

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

2011 Merger

Version 2.0
Standard Operating Procedure
for data transfer

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, Corispe-CAT, MADRID HIV Children, PISCIS, VACH

Copenhagen RCC cohorts

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

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](#)

University of Copenhagen

Faculty of Health Sciences

The Panum Institute

Building 21.1 Blegdamsvej 3B

2200 Copenhagen N

Denmark

Phone: +45 35 45 57 57

Fax: +45 35 45 57 58

<mailto:igr@cphiv.dk>

Prepared by Céline Colin (ISPED, Univ Bordeaux), and Jesper Kjær (CHIP, Univ Copenhagen) based on the HICDEP coding system (authors: Jesper Kjær, Bruno Ledergerber).

TABLE OF CONTENTS

1. Introduction to the COHERE SOP for the 2011 merger	4
2. Scientific projects of COHERE	4
2.1. Project A - Treatment change database.....	4
2.2. Project B - Standard reference distribution of CD4 count response to cART.....	4
2.3. Project C - CD4 dynamics in HIV1 and HIV2	4
2.4. Project D - Effect of NRTI and pegylated interferon in hepatitis C co-infected.....	4
2.5. Project E - Impact of SVR on overall and liver-related survival in hepatitis co-infected patients.....	5
2.6. Project F - Late presenter	5
2.7. Project G - Opportunistic Infection	5
3. Options for data submission	6
3.1. Option 1.....	6
3.2. Option 2.....	6
3.3. Consideration	6
3.4. Guidance for mother-to-child/paediatric cohorts.....	7
4. Timing of the merger	8
5. Eligibility criteria for patients	8
5.1. Project A - Treatment change database.....	8
5.2. Project B - Standard reference distribution of CD4 count response to cART.....	8
5.3. Project C - CD4 dynamics in HIV1 and HIV2	8
5.4. Project D - Effect of NRTI and pegylated interferon in hepatitis C co-infected.....	8
5.5. Project E - Impact of SVR on overall and liver-related survival in hepatitis co-infected patients.....	8
5.6. Project F - Late presenter	8
5.7. Project G - Opportunistic Infection	8
6. Data needed in summary table and HICDEP formats	9
6.1. Summary of tables needed according to option 2	9
6.2. List of all variables according to HICDEP format.....	10
7. COHERE data sections	11
7.1. Demographic, Clinical and Background Information (BAS)	11
7.2. Death and drop-out (LTFU).....	11
7.3. Cross-cohort identification (OVERLAP)	11
7.4. Basic follow-up/visit related data (VIS).....	12
7.5. Antiretroviral drug variables (ART).....	12
7.6. Therapy and prophylaxis against OIs and hepatitis B/C and immunomodulators (MED)	12
7.7. AIDS-defining opportunistic infections (DIS)	12
7.8. Laboratory values – LAB (LAB)	12
7.9. Laboratory values – CD4 (LAB_CD4).....	12
7.10. Laboratory values – CD8 (LAB_CD8)	12
7.11. Laboratory values- HIV-1 RNA (LAB_RNA)	12
7.12. Hepatitis B and C co-infections, toxoplasmosis and CMV (LAB_VIRO)	13
7.13. Background information on resistance tests (LAB_RES)	13
7.14. Level 1: Nucleotide sequence data (LAB_RES_LVL_1).....	13
7.15. Level 2: Amino acid mutations (LAB_RES_LVL_2).....	13
7.16. Level 3: Resistance scores (LAB_RES_LVL_3).....	13
8. COHERE data format	14
8.1. Blank values	14
8.2. Unknown values	14
9. Data file transfers	14
10. Error and discrepancy reporting	14
11. National Regulations	14
12. Details of Variables needed (HICDEP format)	15
12.1. Variables needed for the research analysis (HICDEP format).....	15
12.1.1. Basic clinical, background and demographic information (BAS file).....	15

12.1.2. Death and drop-out (LTFU file)	17
12.1.3. Cross-cohort identification (OVERLAP file).....	19
12.1.4. Basic follow-up/visit related data (VIS file)	20
12.1.5. Antiretroviral drug variables (ART file)	21
12.1.6. Other medication – used for treatment or prophylaxis of OIs, treatment against HBV and HCV and immune-modulators (MED file)	24
12.1.7. Opportunistic infections (DIS file).....	26
12.1.8. Laboratory values (LAB file)	28
12.1.9. Laboratory values (LAB_CD4 file)	29
12.1.10. Laboratory values (LAB_CD8 file)	30
12.1.11. Laboratory values (LAB_RNA file)	31
12.1.12. Viro-/serology tests (LAB_VIRO file).....	32
12.1.13. Main resistance table (LAB_RES file).....	34
12.1.14. Nucleotide sequences (PRO, RT, GP41, GP120) (LAB_RES_LVL_1 file).....	35
12.1.15. Level 2 amino acid mutations (LAB_RES_LVL_2 file).....	36
12.1.16. Level 3 resistance score data and phenotype results (LAB_RES_LVL_3 file).....	37

1. Introduction to the COHERE SOP for the 2011 merger

This document provides guidance on the preparation of data files for the data transfer for the COHERE 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. Scientific projects of COHERE

Seven projects have been endorsed by the COHERE Steering Committee for 2011

2.1. Project A - Treatment change database

Project lead: Jesper Kjaer

The objective of this project is to build a TCE database based on all treatment changes which are based on a recent resistance test, and for which the short term viral load outcome is known, for use by researchers who wish to apply methods to derive interpretation systems and, hopefully eventually lead to a consensus interpretation system for each drug.

