HIV Resistance Collaborative Group

Data Analysis Plan for Resistance Studies

Revised - 31 August, 1999

1. SUMMARY

The RCG has as a mandate the formulation of recommendations on the appropriate use of resistance testing in clinical trials and clinical practice. To this end, the group created the Clinical Validation Subcommittee which is responsible for identifying completed or ongoing studies which could help characterize the correlation between HIV-1 phenotype/genotype and virologic outcomes.

The subcommittee was also mandated to draft definitions of treatment failure, treatment response, and relevant time points for analysis of clinical studies in a uniform manner. Potential standardized methods of data analysis that could be applied to such studies were also to be explored. These areas are addressed in this Draft Data Analysis Plan for Resistance Studies.

2. INTRODUCTION

2.1. Although clinical endpoints (e.g. death or OI) are the least controversial, we recognize that few resistance studies are of sufficient duration to measure this type of endpoint. Hence, we recommend that a virological endpoint based on HIV-1 RNA be used.

2.2. When reporting study results, researchers should note which assay was used to measure HIV-1 RNA, frequency of HIV-1 RNA sampling, as well as when samples for baseline resistance were taken.

3. OBJECTIVE OF THE DAP

3.1. Specify a data analysis plan with a minimum of required analyses and covariates that can be applied to the wide variety of clinical HIV resistance studies to produce a coordinated set of results. Use of a common DAP across studies will facilitate comparison of results from different studies.

3.2. Intervention vs. Non-intervention Studies: The objectives of intervention studies (GART and Viradapt) are different from the objectives of non-intervention studies. In general, intervention studies are designed to investigate the effect of the intervention on patient outcome. Thus, these studies test whether the availability of resistance assay results (possibly with an expert interpretation) has an effect on patient outcome. In contrast, the objective of most non-intervention resistance studies is to investigate whether specific results of resistance tests (e.g. a mutational pattern or inhibitory concentration) are predictive of patient outcome. Thus the analysis methods and the conclusions that may be drawn from the results of these two types of studies may differ.

3.3. Resistance data in an experienced population is the primary focus. Naïve patients and experienced patients should be analyzed separately

3.4. Patients without resistance data should be excluded from all analyses.

3.5. If sample size permits then distinct regimens should be analyzed separately. For example, if the study has a PI-containing regimen and one without a PI, separate analyses would be preferred.

4. NON-INTERVENTION STUDIES

4.1. Study population

The following baseline summaries are suggested for characterization of the study populations.

4.1.1. Summary of Baseline $\log_{10}$ HIV-1 RNA and CD4

4.1.2. Summary of Prior Antiretroviral Treatment (ART) experience to include if possible

For each ART medication, the proportion of patients in the 3 categories below

- Not exposed or exposure less than 1 week
- 1 week – 1 year
- greater than 1 year

Summary of duration of therapy on an ART regimen (i.e., at least three drugs)

If this information is not available, please provide a summary of exposure to the level of available detail.

4.2. Metrics

4.2.1. Endpoints: The primary endpoint is whether or not a patient experiences 'virological failure' by week 24. In general, virological failure will be determined by whether or not a patient's HIV-1 RNA is below a specified threshold (e.g., 400 copies/ml) at 24 weeks post therapy initiation while on original study regimen.

For studies where few patients obtain HIV-1 RNA < 400 copies/ml change from baseline in HIV-1 RNA may be used.

4.2.2. Determination of virological failure

A. **Transient Increases in HIV-1 RNA:** Patients whose HIV-1 RNA is greater than the threshold at week 24, but whose HIV-1 RNA at the next determination is below the threshold despite no change in therapy should NOT be counted as having experienced virological failure by week 24.

B. **Week 24 window**: The window for the HIV-1 RNA measurement at week 24 should be broad. We suggest a window from week 16 to week 32, where 'week' refers to actual time measured from treatment initiation. When more than one HIV-1 RNA is measured in this window, the measurement closest to week 24 should be used. In the event of a tie, use the later of the two measurements.

