# HICDEP: HIV Cohorts Data Exchange Protocol

**Version 1.30 – 28th of March 2008**

Edited by Jesper Kjær and Bruno Ledergerber

Updates maintained by:
Jesper Kjær
CHIP - Copenhagen HIV Programme
University of Copenhagen, Faculty of Health Sciences
The Panum Institute/Building 21.1 Blegdamsvej 3B
2200 Copenhagen N
Denmark

E-mail jkj@cphiv.dk

# Table of contents

# List of tables and figures

# 1 About this document

HIV cohort collaborations have made substantial contributions to the knowledge of HIV epidemiology and management over the last years. So far, most collaborations have incorporated slightly different protocols for data exchange causing unnecessary workload for the people in charge of data extraction.

We were therefore asked to put together this draft consensus protocol for discussion at the 7th International Workshop on HIV Observational Databases, March 29th-30th 2003, Fiuggi, Italy. It is based on our experience with data-exchange protocols for D:A:D, the ART Cohort-Collaboration, the PLATO Collaboration and several previous studies on the safety of stopping OI prophylaxis.

This protocol is based on a relational structure (with some very minor deviations) and currently incorporates 15 data tables and numerous lookup-tables for the codes. It is evident that - depending on the questions addressed - only subsets of tables and fields will have to be extracted for data exchange.

We have not elaborated on database systems (e.g. SQL-Server, Oracle, Access) and their respective file-formats as there are excellent tools for transferring data between most of the popular packages (e.g. StatTransfer from http://www.stattransfer.com). The suggested data structure should work with most formats and software packages.

Please keep in mind that the primary purpose of this document is to provide you with formats for data-exchange but not for an operational database used for data-management on a day-to-day basis. Some considerations with that respect can be found in the appendices.

We plan to update this document on a regular basis and the most recent versions will be made available on the websites of EuroSIDA (http://www.cphiv.dk) and the Swiss HIV Cohort Study (http://www.shcs.ch). Complementary to this document you will during also find the lookup-tables in the HICDEP section on the http://www.cphiv.dk website.

HICDEP is a format under constant improvement and additions are made almost every year, please see Appendix IV – Change Log for the most current updates. Please always use the text files available on the HICDEP website for most current coding lists for ART and MED drugs.

Quality control checks are currently being written and will be available on the above websites before end of 2008. Work to make an extended HICDEP format that enables mother to child and paediatric cohorts to fully utilise HICDEP for data exchange is also under preparation.

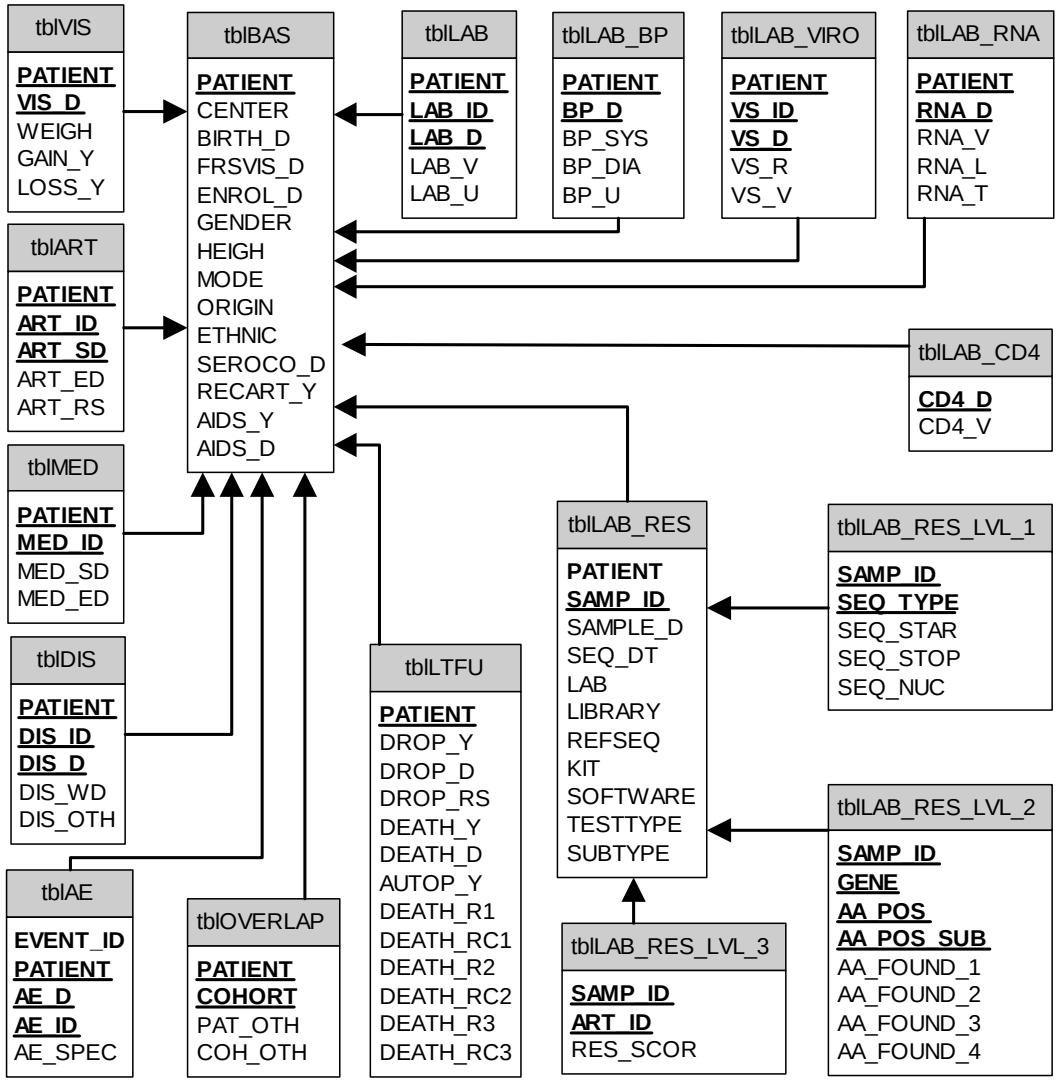
Bruno Ledergerber and Jesper Kjær, March 2008

# 2 General data format

## *2.1 Overview of data tables*

| Table | Content |
|---|---|
| tblAE | holds type and date of adverse events including serious non-AIDS conditions |
| tblART | holds type of antiretroviral drug, start and stop dates and reason for stopping |
| tblBAS | holds demographics, basic clinical information, date of AIDS diagnosis, death and drop-out information |
| tblDIS | holds type and date of CDC-C diseases. |
| tblLAB | holds type, date, value and unit of laboratory tests. |
| tblLAB_BP | holds date, diastolic and systolic values and unit of blood pressure measurements. |
| tblLAB_CD4 | holds date and value of CD4 measurements. |
| tblLAB_RNA | holds date, value, detection limit and type of viral assay. |
| tblLAB_RES | holds background information on the resistance test, laboratory, library, kit, software and type of test |
| tblLAB_RES_LVL_1 | holds nucleoside sequence for the PRO and RT sequences |
| tblLAB_RES_LVL_2 | holds mutations and positions of these. |
| tblLAB_RES_LVL_3 | holds resistance result in relation to antiretroviral drug. |
| tblLAB_VIRO | holds test results for viro-/serological tests (hepatitis etc.) |
| tblLTFU | holds data in death and drop-out |
| tblMED | holds type, start and stop dates for other HIV related medicines. |
| tblOVERLAP | holds information on the patients participation in other cohorts |
| tblVIS | holds visit related information, weight, wasting. |

