

COHERE
**Collaboration of Observational HIV Epidemiological
Research Europe**

Version 1.2
Standard Operating Procedure
for data transfer

1st merger

List of cohort studies to be contacted:

Bordeaux RCC cohorts

ANRS CO1/CO10/CO11 EPF, ANRS CO2 SEROCO, ANRS CO3 AQUITAINE, ANRS CO4 FHDH, ANRS CO6 PRIMO, ANRS CO8 COPILOTE, CASCADE, Co-RIS, GEMES-Haemo, HIV-MIP, MADRID HIV Children, PISCIS, VACH

Copenhagen RCC cohorts

AMACS, ATHENA, CHIC, CHIPS, DANISH HIV Cohort, ECS, EuroSIDA, FRANKFURT HIV, ICC, ICONA, IMIT, ITLR, KOMPNET, MoCHIV, MODENA, NSHPC, SAN RAFFAELE, SHCS, ST PIERRE, Brussels, UCSC

Bordeaux RCC

Institut de Santé Publique,
d'Épidémiologie et de Développement (ISPED)
Université Victor Segalen Bordeaux 2
Casier postal n° 11
146 rue Léo Saignat
F-33076 Bordeaux cedex
France

Phone: +33 557 571 580
Fax: +33 557 571 172

<mailto:Genevieve.Chene@isped.u-bordeaux2.fr>

Copenhagen RCC

Copenhagen HIV Programme

Hvidovre University Hospital
Kettegaards Alle 30, dept. 044
DK-2650 Hvidovre
Denmark

Phone: +45 36 32 30 15
Fax: +45 36 47 33 40

<mailto:jdl@cphiv.dk>

TABLE OF CONTENTS

1. Introduction to the COHERE SOP for 1st merger	3
2. First scientific project of COHERE: “Response to combination Antiretroviral Therapy (cART): variation across ages”	3
3. Timing of the 1st merger	3
4. Eligibility criteria for patients	3
5. Justification of data needed in narrative and HICDEP formats	4
5.1. List of variables/data needed	4
5.1.1. <i>Core data</i>	4
5.1.2. <i>List of variables not defined as core data but needed for consistency checks or analysis</i>	4
5.2. List of variables according to HICDEP format	5
6. COHERE data sections	5
6.1. Demographic, Clinical and Background Information (BAS).....	5
6.2. Death and drop-out (LTFU)	6
6.3. Cross-cohort identification (OVERLAP).....	6
6.4. Basic follow-up/visit related data (VIS).....	6
6.5. Antiretroviral drug variables (ART)	6
6.6. Opportunistic infections (DIS).....	6
6.7. Laboratory values (LAB_CD4)	6
6.8. Laboratory values (LAB_RNA)	7
7. COHERE data format	7
7.1. Blank values	7
7.2. Unknown values	7
8. Data file transfers to the Regional Coordinating Center (Bordeaux or Copenhagen) .	7
9. Error and discrepancy reporting	8
10. National Regulations	8
11. Details of Variables needed (HICDEP format)	9
11.1. Variables needed for the research analysis in year 1 (HICDEP format)	9
11.1.1. Basic clinical, background and demographic information (BAS file).....	9
11.1.2. Death and drop-out (LTFU file)	11
11.1.3. Cross-cohort identification (OVERLAP file).....	13
11.1.4. Basic follow-up/visit related data (VIS file).....	14
11.1.5. Antiretroviral drug variables (ART file)	15
11.1.6. Opportunistic infections (DIS file).....	17
11.1.7. Laboratory values (LAB_CD4 files).....	18
11.1.8. Laboratory values (LAB_RNA files)	19

1. Introduction to the COHERE SOP for 1st merger

This document provides guidance on the preparation of data files for the first data transfer for the COHERE Collaboration, as a pilot phase to show the potential of this collaboration. The COHERE structure, to the extent possible, conforms to the HICDEP (HIV Cohorts Data Exchange protocol). The latest version of HICDEP is available at the CHIP website: www.cphiv.dk/HICDEP.pdf. Changes and additions are always part of the on-going process for projects that extend over time and COHERE is no exception.

