Data Format Specification for Submission of Data for Collaborative Data Analysis Project: *Refinement of Viral Genotype Interpretation Systems for DDI and Abacavir Based on Virologic Response*

Contents

4 - Nov - 03

1. Introduction

This document sets out the data format that we would like for data sets that are to be included in the collaborative analysis detailed in the Ananlysis Plan entitled "Refinement of Viral Genotypic Systems for DDI and Abavacir, based on Virologic Response". If you wish to contribute data for analysis but would have difficulty putting files in this format, you should contact the Forum to discuss how we might arrange for this data manipulation to be done.

2. Inclusion criteria for analysis

The inclusion crieria, as specified in the Analysis Plan, are as follows.

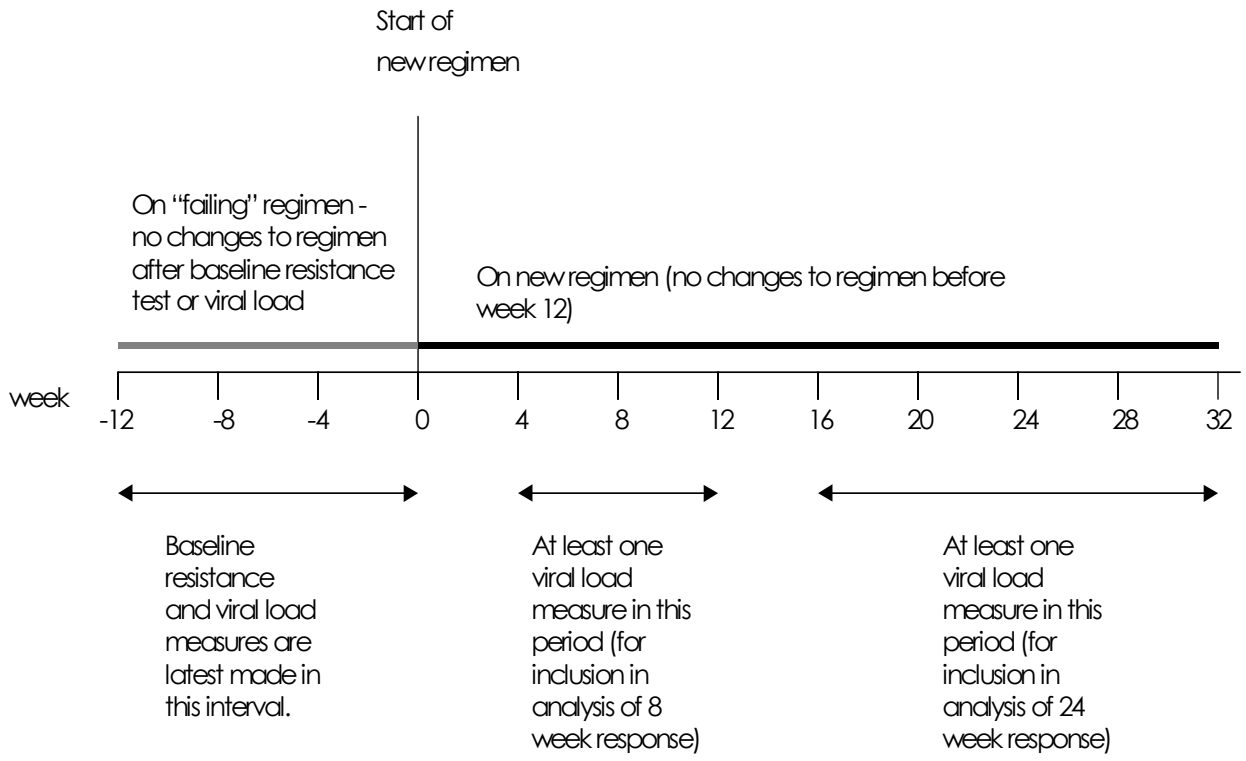
Drug experienced people starting a new regimen including the drug under consideration (being used for the first time) are eligible for inclusion if the following criteria are met.