## *2.2 Diagram*


tblVIS
PATIENT
VIS_D
WEIGH
GAIN_Y
LOSS_Y
tblART
PATIENT
ART_ID
ART_SD
ART_ED
ART_RS
tblMED
PATIENT
MED_ID
MED_SD
MED_ED
tblDIS
PATIENT
DIS_ID
DIS_D
DIS_WD
DIS_OTH
tblAE
EVENT_ID
PATIENT
AE_D
AE_ID
AE_SPEC
tblBAS
PATIENT
CENTER
BIRTH_D
FRSVIS_D
ENROL_D
GENDER
HEIGH
MODE
ORIGIN
ETHNIC
SEROCO_D
RECART_Y
AIDS_Y
AIDS_D
tblOVERLAP
PATIENT
COHORT
PAT_OTH
COH_OTH
tblLAB
PATIENT
LAB_ID
LAB_D
LAB_V
LAB_U
tblLAB_BP
PATIENT
BP_D
BP_SYS
BP_DIA
BP_U
tblLAB_VIRO
PATIENT
VS_ID
VS_D
VS_R
VS_V
tblLAB_RNA
PATIENT
RNA_D
RNA_V
RNA_L
RNA_T
tblLAB_CD4
CD4_D
CD4_V
tblLTFU
PATIENT
DROP_Y
DROP_D
DROP_RS
DEATH_Y
DEATH_D
AUTOP_Y
DEATH_R1
DEATH_RC1
DEATH_R2
DEATH_RC2
DEATH_R3
DEATH_RC3
tblLAB_RES
PATIENT
SAMP_ID
SAMPLE_D
SEQ_DT
LAB
LIBRARY
REFSEQ
KIT
SOFTWARE
TESTTYPE
SUBTYPE
tblLAB_RES_LVL_3
SAMP_ID
ART_ID
RES_SCOR
tblLAB_RES_LVL_1
SAMP_ID
SEQ_TYPE
SEQ_STAR
SEQ_STOP
SEQ_NUC
tblLAB_RES_LVL_2
SAMP_ID
GENE
AA_POS
AA_POS_SUB
AA_FOUND_1
AA_FOUND_2
AA_FOUND_3
AA_FOUND_4

## *2.3 Structure of data*

### 2.3.1 From flat files towards a normalised structure

The data collected in HIV collaborations is presented on the following pages in a set of data files/tables. Typically data would be put into one data file that would hold one line/record per patient where each field is represented as a separate column in that dataset. Often a dataset could contain more than 3000 columns of data.

The implication of going from thousands of fields to fewer fields means that data is in fact transposed from the flat format into the normalised format.

Example:

**Example of a flat file structure**

| PATIENT | ALAT_D | ALAT_V | ALAT_U | ASAT_D | ASAT_V | ASAT_U |
|---|---|---|---|---|---|---|
| 999999 | 01-01-2000 | 15 | U/l | 01-01-2000 | 12 | U/l |

The normalised structure would then be like this:

**Example of a flat file structure that has been normalised**

| PATIENT | TYPE_ID | LAB_DATE | LAB_VAL | LAB_UNIT |
|---|---|---|---|---|
| 999999 | 1 | 01-01-2000 | 15 | U/l |
| 999999 | 2 | 01-01-2000 | 12 | U/l |

The type of measurement is identified through the TYPE_ID field. Here 1 codes for ALAT and 2 codes for ASAT.

| Code | Description |
|---|---|
| 1 | ALAT - Alanin-Aminotransferase |
| 2 | ASAT - Aspartat aminotransferase |

### 2.3.2 Technical considerations

To enable a normalised structure that minimises the number of columns dramatically, the one file solution must be broken into several minor tables. These breakdowns are driven by the different data characteristics.

Each table has a basic structure that includes the patient identifier, a code that represents e.g. drug, adverse event or laboratory test performed. Along with this combination values like date, result, unit etc are present for each record.

A record for a laboratory measurement would include:

**Example of a record for laboratory measurement.**

| Patient identifier | Measurement type identifier | Measured value | Unit of value | Date of determination |
|---|---|---|---|---|

A record for usage of an antiretroviral drug would incl.:

**Example of a record for antiretroviral drug use.**

| Patient identifier | Drug identifier | Start date for usage | End date for usage | Reason for discontinuation |
|---|---|---|---|---|

These issues imply that a set of distinct tables must be generated based on the "nature" of the data. Since laboratory, medication and event data both cannot and should not be mixed at least 3 tables must be designed. Additionally there are other types of information that need their own domains: background information on the patient (height, birth date etc.), visit related data (weight, blood pressure, wasting etc.), and resistance testing (the latter requires more consideration due to the diversity of data present).

In this protocol further separation of data into different tables are presented. These separations are not only based on the rules for the relational model and normalisation, but they are "culturally" related.

For example: antiretroviral treatment medication is kept in one table and other medication in another table; CD4 cell measurements and HIV-RNA measurements are put into separate tables, that are also different from the general laboratory table. These separations are done simply because data in these tables are of distinct importance in analysis and often are gathered more frequently and with more attention than other variables.

## 2.4 Coding conventions

### 2.4.1 Date codes

Although it is best to have precise dates in the format of YEAR-MONTH-DAY (ISO standard), it might be that some cohorts are limited to representing date data at the level of the month only, or information kept on the patient in the charts only defines dates to the month and in some cases only to the year. To solve this a set of date codes are presented here.

**Day unknown**

In this case the date should be coded as the 15$^{th}$ of the month – so that 1999-12-?? Becomes 1999-12-15. This enables the date to be no more than 15 days away from the actual date.

**Month and day unknown**

Best approach to this is to apply something similar, as with unknown dates, this would then mean that 1999-??-?? becomes 1999-07-01.

**Year unknown**

If the year is unknown but the presence of the date value is needed as in case of opportunistic infections or adverse events (see later in this document) a fictive date should be used that couldn't be mistaken with an actual date. An unknown year should be coded as 1911-11-11.

**Specification of precision**

An alternative to the above is to apply an additional field to each date field for which it is known that there might be issues regarding the precision of the dates. The field is then used to specify at which degree of the day, month or year the date is precise:

| Code | Precision of date |
|---|---|
| < | Before this date |
| D | Exact to the date |
| M | Exact to the month |
| Y | Exact to the year |
| > | After this date |

| Code | Precision of date |
|---|---|
| U | Unknown |

### 2.4.2 ICD-10 codes

Homepage for ICD-10: http://www.who.int/whosis/icd10/

The coding system is the official standard for coding of diseases, however there is a wide set of "homebrew" codes used within the HIV field in data coding in general, often it's a 3 or 4 letter codes that is an abbreviation for the AIDS defining disease. ICD-10 doesn't have single codes that represent all single CDC-C events and as a consequence of this a list of 3 to 4 letter codes is the recommended way of coding for all CDC stage C events

ICD-10 codes are however the recommended for codes AE's since it would become impossible for this protocol to maintain a complete list of all possible AE's. ICD-10 is also recommended for causes of death.

### 2.4.3 ATC codes

Homepage for ATC: http://www.whocc.no/atcddd/

ATC is a hierarchical structure for coding medication. The structure and hierarchy are best explained with an example of how a drug code is defined. Here it is on Indinavir:

| | |
|---|---|
| J | ANTIINFECTIVES FOR SYSTEMIC USE<br>(1st level, anatomical main group) |
| J05 | ANTIVIRALS FOR SYSTEMIC USE<br>(2nd level, therapeutic subgroup) |
| J05A | DIRECT ACTING ANTIVIRALS<br>(3rd level, pharmacological subgroup) |
| J05AE | Protease inhibitors<br>(4th level, chemical subgroup) |
| J05AE02 | Indinavir<br>(5th level, chemical substance) |

This hierarchy has some benefits as will be explained later, but one of its limitations is that it's impossible to "read" the code compared to the widely used 3 letter mnemonic codes for antiretroviral drugs.

Example:

| Drug | Code | ATC code |
|---|---|---|
| Indinavir | IDV | J05AE02 |

The difference is that the IDV code is easily readable, where the ATC code is not; going from a flat file structure to a normalised structure the human readable aspect becomes increasingly important. In the flat file format the column names and the possibility of labels makes data more or less readable; in the normalised format only the coding can help. Because of this the 3 letter codes are being presented in this document. However it must be stressed that usage of the ATC coding should be used to diminish the risk of several homebrew and non-compatible coding schemes.

**Currently however, the ATC scheme does not provide sufficient detail on the specific drugs, there is e.g. no official way to code Saquinavir as hard or soft gel. Thus a slight alteration to the set of codes will be presented in the sections of the ART and MED tables. The alterations are designed to extend the existing structure of ATC.**

One of the benefits is that the structure of ATC allows easier statistics on e.g. drug classes

J05AE Protease inhibitors
- J05AE01 Saquinavir
- J05AE02 Indinavir
- J05AE03 Ritonavir
- J05AE04 Nelfinavir
- J05AE05 Amprenavir
- J05AE06 Lopinavir

J05AF Nucleoside and nucleotide reverse transcriptase inhibitors
- J05AF01 Zidovudine
- J05AF02 Didanosine
- J05AF03 Zalcitabine
- J05AF04 Stavudine
- J05AF05 Lamivudine

J05AF06 Abacavir
J05AF07 Tenofovir disoproxil
J05AF30 Combinations[1]
J05AG Non-nucleoside reverse transcriptase inhibitors
J05AG01 Nevirapine
J05AG02 Delavirdine
J05AG03 Efavirenz

Although the codes might be harder to read they provide grouping mechanisms in the way they are coded. Interested readers should go to:

http://www.whocc.no/atcddd/atcsystem.html

…to learn about the structure of ATC. A fully updated database of ATC codes and DDD (Defined Daily Dosage) is available for querying.