Thank you very much for your contribution to this collaborative project!

2. First scientific project of COHERE: “Response to combination Antiretroviral Therapy (cART): variation across ages”

The COHERE Steering Committee has prioritised the project “Response to combination Antiretroviral Therapy (cART): variation across ages” during this pilot phase of the collaboration.

3. Timing of the 1st merger

For the COHERE data mergers, each cohort will be responsible for gathering and computerizing its own data; subsequently it will then be electronically merged into the respective Regional Coordinating Centers (RCCs) database in either Bordeaux or Copenhagen according to the cohorts and ultimately merged as the COHERE main database.

The deadline for data submission for this merger is **1 June 2006**. During 4 weeks after the submission of data, i.e until around **4 July 2006**, we will send out error and discrepancy information in the form of discrepancy report. We will spend the next 5 weeks processing your response to these reports and working closely with you to clean the data. The cleaning of the data should be completed by **10 August 2006**.

4. Eligibility criteria for patients

Please include **all patients** regardless of age (adults and children):

- who have **started cART**, defined as a regimen with at least 3 antiretroviral
- **since 1 January 1998**
- while **being antiretroviral naïve**

5. Justification of data needed in narrative and HICDEP formats

Please submit the data you have for each selected patient.

5.1. List of variables/data needed

5.1.1. Core data

- Demographic characteristics : year of birth, gender, HIV risk transmission group
- Antiretroviral therapy : individual drugs used (start and stop dates)
- All plasma HIV-1 RNA levels and dates (incl. detection limit, if undetectable) (within 3 months prior to starting cART and during follow-up)
- All CD4 cell counts (and percent CD4, if available) and dates (within 3 months prior to starting cART and during follow-up)
- Complete clinical history: AIDS-defining events according to the CDC and death
- Date of beginning of active follow-up (for patients who started being followed in a cohort at some point several months or years after being diagnosed) and all visit dates.

5.1.2. List of variables not defined as core data but needed for consistency checks or analysis

- Ethnicity
- Causes of death
- HIV-status (including first positive HIV serology)
- Date of enrolment into the cohort
- Antiretroviral naïve status at enrolment into the cohort
- Type of HIV-RNA assay

5.2. List of variables according to HICDEP format

Variables/data needed	HICDEP table	HICDEP variables
Birth date	BAS	PATIENT, BIRTH_D, ENROL_D, GENDER, MODE, MODE_OTH, ORIGIN, RACE, SEROCO_D, SEROCO_M, RECART_Y, AIDS_Y, AIDS_D
Sex		
Ethnicity		
Mode of infection		
Death	LTFU	PATIENT, DROP_Y, DROP_D, DROP_RS, DEATH_Y, DEATH_D, DEATH_R1, DEATH_RC1, DEATH_R2, DEATH_RC2
Patient	OVERLAP	PATIENT, COHORT, PAT_OTH, COH_OTH
Cohort		
Latest follow-up date	VIS	PATIENT, VIS_D
Antiretroviral therapy	ART	PATIENT, ART_ID, ART_SD, ART_ED, ART_RS
Reasons for stopping regimen		
AIDS defining events	DIS	PATIENT, DIS_ID, DIS_D
CD4 cell counts	LAB_CD4	PATIENT, CD4_D, CD4_V, CD4_U
Plasma HIV-1 RNA	LAB_RNA	PATIENT, RNA_D, RNA_V, RNA_L, RNA_T

See Section 11 "Details of Variables needed (HICDEP format)" for more details.

6. COHERE data sections

6.1. Demographic, Clinical and Background Information (BAS)

Each patient should appear once in this table.

Please make sure that the enrolment date, ENROL_D, is the date that the patient enrolled in the local cohort.