- Virologically failed the previous regimen (according to the clinician's judgement)
- An available genotypic resistance test on the previous regimen (measured < 12 weeks before start of new regimen) while on the previous regimen.
- An available viral load measure while on the previous failing regimen (which must also be < 12 weeks before start of new regimen). This is the baseline viral load. This viral load should be at least 500 copies/mL.
- At least one viral load measured between 4-12 weeks (the 8 week viral load) or between 16-32 weeks (the 24 week viral load) from the start of the new regimen (containing the drug under consideration)

- There are no changes in therapy between the time of the baseline viral load or resistance test and the start of the new regimen, nor between the time of the start of the new regimen and week 12.

- There is no evidence of inadequate adherence to the new regimen.

This is illustrated in the following figure.



3. File format specifications

Please provide either files 1-3 or files A-D, plus files R1 & R2.

Note, if files A-D are provided (rather than files 1-3) then it is not necessary to select patients who are eligible for the analysis. This can be done centrally. Further, if the option of providing files A-D is taken, files may be submitted according to the "HICDEP" protocol developed between HIV cohorts.

Files should preferably be in SAS format.

File 1 – fixed variables

Variable	Name	format / coding
Unique id	id	
Date of birth	birth_d	ddmmmyyyy
HIV Exposure (judged most likely route)	risk	1=msm, 2=idu, 3=heter, 4=other/unknown
Gender	gender	1=male, 2=female
Ethnicity	ethnic	1=white, 2=African American, 3=hispanic 4=black African, 5=other/unknown
Date of start of new regimen (baseline) (ie date of start of abacavir)	t0_d	ddmmmyyyy
Viral load at baseline (<12 weeks from baseline, value taken while on previous regimen and closest to baseline date)	vl0	
Assay lower limit of above viral load	vl0_ll	
Date of above viral load	vl0_d	ddmmmyyyy
CD4 count at baseline (<12 weeks from baseline, value taken while on previous regimen and closest to baseline date)	cd40	
Date of above CD4 count	cd40_d	ddmmmyyyy
Previous CDC C diagnosis before baseline	cdc_c	1=yes, 2=no
Minimum viral load ever, pre-baseline	vlmin	If undetectable, code as “0”.
Assay lower limit of above viral load	vlmin_ll	
Maximum viral load ever, pre-baseline	vlmax	
Assay lower limit of above viral load	vlmax_ll	
Minimum CD4 count ever, pre-baseline	cd4min	

Date of resistance test (<12 weeks from baseline, taken while on previous regimen and, if more than one, that closest to baseline date)	res_d	ddmmmyyyy
Patient known to be poorly adherent to new regimen during first 32 weeks ?	adh	1 = yes, 2 = no
Week 8 viral load (closest value to 8 weeks, within 4-12 week window)	vl8	
Date of above viral load	vl8_d	ddmmmyyyy
Assay lower limit of above viral load	vl8_ll	
Week 24 viral load (closest value to 24 weeks, within 16-32 week window)	vl24	
Date of above viral load	vl24_d	ddmmmyyyy
Assay lower limit of above viral load	vl24_ll	
Date of first change in drug regimen between week 12 and 32 ?	first_ch	ddmmmyyyy (blank if no change)
If regimen changed between weeks 12-32, most recent viral load before change	vlch	
Date of above viral load	vlch_d	ddmmmyyyy
Assay lower limit of above viral load	vlch_ll	
Viral load assay method (must be the same method for baseline and response-defining viral load, for any given patient)	vlmeth	1 = Roche 2 = Chiron 3 = NASBA 4 = Other (specify below)
Viral load assay method Specify details.	vlmeth_sp	text

File 2 – drugs ever used prior to baseline

(one line for each drug)

Variable	Name	format / coding
Unique id	id	
Drug	p_drug	see drug codes below
Was drug in previous failing regimen immediately preceeding baseline ?	imm_prec	1=yes, 0=no

File 3 – drugs in new regimen

(one line for each drug in new regimen (include abacavir))

Variable	Name	format / coding
Unique id	id	
Drug	drug	see drug codes below
Date started drug	start_d	ddmmmyyyy
Drug was added at time zero, with abacavir	start_0	1=yes, 0=no