### 2.4.4 Other codes

It is often necessary to code for values like “Yes”, “No” and “Unknown”, this document suggests that the following codes should be used:

**No, Yes and Unknown – general coding convention.**

| Code | Description |
|---|---|
| 0 | No |
| 1 | Yes |
| 9 | Unknown |

Unknown should be used to identify the difference between a value that has not yet been collected (Empty) and a value that cannot be collected (Unknown). Empty values should be required where Unknown values make little sense to keep querying for a value.

Example – weight:

[1] The code does currently not distinguish between Trizivir and Combivir

Depending on the unit in which weight is measured, a different value for Unknown should be applied. In the case of kg the “Unknown” code should be 999 and not just 9 or 99, the last two could be actual values.

Blank values, for SAS users also known as “.” and for database programmers known as NULL, should be used wherever specified in this protocol. However, sometimes it might be more correct just to omit the record if no value has been recorded, test has not been performed etc.

# 3 Specific descriptions

On the following pages the specific tables structure is described in detail, and a list of suggested codes, both standard and human readable, are presented.

All codes apart from trivial no, yes or unknown codes are presented as lookup tables, the usage of these are described in the Appendix II - Considerations for using the format to create a database - Lookup tables chapter which shows how these can be implemented in a database.

Along with the basic structure described in each “Core fields” section, additional fields containing additional or more specific data are described in the “Additional fields” sections. These fields were taken from several cohort collaborations but with the required changes that were needed for the specific data structures. This is presented to the reader to show that the core structure is not a fixed proposal but rather a basic structure, which can be altered by adding fields.

Fields shaded in light grey are strictly mandatory fields that define the uniqueness of each record. Issues regarding duplicates are discussed at the very end of this document.

## *3.1 tblBAS - Basic clinical, background and demographic information*

### 3.1.1 Core fields

**tblBAS**

| *Explanation of Variable* | Code to identify patient (Cohort Patient ID) | Code for Clinic/Centre/Hospital where patient is seen. | Birth date Leave BLANK if not able to give this information | First seen at clinic | Date of enrolment into the cohort | Gender/sex |
|---|---|---|---|---|---|---|
| *FIELD NAME* | **PATIENT** | **CENTER** | **BIRTH_D** | **FRSVIS_D** | **ENROL_D** | **GENDER** |
| *Format of data* | Character (or numeric if possible) | Character | yyyy-mm-dd | yyyy-mm-dd | yyyy-mm-dd | Numeric:<br>1=Male<br>2=Female<br>9=Unknown |

**tblBAS continued**

| *Explanation of Variable* | Height of patient at visit/most current | Mode of infection | Nationality or region of origin of patient | Ethnicity of patient | Date of seroconversion | Has the patient received antiretroviral treatment |
|---|---|---|---|---|---|---|
| *FIELD NAME* | **HEIGH**[2] | **MODE** | **ORIGIN** | **ETHNIC** | **SEROCO_D** | **RECART_Y** |
| *Format of data* | Numeric (metric):<br>999=Unknown | *See below for coding* | *Character – 1 – 3 letter/numeric codes See below for coding* | *See below for coding* | yyyy-mm-dd | Numeric:<br>0=No<br>1=Yes<br>9=Unknown |

**tblBAS continued**

| *Explanation of variable* | Has patient been given an AIDS diagnosis? | IF YES, date of AIDS diagnosis |
|---|---|---|
| *FIELD NAME* | **AIDS_Y** | **AIDS_D** |

[2] Please note that this field would be more appropriate to include in the tblVISIT table if data is collected for children.

| | | |
|---|---|---|
| *Format of data* | Numeric:<br>0=No<br>1=Yes<br>9=Unknown | yyyy-mm-dd |

### 3.1.2 Coding options / lookup tables

BIRTHDAY – depending on local laws it might be needed to code this as the year only, it is however strongly suggested to use a date value rather than an age or year in numeric form if possible.

**tblBAS_CODE_MODE - lookup table for mode of infection**

| Code | Mode of infection |
|---|---|
| 1 | homo/bisexual |
| 2 | injecting drug user |
| 3 | (1 + 2) |
| 4 | haemophiliac |
| 5 | transfusion, non-haemophilia related |
| 6 | heterosexual contact |
| 7 | (6 + 2) |
| 8 | Perinatal |
| 90 | other, (specify) |
| 99 | Unknown |

**tblBAS_CODE_ORIGIN - lookup table for origin**

| Code | Region codes for origin |
|---|---|
| 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 |

In case of a need for a more detailed level of origin (nationality) codes should be the ISO 2 or 3 letter codes, the list can be found several places on the Internet; this site has them ready for a database import:

http://www.din.de/gremien/nas/nabd/iso3166ma/codlstp1/db_en.html

The United Nations Statistics Division has the list of 3 letter codes, and also region classification and numeric representation as it is used by the Statistics Division of the United Nations Secretariat:

http://unstats.un.org/unsd/methods/m49/m49.htm

**tblBAS_CODE_ETHNIC - lookup table for ethnicity**

| Code | Ethnicity of patient |
|---|---|
| 10 | White |
| 20 | Black |
| 21 | Black African |
| 22 | Black Caribbean |
| 30 | Hispanic |
| 40 | Asian |
| 50 | American |
| 60 | Indigenous |
| 1020 | 1+2 |
| 1040 | 1+4 |
| 2030 | 2+3 |
| 3040 | 3+4 |
| 98 | Prohibited |
| 99 | Unknown |

### 3.1.3 Additional fields

For mode of infection, origin and death a set of other fields are often used to capture what cannot be coded. These fields are represented here as optional fields as it is the intention that the suggested codes applied to the MODE, ORIGIN, DEATH_R1-3 and ICD10_1-3 should be able to cover all possible values.

| Mode of infection OTHER | Origin of patient OTHER | Reason for death – other - description |
|---|---|---|
| **MODE_OTH** | **ORI_OTH** | **DEATH_OT** |
| Characters | Character | Character |

## *3.2 tblLTFU – Death and drop-out*

### 3.2.1 Core fields

**tblLTFU continued**

| *Explanation of variable* | Code to identify patient (Cohort Patient ID) | Has the patient DROPPED OUT? | IF YES, Date of Last Visit | IF YES, Reason for DROP |
|---|---|---|---|---|
| *FIELD NAME* | **PATIENT** | **DROP_Y** | **DROP_D** | **DROP_RS** |
| *Format of data* | Character (or numeric if possible) | Numeric:<br>0=No<br>1=Yes | yyyy-mm-dd | *See below for coding* |

**tblLTFU continued**

| *Explanation of Variable* | Has the patient died? | Date of Death | Was an autopsy Performed? |
|---|---|---|---|
| *FIELD NAME* | **DEATH_Y** | **DEATH_D** | **AUTOP_Y** |
| *Format of data* | Numeric:<br>0=No<br>1=Yes | yyyy-mm-dd | 0=No<br>1=Yes<br>9=Unknown |

**tblLTFU continued**

| *Explanation of Variable* | Cause of death | Coding of causal relation of the code given in DEATH_R1 to the death | Cause of death | Coding of causal relation of the code given in DEATH_R2 to the death | Cause of death | Coding of causal relation of the code given in DEATH_R3 to the death |
|---|---|---|---|---|---|---|
| *FIELD NAME* | **DEATH_R1** | **DEATH_RC1** | **DEATH_R2** | **DEATH_RC2** | **DEATH_R3*** | **DEATH_RC3*** |
| *Format of data* | *See below for coding* | Character with codes:<br>I = Immediate cause<br>U = Underlying cause/condition<br>C = Contributing cause<br>N = Not available | *See below for coding* | Character with codes:<br>I = Immediate cause<br>U = Underlying cause/condition<br>C = Contributing cause<br>N = Not available | *See below for coding* | Character with codes:<br>I = Immediate cause<br>U = Underlying cause/condition<br>C = Contributing cause<br>N = Not available |

List of DEATH_R# and DEATH_RC# should be continued for as many reasons that are recorded.

The DEATH_RC# fields should enable cohorts to transfer data in accordance with the Coding of Death project (CoDe). Please visit the CoDe website for more information (http://www.cphiv.dk/CoDe/). You are also welcome to contact the CoDe group for electronic sample forms for detailed collection of data used for the CoDe review process.

CoDe defines 1 immediate, 2 contributing and 1 underlying cause of death.