- For studies of duration shorter than 24 weeks but at least 16 weeks, the last week for which there is sufficient data should be used.

- For studies of duration shorter than 16 weeks but at least 8 weeks, change from baseline in HIV-1 RNA may be used as an endpoint and an analysis similar in spirit to that described here should be used.

C. **HIV-1 RNA failure threshold**: For each study an HIV-1 RNA threshold should be set for the determination of virological failure. If a threshold was specified in the original study protocol that threshold should be used. Otherwise, a threshold of < 400 copies/ml should be used in the primary analysis. A secondary analysis using <50 copies/ml should also be provided if the researcher has these values available and if the results are substantially different from a <400 copies/ml analysis. If not substantially different, a statement that a similar analysis using the lower threshold yields similar results would be sufficient.

D. **Patients who withdraw early from study medication (permanently discontinue any component of the new drug regimen):** We request two analyses of each dataset. The two analyses represent the extremes for handling early patient withdrawals. In the DAC ('dropouts as censored') analysis, all data from patients who withdraw early from study medication without evidence of virological failure are treated as censored and dropped from the analysis. In a second analysis, the DAF ('dropouts as failures') analysis, all patients who

withdraw early are counted as failures. Details are given in flowcharts (See Appendix 1) as well as discussed below.

More specifically, for the DAC analysis and a given patient who withdrew prior to the beginning of the 24-week window:

- if the patient's last HIV-1 RNA value measured on study regimen is < threshold then he/she is censored (i.e., treated as missing) in the DAC.

- if the patient's last measured HIV-1 RNA value on study regimen is > threshold then he/she is scored as a failure in the DAC analysis if any of the following 3 conditions apply. Otherwise, he/she is scored as censored

  - Patient has at least two HIV-1 RNA measurements ≥ baseline HIV-1 RNA (confirmed non-responder).

  - Reduction in HIV-1 RNA from baseline is < 0.5 log10 HIV-1 RNA for all values measured between and including weeks 4 through week 8 (substantial evidence of lack of response).

  - HIV-1 RNA nadir is below the threshold (evidence of rebound).

E. Patients on study regimen for at least 16 weeks who have no HIV-1 RNA values in the week 24 window: Exclude from both the DAC and DAF analyses. We suggest that investigators report the proportion of patients who were on study regimen for at least 24 weeks and do not have an HIV-1 RNA value in the week 24 window

F. Reporting proportion of withdrawals due to virological failure:

We suggest that investigators report

- the definition of virological failure used in their study

- the number of patients who withdrew early

- the proportion of these who withdrew due to virological failure if that information is available

- the proportion of these who withdrew for other reasons, e.g. adverse experiences, lost to follow-up, etc, if that information is available.

4.2.3. Resistance Measures

4.2.3.1. Genotypic Measures: For each drug, CVSG will construct a list of positions and substitutions that are believed to increase the drug's *in vitro* IC50 (Table 1).

TABLE 1. Algorithm for Mutations Associated with Resistance to Specific Antiretroviral Drugs (Resistance Collaborative Group, unpublished)

(To be used for determination of genotypic sensitivity scores for each drug in a regimen for the RCG DAP, as follows: 0 if a mutation listed for that drug is present; 1 if none of the mutations listed for that drug are present. Additional criteria for scoring special cases are listed below the table.)