2.2. Project B - Standard reference distribution of CD4 count response to cART

Project leads: Rodolphe Thiebaut, Maria Dorucci and Carlo Torti

This project aims at estimating the distribution of CD4 change at 6 and 12 months according to major predictive factors at initiation of cART among patients who achieved and maintained viral suppression at 6 and 12 months after starting cART in 2005-2008. Among this population, given the importance of the initial immunological response to cART in prediction of a better prognosis (ART-CC Lancet 2003), the project will explore the association between CD4 evolution at 12 months and the time to reach a CD4 count within a "normal" range, i.e. $> 500/\text{mm}^3$ as well as the time to reach a good immunological response defined by CD4 count > 500 , CD4 percentage $> 29\%$ and CD4/CD8 ratio > 1 .

2.3. Project C - CD4 dynamics in HIV1 and HIV2

Project lead: Sophie Matheron

The objective of this project is to compare immunological outcome in HIV-1 and HIV-2 infected patients while controlling for plasma HIV RNA in three circumstances:

- Natural history in absence of antiretroviral therapy:
 - in seroincident patients with known or well estimated date of infection,
 - in seroprevalent patients
- Response to first line cART
 - To study the influence of HIV-2 subtypes on natural history and response to therapy.

2.4. Project D - Effect of NRTI and pegylated interferon in hepatitis C co-infected

Project lead: Colette Smit

The objective of this project is to study the effect of different NRTI-based regimens on the response to anti-HCV treatments, among all HCV-HIV co-infected patients,

receiving interferon or pegylated-interferon (PEG-IFN) with ribavirin and who are treated with cART.

2.5. Project E - Impact of SVR on overall and liver-related survival in hepatitis co-infected patients

Project lead: Antonella d'Arminio Monforte

The objective of this project is to assess the impact of SVR on overall and liver-related survival.

2.6. Project F - Late presenter

Project lead: Ole Kirk

The objective of this project is to standardise definitions, and carry out an analysis to characterising epidemiology for those persons who are presenting late with HIV over time compared to those presenting early for care, with a special focus on changes in the characteristics and proportion of late presenters in different parts of Europe.

2.7. Project G - Opportunistic Infection

Project leads: Hansjakob Furrer and Jose Miro

The objective of this project is to validate existing guidelines for discontinuation of

- Primary prophylaxis against toxoplasmosis and non-tuberculous mycobacteria (NTM)
- Secondary prophylaxis (maintenance therapy): Pneumocystis jirovecii pneumonia (PcP), toxoplasmosis, cytomeglovirus disease, cryptococcosis, disseminated infections with non-tuberculous mycobacteria (NTM), leishmaniasis, recurrent salmonella bacteremia and other rare infections.

3. Options for data submission

Considering the large number of projects for the 2011 merger, 2 options are envisaged for alignment of eligibility criteria:

3.1. Option 1

Submit data in one unique dataset for all projects:

- Eligibility criteria: all patients as soon as they are diagnosed for HIV
- Data: all available data corresponding to the SOP (tables 1 to 16). For more details, see Table 6.1, paragraph 6.
- Projects covered:
 - Project A - Treatment change database
 - Project B - Standard reference distribution of CD4 count response to cART
 - Project C - CD4 dynamics in HIV1 and HIV2
 - Project D - Effect of NRTI and pegylated interferon in hepatitis C co-infected
 - Project E - Impact of SVR on overall and liver-related survival in hepatitis co-infected patients
 - Project F - Late presenter
 - Project G - Opportunistic infection

3.2. Option 2

Submit data in two separate datasets for:

Dataset 1:

- Eligibility criteria: all patients regardless of age (adults and children) and regardless of when cART was initiated or naïve with follow-up after 1 January 1997.
- Data: all variables needed for the projects listed below (tables 1 to 12, i.e. all tables except resistance data). For more details, see Table 6.1, paragraph 6.
- Projects covered:
 - Project B - Standard reference distribution of CD4 count response to cART
 - Project C - CD4 dynamics in HIV1 and HIV2
 - Project D - Effect of NRTI and pegylated interferon in hepatitis C co-infected
 - Project E - Impact of SVR on overall and liver-related survival in hepatitis co-infected patients
 - Project G - Opportunistic infection

Dataset 2:

- Eligibility criteria: all patients as soon as they are diagnosed for HIV
- Data: all variables needed for the projects listed below (tables 1 to 5, 7 to 16, i.e. all tables except "other medications"). For more details, see Table 6.1, paragraph 6.
- Projects covered:
 - Project A - Treatment change database
 - Project F - Late presenter

3.3. Consideration

Even if you submit data for option 1 or 2, you can still opt out of individual project if you don't want to participate. Data will only be used for the specified purpose.

3.4. Guidance for mother-to-child/paediatric cohorts

Please find below the list of relevant projects endorsed as being relevant to mother-to-child/paediatric cohort datasets. For the 2011 merger, mother-to-child/paediatric-only cohorts can submit data to:

- Project A - Treatment change database
- Project B - Standard reference distribution of CD4 count response to cART
- Project G - Opportunistic infection

Mother-to-child cohorts can also submit data to:

- Project F - Late presenters (which does not include paediatrics)

The following projects are not relevant to mother-to-child/ paediatric-only cohorts:

- Project C - CD4 dynamics in HIV1 and HIV2
- Project D - Effect of NRTI and pegylated interferon in hepatitis C co-infected
- Project E -Impact of SVR on overall and liver-related survival in hepatitis co-infected patients

4. Timing of the 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 **15 March 2011**. During 6 weeks after the submission of data, i.e. until **30 April 2011**, we will send out error and discrepancy information in the form of discrepancy report. We will spend the next few 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 **31 May 2011**.

5. Eligibility criteria for patients

5.1. Project A - Treatment change database

All patients with a resistance test

5.2. Project B - Standard reference distribution of CD4 count response to cART

All patients starting cART with plasma HIV RNA < 50 copies/ml at 6 months and with at least one available measure of CD4 count 6 months after cART initiation.