If an AIDS diagnosis has been given to the patient, please report the date of AIDS diagnosis (AIDS_D).

Some cohorts are prohibited from reporting certain types of data such as origin or race: for ORIGIN, please leave these fields *blank*; for RACE, use the code "98".

The BAS table in the Appendix describes the coding of these variables in more detail.

Please submit the date of 1st HIV-1 positive test and provide the correct code in SEROCO_M (code = 4). In case that you do not have the date for the 1st HIV-1 positive test please submit the date of seroconversion in SEROCO_D and indicate the correct code in SEROCO_M to specify the source of this date.

6.2. Death and drop-out (LTFU)

All of the death and drop-out variables are incorporated in this table.

A patient is considered as drop-out if he/she has left the cohort, withdrawn consent, or if there is no new information on the patient during the preceding twelve months. Patients without a visit date, death date or drop-out date will be considered lost to follow-up.

When cohorts have recorded the “underlying” cause of death, they should only report the underlying cause of death in the variable “DEATH-R1” and “DEATH-RC1”. The underlying cause of death is defined by the disease or injury which initiated the train of morbid events leading to death (International Classification of Diseases-10th revision).

When cohorts have recorded cause(s) of death but cannot differentiate between the “immediate”, “contributing” or “underlying” cause(s), they should report all available data “DEATH-R1” and “DEATH-R2”. When submitting, each cohort should identify whether they have recorded the underlying cause of death or not in the variables “DEATH-RC1” and “DEATH-RC2”.

The LTFU table in the Appendix describes the coding of these variables in more detail.

6.3. Cross-cohort identification (OVERLAP)

Please submit the OVERLAP: Cross-Cohort Identification table even if you don't have patients participating in other COHERE cohorts (in this case, leave the table empty).

Patients who are known to be in other cohorts participating in COHERE should be entered in this table, once for each cohort. Two fields are provided for this information: The COH_OTH field contains a 20-character name identifying the other cohort and the PAT_OTH field is for the unique patient identifier used in this cohort.

6.4. Basic follow-up/visit related data (VIS)

See the VIS table in the Appendix for details.

6.5. Antiretroviral drug variables (ART)

Each antiretroviral treatment is identified by its Anatomical Therapeutic Chemical (ATC) code, which can be up to 12 characters. If the patient has been given ART, enter the proper ATC code in the ART_ID field followed by ART_SD (start date) and ART_ED (stop date).

The ART table in the Appendix describes the coding of these variables in more detail.

6.6. Opportunistic infections (DIS)

All DIS_ID (code to identify the event) and DIS_D (date of AIDS-defining opportunistic events) should be reported.

See the DIS table in the Appendix for details.

6.7. Laboratory values (LAB_CD4)

See the LAB_CD4 table in the Appendix for details.

6.8. Laboratory values (LAB_RNA)

See the LAB_RNA table in the Appendix for details.

The RNA_V (HIV-RNA measurement value (copies/ml)) should be coded as -1 only if the value is strictly inferior to the RNA_L (lower limit of HIV-RNA assay).

7. COHERE data format

Please submit your data using the HICDEP formats described in the tables in the section 11. The HICDEP format is based on a relational structure and currently incorporates 15 data tables and numerous lookup-tables for the codes.

7.1. Blank values

When a variable is not applicable or not used, leave the field *blank*. If data is missing whereas a response is required or available, the cohort validation programs should detect this and this information will become part of the database including errors and discrepancies.

7.2. Unknown values

The category “unknown” indicates that the information needed is unknown or purposely left as missing. The codes 9, 99 and 999 are used to designate this category. Please see the tables in the Appendix for the specific coding.