Nucleoside Reverse Transcriptase Inhibitors	
Zidovudine (ZDV)	K70R; T215Y or F; M41L; D67N; L210W; K219Q
Stavudine (d4T)	See MNR-1 and MNR-2, below (Note : V75T may be selected in vitro by d4T, but it is rare in vivo and not clearly associated with d4T failure. Therefore, it should not be counted in a genotypic sensitivity score as indicative of d4T resistance.)
Didanosine (ddI)	L74V; K65R; M184V or I
Zalcitabine (ddC)	K65R; T69D; L74V; M184V or I
Lamivudine (3TC)	M184V or I
Abacavir (ABC)	Any 3 or more of the following: M184V; K65R; L74V; Y115F; T215Y or F ; M41L; D67N; K70R; L210W; K219Q
Multi-Nucleoside Resistance-1 (MNR-1)[1]	Q151M; Secondary: A62V; V75I; F77L; F116Y Note: In some cases, higher levels of resistance may be seen when the secondary mutations listed above are added to the Q151M mutation, although those secondary mutations may not cause multinucleoside resistance by themselves and therefore should not be counted in a genotypic sensitivity score as indicative of multinucleoside resistance unless Q151M is also present.)
Multi-Nucleoside Resistance-2 (MNR-2)[2]	3 amino acids encoded by an insertion between RT codons 69 (69Ins) and 70; Secondary: A62V; M41L; D67N; K70R; L210W; T215Y or F; K219Q (Note: In some cases, higher levels of resistance may be seen when the secondary mutations listed above are added to the 69Ins mutation, although those secondary mutations may not cause multinucleoside resistance by themselves and therefore should not be counted in a genotypic sensitivity score as indicative of multinucleoside resistance unless 69Ins is also present.)
Nucleotide Reverse Transcriptase Inhibitor	
Adefovir (ADV)	K65R; K70E (Note: M184V causes increased susceptibility to ADV)
Multi-Nucleoside Resistance-2 (MNR-2)[2]	See Nucleoside Reverse Transcriptase Inhibitor Section above for details

Non-Nucleoside Reverse Transcriptase Inhibitors	
Nevirapine (NVP)	K103N; V106A; V108I; Y181C or I; Y188C or L or H; G190A or S
Delavirdine (DLV)	K103N; Y181C; P236L
Efavirenz (EFV)	K103N; Y188L; G190S or E Secondary: L100I; K101E or Q; V108I; Y188H; P225H (Note: In some cases, higher levels of resistance may be seen when the secondary mutations listed above are added to the K103N mutation, although those secondary mutations may not cause resistance by themselves and therefore should not be counted in a genotypic sensitivity score as indicative of resistance unless K103N is also present.)
Protease Inhibitors	
Indinavir (IDV)	V32I; V82A or T or F; I84V; L90M
Ritonavir (RTV)	V32I; V82A or T or F or S; I84V; L90M
Saquinavir (SQV)	G48V; V82A or T; I84V; L90M
Nelfinavir (NFV)	D30N; V82F; I84V; L90M
Amprenavir (APV)	V32I; I50V; I84V

[1]This MNR profile is associated with resistance to ZDV, d4T, 3TC, ddI, ddC, and abacavir.
[2]This MNR profile is associated with resistance to ZDV, d4T, 3TC, ddI, ddC, abacavir, and adefovir.
Further criteria for determining genotypic sensitivity scores :

1. Adefovir susceptible wild type virus counts as 1.0 and adefovir hyper-susceptible RT M184V counts as 1.5.
2. Primary analysis should not allow RT 184V reversal of ZDV resistance to be considered. In other words, ZDV resistance would be the interpretation whether or not RT 184V is also present. A second analysis using an interpretation of ZDV resistance reversal by 184V could be done at the investigators' option.
3. In the absence of any of the mutations listed for d4T, ddI, or 3TC in the table above, a genotypic sensitivity score of 0.75 will be assigned for d4T, ddI, or 3TC if three or more of the following mutations are present: RT M41L; D67N; K70R; L210W; T215Y or F; K219Q. This is based on data suggesting that there may be a cross-class effect of these mutations on responses to other NRTIs and that they may sometimes be selected *in vivo* by d4T , even though phenotypic cross-resistance has not been clearly defined *in vitro* (references available on request).

Two measures of genotypic information will be explored.