### 3.2.2 Coding options / lookup tables

**tblLTFU_CODE_DEATH - lookup table for cause of death**

| Code | Cause of Death |
|---|---|
| 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 |

**tblLTFU_CODE_DROP - lookup table for reason for drop out**

| Code | Reason for Drop Out |
|---|---|
| 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 |

* All patient data *may* be required to be deleted, patient id and reason for drop out should of cause remain in this table.

### 3.2.3 Additional fields

| Cause of death as ICD-10 **if available** | Cause of death as ICD-10 **if available** | Cause of death as ICD-10 **if available** |
|---|---|---|
| **ICD10_1** | **ICD10_2** | **ICD10_3** |
| Character | Character | Character |

List of ICD10_# inplace of or together with DEATH_R# and together DEATH_RC# and should be continued for as many reasons that are recorded. CoDe defines 1 immediate, 2 contributing and 1 underlying cause of death.

| Last date known to be alive |
|---|

| L_ALIVE |
| --- |
| yyyy-mm-dd |

## *3.3 tblOVERLAP - Cross-cohort identification*

### 3.3.1 Core fields

**tblOVERLAP table**

| *Explanation of variable* | Code to identify patient (Cohort Patient ID) | Code/name of the cohort | Unique patient identifier in other cohorts | Name of the cohort |
|---|---|---|---|---|
| *FIELD NAME* | **PATIENT** | **COHORT** | **PAT_OTH** | **COH_OTH** |
| *Format of data* | Character (or numeric if possible) | Character | Character | Character |

Table OVERLAP holds the identifiers of patients in overlapping (super-) cohorts.

Patients of an “original”-cohort who also participate in a “super”-cohort should be analysed within the “original”-cohort only. To suppress these patients from the datasets of the “super”-cohorts the identifier used in the “super”-cohort is needed. It is suggested that “original”-cohorts report id’s from the “super”-cohorts, since the “super”-cohorts might not even know the other ID’s. Often this information is only available at centre level.

A record should be present for each cohort that the patient is participating in (apart from it’s own “original”-cohort).

## *3.4 tblVIS - Basic follow-up/visit related data*

### 3.4.1 Core fields

**tblVIS table**

| *Explanation of variable* | Code to identify patient (Cohort Patient ID) | Date of patient visit | Weight of patient at visit | Is the patient gaining fat in the abdomen, neck, breast or other defined locations? | Is the patient experiencing loss of fat from extremities, buttocks or face? |
|---|---|---|---|---|---|
| *FIELD NAME* | **PATIENT** | **VIS_D** | **WEIGH** | **GAIN_Y** | **LOSS_Y** |
| *Format of data* | Character (or numeric if possible) | yyyy-mm-dd | Numeric (metric: kg): 999=Unknown | Numeric: 0=No 1=Yes 9=Unknown | Numeric: 0=No 1=Yes 9=Unknown |

Depending on the collaboration this data might be collected in intervals of a year, e.g. from July last to July this year. In that case all visit dates or a fixed number of visit dates for that period should be gathered, if the patient did not have a visit in the defined period, a record with the PATIENT id and empty fields for VIS_D etc. should be included.

## *3.5 tblART - Antiretroviral treatment*

### 3.5.1 Core fields

**tblART table**

| *Explanation of variable* | Code to identify patient (Cohort Patient ID) | Code representing the antiretroviral treatment | Date of Initiation of treatment | Date of stopping treatment | Reason for stopping treatment |
|---|---|---|---|---|---|
| *FIELD NAME* | **PATIENT** | **ART_ID** | **ART_SD** | **ART_ED** | **ART_RS** |
| *Format of data* | Character (or numeric if possible) | *See below for coding* | yyyy-mm-dd | yyyy-mm-dd | *See below for coding* |

### 3.5.2 Coding options / lookup tables

Updated list available at: http://www.cphiv.dk/HICDEP/Files/tabid/160/Default.aspx

**tblART_CODE_DRUG lookup table**

| Code (Extended ATC codes) | Anti-Retroviral Drugs |
|---|---|
| J05AF06 | Abacavir (1592U89) (ZIAGEN) |
| J05AF08 | Adefovir (PREVEON) |
| J05AF-ALO | Alovudine |
| J05AF-AMD | Amdoxovir (DADP) |
| J05AE05 | Amprenavir (AGENERASE) |
| J05A | ART unspecified |
| J05AE08 | Atazanavir (Reyataz) |
| J05AR06 | Atripla (Emtricitabine/Tenofovir/Efavirenz) |
| J05A-BEV | Beviramat |
| J05AG-CPV | Capravirine |

| Code (Extended ATC codes) | Anti-Retroviral Drugs |
|---|---|
| J05AR01 | Combivir (Zidovudine/Lamivudine) |
| J05AE10 | Darunavir (TMC-114, Prezista) |
| J05AG02 | Delavirdine (U-90152) (RESCRIPTOR) |
| J05AF02 | Didanosine (ddI) (VIDEX) |
| J05AR05 | Douvir-N (Zidovudine/Lamivudine/Nevirapine) |
| J05AG-DPC083 | DPC 083 |
| J05AG-DPC961 | DPC 961 |
| J05AG03 | Efavirenz (DMP-266) (STOCRIN, SUSTIVA) |
| J05AX-EVG | Elvitegravir (Gilead) |
| J05AG-EMV | Emivirine (MKC442) |
| J05AF09 | Emtricitabine |
| J05AX07 | Enfurvirtide (Fuzeon, T-20) |
| J05AF10 | Entecavir |
| J05AG-ETV | Etravirine (TMC 125) |
| J05AE07 | Fos-amprenavir (Telzir, Lexiva) |
| J05AF-FOZ | Fozivudine tidoxi |
| L01XX05 | Hydroxyurea/Hydroxycarbamid (Litalir) |
| J05AE02 | Indinavir (CRIXIVAN) |
| J05AR02 | Kivexa (Lamivudine/Abacavir) |
| J05AF05 | Lamivudine (3TC, EPIVIR) |
| J05AF-LDN | Lodenosine (trialdrug) |
| J05AE06 | Lopinavir/Ritonavir (Kaletra) |
| J05AG-LOV | Loviride |
| J05AX09 | Maraviroc (Pfizer) |
| J05AE-MOZ | Mozenavir (DMP-450) |
| J05AE04 | Nelfinavir (VIRACEPT) |
| J05AG01 | Nevirapine (VIRAMUN) |
| J05AG | NNRTI unspecified |
| J05AF | NRTI unspecified |

| Code (Extended ATC codes) | Anti-Retroviral Drugs |
|---|---|
| J05A-PBT | Participant in Blinded Trial |
| J05AE | PI unspecified |
| J05AX08 | Raltegravir (Merck) |
| J05AF-RVT | Reverset |
| J05AG-RPV | Rilpivirine (TMC-278) |
| J05AE03 | Ritonavir (NORVIR) |
| J05AE03-H | Ritonavir high dose (NORVIR) |
| J05AE03-L | Ritonavir low dose (NORVIR) |
| J05AE01 | Saquinavir (gel, not specified) |
| J05AE01-SQH | Saquinavir hard gel (INVIRASE) |
| J05AE01-SQS | Saquinavir soft gel (FORTOVASE) |
| J05AF04 | Stavudine (d4T) (ZERIT) |
| J05AF11 | Telbivudine |
| J05AF07 | Tenofovir (VilREAD) |
| J05AE09 | Tipranavir (Aptivus) |
| J05AR04 | Trizivir (Zidovudine/Lamivudine/Abacavir) |
| J05AR03 | Truvada (Tenofovir/Emtricabine) |
| J05AX-VIC | Vicriviroc (Schering) |
| J05AF03 | Zalcitabine (ddC) (HIVID) |
| J05AF01 | Zidovudine (AZT, RETROVIR) |

J05A DIRECT ACTING ANTIVIRALS
J05AE Protease inhibitors
J05AF Nucleoside and nucleotide reverse transcriptase inhibitors
J05AG Nucleosides and nucleotides reverse transcriptase inhibitors

A set of extended ATC codes are being presented here in order to code both more specific on subtypes of the drugs, e.g. saquinavir hard and soft gel, but also to enable coding of drugs that are at their trial stage and have not yet been assigned an ATC code, to do this the drug will be assigned the code elements as far down the levels as possible. Given two examples to illustrate this:

Saquinavir - Hard Gel:
J05AE01-SQH
Saquinavir - Hard Gel:
J05AE01-SQS
Saquinavir – not specified:
J05AE01

This will ensure the fidelity needed to distinguish between hard and soft gel and not specified, but also for analysis easily include all records which coding starts with J05AE01, regardless if the drug is hard or soft gel.