The date of 1911-11-11 is to be used, whenever the use of a drug, a treatment episode etc, is known to have occurred but the date is unknown. Similarly, for other types of variables, there is most often “yes/no” question, followed by the “date” question (for example: “Has the patient an AIDS diagnosis?” and then: “If yes, date of AIDS diagnosis”). For these types of questions, if the event is known to have occurred but the date is unknown, code the date as: 1911-11-11. Then the COHERE validation programs will detect a “yes AIDS diagnosis” – “unknown date of AIDS diagnosis”. If the only information available regarding a date is the year, then it should be entered as July 1, XXXX (XXXX-07-01). If the month and the year are given, the date should be entered with the day being the 15th.

8. Data file transfers to the Regional Coordinating Center (Bordeaux or Copenhagen)

Please submit your data using Access (version 97 or 2000), SAS (version 8 or 9), STATA (version 6), ASCII semicolon separate files or XML format. Both Regional Coordinating Center will take care of the final transformation from your preferred data format with StatTransfer.

For security purposes, cohort data files to be transferred to the RCC and between the two RCCs will be encrypted and compressed with ZIP archive using the AES encryption algorithm. The encryption password (minimum 10 characters long, including upper/lowercase, numbers and special characters) will be communicated to the RCCs by fax or by telephone. These zip-files can be uploaded onto the servers of the RCCs using the secure file transfer protocol (ftps) or send on a CD-ROM by registered mail.

9. Error and discrepancy reporting

Within eight weeks of data submission, we will e-mail a report to the cohort data managers in order for them to check and correct their data and to replace “missing” values.

The cohort data managers should enter the corrected data into their own database and then send the revised tables to the COHERE data manager. These revised tables will then be re-checked and then, if there are no further problems, added to the rest of the cohort’s data.

10. National Regulations

As the COHERE collaboration will be an academic collaboration between an anticipated number of 800 centers in over 30 European countries, it is the responsibility of each investigator/sponsor to follow current national regulations, regarding data extraction and data transfer.

11. Details of Variables needed (HICDEP format)

11.1. Variables needed for the research analysis in year 1 (HICDEP format)

11.1.1. Basic clinical, background and demographic information (BAS file)

Table 1 below details the baseline data that should be included in BAS file.

Table 1 – Variables to be included in BAS file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort patient ID). Unique and anonymous
BIRTH_D	Date (for example yyyy-mm-dd)	Birth date
ENROL_D	Date (for example yyyy-mm-dd)	Date of enrolment into the cohort
GENDER	Numeric with codes : 1 = Male 2 = Female 9 = Unknown	Gender/Sex
MODE	Numeric with codes : 1 = homo/bisexual 2 = injecting drug user 3 = homo/bisexual and injecting drug user 4 = haemophiliac 5 = transfusion, non-haemophilia related 6 = heterosexual contact 7 = heterosexual contact and injecting drug user 8 = perinatal 90 = other (specify) 99 = unknown	Mode of infection
MODE_OTH	Characters	Mode of infection OTHER
ORIGIN	Numeric with codes : 10 = Africa 11 = Northern Africa 12 = Sub-Saharan Africa 20 = Asia 30 = Oceania (not Australia) 40 = Australia & New Zealand 50 = Americas 51 = North America 52 = Central & South America 60 = Middle East 70 = Europe 71 = Western Europe 72 = Eastern Europe 99 = Unknown	Nationality or region of origin of patient
RACE	Numeric with codes : 10 = White 20 = Black 21 = Black African 22 = Black Caribbean 30 = Hispanic 40 = Asian 50 = American 60 = Indigenous 1020 = White/Black 1040 = White/Asian 2030 = Black/Hispanic 3040 = Hispanic/Asian 98 = Prohibited 99 = Unknown	Race of patient