The first, a genotypic sensitivity score, is based on the number of drugs in the study regimen to which the patient has genotypic sensitivity as defined in Table 1. For each drug in the regimen, the genotypic sensitivity is 0 if genotypic resistance is present based on the algorithm and 1 if the patient's genotype is sensitive. There are exceptional cases, described in Table 1, in which scores of 1.5 and 0.75 are assigned. Also note that a mixture of mutations at the same position, e.g. M184V and M184I, should be handled as a single mutation.

If the result at a codon is indeterminate/unknown/missing then for the purpose of this analysis the unknown genotype will be considered to be one which does not confer resistance. It is expected that this will be infrequent and the investigator should comment upon the amount of missing data.

The overall genotypic sensitivity score is defined as the sum of the genotypic sensitivities over all the drugs in the regimen.

The second measure consists of one or more of the following variables - the number of PI mutations, the number of NRTI mutations (including the nucleotide RTI, adefovir), and the number of NNRTI mutations from all drugs in Table 2, i.e., not limited to the drugs in the patient's regimen. Because the number of mutations to a class not present in the study regimen is unlikely to be predictive of virological failure, only the classes of drugs in the study regimen will be modeled; hence, this measure is comprised of at most three variables.

Table 2. Mutations Listed per Drug Class (To be used for determination of total number of mutations in each drug class)

Drug Class	
NRTI[1]	M41L A62V D67N K65R 69Ins T69D K70R L74V V75I F77L Y115F F116Y Q151M M184V or I L210W T215Y or F K219Q
Nucleotide RTI[2]	K65R K70E 69Ins
NNRTI[3]	A98G L100I K101E or Q K103N V106A V108I Y181C or I Y188C or L or H G190A or S or E P225H P236L

PI[4]	D30N V32I G48V I50V V82A or F or T or S I84V L90M To be counted only if present with one or more of the mutations listed above : L10F or I or R or V K20M or R L24I L33F M36I M46I or L I47V I54V or L A71V or T G73S or A V77I N88D

[1] NRTI: nucleoside reverse transcriptase inhibitor, includes: zidovudine (ZDV), stavudine (d4T), lamivudine (3TC), didanosine (ddI), abacavir (ABC) and zalcitabine (ddC)
[2] Nucleotide RTI includes adefovir (ADV)
[3] NNRTI: non-NRTI, includes: nevirapine (NVP), delavirdine (DLV), and efavirenz (EFV)
[4] PI: protease inhibitor, includes: indinavir (IDV), saquinavir (SQV), nelfinavir (NFV), ritonavir (RTV), and amprenavir (APV)

4.2.3.2. Phenotypic Measures: In general, we suggest that phenotypic measures of resistance based upon the fold-resistance measure and handled as an ordered categorical variable.

Two metrics of phenotypic information will be explored.

1. Using the Minimum cut off for the assay

   - S – less than or equal to the minimum cut off for the assay
   - R – greater than the minimum cut off for the assay

2. Using 10-fold as a cut off

- S – less than or equal to 10-fold
- R – Greater than 10-fold

Two measures of phenotypic information will be explored.

The first, a phenotypic sensitivity score, is based upon the number of drugs in the study regimen to which the patient has phenotypic sensitivity. For each drug in the regimen, the phenotypic sensitivity score is defined as 1 if sensitive (S), or 0 if resistant (R).

The overall phenotypic sensitivity score is defined as the sum of the phenotypic sensitivities over all the drugs in the regimen.

