

# 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 define with increased precision and reliability the relationships between HIV resistance and virological response to treatment.*

### **Our approach, methods and aims**

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

The principle aim of the RDI is to provide a treatment decision-making aid free of charge over the Internet such that physicians entering the genotype and other baseline data for a patient will receive a report containing predictions of virological responses to a range of alternative antiretroviral combinations.

### **Summary of progress to date**

To date, data from around 65,000 patients has been contributed by approximately 20 institutions in countries around the world: including USA, Canada, Australia, Japan, Italy, Spain, The Netherlands, Germany, UK and South Africa. Computational models have been developed that are predicting virological response with increasing accuracy. Data have been presented at numerous international conferences.

## Data required

---

The data required by the RDI for modelling the relationship between genotype, other historical and baseline factors and virological response to combination therapy are:

- Genotypes (preferably protease and RT sequences or, at minimum, all changes from wild type sequence) and dates that the samples were obtained
- Viral loads and dates of samples
- Names of antiretroviral drugs used (at any dose) in treatment and their start and stop dates – as complete a history of treatment as possible
- CD4 counts and dates

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
- Measures or estimates of adherence and dates
- Viral clade
- Age
- Sex

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 can be provided to the RDI in one of two ways:

1. As an email attachment (the most common route to date)
2. By uploading the files to the RDI https server. This requires a password (details provided 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 10G 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.

## 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 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.

## 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 such as genotype 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](mailto:andrewrevell@hivrdi.org)

Tel: +44 207 226 7314