5.3. Project C - CD4 dynamics in HIV1 and HIV2

- All HIV infected patients with at least one available measure of CD4 count before cART
- All patients starting a first cART with at least one available measure of CD4 count after cART initiation

5.4. Project D - Effect of NRTI and pegylated interferon in hepatitis C co-infected

All patients regardless of age (adults and children):

- with at least one test regarding HCV infection
- who have started cART, defined as a regimen with at least 3 antiretroviral
- since 1 January 1998
- while being antiretroviral naïve

5.5. Project E - Impact of SVR on overall and liver-related survival in hepatitis co-infected patients

All patients regardless of age (adults and children):

- Regardless of when cART was initiated or naïve with follow-up after 1 January 1997
- With at least one test regarding HCV infection

5.6. Project F - Late presenter

All patients as soon as they are diagnosed for HIV

5.7. Project G - Opportunistic Infection

All patients regardless of age (adults and children):

- Regardless of when cART was initiated or naïve with follow-up after 1 January 1997

6.2. List of all variables according to HICDEP format

Variables/data needed	HICDEP table	HICDEP variables
Patient identifier	BAS	PATIENT, CENTER, BIRTH_D, FRSVIS_D, ENROL_D, GENDER, MODE, MODE_OTH, ORIGIN, ETHNIC, SEROCO_D, SEROCO_M, RECART_Y, AIDS_Y, AIDS_D, HEIGH, ALCO_Y
Birth date		
Sex		
Mode of infection		
Region of origin		
Ethnicity		
HIV-test results		
Date of enrolment into the cohort		
Has the patient received ART?		
Has the patient been given an AIDS diagnosis. If yes, date of AIDS diagnosis		
Has the patient ever been abusing alcohol?		
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, WEIGH, LOSS_Y, GAIN_Y, HEIGH
Antiretroviral therapy	ART	PATIENT, ART_ID, ART_SD, ART_ED, ART_RS
Reasons for stopping regimen		
OI/HBV/HCV therapy/prophylaxis and Immune-modulators	MED	PATIENT, MED_ID, MED_SD, MED_ED
AIDS defining events	DIS	PATIENT, DIS_ID, DIS_D, DIS_WD, DIS_OTH
Laboratory values	LAB	PATIENT, LAB_ID, LAB_FA, LAB_ST, LAB_D, LAB_V, LAB_U
CD4 cell counts and %	LAB_CD4	PATIENT, CD4_D, CD4_V, CD4_U
CD8 cell counts and %	LAB_CD8	PATIENT, CD8_D, CD8_V, CD8_U
Plasma HIV-1 RNA	LAB_RNA	PATIENT, RNA_D, RNA_V, RNA_L, RNA_T, RNA_UL
Hepatitis B and C co-infections, toxoplasmosis and CMV	LAB_VIRO	PATIENT, VS_ID, VS_D, VS_R, VS_V, VS_U, VS_LL, VS_UL, VS_T
Resistance test, background information	LAB_RES	PATIENT, SAMP_ID, SAMPLE_D, SEQ_DT, LAB, TESTTYPE, KIT, SOFTWARE, LIBRARY, REFSEQ, SUBTYPE
Sequence data in nucleotide format	LAB_RES_LVL_1	SAMP_ID, SEQTYPE, SEQ_STAR, SEQ_STOP, SEQ_NUC
Amino acid mutations	LAB_RES_LVL_2	SAMP_ID, GENE, AA_POS, AA_POS_SUB, AA_FOUND_1, AA_FOUND_2, AA_FOUND_3, AA_FOUND_4
Resistance scores for antiretroviral drugs	LAB_RES_LVL_3	SAMP_ID, ART_ID, RES_SCOR

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

7. COHERE data sections

7.1. 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 event has been diagnosed, please report the date of AIDS diagnosis (AIDS_D).

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.

7.2. 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.

7.3. 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.

Please note that the data submitted is requested to be on all patients that fulfil the inclusion criteria. Do not remove patient data for which you know overlaps with other patients in other cohorts!

An algorithm has been developed to identify overlaps between cohorts based on similarity between patient characteristics (probabilistic linkage). Further optimisation and validation is however needed. For this to be done we need as many known overlapping patients between cohorts and the detailed data as possible. Identified overlaps NOT listed in the OVERLAP file will be queried back to the cohorts for verification.

COHERE RCCs has a set of agreements registered between overlapping cohorts and hospital clinics that participate in more than one cohort as to who will include their data into the analysis. Based on these agreements the RCCs will perform inclusion and exclusion of these patients centrally at the RCCs.

A central registry of overlaps and inclusion/exclusion rules on patient, site, country and cohort level will be maintained by the two RCCs. Following this merger the algorithm will be presented, published and thereby made publicly available for other collaborations to use.

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

See the VIS table in the Appendix for details.

7.5. 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.

7.6. Therapy and prophylaxis against OIs and hepatitis B/C and immunomodulators (MED)

Each prophylaxis treatment is identified by its Anatomical Therapeutic Chemical (ATC) code, which can be up to 12 characters. If the patient has been given therapy and/or prophylaxis against OIs, against hepatitis B and/or C infection and immuno-modulators, enter the proper ATC code in the MED_ID field followed by MED_SD (start date) and MED_ED (stop date).

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

7.7. AIDS-defining 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.

7.8 Laboratory values – LAB (LAB)

See the LAB table in the appendix for details.

7.9. Laboratory values – CD4 (LAB_CD4)

See the LAB_CD4 table in the Appendix for details.

7.10. Laboratory values – CD8 (LAB_CD8)

See the LAB_CD8 table in the Appendix for details.

7.11. Laboratory values- HIV-1 RNA (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).

7.12. Hepatitis B and C co-infections, toxoplasmosis and CMV (LAB_VIRO)

For every test listed in VS_ID please provide VS_R (result) and VS_V and VS_U where relevant. All available tests should be submitted.