The second measure consists of one or more of the following variables - the number of PI drugs, the number of NRTI drugs (including the nucleotide RTI, adefovir), and the number of NNRTI drugs in the study regimen to which the patient has phenotypic sensitivity (PI phenotypic sensitivity, NRTI phenotypic sensitivity, and NNRTI phenotypic sensitivity). Because the number of drugs in a class not present in the study regimen to which the patient has phenotypic sensitivity is unlikely to be predictive of virological failure, only the classes of drugs in the study regimen will be modeled; hence, this measure is comprised of at most three variables

4.2.4. Confounding Variables Measured Prior To Therapy Initiation

1. Baseline $\log_{10}$ HIV-1 RNA

2. Potent PI or NNRTI Y/N – Y if naive to PIs and a PI is in the regimen OR if NNRTI-naïve and an NNRTI is in the regimen; N otherwise. To establish consistency, please code Yes as 1 and No as 0.

3. Number of new drugs (less than 1 week of prior exposure) in the regimen

   - In ddI-naïve subjects, ddI + hydroxyurea (HU) counted as ONE new drug
   - In ddI-experienced subjects, ddI + hydroxyurea (HU) counted as 0.5 new drug
   - A "mini-dose" of ritonavir (i.e., 100 or 200 mg BID) should NOT be counted as a new drug

4.3. Method of Analysis:

The recommended method of analysis will be logistic regression of the binary response virological failure by week 24, i.e. logit(Pr(Failure)), with a common set of covariates. All variables in the model are handled as continuous (not categorical).

In addition to the analyses requested below, investigators are encouraged to report results using other models and/or analysis techniques. It would be helpful when doing so if the investigator provided an explanation as to why the particular method was used and why the results may be different from those obtained using the method described above.

For studies where few patients obtain HIV-1 RNA < 400 copies/ml change in HIV-1 RNA from baseline may be used and linear regression for the models below should be performed.

Note: Patients with missing data should be excluded from all analyses.

4.3.1. Models for studies with Genotypic data: The basic strategy will be to perform at least 6 model fits (A – F) with the following covariates

A. Baseline $\log_{10}$ HIV-1 RNA

B. New Drug Covariates

C. Overall Genotypic Sensitivity Score

D. # of PI mutations, # of NRTI mutations, # NNRTI mutation

E. Baseline $\log_{10}$ HIV-1 RNA, New Drug Covariates, Overall Genotypic Sensitivity Score

F. Baseline $\log_{10}$ HIV-1 RNA, New Drug Covariates, # of PI mutations, # of NRTI mutations, # NNRTI mutation

Where

New Drug Covariates are

- Potent PI or NNRTI Y/N – Y if naive to PIs and PI in regimen OR if NNRTI-naïve and NNRTI in regimen; N otherwise
- Number of new drugs in the regimen

4.3.2. Models for studies with Phenotypic data: The basic strategy will be to perform at least 10 model fits (A – J) with the following covariates

A. Baseline $\log_{10}$ HIV-1 RNA

B. New Drug Covariates

C. Overall Phenotypic Sensitivity Score (based on assay minimum cutoff)

D. Overall Phenotypic Sensitivity Score (based on 10-fold cutoff)

E. PI phenotypic sensitivity, NRTI phenotypic sensitivity, and NNRTI phenotypic sensitivity (based on assay minimum cutoff)

F. PI phenotypic sensitivity, NRTI phenotypic sensitivity, and NNRTI phenotypic sensitivity (based on 10-fold cutoff)

G. Baseline $\log_{10}$ HIV-1 RNA, New Drug Covariates, Overall Phenotypic Sensitivity Score (based on assay minimum cutoff)

H. Baseline $\log_{10}$ HIV-1 RNA, New Drug Covariates, Overall Phenotypic Sensitivity Score (based on 10-fold cutoff)

I. Baseline $\log_{10}$ HIV-1 RNA, New Drug Covariates, PI phenotypic sensitivity, NRTI phenotypic sensitivity, and NNRTI phenotypic sensitivity (based on assay minimum cutoff)

J. Baseline $\log_{10}$ HIV-1 RNA, New Drug Covariates, PI phenotypic sensitivity, NRTI phenotypic sensitivity, and NNRTI phenotypic sensitivity (based on assay 10-fold cutoff)

4.3.3. Models for Studies with Genotypic and Phenotypic Data

For studies with both Genotypic and Phenotypic data, all models in 4.3.1 and 4.3.2 above should be performed.