**tblART_CODE_RS lookup table for reason of stopping treatment**

| Code | Coding for Reason of Stopping Treatment |
|---|---|
| 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 …etc.) |
| 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 related) |
| 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, not specified above |
| 94.1 | Non-compliance |
| 95 | Physician's decision, not specified above |
| 96 | Pregnancy |
| 97 | Study treatment |
| 98 | Other causes, not specified above |
| 99 | Unknown |

Although most reasons mentioned above could be coded in ICD-10 codes, it is still certain that many reasons will never be defined in terms of ICD-10 (STI, physicians and patients decision), so to avoid a mixture of two different coding systems it is advised to go with the above list and build upon that for now.

### 3.5.3 Additional fields

Depending on the aim of the study it might be needed to gather both the dosage and the frequency of the dosage taken. However many cohorts do not collect this data and thus these fields are optional.

**ART_DO and ART_FR - ART dosage and frequency fields**

| Dosage (mg or mL) | Frequency |
|---|---|
| **ART_DO** | **ART_FR** |
| Numeric | *See below for coding* |

**tblART_CODE_FR Coding of frequency**

| Code | Code for frequency |
|---|---|
| 1 | 1 daily dose/qd |
| 2 | 2 daily doses/bid |
| 3 | 3 daily doses/tid |
| 4…and on | Code gives number of daily doses |

## *3.6 tblMED - Other medication*

### 3.6.1 Core fields

**tblMED table**

| *Explanation of variable* | Code to identify patient (Cohort Patient ID) | Code representing the treatment | Date of Initiation of Treatment | Date of stopping treatment |
|---|---|---|---|---|
| *FIELD NAME* | **PATIENT** | **MED_ID** | **MED_SD** | **MED_ED** |
| *Format of data* | Character (or numeric if possible) | *See below for coding* | yyyy-mm-dd | yyyy-mm-dd |

### 3.6.2 Coding options / lookup tables

Updated list available at: http://www.cphiv.dk/HICDEP/Files/tabid/160/Default.aspx

**tblMED_CODE Coding of other HIV-related drugs**

| Codes (Extended ATC codes) | Other Hiv-Related Drugs |
|---|---|
| C09 | ACE inhibitors |
| J05AB01 | Aciclovir (ZIVORAX) |
| J01GB06 | Amikacine (AMIKINE) |
| J02AA01 | Amphotericin B (FUNGIZON) |
| A14A | Anabolic steroids/appetite stimulants |
| B01AC | Anti-platelets |
| P01AX06 | Atovaquone (WELLVONE, MEPRONE) |
| J01FA10 | Azithomycine (ZITHROMAX) |
| L01DC01 | Bleomycine |
| L01AD02 | CCNU (LOMUSTINE) |

| Codes (Extended ATC codes) | Other Hiv-Related Drugs |
|---|---|
| J05AB12 | Cidofovir (VISTIDE) |
| J01MA02 | Ciprofloxacine (CIPROXINE, CILOXAN) |
| J01FA09 | Clarithromycin (KLACID) |
| J01FF01 | Clindamycine (DALACIN) |
| J04BA01 | Clofazimine (LAMPREN) |
| J01RA02 | Cosoltrime (MADERAN) |
| J01EE | Cotrimoxazole - Comb. of sulfonamides and trimethoprim (BACTRIM, EUSAPRIM, NOPIL) |
| L01AA01 | Cyclophosphamide (ENDOXAN) |
| L01AX04 | Dacarbazine (DTIC - Dome) |
| J04BA02 | Dapsone |
| L01DB01 | Doxorubicine, Adriamycine (ADRIBLASTIN, CAELYX, DOXIL) |
| J04AK02 | Ethambutol (EMB, MYAMBUTOL) |
| L01CB01 | Etoposide (VEPESIDE, EXITOP 100) |
| J05AB09 | Famciclovir |
| J02AC01 | Fluconazole (DIFLUCAN) |
| V03AF03 | Folinate of calcium (LEUCOVORINE) |
| J05AD01 | Foscarnet (FOSCAVIR) |
| J05AB06 | Ganciclovir (CYMEVENE) |
| L03AA02 | G-CSF/Filgastrim (NEUPOGEN) |
| J02AB | Imidazoles (DAKTARIN, NIZORAL, PEVARYL …) |
| A10A | Insulin or derivatives hereof |
| L03AB | Interferons |
| L03AC-IL2 | Interleukin 2 (PROLEUKIN) |
| J04AC01 | Isoniazide (RIMIFON) |
| J02AC02 | Itraconazole (SPORANOX) |
| J02AB02 | Ketoconazole |
| J01MA12 | Levofloxacin (TAVANIC) |
| C10 | Lipid lowering agents |
| L01BA01 | Methotrexate |
| J01AA08 | Minocycline (MINOCIN) |
| J01MA14 | Moxifloxacin |

| Codes (Extended ATC codes) | Other Hiv-Related Drugs |
|---|---|
| L01CA02 | Oncovin (VINCRISTINE) |
| A10B | Oral antidiabetic agents |
| C-HYP | Other anti-hypertensive agents [C02, C03, C04, C07, C08] |
| L03AB11 | Peginterferon alfa-2a (PEGASYS) |
| L03AB-AL2 | Peginterferon alfa-2a/alfa-2b (PEGINTRON, PEGASYS) |
| L03AB10 | Peginterferon alfa-2b (PEGINTRON) |
| P01CX01 | Pentamidine aerosol (PENTACARNET) |
| L01XB01 | Procarbazine (NATULAN) |
| J04AK01 | Pyrazinamide (PYRAZINAMID) |
| P01BD01 | Pyrimethamine (DARAPRIM) |
| P01BD51 | Pyrimethamine/Sulfadoxine (FANSIDAR) |
| J05AB04 | Ribavirin |
| J04AB04 | Rifabutin (MYCOBUTIN) |
| J04AB02 | Rifampin (RIMATICIN) |
| J04AM02 | RIFATER |
| J01EC02 | Sulfadiazine |
| J01EE01 | Sulfamethoxazole and trimethoprim (Bactrim) |
| J01EE03 | Sulfametrole and trimethoprim - Cosoltrime (MADERAN) |
| J01EA01 | Trimethoprim (MONOTRIM, NOPIL) |
| J05AB11 | Valaciclovir (VALTEX) |
| L01CA01 | Vinblastin (VELBE) |
| J02AC03 | Voriconazole |

J01 ANTIBACTERIALS FOR SYSTEMIC USE
J02 ANTIMYCOTICS FOR SYSTEMIC USE
J04 ANTIMYCOBACTERIALS
J05 ANTIVIRALS FOR SYSTEMIC USE
L01 ANTINEOPLASTIC AGENTS
L03 IMMUNOSTIMULANTS

### 3.6.3 Additional fields

Please see same section under tblART - Antiretroviral treatment

## *3.7 tblDIS - Opportunistic infections*

### 3.7.1 Core fields

**tblDIS table**

| *Explanation of variable* | Code to identify patient (Cohort Patient ID) | Code to identify event | Date of event | Means of diagnosis | Other location, only to be filled out if code alone is not sufficient |
|---|---|---|---|---|---|
| *FIELD NAME* | **PATIENT** | **DIS_ID** | **DIS_D** | **DIS_WD** | **DIS_OTH*** |
| *Format of data* | Character (or numeric if possible) | *See below for coding* | yyyy-mm-dd | *See below for coding* | Character |

* DIS_OTH might be part of the records unique identification (fields shaded in grey)

### 3.7.2 Coding options / lookup tables

**tblDIS_CODE Coding of severe opportunistic infections**

| Code | Severe Opportunistic Infections |
|---|---|
| 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 |
| HIST | Histoplasmosis, extrapulm. |
| WAST | HIV Wasting Syndrome |

| Code | Severe Opportunistic Infections |
|---|---|
| ISDI | Isosporiasis diarrhoea (duration > 1 month) |
| LEIS | Leishmaniasis, visceral |
| MCDI | Microsporidosis diarrhoes (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-tyhpoid) (recurrent) |
| TOX | Toxoplasmosis, brain |
| FBLS | Focal Brain lesion |
| **Code** | **Malignancies** |
| 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 |

**tblDIS_CODE_DIAG Coding of means of diagnosis**

| Code | Means of diagnosis |
|---|---|
| 1 | Definitive diagnosis |
| 2 | Presumptive diagnosis |
| 3 | Diagnosis from autopsy |
| 4 | Diagnosis from registry |

### 3.7.3 Additional fields

Please see same section under tblAE - Adverse Events for specification on optional fields.

