INFORMATION FOR DATA CONTRIBUTORS

This document is designed for institutions considering the provision of data to the HIV Resistance Response Database Initiative. It provides information on the following topics: the data required; practicalities of transfer; security and access; use of the data; acknowledgement; publications, data ownership and intellectual property.

The RDI

The HIV Resistance Response Database Initiative (RDI) is a not-for-profit initiative, registered in the UK, with the following mission:

To improve the clinical management of HIV infection by developing a large clinical database and bioinformatic techniques that predict accurately any individual's response to any combination of HIV drugs.

Our approach and progress
The approach pioneered by the RDI involves relating clinical and laboratory parameters directly to the virological response of patients to treatment in clinical practice, using a substantial database of clinical data and computational modelling.

Summary of progress to date
Data from around 250,000 patients have been contributed by more than 50 institutions and cohorts covering approximately 50 countries around the world. Computational models have been developed that are predicting virological response with typical accuracy of 80%. In October 2010, the RDI launched its HIV Treatment Response Prediction System, **HIV-TRePS**, enabling healthcare professionals anywhere in the world to use the models to obtain predictions of the probability of any combination of antiretroviral drugs producing a virological response. In July 2011, the RDI developed and made available new models that do not require a genotype for their predictions, relying instead on viral load, CD4 count and treatment history variables. This version of TRePS is designed specifically for resource-limited settings where genotyping may not be accessible. HIV-TRePS has been used in approximately 80 countries worldwide.

Data required for models

The data required by the RDI for modelling virological response to combination therapy are:

- Viral loads (absolute or $\log_{10}$ number of copies HIV RNA/mL and the dates of the samples
- CD4 counts (cells/mL) and dates of the samples
- Names of antiretroviral drugs and formulations used (fixed dose combinations differentiated from single agents), at any dose, in treatment and their start and stop dates, including as complete a history of treatment as possible
- Where available, genotypes (preferably protease and RT sequences or, at minimum, all changes from wild type sequence) and dates that the samples were obtained.

In addition to these essentials the following data are collected wherever possible and are playing an increasingly important role in our research:

- Any additional historical drug exposure information
- Any measures or estimates of adherence and dates
- Date or year of birth or current age (subject to local data protection requirements)
- Sex
- Other demographics, if available (subject to local data protection requirements)
- Integrase sequence or mutation data

The preference is for long-term longitudinal data covering multiple treatment changes per patient.

QA standards and minimum data requirements
Clinical data should 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

Sequence Data specification
Sequence data will be fully quality assured and should 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
Data can be accepted by the RDI in almost any format. The most common formats in which data have been provided are:

- MS Excel spreadsheets
- CSV text files
- MS Access files
- Oracle tables
- SAS files

Data transfer

Data are usually provided to the RDI as an email attachment. Other arrangements can be made on request.

Use of the data

The data will be used by the RDI in pursuit of its stated mission. On receipt, 'cleaning' and completion of any data clarification work, the data are imported into the RDI Oracle database. Data are then extracted as individual 'treatment change episodes' (TCEs) for modelling purposes.

The initiative may involve specific research studies conducted in collaboration with one or more third parties (collaborating research institutions). Collaborating third parties will not have access to any of the raw data held by the RDI.

Security, storage and access

Patient anonymity
All data provided to the RDI must first be stripped of all patient identifiers and allocated a unique code by the source clinic (to enable data clarification queries after transfer).

Secure data storage
All raw data is stored on a secure password-protected stand-alone Oracle 12c database, which is housed behind an SPI True Firewall. This is regularly backed-up, including whenever new data are added. All web-based data (computational models, TCEs, the RDI web site and user interfaces) are stored in MySQL databases, on a secure server with state-of-the-art facilities. This includes 18,000 MBit fiber optic connectivity with nine leading carriers to ensure our databases and online applications are amongst the fastest and most secure.

Access
No one other than RDI personnel is able to access the data in the RDI databases. Employees and consultants of the RDI are legally prohibited from providing data to any third parties. Researchers or clinicians using the proposed online tool will not be able to access, view or download data, either in raw form or as TCEs. The RDI is fully compliant with the General Data Protection Regulation (2018)

Acknowledgement

Organisations contributing data will be acknowledged on the RDI website, RDI publications, acknowledgement slides and posters where research studies have been conducted utilising those data specifically.

Publications and presentations

The RDI encourages academic institutions not only to contribute data but also to suggest collaborative studies that can be conducted using those data or the RDI database. Members of collaborating institutions are encouraged to present the results of such studies. The RDI is also keen to develop and publish joint publications with members of collaborating institutions. Results of any studies using specific datasets will only be presented with the prior review and approval of the collaborating organisation that provided the data. Where journal authorship policies permit, contributors of substantial datasets used in the RDI studies are included as authors on RDI publications.

Ownership and Intellectual Property

Data provided to the RDI remains the property of the contributing organisation and ownership is flagged in the database. Data can be withdrawn from the database by the organisation at any time.

The models developed by the RDI, for example linking baseline variables to virological response, are the property of the RDI. The RDI is committed to providing open and free access to use these models to clinicians, healthcare workers and researchers for use as a research tool.

Contact for further information

Andrew Revell PhD Executive Director andrewrevell@hivrdi.org Tel: +44 207 226 7314