Table 1 (continued) – Variables to be included in BAS file

Name	Format and definition	Description
SEROCO_D	Date (for example yyyy-mm-dd)	Date of seroconversion or date of 1 st HIV diagnosis
SEROCO_M	Numeric with codes : 1=midpoint between last neg. and first pos. HIV-1 test 2=lab evidence of seroconversion 3=seroconversion illness 4=first pos HIV-1 test 9=other	Source of the SEROCO_D
RECART_Y	Numeric with codes : 0 = No 1 = Yes 9 = Unknown	Has the patient received antiretroviral treatment?
AIDS_Y	Numeric with codes : 0 = No 1 = Yes 9 = Unknown	Has patient been given an AIDS diagnosis?
AIDS_D	Date (for example yyyy-mm-dd)	If yes, date of AIDS diagnosis

Example :

PATIENT	BIRTH_D	ENROL_D	GENDER	MODE	MODE_OTH	ORIGIN	RACE
991	1928-07-08	1998-12-09	1	6		71	10
992	1949-05-26	2001-06-29	2	6		71	10
993	1937-09-07	2000-06-02	1	1		12	20

PATIENT	SEROCO_D	SEROCO_M	RECART_Y	AIDS_Y	AIDS_D
991	1998-07-21	3	0	0	
992	2000-04-12	3	0	1	1999-03-30
993	1999-09-30	3	0	0	

11.1.2. Death and drop-out (LTFU file)

Table 2 below details the information to be included in LTFU file.

Table 2 - Variables to be included in LTFU file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort patient ID). Unique and anonymous
DROP_Y	Numeric with codes : 0 = No 1 = Yes	Has the patient dropped out?
DROP_D	Date (for example yyyy-mm-dd)	If yes, date of last visit
DROP_RS	Numeric with codes : 1 = Patient lost to follow-up / not known to be dead 2 = Patient has not had visit within required amount of time 3 = Patient moved away 4 = Patient moved and is followed by another centre 5 = Patients decision 6 = Consent withdrawn 7 = Incarceration / jail 8 = Institutionalisation (drug treatment, psychological...etc) 9 = Other	If yes, reason for drop
DEATH_Y	Numeric with codes : 0 = No 1 = Yes	Has the patient died?
DEATH_D	Date (for example yyyy-mm-dd)	Date of death
DEATH_R1	Numeric with codes : 1 = Myocardial infarction 2 = Stroke 3 = Other cardiovascular diseases 4 = Symptoms caused by mitochondrial toxicity 4.1 = Lactic acidosis 5 = Complications due to diabetes mellitus 6 = Pancreatitis 7 = Complications due to hepatitis 7.1 = Hepatitis related 7.2 = Liver failure not related to hepatitis or mitochondrial toxicity 8 = HIV-related 8.1 = AIDS defining event 8.2 = Invasive bacterial infection 9 = Renal failure 10 = Bleeding (haemophilia) 20 = non-AIDS defining cancer 90 = Other 91 = Suicide 92 = Drug overdose 93 = accident 99 = unknown, fatal case with no information	Cause of death
DEATH_RC1	Character with codes: I = Immediate cause U = Underlying cause/condition C = Contributing cause N = Not available	Coding of causal relation of the code given in DEATH_R1 to the death

Table 2 (continued) - Variables to be included in LTFU file

Name	Format and definition	Description
DEATH_R2	Numeric with codes : 1 = Myocardial infarction 2 = Stroke 3 = Other cardiovascular diseases 4 = Symptoms caused by mitochondrial toxicity 4.1 = Lactic acidosis 5 = Complications due to diabetes mellitus 6 = Pancreatitis 7 = Complications due to hepatitis 7.1 = Hepatitis related 7.2 = Liver failure not related to hepatitis or mitochondrial toxicity 8 = HIV-related 8.1 = AIDS defining event 8.2 = Invasive bacterial infection 9 = Renal failure 10 = Bleeding (haemophilia) 20 = non-AIDS defining cancer 90 = Other 91 = Suicide 92 = Drug overdose 93 = accident 99 = unknown, fatal case with no information	Cause of death
DEATH_RC2	Character with codes: I = Immediate cause U = Underlying cause/condition C = Contributing cause N = Not available	Coding of causal relation of the code given in DEATH_R2 to the death

Example :

PATIENT	DROP_Y	DROP_D	DROP_RS	DEATH_Y	DEATH_D	DEATH_R1	DEATH_RC1	DEATH_R2	DEATH_RC2
991	0			0					
992	1	2002-09-13	1	0					
993	0			1	2002-10-14	8.1	I		

IMPORTANT: Please append as many DEATH_R# and DEATH_R#_T columns as you need to submit all your registered causes of death.