## *3.8 tblLAB - Laboratory values*

### 3.8.1 Core fields

**tblLAB table**

| *Explanation of variable* | Code to identify patient (Cohort Patient ID) | Code representing the measurement | Date of measurement/sample | Value of measurement? | What unit of measurement was used ? |
|---|---|---|---|---|---|
| *FIELD NAME* | **PATIENT** | **LAB_ID** | **LAB_D** | **LAB_V** | **LAB_U** |
| *Format of data* | Character (or numeric if possible) | *See below for coding* | yyyy-mm-dd | Numeric<br>-1 = undetectable or detection limit as negative value | *See below for coding* |

### 3.8.2 Coding options / lookup tables

**tblLAB_CODE Coding of measurement type**

| Code | Measurement |
|---|---|
| ALB | Albumine |
| ALP | Alk.P.tase |
| ALT | Alanin-Aminotransferase |
| AMY | Amylase |
| APT | Alk. Phosphatate |
| AST | Aspartat aminotransferase |
| BIL | Total Bilirubin |
| CD3 | CD3 |
| CD3P | % CD3 of leukocytes |
| CD8 | CD8 |
| CD8P | % CD8 of leukocytes |
| CHOL | Total Cholesterol |

| Code | Measurement |
|---|---|
| CL- | Cl- |
| CRE | Creatinine |
| GLUC | Glucose |
| GLYCE | Glycemia |
| HAEM | Haemoglobin |
| HDL | Serum HDL |
| HEMA | Hematocrit |
| INR | Quick/INR |
| LACT | Lactate |
| LEUK | Leukocytes |
| LYM | Lymphocytes |
| LYMP | % Lymphocytes of leukocytes |
| MCV | MCV |
| NA+ | Na+ |
| NEU | Neutrophils |
| PHA | PH arterial |
| PHV | PH venous |
| PLT | Platelet count |
| PP | PP factor (II, VII, X) |
| THR | Thrombocytes |
| TRIG | Serum Triglyceride |
| URA | Uric acid |
| WBC | WBC count |

**tblLAB_CODE_UNITS Coding of units**

| …in case of measurement of: | Unit codes/strings |
|---|---|
| Alanin-Aminotransferase | 5=IU/L (u/L) |
| | 11=μkat/L |
| Albumine | 2=gm/L |
| Alk. Phosphatate | 5=IU/L (u/L) |
| Amylase | 5=IU/L (u/L) |
| | 11=μkat/L |
| Creatinine | 6=μmol/L |
| Glucose | 1=mmol/L |
| Haemoglobin | 1=mmol/L |

| …in case of measurement of: | Unit codes/strings |
|---|---|
| | 2=gm/L |
| | 3=gm/dL |
| Lactate | 1=mmol/L |
| Lactate | 4=mg/dL |
| Lymphocyte count | 8=1E+9/L |
| | 9=1E+6/L |
| | 10=cells/µL |
| Platelet count | 8=1E+9/L |
| | 9=1E+6/L |
| | 10=cells/µL |
| Quick/INR | 7=INR |
| Serum HDL | 1=mmol/L |
| | 2=gm/L |
| | 3=gm/dL |
| | 4=mg/dL |
| | 5=IU/L (u/L) |
| Serum Triglyceride | 1=mmol/L |
| | 2=gm/L |
| | 4=mg/dL |
| Total Bilirubin | 6=µmol/L |
| Total Cholesterol | 1=mmol/L |
| | 2=gm/L |
| | 3=gm/dL |
| | 4=mg/dL |
| WBC count | 8=1E+9/L |
| | 9=1E+6/L |
| | 10=cells/µL |

It is recommended to use the string codes from the above table since this makes the data human readable.

### 3.8.3 Additional fields

Other detailed information regarding the patient and the measurement would be relevant, like the proposed fasting information shown below.

**LAB_FA field**

| Was the blood sample taken while fasting? |
|---|
| **LAB_FA** |
| Numeric:<br>0=No<br>1=Yes<br>9=Unknown |

**LAB_ST field**

| Specimen type |
|---|
| **LAB_ST** |
| Character<br>WB=Whole Blood<br>P=Plasma<br>S=Serum |

Depending on the set of measurements collected and the mandatory fields applicable to these individual measurements, it might be useful to separate the LAB table into several sub tables. This is already shown for the CD4 and RNA measurements: the level of detail needed for CD4 is less than for the LAB variables in general (no unit since it's always the same), while for RNA the data required is more detailed (assay and detection limit).

## *3.9 tblLAB_CD4 - Laboratory values*

### 3.9.1 Core fields

**tblLAB_CD4 table**

| Explanation of variable | Code to identify patient (Cohort Patient ID) | Date of measurement | Value of CD4 measurement |
|---|---|---|---|
| **FIELD NAME** | **PATIENT** | **CD4_D** | **CD4_V** |
| Format of data | Character (or numeric if possible) | yyyy-mm-dd | Numeric (per microliter)<br>-1 = undetectable or detection limit as negative value |

If needed the above structure could be used to form a CD8 table: tblLAB_CD8.

### *3.10* Additional fields

The above table is assumed to contain absolute CD4 cell counts per mL as standard. In case CD4 % should be collected as well, please append the following field to the table:

| What is the unit of measurement value? |
|---|
| **CD4_U** |
| Numeric with codes<br>(or full string):<br>1=cells/µl<br>2=% |

## *3.11 tblLAB_RNA - Laboratory values*

### 3.11.1 Core fields

**tblLAB_RNA table**

| *Explanation of variable* | Code to identify patient (Cohort Patient ID) | Date of Measurement/Sample | HIV-RNA measurement value | Lower Limit of HIV-RNA Assay | IF AVAILABLE, What type of VIRAL ASSAY was used for this measurement? |
|---|---|---|---|---|---|
| *FIELD NAME* | **PATIENT** | **RNA_D** | **RNA_V** | **RNA_L** | **RNA_T** |
| *Format of data* | Character (or numeric if possible) | yyyy-mm-dd | Numeric<br>-1 = undetectable/below level of detection or detection limit as negative value | Numeric | *See below for coding* |

### 3.11.2 Coding options / lookup tables

**tblLAB_RNA_CODE_ASSAY Coding of viral assays**

| Code | Viral assay used |
|---|---|
| 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 |

| Code | Viral assay used |
|---|---|
| 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 |
| 65 | Cobas 1.5 |
| 66 | Cobas 1.5 ultra-sensitive |
| 90 | Other |
| 99 | Unknown |

### 3.11.3 Additional fields

**RNA_UL Upper limit of assay**

| IF AVAILABLE, Upper Limit of assay |
|---|
| **RNA_UL** |
| Numeric |

## *3.12 tblLAB_BP - Laboratory values – Blood pressure*

### 3.12.1 General format

**tblLAB_BP Blood pressure table**

| *Explanation of variable* | Code to identify patient (Cohort Patient ID) | Date of Measurement/Sample | If the patients Blood Pressure was taken, what was the Systolic Blood Pressure | If the patients Blood Pressure was taken, what was the Diastolic Blood Pressure | What unit of measurement was used ? |
|---|---|---|---|---|---|
| *FIELD NAME* | **PATIENT** | **BP_D** | **BP_SYS** | **BP_DIA** | **BP_U** |
| *Format of data* | Character (or numeric if possible) | yyyy-mm-dd | Numeric | Numeric | *See below for coding* |

### 3.12.2 Coding options / lookup tables

**tblLAB_BP_CODE Coding of unit for blood pressure**

| Code | Unit for blood pressure |
|---|---|
| 1 | mmHg |
| 2 | cmHg |
| 3 | Kpa |

## *3.13 tblLAB_VIRO - Laboratory values – viro-/serology*

### 3.13.1 General format

**tblLAB_VIRO Viro-/serology table**

| *Explanation of variable* | Code to identify patient (Cohort Patient ID) | Viral test | Measurement date | Measurement result | Measurement value (HCV-RNA & HBV-DNA only) (copies/ml) | Measurement unit |
|---|---|---|---|---|---|---|
| *FIELD NAME* | **PATIENT** | **VS_ID** | **VS_D** | **VS_R** | **VS_V** | **VS_U** |
| *Format of data* | Character (or numeric if possible) | *See below for coding* | yyyy-mm-dd | 0= negative<br>1= positive<br>9= unknown/borderline | Numeric | *See below for coding* |