4.3.4. Summarizing Results:

A natural summary statistic for the results from logistic regression is the odds-ratio. We suggest that investigators report the estimated odds-ratio and the 95% confidence interval for all covariates in each model

A p-value for the effect of the resistance measures should also be supplied. P-value could be calculated using a maximum likelihood test from a nested model or the Wald-statistic.

Optionally, plots of the cumulative proportion failing over time stratified by the resistance measure (e.g. Kaplan Meier Plots), incorporating the censoring information from the DAC dataset, would be extremely useful.

4.4. Multiplicity Adjustments: No formal adjustments for multiplicities.

5. INTERVENTION STUDIES:

The analysis is similar to that for non-intervention studies with the exception that Trt should be modeled alone and the models described above in section 4.3.1 and 4.3.2 should be done with and without Trt in each, requiring double the number of analyses plus one.

Appendix 1 Virological Endpoint Flowcharts

Dropouts as Failures

<table>
<tr><td colspan="4">Exclude HIV-1 RNA measurements taken after discontinuation of original study regimen</td></tr>
<tr><td colspan="4">On original study regimen ≥ 16 weeks (112 days)?</td></tr>
<tr><td colspan="3">Yes</td><td>No</td></tr>
<tr><td colspan="3">At least 1 HIV-1 RNA measurement while on original regimen in weeks 16-32 (days 106 – 224)?</td><td rowspan="6">Failure</td></tr>
<tr><td colspan="2">Yes</td><td>No</td></tr>
<tr><td colspan="2">Week 24 HIV-1 RNA[1] < threshold?</td><td rowspan="5">Censored</td></tr>
<tr><td>Yes</td><td colspan="2">No</td></tr>
<tr><td rowspan="3">Success</td><td colspan="2">Next HIV-1 RNA measurement < threshold?</td></tr>
<tr><td>Yes</td><td>No</td></tr>
<tr><td>Success</td><td>Failure</td></tr>
</table>

Dropouts as Censored

<table>
<tr><td colspan="6">Exclude HIV-1 RNA measurements taken after discontinuation of original study regimen</td></tr>
<tr><td colspan="6">On original study regimen ≥ 16 weeks (112 days)?</td></tr>
<tr><td colspan="3">Yes</td><td colspan="3">No</td></tr>
<tr><td colspan="3">At least 1 HIV-1 RNA measurement while on original regimen in weeks 16-32 (days 106 – 224)?</td><td colspan="3">Last HIV-1 RNA measurement while on original regimen < threshold?</td></tr>
<tr><td colspan="2">Yes</td><td>No</td><td colspan="2">No</td><td>Yes</td></tr>
<tr><td colspan="2">Week 24 HIV-1 RNA [1]< threshold?</td><td rowspan="5">Censored</td><td colspan="2">At least 2 HIV-1 RNA measurements > baseline?
OR
< 0.5 log reduction for all measurements made during weeks 4 – 8 (days 22 – 56)?
OR
Nadir < threshold?</td><td rowspan="5">Censored</td></tr>
<tr><td>Yes</td><td colspan="2">No</td><td>Yes</td><td>No</td></tr>
<tr><td>Success</td><td colspan="2">Next HIV-1 RNA measurement < threshold?</td><td>Failure</td><td>Censored</td></tr>
<tr><td rowspan="2"></td><td>Yes</td><td>No</td><td colspan="3" rowspan="2"></td></tr>
<tr><td>Success</td><td>Failure</td></tr>
</table>

Note: Week x – y comprises the first day of week x through the last day of week y

[1]Week 24 HIV-1 RNA is the measurement taken closest to the midpoint of week 24 (day 165) while on the original study regimen using the later of two measurements in the event of a tie.