11.1.3. Cross-cohort identification (OVERLAP file)

Table 3 below details the information to be included in OVERLAP file.

Table 3 - Variables to be included in OVERLAP file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort patient ID). Unique and anonymous
COHORT	Character	Code/name of the cohort
PAT_OTH	Character	Unique patient identifier in other cohorts
COH_OTH	Character	Name of the cohort

Example :

PATIENT	COHORT	PAT_OTH	COH_OTH
991	FHDH	712	COPILOTE

11.1.4. Basic follow-up/visit related data (VIS file)

Table 4 below details the information to be included in VIS file.

Table 4 - Variables to be included in VIS file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort patient ID). Unique and anonymous
VIS_D	Date (for example yyyy-mm-dd)	Date of patient visit

Example :

PATIENT	VIS_D
991	1998-12-14
991	1999-04-25
991	2000-05-02
991	2001-03-21
991	2002-02-11
991	2003-03-14
991	2004-01-05
992	2001-07-14
992	2002-09-13
993	2000-08-12
993	2001-09-03
993	2002-08-16

11.1.5. Antiretroviral drug variables (ART file)

Table 5 below details the data on antiretroviral regimens that should be included in the ART file.

Table 5 – Variables to be included in ART file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort patient ID). Unique and anonymous
ART_ID	Character with codes : J05A=ART unspecified Protease inhibitors J05AE=PI unspecified J05AE01=Saquinavir (gel, not specified) J05AE01-SQH=Saquinavir hard gel (INVIRASE) J05AE01-SQS=Saquinavir soft gel (FORTOVASE) J05AE02=Indinavir (CRIXIVAN) J05AE03=Ritonavir (NORVIR) J05AE03-H=Ritonavir high dose (NORVIR) J05AE03-L=Ritonavir low dose (NORVIR) J05AE04=Nelfinavir (VIRACEPT) J05AE05=Amprenavir (141W94) (AGENERASE) J05AE06=Lopinavir/Ritonavir (ABT-378/r, Kaletra) J05AE07=Fosemprenavir J05AE08=Atazanavir J05AE09= Tipranavir J05AE-GW4=GW433908/VX-275 (Drug phase III) (PROGENERASE) J05AE-TMC=TMC 114 (Tibotec) Nucleoside and nucleotide reverse transcriptase inhibitors J05AF=NRTI unspecified J05AF01=Zidovudine (AZT, RETROVIR) J05AF02=Didanosine (ddI) (VIDEX) J05AF03=Zalcitabine (ddC) (HIVID) J05AF04=Stavudine (d4T) (ZERIT) J05AF05=Lamivudine (3TC, EPIVIR) J05AF06=Abacavir (1592U89) (ZIAGEN) J05AF07=Tenofovir (VilREAD) J05AF08=Adefovir (PREVEON) J05AF09=Emtricitabine J05AF30-COM=Zidovudine/Lamivudine - COMBIVIR (AZT/3TC, RETROVIR/EPIVIR) J05AF30-TZV=Trizivir J05AF-FOZ=Fozivudine tidoxi J05AF-LDN=Lodanosine (trialdrug) J05AF30-KIV=Kivexa J05AF30-TRU=Truvada J05AF-RVT=Reverset Non-nucleoside reverse transcriptase inhibitors J05AG=NNRTI unspecified J05AG01=Nevirapine (VIRAMUN) J05AG02=Delavirdine (U-90152) (RESCRIPTOR) J05AG03=Efavirenz (DMP-266) (STOCRIN, SUSTIVA) J05AG-LOV=Loviride J05AG-TMC=TMC 125 (Tibotec) Other J05AX07= Enfuvirtide (FUZEON - T-20) L01XX05=Hydroxyurea/Hydroxycarbamid (LITALIR) J05A-MK=MK-0518 J05A-PBT=Participant in Blinded Trial	Code representing the antiretroviral treatment