7.13. Background information on resistance tests (LAB_RES)

In this table any relevant information about the actual resistance test should be provided, sectioned into sample information:

- SAMP_ID: sample id – should be unique for all samples in the cohort,
- SAMPLE_D: date the sample was drawn,
- SEQ_DT: please provide date and time for sequencing if the result obtained is a genotypic test
- LAB: text to give name of the laboratory,
- TESTTYPE: either genotype or phenotype – or both at the same time,
- KIT: name of the commercial kit used for the sequencing/phenotype test or if in house please specify this.
- SOFTWARE and LIBRARY to provide information on the software and the library/version that was used to get the resistance scores if that is provided in LAB_RES_LVL_3.
- REFSEQ to inform which ref. sequence was use in the phenotype test or which sequence was used as wildtype reference in the genotype tests.
- Subtype should, if available, be reported in SUBTYPE.

7.14. Level 1: Nucleotide sequence data (LAB_RES_LVL_1)

This is the preferred level at which resistance tests should be reported if the full sequence is available, please dissect the sequence data into the genes for the enzymes that are targeted with treatment: protease, reverse-transcriptase etc. If available please also state start and end of the sequencing attempt – hence primers used for protease sequences often cut-off the first 9 nucleotides, for RT this is often the first ~112 nucleotides.

7.15. Level 2: Amino acid mutations (LAB_RES_LVL_2)

List all identified mutations here that are recorded for the test, do not submit data here if the full sequence is available nor if the test was a phenotype test. The format allows for submission of mixtures/ambiguous nucleotides that code for more than one amino acid. Up to 4 fields are available for this – in need of more please add additional fields. Insertions should be coded with a sub-position variable: a, b, c etc – please see example in 12.1.15. Deletions are to be added with a position and no amino acid code or a “-“ in AA_FOUND_1 (and blank for the other AA_FIELD_# variables).

7.16. Level 3: Resistance scores (LAB_RES_LVL_3)

Together with filling in software/library in LAB_RES this table can be used to submit resistance score data from genotypic resistance algorithms. Please do only use this if none of the two other levels (1 or 2) is available. Providing IC 50 FC phenotype data in case of no genotype data is also allowed, if so, then TESTTYPE in LAB_RES should indicate that this is a phenotype based resistance test.

8. COHERE data format

Please submit your data using the HICDEP formats described in the tables in the section 12. The HICDEP format is based on a relational structure and currently incorporates 17 data tables and numerous lookup-tables for the codes.

8.1. Blank values

When a variable is not applicable or missing and there is no code for the reason, then leave it blank. The cohort validation consistency check programs will detect and report back on any invalid blanks i.e. when a response is required or when there is a code you could use such as "unknown" or "unavailable" or "missing".

8.2. 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.

9. Data file transfers

Please submit your data using Access (version 97 or higher), 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.

10. Error and discrepancy reporting

Within six 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.

11. National Regulations

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

12. Details of Variables needed (HICDEP format)

12.1. Variables needed for the research analysis (HICDEP format)

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

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

Projects: ALL

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
CENTER	Character	Code for Clinic/Centre/Hospital where patient is seen.
BIRTH_D	Date (for example yyyy-mm-dd)	Birth date
FRSVIS_D	Date (for example yyyy-mm-dd)	First seen at clinic
ENROL_D	Date (for example yyyy-mm-dd)	Date of enrolment into the cohort
HEIGH	Numeric (metric - meters): 999=Unknown	Height of patient at visit/most current – use HEIGH variable in table VIS for children also.
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
ETHNIC	Numeric with codes: 10 = White 20 = Black 21 = Black African 22 = Black Caribbean 30 = Hispanic 40 = Asian 50 = American	Ethnicity of patient

Name	Format and definition	Description
	60 = Indigenous 1020 = White/Black 1040 = White/Asian 2030 = Black/Hispanic 3040 = Hispanic/Asian 102040 = White/Black/Asian 97 = other 98 = Prohibited 99 = Unknown	
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: 0=No 1=Yes 9 = Unknown	Has the patient received antiretroviral treatment?
AIDS_Y	Numeric with codes : 0 = No 1 = Yes 9 = Unknown	Has the patient been given an AIDS diagnosis?
AIDS_D	Date (for example yyyy-mm-dd)	If yes, date of AIDS diagnosis
ALCO_Y*	Numeric with codes : 0 = No 1 = Yes 9 = Unknown	Has the patient ever been abusing alcohol?

Example :

PATIENT	CENTER	BIRTH_D	FRSVIS_D	ENROL_D	GENDER	MODE	MODE_OTH
991	Bordeaux	1928-07-08	1998-11-15	1998-12-09	1	6	
992	Copenhagen	1949-05-26	2001-06-29	2001-06-29	2	6	
993	Copenhagen	1937-09-07	1999-09-07	2000-06-02	1	1	

ORIGIN	ETHNIC	SEROCO_D	SEROCO_M	RECART_Y	AIDS_Y	AIDS_D	ALCO_Y
71	10	1998-07-21	3	1	0	1999-08-12	0
71	10	2000-04-12	3	1	1	1999-09-13	0
12	20	1999-09-30	3	1	0		1

* Additional field according to HICDEP 1.5, needed for the HCV project

12.1.2. Death and drop-out (LTFU file)

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

Projects: ALL

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: 0 = Patient was not infected (mainly for children) 1 = Patient lost to follow-up / not known to be dead 2 = Patient has not had visit within required amount of time 2.1 = Patient did not respond to several invitations 3 = Patient moved away 3.1 = Patient moved to another country 4 = Patient moved and is followed by another centre 4.1 = Paediatric patient transferred to adult care 5 = Patients decision 5.1 = Patients caretaker wanted to discontinue (for children) 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 50 = sudden infant death 51 = neonatal death (including prematurity/ other complications) 55 = child abuse 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

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 50 = sudden infant death 51 = neonatal death (including prematurity/ other complications) 55 = child abuse 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_RC# columns as you need to submit all your registered causes of death.

