

Data guidelines

The success of the RDI initiative in modelling the relationship between genotype and virological response clearly depends on the nature of the data in the database. The database is being populated with data from patients with comprehensive longitudinal follow-up. The basic unit of data in the database is the treatment change episode (TCE). Each TCE represents a single patient who has their antiretroviral treatment regimen changed (at least one drug changed). The bio-mathematical modelling requires that as a minimum we have the viral load and genotype at the point of treatment change, the drugs used in the new regimen, and the viral load at one or more subsequent time points, ideally at 24 weeks but at least 12 weeks or more after the treatment change. Additional information being collected includes the treatment history of the patient and the reason for treatment change (virologic failure, tolerability, etc). In particular it is important to indicate if treatment is being initiated following an interruption to treatment that is known to be of three days or more duration because of the possibility of undetectable resistant minority quasi-species at baseline.

All data entered into the database will be stripped of any source patient identifiers and each data set assigned a unique identifier.

In summary the data being collected are as follows:

Mandatory Fields:		
Unique patient identifier	Genotype test date	Drugs in new regimen
Patient sex	Reference sequence	Response viral load(s)
Paediatric/adult variable: Is patient >12 years YES or NO	Protease start/end index	Follow-up viral load date(s)
Viral load (HIV RNA) assay used	RT start/end index	
Baseline viral load	Generated protease sequence	
Viral load date	Generated RT sequence	
Optional Fields:		
CD4 Count	Drug adherence indication	Reason for treatment change (0=failure, 1=first treatment, 2=post-STI, 3= tolerability or other reason)
CD4 date	Plasma drug levels	
Phenotypic test results	Treatment history	
Phenotype test method and date		

QA standards and minimum data requirements

There will be strict quality assurance of the database with clinical data having to conform to one of two standards:

1. Clinical trials standards - compliance with Good Clinical Practice (GCP) standards
2. Cohort standards - compliance with the principles of the Declaration of Helsinki

Data meeting either of these standards will be flagged in the database as 1 or 2 respectively.

Sequence Data

Sequence data will also be fully quality assured and must comply with the following minimal length and standard criteria:

Protease: A bi-directional sequence of the protease gene from codons 10 to 99.

RT: A bi-directional sequence of the reverse transcriptase gene from codons 41 to codon 235.

Data format and transfer

For ease of transfer, the preferred format is as an Excel spreadsheet. The data can also be provided as an Access spreadsheet or as a CSV file (Comma Separated Variable). The exact arrangement of the data within these formats is not critical and the data will undergo thorough review and re-organization as part of the input process.