### 3.13.2 Coding options/lookup tables

**tblLAB_VIRO_CODE Coding of viro-/serology test**

| Code | Viral test |
|---|---|
| BVA | Bacterial vaginosis unspecified method |
| BVAC | Bacterial vaginosis - clinical |
| BVAG | Bacterial vaginosis - gram stain |
| CHLA | Chalmydia |
| CHLA | Chalmydia |
| CMVA | CMV anitbodies |
| GONO | Gonorrhoe |
| HBV | Marker for hepatitis B infection (=HBVAC) - test unknown |
| HBVAC | HBV antibody (core) |
| HBVAE | HBV antibody (envelope) |
| HBVAS | HBV antibody (surface) |
| HBVD | HBV-dna |

| | |
|---|---|
| HBVGE | HBV antigen (envelope) |
| HBVGS | HBV antigen (surface) |
| HCV | Marker for hepatitis C infection - test unknown |
| HCVA | HCV antibody |
| HCVG | HCV antigen |
| HCVR | HCV-rna |
| HIV-1 | HIV-1 test |
| HIV-2 | HIV-2 test |
| HIVAE | HIV antibodies ELISA |
| HIVAWB | HIV antibodies Western blot |
| MYCO | Mycoplasma |
| P24AG | P24 Ag |
| RUB | Rubella |
| STR | Streptococcus, group B |
| TOXA | Toxo antibodies |
| UREP | Ureaplasma |

**tblLAB_VIRO_CODE_UNITS Coding of units**

| **Test measurement unit** |
|---|
| 1=copies/mL |
| 2=IU/mL |
| 3=Geq (millions of genome equivalent) |

## *3.14 Additional fields*

**VS_LL Lower limit of assay**

| IF AVAILABLE, Lower Limit of assay |
|---|
| **VS_LL** |
| Numeric |

**VS_UL Upper limit of assay**

| IF AVAILABLE, Upper Limit of assay |
|---|
| **VS_UL** |
| Numeric |

**VS_T Type of assay**

| IF AVAILABLE, What type of ASSAY was used for this measurement? |
|---|
| **VS_T** |
| *See below for coding* |

**tblLAB_VIRO_CODE_ASSAY Coding of assay**

| Code | Viral test used |
|---|---|
| 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 |

## *3.15 tblLAB_RES - Resistance testing*

### 3.15.1 Core fields

**tblLAB_RES table - Test background information (main table)**

| *Explanation of Variable* | Code to identify patient (Cohort Patient ID) | The assigned sample ID | Date of the actual sample taken (NOT the test date) | Date and time when the sequencing was performed | Name of laboratory where the test was performed | Library/algorithm used to identify resistance mutations |
|---|---|---|---|---|---|---|
| *FIELD NAME* | **PATIENT** | **SAMP_ID** | **SAMPLE_D** | **SEQ_DT** | **LAB** | **LIBRARY** |
| *Format of data* | Character (or numeric if possible) | Character (or numeric if possible) | yyyy-mm-dd | yyyy-mm-dd hh:mm | Character | Character |

**tblLAB_RES table continued**

| *Explanation of Variable* | Name/identifier of reference HIV strain used to find mutations | Vendor and version/name of the kit used for the test | Software and version used to determine resistance | Type of test | Subtype of HIV-RNA |
|---|---|---|---|---|---|
| *FIELD NAME* | **REFSEQ** | **KIT** | **SOFTWARE** | **TESTTYPE** | **SUBTYPE** |
| *Format of data* | Character | Character | Character | Numeric | Character |

**tblLAB_RES_LVL_1 Level 1 - Nucleotide sequences (PRO, RT, GP41, GP120) (no entry if the test was a phenotype test)**

| *Explanation of Variable* | The assigned sample ID | Type of nucleotide sequence if available | Start position for the sequence | Stop position for the sequence | Nucleotide sequence if available |
|---|---|---|---|---|---|
| *FIELD NAME* | **SAMP_ID** | **SEQTYPE** | **SEQ_STAR** | **SEQ_STOP** | **SEQ_NUC** |
| *Format of data* | Character (or numeric if possible) | Character<br>PRO = PRO sequence<br>RT = RT sequence<br>GP41 = GP41 sequence<br>GP120 = GP120 sequence | Numeric | Numeric | Character |

**tblLAB_RES_LVL_2 Level 2 - Mutations**

| *Explanation of Variable* | The assigned sample ID | Type of sequence/gene (PRO, RT, GP41, GP120) | Position of the mutation in the sequence | Subposition used to code insertions | Mutation (Amino acid) found in the sequence | Mutation (Amino acid) found in the sequence (if more than 1) | Mutation (Amino acid) found in the sequence (if more than 2) | Mutation (Amino acid) found in the sequence (if more than 3) |
|---|---|---|---|---|---|---|---|---|
| *FIELD NAME* | **SAMP_ID** | **GENE** | **AA_POS** | **AA_POS_SUB** | **AA_FOUND_1** | **AA_FOUND_2** | **AA_FOUND_3** | **AA_FOUND_4** |
| *Format of data* | Character (or numeric if possible) | Character<br>PRO = PRO sequence<br>RT = RT sequence<br>GP41 = GP41 sequence<br>GP120 = GP120 sequence | Numeric | Character:<br>a = first<br>b = second<br>etc. | Character<br><br>Empty = Amino acid has been deleted | Character | Character | Character |

AA_FOUND could be extended (AA_FOUND_#) if mixtures with more than 4 amino acids are found.

**tblLAB_RES_LVL_3 Level 3 - Resistance test result**

| *Explanation of variable* | The assigned sample ID | Drug code of antiretroviral | Score of resistance or recommendation given from the test. |
|---|---|---|---|
| *FIELD NAME* | **SAMP_ID** | **ART_ID** | **RES_SCOR**[3] |
| *Format of data* | Character (or numeric if possible) | Character – see codes for ART | Character |

**Resistance should be reported at lowest level of interpretation possible – so if the nucleotide sequence is available this should be reported rather than the list of mutations or resistance scores. However, the resistance test results should be captured if they have been part of the physician's treatment decisions for the patient.**

The above four tables are designed to capture several possible formats the clinics and cohorts might have recorded resistance test data in.

[3] These scorings and recommendation will have to be unified towards a common scoring system in order to use these for analysis. This step is however a final step that should be carried out after the collection and merging of data.

Once this data is gathered it should like all other tables be quality assessed. For the full nucleotide sequences a short guide on "Sequence Quality Control" can be found here:

**http://hiv-web.lanl.gov/content/hiv-db/CONTAM/contam_main.html**

**Coding options/lookup tables**

**TESTTYPE Resistance test - type**

| Code | Code for TESTTYPE |
|---|---|
| 1 | Genotype |
| 2 | Phenotype |
| 9 | Other |

### 3.15.2 Additional fields

As shown with the core fields, the SAMP_ID is the link between the 3 levels of data and the test background information table. The sample identifier, however, must be unique for the format to work. This might not always be the case. If needed SAMPLE_D could be used as an additional part of the key, or just SAMPLE_D along with the PATIENT key[4].

Some prior assessment of the assigned sample identifiers has to be done in order to avoid duplicates.

In a running database the duplicate issues are easily resolved by adding a unique auto-generated key as the identifier between 3 levels of data and the test background information table SAMP_ID.

Along with the SAMP_ID it might be necessary to store the ID assigned to the sample at both the testing laboratory but also the centres laboratory in order to track the sample. Each of these could also be used as the SAMP_ID value.

[4] However this raises the issue about several aliquots from the same day will look like duplicates in the tables.

**SAMP_LAB, SAM_INT - different sample ID's**

| The assigned sample ID at the lab where the resistance test is preformed. | The assigned sample ID from the centre. |
|---|---|
| **SAMP_LAB** | **SAMP_INT** |
| Character (or numeric if possible) | Character (or numeric if possible) |

In cases where the amino acid sequence is collected rather than the nucleotide sequence, the field SEQ_NUC in tblLAB_RES_LVL_1 might be replaced with SEQ_AA, which is the nucleotide sequence, expressed in an amino acid sequence.

**SEQ_AA Amino acid sequence**

| Amino acid sequence if available (empty if test was phenotype) |
|---|
| **SEQ_AA** |
| Character |

However using the amino acid sequence does not give the same detail of data as the nucleoside sequence: wobbles in the nucleoside sequence can either complicate the reading and alignment of the amino acid sequence or the wobbles can be lost and silent mutations are lost.

For phenotype test results it will be necessary to extend the tblLAB_RES_LVL_3 table with a field to store the cut-off value.

**RES_CUT Cut-off value for phenotype test result**

| Cut-off value for phenotype test result |
|---|
| **RES_CUT** |
| Character |

However using the amino acid sequence does not give the same detail of data as the nucleoside sequence: wobbles in the nucleoside sequence can either complicate the reading and alignment of the amino acid sequence or the wobbles can be lost and silent mutations are lost.