Table 5 (continued) – Variables to be included in ART file

Name	Format and definition	Description
ART_SD	Date (for example yyyy-mm-dd)	Date of initiation of treatment
ART_ED	Date (for example yyyy-mm-dd)	Date of stopping treatment
ART_RS	Numeric with codes : 1 = Treatment failure (i.e virological, immunological, and/or clinical failure) 1.1 = Virological failure 1.2 = Partial virological failure 1.3 = Immunological failure – CD4 drop 1.4 = Clinical progression 2 = Abnormal fat redistribution 3 = Concern of cardiovascular disease 3.1 = Dyslipidaemia 3.2 = Cardiovascular disease 4 = Hypersensitivity reaction 5 = Toxicity, predominantly from abdomen/G-I tract 5.1 = Toxicity – GI tract 5.2 = Toxicity – Liver 5.3 = Toxicity – Pancreas 6 = Toxicity, predominantly from nervous system 7 = Toxicity, predominantly from kidneys 8 = Toxicity, predominantly from endocrine system 8.1 = Diabetes 9 = Haematological toxicity (anemia,...) 10 = Hyperlactataemie/lactic acidosis 88 = Death 90 = Side effects – any of the above but unspecified 90.1 = Comorbidity 91 = Toxicity, not mentioned above 92 = Availability of more effective treatment (not specifically failure or side effect) 92.1 = Simplified treatment available 92.2 = Treatment to complex 92.3 = Drug interaction 93 = Structured Treatment Interruption (STI) 93.1 = Structured Treatment Interruption (STI) – at high CD4 94 = Patient's wish/decision 94.1 = Non-compliance 95 = Physician's decision 96 = Pregnancy 97 = Study treatment 98 = Other causes 99 = Unknown	Reason for stopping treatment

Example :

PATIENT	ART_ID	ART_SD	ART_ED	ART_RS
991	J05AF08	2000-10-21	2000-12-12	1
991	J05AF04	2001-03-03		
991	J05AF02	2000-10-21		
992	J05AE02	2002-04-12	2002-05-18	3.1
992	J05AE03	2002-04-12	2002-05-18	3.1

11.1.6. Opportunistic infections (DIS file)

Table 6 below details the data on AIDS-defining opportunistic events diagnosed during follow-up that should be included in DIS file.

Table 6 – Variables to be included in DIS file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort patient ID). Unique and anonymous
DIS_ID	Character with codes : DEM=AIDS dementia complex BCNE=Bacterial pneumonia, recurrent (>2 episodes within 1 year) CANO=Candidiasis, oesophageal, bronchi, trachea, or lungs COCC = Coccidioidomycosis, disseminated or extrapulmonary CRCO=Cryptococcosis, extrapulm. CRSP=Cryptosporidiosis (duration > 1 month) CMVR=Cytomegalovirus (CMV) chorioretinitis CMVO=CMV - other location HERP=Herpes simplex virus ulcers (duration > 1 month) or pneumonitis/esophagitis/bronchitis HIST=Histoplasmosis, extrapulm. or disseminated WAST=HIV Wasting Syndrome ISDI=Isosporiasis diarrhoea (duration > 1 month) LEIS=Leishmaniasis, visceral MCDI=Microsporidiosis diarrhoea (dur. > 1 month) MC=Mycobact. avium complex (MAC) or Kanasii, extrapulm. MCP=Mycobact. tuberculosis pulm. MCX=Mycobact. tuberculosis extrapulm. MCPO=Mycobact. pulm., other MCXO=Mycobact. extrapulm., other PCP=Pneumocystis carinii pneumonia (PCP) LEU=Progressive multifocal leucoencephalopathy SAM=Salmonella bacteriaemia (non-typhoid) (recurrent) TOX=Toxoplasmosis, brain FBLS=Focal Brain lesion KS=Kaposi Sarcoma HG=Hodgkins Lymphoma NHG=Non-Hodgkin Lymphoma - not specified NHGB=Non-Hodgkin Lymphoma - Burkitt (Classical or Atypical) NHGI=Non-Hodgkin Lymphoma - Diffuse large B-cell lymphoma (Immunoblastic or Centroblastic) NHGU=Non-Hodgkin Lymphoma - Unknown/other histology NHGP=Non-Hodgkin Lymphoma - Primary Brain Lymphoma CRVC=Cervical Cancer	Code to identify opportunistic event
DIS_D	Date (for example yyyy-mm-dd)	Date of event