File A – fixed variables

Variable	Name	format / coding
Unique id	id	
Date of birth	birth_d	dd mm yyyy
HIV Exposure (judged most likely route)	risk	1=msm, 2=idu, 3=heter, 4=other/unknown
Gender	gender	1=male, 2=female
Ethnicity	ethnic	1=white, 2=African American, 3=hispanic 4=black African, 5=other/unknown
Date of start of new regimen (baseline) (ie date of start of abacavir)	t0_d	dd mm yyyy

File B - all antiretroviral data

Variable	Name	format / coding
Unique id	id	
Drug	drug	See below for coding*
Start	start_d	ddmmmyyyy
Stop	stop_d	ddmmmyyyy

File C – all viral load data

Variable	Name	format / coding
Unique id	id	
Date viral load	vl_d	ddmmmyyyy
Viral load value	vl	
lower limit of assay	llim	

File D – all CD4 count data

Variable	Name	format / coding
Unique id	id	
Date CD4 count	cd4l_d	ddmmmyyyy
CD4 count	cd4	

File R1 – genotypic resistance test data

Can include results from more than one resistance test per patient. One line of data per codon, per gene, per patient.

Can provide data either

(i) for all codons

or

(ii) only for codons, between 1 and 99 for PR and 1 and 230 for RT, where an amino acid differs from HXB2 or where amino acid is unknown.

Variable	Name	format / coding
Unique id	id	
Date of sample for test	sample_d	ddmmmyyyy
Gene	gene	1=rt, 2=pr
Codon number	codon	note: use letters for insertions (eg at codon 69, 69a, 69b, etc..)
Amino acid 1	aa1	standard amino acid letter code (0 = deleted, X = not sequenced/unknown * = start / stop codons)
Amino acid 2	aa2	standard amino acid letter code
Amino acid 3	aa3	standard amino acid letter code
Amino acid 4 (2-4 are used when there is a mixture)	aa4	standard amino acid letter code

Example (option (i))

Id	sample_d	gene	codon	aa1	aa2	aa3	aa4
.	.	.	.	.			
.	.	.	.	.			
073	12APR2000	2	81	P			
073	12APR2000	2	82	V			
073	12APR2000	2	83	N			
073	12APR2000	2	84	V			
073	12APR2000	2	85	I			
073	12APR2000	2	86	G			
073	12APR2000	2	87	R	I		
073	12APR2000	2	88	N			
073	12APR2000	2	89	L			
073	12APR2000	2	90	L			
.	.	.	.	.			
.	.	.	.	.			

Using option (ii), only the lines where an amino acid differes from HXB2 would be included…

Id	sample_d	gene	codon	aa1	aa2	aa3	aa4
.	.	.	.	.			
.	.	.	.	.			
073	12APR2000	2	84	V			
073	12APR2000	2	87	R	I		
.	.	.	.	.			
.	.	.	.	.			

File R2 – genotypic resistance test method

(one line per test, per person)

Variable	Name	format / coding
Unique id	id	
Date of test	test_d	ddmmmyyyy
Method	r_method	text
Format	format	amino acid provided for all codons in R1 (ie option (i)) = 1
		amino acid provided only for where amino acid differs from HXB2 or where amino acid is unknown (ie option (ii)) = 2

Coding for antiretrovirals

1	zidovudine
2	ddC
3	ddI
4	d4T
5	3TC
6	abacavir
7	combivir
8	trizivir
9-18	use these code numbers for other NRTI's
20	nevirapine
21	efavirenz
22	delavirdine
29-39	use these code numbers for other NNRTI's
40	saquinavir hard gel (invirase)
41	indinavir
42	ritonavir
43	nelfinavir
44	saquinavir soft gel (fortovase)
45	amprenavir
46	lopinavir/r
49-59	use these code numbers for other PI's
60	T20

4. Naming of files

Two letter code for trial / dataset - XX

Three letter code for drug (abacavir - ABA or didanosibe - DDI)

File id - Y

Separated by underscores

XX_ABA_Y

Eg. For narval trial (NA), for abacavir, file 1 would be NA_ABA_1, file R2 would be NA_ABA_R2, etc.