous set.

PROCESSING OF EXTERNAL DATA

- Analysis files subsequently posted by the Canadian coordinating center are derived from the merged database. Frequency distributions generated for all the fields in the analysis files are compared to each coordinating center's data for consistency.

INTERNAL/EXTERNAL DATA UPDATE FREQUENCY

- Internal updates occur at the coordinating center level on a weekly basis, or more frequently, as needed.
- Updated analysis files are generated and posted once a month.

DATA CHECKS AND QUERY GENERATION

A. 'Country Level' Database

Validation of Computerized Data Checking

Computerized data checking routines were developed and their accuracy was checked prior to receiving patient data. Pictitious data with built-in errors were checked to ensure that all programs are properly checking queries.

COMPUTER CHECKING

- Computerized data checking routines are run to detect missing and inconsistent data. Certain forms such as Serious Adverse Events and AIDS Event/Death undergo stringent visual inspection. The process is designed to perform checks on one data set at a time. Form-specific data checks as well as cross-form consistency checking is performed at different stages of the data management system.
- When a query is generated, information identifying the error, the offending field and the record in the data set being checked is written to the data check dataset.
- The data-check dataset contains computer-generated queries and manually entered changes to the data. If more than one check is generated for a particular field, only one check is recorded in the data check data set. If a manually entered change is already present for a particular field, it is assumed that it corrects any problems with the particular field, and all generated checks for that field are not recorded in the data check dataset.

STANDARD DATA QUERIES

- Standard checks that may be selected for any field include: (a) a value not in a predefined list; (b) a value less than a predefined minimum; (c) a value greater than a predefined maximum; (d) a value of zero; (e) a blank or missing value.
- Fields may be coded with a special value, e.g. 99, to indicate that the data for that field cannot be obtained. In this case, the standard checks for that particular field are ignored. A field dependency can be coded for each check, whereby the check is only performed if the dependent field meets certain criteria. Criteria can include any of the standard comparison operators (less than, greater than, etc.) and the list comparison operator IN. Each check is run independent of the others. Multiple checks can be specified for the same field. For example, both check for zero value and a blank or missing value may be specified for the same field, since the two checks do not overlap.

CUSTOM DATA CHECKS

- Custom checks are optional. The code for custom checks can involve any type of comparison, calculation, or manipulation that can be done inside a SAS data step, including cross-form checking.

DOCUMENTATION OF REQUESTS FOR CLARIFICATION

- The date on which an error was first queried is automatically recorded so that delay in response to error requests can be readily documented.

QUALITY ASSURANCE PROCEDURES

- Random audits on selected forms check against data scanned into the database.
- The quality of CRF is documented in terms of the number of errors per form by site.

QUALITY ASSURANCE PRIOR TO DATABASE CLOSURE FOR REPORT GENERATION

- All databases are cleaned and rechecked for discrepancies and then frozen/closed.
- To assure data management uniformity certain procedures are in place:
 - Automated data checking program to generate query reports
 - Distribution of computerized query reports to sites for resolution
 - Verification of site personnel initiated on CRF and edited query reports.

B. 'Tri-national Level' Merged Database

Merged Database Checks

- The Canadian coordinating center runs a set of data checks on the merged database.
- Queries generated by scanning are archived in a secure database at the coordinating center.
- The coordinating centers act on the queries, attach comments regarding the status of the queries and return the ASCHI file to the Canadian coordinating center.
- After all the merged database queries have been assigned a status, a file containing pending merged database queries is posted together with the analysis files.

Using the Database

- Database completeness and consistency will be based on the performance of the data checking program and on coordinating center standards.
- Quality assurance checks will be done on (a) all CRF variables of each type, (b) every treatment allocation code, (c) every categorical primary outcome.
- The TSC and TMC will be informed when the study database is complete. When the status of the database meets the criteria of the subsequent errors or inconsistencies are found, with the biosponsors' acknowledgement, the database will be reopened to enter corrections and/or additions. The TSC and TMC will once again be notified.
- Copies of the closed database prior to responding will be maintained and subsequent edits will be kept in a separate database. Once the database is closed, the database is archived at the coordinating center.