Example :

PATIENT	DIS_ID	DIS_D
991	ISDI	2000-10-21
991	SAM	1999-08-12
991	TOX	2001-07-14
992	PCP	1999-09-13

11.1.7. Laboratory values (LAB_CD4 files)

Table 7 below detail the laboratory data that should be included in LAB_CD4 files.

Table 7 – Variables to be included in LAB_CD4 file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort patient ID). Unique and anonymous
CD4_D	Date (for example yyyy-mm-dd)	Date of measurement
CD4_V	Numeric -1 = undetectable or detection limit as negative value	Value of CD4 measurement
CD4_U	Numeric with codes: 1 = cells/ μ l 2 = %	CD4 cell count or CD4 percent

Example :

PATIENT	CD4_D	CD4_V	CD4_U
991	1998-04-13	15	1
991	1998-04-13	85	2
991	1999-08-12	50	1
991	2001-07-14	100	1
991	2001-09-13	140	1
992	2000-05-18	197	1
992	2000-05-18	46	2
992	2001-03-30	213	1

11.1.8. Laboratory values (LAB_RNA files)

Table 8 below detail the laboratory data that should be included in LAB_RNA files.

Table 8 – Variables to be included in LAB_RNA file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort patient ID). Unique and anonymous
RNA_D	Date (for example yyyy-mm-dd)	Date of measurement
RNA_V	Numeric -1 = undetectable/below level of detection or detection limit as negative value	HIV-RNA measurement value (copies/ml)
RNA_L	Numeric	Lower limit of HIV-RNA assay
RNA_UL	Numeric	Upper limit of HIV-RNA assay
RNA_T	Numeric with codes : 5 = Roche TaqMan 10 = Roche 1.0 15 = Roche 1.5 ultra-sensitive 19 = Any Roche (unspecified) 20 = NASBA 21 = NASBA ultra-sensitive 29 = Any NASBA (unspecified) 31 = Chiron b-DNA 1.0 32 = Chiron b-DNA 2.0 33 = Chiron b-DNA 3.0 39 = Any Chiron (unspecified) 40 = Abbott ultra-sensitive 50 = Monitor 1.0 51 = Monitor 1.0 ultra-sensitive 55 = Monitor 1.5 56 = Monitor 1.5 ultra-sensitive 65 = Cobas 1.5 66 = Cobas 1.5 ultra-sensitive 90 = Other 99 = Unknown	IF AVAILABLE, what type of viral assay was used for this measurement?

Example :

PATIENT	RNA_D	RNA_V	RNA_L	RNA_UL	RNA_T
991	1998-04-13	12586	50		51
991	1999-08-12	4623	50		51
991	2001-07-14	200	50		51
991	2001-09-13	742	50		51
992	2000-05-18	500	50		15
992	2001-03-30	50	50		15
992	2002-01-14	-1	50		15