12.1.3. Cross-cohort identification (OVERLAP file)

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

Projects: ALL

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

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

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

Projects: ALL

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
WEIGH	Numeric (metric - kilograms) 999=Unknown	Weight of the patient at visit
LOSS_Y*	Numeric with codes : 0 = No 1 = Yes 9 = Unknown	Is the patient experiencing loss of fat from extremities, buttocks or face?
GAIN_Y*	Numeric with codes : 0 = No 1 = Yes 9 = Unknown	Is the patient gaining fat in the abdomen, neck, breast or other defined locations?
HEIGH#	Numeric (metric - meters): 999=Unknown	Height of patient at visit – relevant for children only – use HEIGH variable in table BAS for adults.

Example :

PATIENT	VIS_D	WEIGH	LOSS_Y	GAIN_Y	HEIGH
991	1998-12-14	76	1	1	
991	1999-04-25	80.5	1	1	
991	2000-05-02	81	1	1	
991	2001-03-21	82	0	1	
991	2002-02-11	90	0	1	
991	2003-03-14	85	0	0	
991	2004-01-05	86	0	0	
992	2001-07-14	100	0	0	
992	2002-09-13	87	0	0	
993	2000-08-12	65	0	0	
993	2001-09-03	999	0	0	
993	2002-08-16	75	1	1	

* Mandatory fields according to HICDEP 1.5 format, may be left blank for this merger

Additional field according to HICDEP 1.5, needed for children only

12.1.5. Antiretroviral drug variables (ART file)

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

Projects: ALL

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 (AGENERASE) J05AE06=Lopinavir/Ritonavir (Kaletra) J05AE07=Fos-amprenavir (Telzir, Lexiva) J05AE08=Atazanavir (Reyataz) J05AE09=Tipranavir (Aptivus) J05AE10=Darunavir (TMC-114, Prezista) J05AE-MOZ=Mozenavir (DMP-450) 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 J05AF10=Entecavir J05AF11=Telbivudine J05AF12=Clevudine J05AF-ALO=Alovudine J05AF-AMD=Amdoxovir (DADP) J05AF-FOZ=Fozivudine tidoxi J05AF-LDN=Lodenoisine (trialdrug) 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) J05AG04=Etravirine J05AG-CPV=Capravirine J05AG-DPC083=DPC 083 J05AG-DPC961=DPC 961 J05AG-EMV=Emivirine (MKC442) J05AG-LOV=Loviride J05AG-RPV=Rilpivirine (TMC-278) Combination drugs J05AR01=Combivir (Zidovudine/Lamivudine) J05AR02=Kivexa (Lamivudine/Abacavir) J05AR03=Truvada (Tenofovir/Emtricitabine) J05AR04=Trizivir (Zidovudine/Lamivudine/Abacavir)	Code representing the antiretroviral treatment

Name	Format and definition	Description
ART_ID	J05AR05=Douvir-N (Zidovudine/Lamivudine/Nevirapine) J05AR06=Atripla (Emtricitabine/Tenofovir/Efavirenz) Integrase Inhibitors J05AX-EVG=Elvitegravir (Gilead) J05AX08=Raltegravir (Merck) Maturation inhibitors J05A-BEV=Beviramat Fusion inhibitors J05AX07=Enfuvirtide (Fuzeon, T-20) J05AX09=Maraviroc (Pfizer) J05AX-VIC=Vicriviroc (Schering) Other L01XX05=Hydroxyurea/Hydroxycarbamid (Litalir) J05A-PBT=Participant in Blinded Trial	Code representing the antiretroviral treatment
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 (anaemia...) 10 = Hyperlactataemie/lactic acidosis 70 = Pregnancy – toxicity concerns 75 = Pregnancy – prevention of mother to child transmission 76 = Post-partum prophylaxis 77 = Dose change for height/ weight 88 = Death 90 = Side effects – any of the above but unspecified 90.1 = Co morbidity 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

12.1.6. Other medication – used for treatment or prophylaxis of OIs, treatment against HBV and HCV and immune-modulators (MED file)

Table 6 below details the baseline data that should be included in MED file.

Projects: Hepatitis, OI

Table 6 – Variables to be included in MED file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort patient ID). Unique and anonymous
MED_ID	Character with codes: J01AA08=Minocycline (MINOCIN) J01EA01=Trimethoprim (MONOTRIM) J01EC02=Sulfadiazine J01EE01=Sulfamethoxazole and trimethoprim (Bactrim) J01EE03=Sulfametrole and trimethoprim - Cosoltrime (MADERAN) J01EE=Cotrimoxazole - Comb. of sulfonamides and trimethoprim (BACTRIM, EUSAPRIM, NOPIL) J01FA09=Clarithromycine (KLACID) J01FA10=Azithomycine (ZITHROMAX) J01FF01=Clindamycine (DALACIN) J01GB06=Amikacine (AMIKLINE) J01MA02=Ciprofloxacin (CIPROXINE, CILOXAN) J01MA12=Levofloxacin (TAVANIC) J01MA14=Moxifloxacin J01RA02=Cosoltrime (MADERAN) J02AA01=Amphotericin B (FUNGIZON) J02AB02=Ketoconazole J02AB=Imidazoles (DAKTARIN, NIZORAL, PEVARYL) J02AC01=Fluconazole (DIFLUCAN) J02AC02=Itraconazole (SPORANOX) J02AC03=Voriconazole J04AB02=Rifampin (RIMATICIN) J04AB04=Rifabutin (MYCOBUTIN) J04AC01=Isoniazide (RIMIFON) J04AK01=Pyrazinamide (PYRAZINAMID) J04AK02=Ethambutol (EMB, MYAMBUTOL) J04AM02=RIFATER J04BA01=Clofazimine (LAMPREN) J04BA02=Dapsone J05AB01=Aciclovir (ZOVIRAX) J05AB04=Ribavirin J05AB06=Ganciclovir (CYMEVENE) J05AB09=Famciclovir J05AB11=Valaciclovir (VALTEX) J05AB12=Cidofovir (VISTIDE) J05AD01=Foscarnet (FOSCAVIR) L03AA02=G-CSF/Filgrastim (NEUPOGEN) L03AB-AL2=Peginterferon alfa-2a/alfa-2b (PEGINTRON, PEGASYS) L03AB10=Peginterferon alfa-2b (PEGINTRON) L03AB11=Peginterferon alfa-2a (PEGASYS) L03AB=Interferons L03AC-IL2=Interleukin 2 (PROLEUKIN) P01AX06=Atovaquone (WELLVONE, MEPRONE) P01BD01=Pyrimethamine (DARAPRIM) P01BD51=Pyrimethamine/Sulfadoxine (FANSIDAR)	Code representing the drug, any missing code can be found at: http://www.whocc.no/atcddd/indexdatabase

Name	Format and definition	Description
MED_ID	P01CB=Antimony compounds P01CX01=Pentamidine aerosol (PENTACARNET) V03AF03=Folate of calcium (LEUCOVORINE)	Code representing the drug, any missing code can be found at: http://www.whooc.no/atcddd/indexdatabase
MED_SD	Date (for example yyyy-mm-dd)	Date of initiation of treatment
MED_ED	Date (for example yyyy-mm-dd)	Date of stopping treatment

Example:

PATIENT	MED_ID	MED_SD	MED_ED
991	L03AC-IL2	2000-10-21	2000-12-12
991	J05AB04	2001-03-03	
991	J02AA01	2000-10-21	
992	J05AB12	2002-04-12	2002-05-18
992	J01EE	2002-04-12	2002-05-18

12.1.7. Opportunistic infections (DIS file)

Table 7 below details the data on AIDS-defining opportunistic events diagnosed during follow-up that should be included in DIS file.

Projects: ALL

Table 7 – 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 = <u>Coccidioidomycosis</u> , 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 MCO=Mycobact., 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 <hr/> Paediatric specific CDC stage C codes: SRBI=Serious recurrent/ multiple bacterial infections CMVP=Cytomegalovirus (CMV) disseminated with onset >1 month, paediatrics ENC=Encephalopathy	Code to identify opportunistic event
DIS_D	Date (for example yyyy-mm-dd)	Date of event
DIS_WD	Numeric with codes: 1=Definitive diagnosis 2=Presumptive diagnosis 3=Diagnosis from autopsy 4=Diagnosis from registry 9=Unknown/unavailable	Way/means of diagnosis
DIS_OTH	Character	Other location, only to be filled out if code alone is not sufficient

Example :

PATIENT	DIS_ID	DIS_D	DIS_WD	DIS_OTH
991	ISDI	2000-10-21	1	
991	SAM	1999-08-12	2	
991	TOX	2001-07-14	1	
992	PCP	1999-09-13	9	

12.1.8. Laboratory values (LAB file)

Table 8 below details the laboratory data that should be included in LAB file.

Projects: hepatitis, late presenter

Table 8 – Variables to be included in LAB file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort patient ID). Unique and anonymous
LAB_ID	Character with codes: ALB=Albumine ALT=Alanin-Aminotransferase AST=Aspartat aminotransferase BIL=Total Bilirubin GGT=gGT HAEM=Haemoglobin INR=Quick/INR PLT=Platelet count PTR=Prothrombin rate	Code representing the measurement
LAB_FA	Numeric with codes: 0=No 1=Yes 9=Unknown	Was the blood sample taken while fasting?
LAB_ST	Character with codes: WB=Whole Blood P=Plasma S=Serum	Specimen type
LAB_D	Date (for example yyyy-mm-dd)	Date of measurement
LAB_V	Numeric -1 = undetectable or detection limit as negative value	Value of measurement
LAB_U	Numeric with codes: 1=mmol/L 2=gm/L 3=gm/dL 4=mg/dL 5=IU/L (u/L) 6=mmol/L 7=INR 8=1E+9/L 9=1E+6/L 10=cells/ μ L 11=mkat/L	Unit of measurement

Example :

PATIENT	LAB_ID	LAB_FS	LAB_ST	LAB_D	LAB_V	LAB_U
991	ALB	9	P	1998-04-13		
991	ALT	0	P	1998-04-13		
991	ALT	1	P	1999-08-12		
991	BIL	1	P	2001-07-14		
991	INR	1	WB	2001-09-13		
992	PLT	0	P	2000-05-18		
992	ALB	0	P	2000-05-18		
992	INR	9	WB	2001-03-30		

12.1.9. Laboratory values (LAB_CD4 file)

Table 9 below details the laboratory data that should be included in LAB_CD4 file.

Projects: ALL

Table 9 – 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 = % 3 = total lymphocytes/ μ l	CD4 cell count, CD4 percent or total lymphocyte count

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

12.1.10. Laboratory values (LAB_CD8 file)

Table 10 below details the laboratory data that should be included in LAB_CD8 file.

Projects: Standard reference distribution of CD4 count response to cART and CD4 dynamics in HIV1 and HIV2

Table 10 – Variables to be included in LAB_CD8 file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort patient ID). Unique and anonymous
CD8_D	Date (for example yyyy-mm-dd)	Date of measurement
CD8_V	Numeric -1 = undetectable or detection limit as negative value	Value of CD8 measurement
CD8_U	Numeric with codes: 1 = cells/ μ l 2 = % 3 = total lymphocytes/ μ l	CD8 cell count, CD8 percent or total lymphocyte count

Example :

PATIENT	CD8_D	CD8_V	CD8_U
991	1998-04-13	250	1
991	1998-04-13	85	2
991	1999-08-12	346	1
991	2001-07-14	550	1
991	2001-09-13	678	1
992	2000-05-18	197	1
992	2000-05-18	46	2
992	2001-03-30	1010	1

12.1.11. Laboratory values (LAB_RNA file)

Table 11 below details the laboratory data that should be included in LAB_RNA file.

Projects: ALL

Table 11 – 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_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 41 = Abbott LCx 50 = Monitor 1.0 51 = Monitor 1.0 ultra-sensitive 55 = Monitor 1.5 56 = Monitor 1.5 ultra-sensitive 59 = Monitor unspecified 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?
RNA_UL*	Numeric	Upper limit of HIV-RNA assay

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

* Additional field according to HICDEP 1.5, by standard part of COHERE mergers.

12.1.12. Viro-/serology tests (LAB_VIRO file)

Table 12 below details the viro-/serology data that should be included in LAB_VIRO file.

Projects: hepatitis, late presenter

Table 12 – Variables to be included in LAB_VIRO file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort patient ID). Unique and anonymous
VS_ID	HBV=Marker for hepatitis B infection - test unknown HBVAC=HBV IgG antibody (core, HBcIgG) HBVAE=HBV antibody (envelope) HBVAS=HBV antibody (surface, HBsAb) HBVD=HBV-dna HBVGE=HBV antigen (envelope, HBeAG) HBVGS=HBV antigen (surface, HBsAg) HCV=Marker for hepatitis C infection - test unknown HCVA=HCV antibody IgG * HCVR=HCV-rna TOXA=Toxoplasma antibodies CMVA=CMV antibodies HDVA=Hepatitis delta antibody HIV-1=HIV-1 test HIV-2=HIV-2 test	Code to identify viro-/serology test
VS_D	Date (for example yyyy-mm-dd)	Date of measurement
VS_R	Numeric 0= negative 1= positive 9= unknown/borderline	Result of test
VS_V	Numeric -1 = undetectable/below level of detection or detection limit as negative value	Measurement value (where relevant)
VS_U	Numeric with codes: 1=copies/mL 2=IU/mL 3=Geq (millions of genome equivalent)	Unit of measurement (where relevant) please communicate if unit of relevance is missing
VS_LL*	Numeric	IF AVAILABLE, Lower Limit of assay
VS_UL*	Numeric	IF AVAILABLE, Upper Limit of assay
VS_T*	Numeric with codes: 1=Roche qualitative (Amplicor) [HCV and HBV] 2=Roche quantitative test for HBV (Cobas Amplicor HBV monitor) 3=Bayer Bdna quantitative [HCV] 4=Bayer Bdna quantitative [HBV] 5=Roche Taqman 9=Other	IF AVAILABLE, What type of ASSAY was used for this measurement?

* Two step test: Screening with EIA, and confirmation by testing for a panel of specific antibodies (recombinant immunoblot assay): Only patients with a positive screening and confirmation test should be coded as positive.

* Mandatory fields according to HICDEP 1.5 format, may be left blank.

Example:

PATIENT	VS_ID	VS_D	VS_R	VS_V	VS_U	VS_LL	VS_UL	VS_T
991	TOXA	1998-04-13	0					
991	CMVA	1998-04-13	1					
991	HCVA	1999-08-12	0					
991	HCVR	2001-07-14	1	100000000	1	617		1
991	HBVD	2001-09-13	1	38000	2	617		1
992	HBV	2000-05-18	0					
992	HCV	2000-05-18						
992	CMVA	2001-03-30	1					

12.1.13. Main resistance table (LAB_RES file)

Table 13 below details the background data for the resistance test that should be included in LAB_RES file.

Project: treatment change database

This table is required to be submitted. Sequence and/or amino acid data is as minimum required in LAB_RES_LVL1 or LAB_RES_LVL1 and LAB_RES_LVL2

Table 13 – Variables to be included in LAB_RES file

Name	Format and definition	Description
PATIENT	Character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
SAMP_ID	Character (or numeric if possible)	The assigned UNIQUE sample ID
SAMPLE_D	Date (for example yyyy-mm-dd)	Date of the actual sample taken (NOT the test date)
SEQ_DT	Date - time (for example yyyy-mm-dd hh:mm)	Date and time when the sequencing was performed
LAB	Character	Name of laboratory where the test was performed
TESTTYPE	Numeric with codes: 1 = Genotype 2 = Phenotype 3 = Genotype and phenotype	Type of test
KIT	Character	Vendor and version/name of the kit used for the test
SOFTWARE	Character	Software and version used to determine resistance
LIBRARY	Character	Library/algorithm used to identify resistance mutations
REFSEQ	Character	Name/identifier of reference HIV strain used to find mutations
SUBTYPE	Character: A B CRF01_AE ...etc and mixtures of these if relevant	Subtype of HIV-1 virus

Example:

PATIENT	SAMP_ID	SAMPLE_D	SEQ_DT	LAB	TESTTYPE
AAA	1	2002-12-01	2003-12-07	LAB 1	Genotype
AAA	2	2004-01-14	2004-01-16	LAB B	Genotype
999	3	2006-05-13	2006-05-27 12:14	LAB Z	Phenotype
1111	4	2007-06-17	2007-06-23	LAB Y	Genotype

KIT	SOFTWARE	LIBRARY	REFSEQ	SUBTYPE
VisibleGenetics			HXB2	CRF_AE
VisibleGenetics			HXB2	CRF_AE
PhenoSense			NL4-3	C
In house	Stanford HIVdb	4.3.6	HXB2	B

12.1.14. Nucleotide sequences (PRO, RT, GP41, GP120) (LAB_RES_LVL_1 file)

Table 14 below details the information to be included in LAB_RES_LVL_1 file (no entry if the test was a phenotype test).

Project: treatment change database

This table is required to be submitted; yet SEQ_NUC is allowed blank if LAB_RES_LVL_2 is used to report back single amino acid mutations. The table is necessary to distinguish between tests with no amino acid mutations identified and tests for which the sequence could not be or only partly obtained (regardless of PCR or sequencing issues)

Table 14 - Variables to be included in LAB_RES_LVL_1 file

Name	Format and definition	Description
SAMP_ID	Character (or numeric if possible)	The assigned UNIQUE sample ID
SEQTYPE	Character with codes: PRO = PRO sequence RT = RT sequence GP41 = GP41 sequence GP120 = GP120 sequence	Type of nucleotide sequence
SEQ_STAR	Numeric 0=no sequence – hence no start position	Start position for the sequence
SEQ_STOP	Numeric 0=no sequence – hence no stop position	Stop position for the sequence
SEQ_NUC	Character/String – IUPAC letter codes only Empty/blank/null=No sequence submitted or no sequence obtained	Nucleotide sequence if available

Example:

SAMP_ID	SEQTYPE	SEQ_STAR	SEQ_STOP	SEQ_NUC
1	PRO	10	297	CCTCAGAT.....
1	RT	112	741	TGTACAGT....
2	PRO	10	200	CCTCAGAT.....
2	RT	0	0	

12.1.15. Level 2 amino acid mutations (LAB_RES_LVL_2 file)

Table 15 below details the information to be included in LAB_RES_LVL_2 file (no entry if the test was a phenotype test or sequence has been reported in LAB_RES_LVL_1).

Project: treatment change database

Table 15 - Variables to be included in LAB_RES_LVL_2 file

Name	Format and definition	Description
SAMP_ID	Character (or numeric if possible)	The assigned UNIQUE sample ID
GENE	Character with codes: PRO = PRO sequence RT = RT sequence GP41 = GP41 sequence GP120 = GP120 sequence	Type of sequence/gene (PRO, RT, GP41, GP120)
AA_POS	Numeric	Position of the mutation in the sequence
AA_POS_SUB	Character a, b, c etc	Subposition used to code insertions
AA_FOUND_1	Amino acid 1-letter code	Mutation (Amino acid) found in the sequence
AA_FOUND_2	Amino acid 1-letter code	Mutation (Amino acid) found in the sequence (if more than 1)
AA_FOUND_3	Amino acid 1-letter code	Mutation (Amino acid) found in the sequence (if more than 2)
AA_FOUND_4	Amino acid 1-letter code	Mutation (Amino acid) found in the sequence (if more than 3)

Example:

SAMP_ID	GENE	AA_POS	AA_POS_S UB	AA_FOUND _1	AA_FOUND _2	AA_FOUND _3	AA_FOUND _4
2	PRO	10		W			
2	RT	69	a	T			
2	RT	69	b	S			
2	RT	69	c	S			
2	RT	184		I	V		

12.1.16. Level 3 resistance score data and phenotype results (LAB_RES_LVL_3 file)

Table 16 below details the information to be included in LAB_RES_LVL_3 file (no entry if the test was reported in either LAB_RES_LVL_1 or LAB_RES_LVL_2).

Project: treatment change database

Table 16 - Variables to be included in LAB_RES_LVL_3 file

Name	Format and definition	Description
SAMP_ID	Character (or numeric if possible)	The assigned UNIQUE sample ID
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 (AGENERASE) J05AE06=Lopinavir/Ritonavir (Kaletra) J05AE07=Fos-amprenavir (Telzir, Lexiva) J05AE08=Atazanavir (Reyataz) J05AE09=Tipranavir (Aptivus) J05AE10=Darunavir (TMC-114, Prezista) J05AE-MOZ=Mozenavir (DMP-450) Nucleoside and nucleotide reverse transcriptase inhibitors J05AF=NRTI unspecified J05AF01=Zidovudine (AZT, RETROVIR) J05AF02=Didanosine (ddl) (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 J05AF10=Entecavir J05AF11=Telbivudine J05AF12=Clevudine J05AF-ALO=Alovudine J05AF-AMD=Amdoxovir (DADP) J05AF-FOZ=Fozivudine tidoxi J05AF-LDN=LodenoSine (trialdrug) 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) J05AG04=Etravirine J05AG-CPV=Capravirine J05AG-DPC083=DPC 083 J05AG-DPC961=DPC 961 J05AG-EMV=Emvirine (MKC442) J05AG-LOV=Loviride J05AG-RPV=Rilpivirine (TMC-278) Combination drugs J05AR01=Combivir (Zidovudine/Lamivudine) J05AR02=Kivexa (Lamivudine/Abacavir)	Drug code of antiretroviral

	J05AR03=Truvada (Tenofovir/Emtricitabine) J05AR04=Trizivir (Zidovudine/Lamivudine/Abacavir) J05AR05=Douvir-N (Zidovudine/Lamivudine/Nevirapine) J05AR06=Atripla (Emtricitabine/Tenofovir/Efavirenz) Integrase Inhibitors J05AX-EVG=Elvitegravir (Gilead) J05AX08=Raltegravir (Merck) Maturation inhibitors J05A-BEV=Beviramat Fusion inhibitors J05AX07=Enfuvirtide (Fuzeon, T-20) J05AX09=Maraviroc (Pfizer) J05AX-VIC=Vicriviroc (Schering) Other L01XX05=Hydroxyurea/Hydroxycarbamid (Litalir) J05A-PBT=Participant in Blinded Trial	
RES_SCOR	Resistance score (Numeric or character) or IC50 FC (numeric)	Score of resistance or recommendation given from the test.

Example:

SAMP_ID	ART_ID	RES_SCOR
3	J05AE05	2
3	J05AF07	45
3	J05AG01	1
4	J05AE01	Susceptible
4	J05AG03	Intermediate resistance
4	J05AF05	High-level resistance