## 3.16 tblAE - Adverse Events

### 3.16.1 Core fields

**tblAE table**

| *Explanation of variable* | Unique Event Identifier(foreign key to the different event tables) | Code to identify patient (Cohort Patient ID) | Date of event | Code to identify event | Code to further specify event |
|---|---|---|---|---|---|
| *FIELD NAME* | **EVENT_ID** | **PATIENT** | **AE_D** | **AE_ID** | **AE_SPEC** |
| *Format of data* | Numeric | Character (or numeric if possible) | yyyy-mm-dd | *See below for coding* | *See below for coding* |

**tblAE table continued (verification status fields)**

| *Explanation of variable* | The source documentation is available | The date for source documentation verification | The monitor has verified the source documentation | The date for monitor verification | Final verification / approval | Final verification date | Final verification by – signature/name |
|---|---|---|---|---|---|---|---|
| *FIELD NAME* | **SRCDOC_Y** | **SRCDOC_D** | **VERIFY_Y** | **VERIFY_D** | **APPROV_Y** | **APPROV_D** | **APPROV_S** |
| *Format of data* | Numeric:<br>1=Yes<br>0=No | yyyy-mm-dd | Numeric:<br>1=Yes<br>0=No | yyyy-mm-dd | Numeric:<br>1=Yes<br>0=No | yyyy-mm-dd | Character |

**Please see the HICDEP website for examples of detailed AE tables for the events listed in 3.16.2 below. Data format is available in the HICDEP DAD event forms document.**

### 3.16.2 Coding options / lookup tables

**Examples of AE codes on adverse events collected in DAD**

| Code (AE_ID) | Adverse Event |
|---|---|
| AMI | Acute myocardial infarction |
| CLD | Chronic liver disease |
| COR | (Possible) Coronary Death |
| DIA | Diabetes mellitus |
| ESRD | End stage renal disease |

| Code (AE_ID) | Adverse Event |
|---|---|
| FAT | Fatal case with insufficient data |
| ICP | Invasive Cardiovascular Procedures |
| NADM | Non-AIDS defining malignancies |
| STR | Stroke (infarction or haemorrhagia) |

| Code (AE_ID) | Code (AE_SPEC) | Description |
|---|---|---|
| AMI | DAMI | Definitive Myocardial infarction |
| | PAMI | Possible Myocardial infarction |
| ICP | ANG | Invasive Cardiovascular Procedures: Coronary angioplasty/stenting |
| | BYP | Invasive Cardiovascular Procedures: Coronary artery by-pass grafting |
| | END | Invasive Cardiovascular Procedures: Carotic endarterectomy |
| NADM | ALL | Leukemia: Acute lymphoid |
| | AML | Leukemia: Acute myeloid |
| | ANUS | Anus cancer |
| | BLAD | Bladder cancer |
| | BRCA | Breast cancer |
| | CERV | Cervical dysplasia/carcinoma in situ |
| | CLL | Leukemia: Chronic lymphoid |
| | CML | Leukemia: Chronic myeloid |
| | COLO | Colon cancer |
| | COTC | Connective tissue cancer |
| | HDL | Hodgkin lymphoma |
| | KIDN | Kidney cancer |
| | LEUK | Leukemia: unspecified |
| | LIPC | Lip cancer |
| | LIVR | Liver cancer |
| | LUNG | Lung cancer |
| | MALM | Malignant melanoma |
| | MEAC | Metastasis: of adenocarcinoma |
| | MEOC | Metastasis: of other cancertype |

| Code (AE_ID) | Code (AE_SPEC) | Description |
|---|---|---|
| | MESC | Metastasis: of squamuos cell carcinoma |
| | META | Metastasis: unspecified |
| | MULM | Multiple myeloma |
| | PENC | Penile cancer |
| | PROS | Prostate cancer |
| | RECT | Rectum cancer |
| | STOM | Stomach cancer |
| | TESE | Testicular seminoma |
| | UTER | Uterus cancer |
| STR | SHAE | Stroke: Haemorrhagia |
| | SINF | Stroke: Infarction |
| | SUNK | Stroke: Unknown |

### 3.16.3 Additional fields

**AE_Y Event – No and yes**

| Has the patient had an event. |
|---|
| **AE_Y** |
| Numeric:<br>0=No<br>1=Yes<br>9=Unknown |

AE_Y: This field should be more or less obsolete if the date codes are applied, otherwise it could be used to state that an event had occurred but the date (if left blank) was not known or, if coded 9 (Unknown), that the centre was not aware if an event has occurred or not.

**AE_NAME Full name of event**

| Full name of the event |
|---|
| **AE_NAME** |

| Character |
|---|

AE_NAME: The full name as it might have been entered into the database or presented on a case report form.

**AE_DESCRIP Full description of event**

| Full description of the event |
|---|
| **AE_DESCRIP** |
| Character |

AE_DESCRIP: The actual text description of the event, especially in clinical trials where the physician's full diagnose might be required.

**AE_R_Y Relation to treatment**

| Relation to treatment |
|---|
| **AE_R_Y** |
| Numeric:<br>0=Not related<br>1=Definite<br>2=Remote/Unlikely<br>3=Possible<br>4=Probable |

# 4 Acknowledgement

We want to thank the following collaborations and cohorts for having given us insight into their data exchange protocols:

ART-CC
CASCADE
D:A:D
EuroSIDA
PLATO
SHCS

We would also like to thank the following persons for contributing to this document:

Allen Sawitz
Andrew Phillips
Caroline Sabin
Céline Colin
Martin Rickenbach
Morten Svenson
Nina Friis-Møller
Olivia Keiser
Rikke Salbøl Brandt
Sarah Walker
Stephen Hart
Zoé Fox

# Appendix I - Considerations for data management

## *Uniqueness of PATIENT ID*

Each table in this document is shown with the PATIENT field as part of the unique identifier for each record, in many cases it might be necessary to specify both PATIENT and COHORT as identifier since the PATIENT id might not be unique across cohorts, this should however be implemented only if needed.

| Code to identify patient (Cohort Patient ID) | Code/name of the cohort |
|---|---|
| **PATIENT** | **COHORT** |
| Character (or numeric if possible) | Character |

Another solution, which might be worth considering when building the final dataset for the analysis, is to concatenate the patient id and the cohort id into a single unique new patient id.

## *Duplicate records*

Transposition of the flat format data, where there is one record per patient, into a normalised structure that has multiple rows per patient requires that the combination of PATIENT, TYPE_ID and LAB_DATE is unique for each row in the table.

The normalised structure and the relational model does not allow for duplicate records to exist:

**Example of a normalised structure with duplicate records**

| PATIENT | TYPE_ID | LAB_DATE | LAB_VAL | LAB_UNIT |
|---|---|---|---|---|
| 999999 | 1 | 01-01-2000 | 15 | U/l |
| 999999 | 1 | 01-01-2000 | 15 | U/l |

The typical solution to this is to generate an auto-incremented value – RECORD_ID - for each record in the table:

**Example of a normalised structure with duplicate records resolved through auto-incrementing the RECORD_ID field**

| RECORD_ID | PATIENT | TYPE_ID | LAB_DATE | LAB_VAL | LAB_UNIT |
|---|---|---|---|---|---|
| 1 | 999999 | 1 | 01-01-2000 | 15 | U/l |
| 2 | 999999 | 1 | 01-01-2000 | 15 | U/l |

But it can, and should be dealt with in the design of the study. The presence of duplicate records like in the example shown above might not make any sense; if they did occur it then would be on account of a recording error. If, however, it was important to keep records that show measurements taken the same day but at different times, the format of the LAB_DATE shouldn't be date (YYYY-MM-DD) but instead should be date-time (YYYY-MM-DD hh:mm:ss):

**Example of a normalised structure with duplicate records resolved through change of date data type from date to date-time**

| PATIENT | TYPE_ID | LAB_DATE | LAB_VAL | LAB_UNIT |
|---|---|---|---|---|
| 999999 | 1 | 01-01-2000 10:00:00 | 15 | U/l |
| 999999 | 1 | 01-01-2000 14:00:00 | 15 | U/l |

Different and more sophisticated methods to make records unique will be presented later in this document under the definition of the LAB table.

# Appendix II - Considerations for using the format to create a database

## *Administrative fields*

Sometimes it might be needed to have a fixed value that shows from which visit or merger a value originates, this does not only apply to the VIS table but could be applied to all tables. This however does depend on the nature of the database and needs for data management, the field below should be considered an administrative support field for data management.

**VISIT field**

| Visit number |
|---|
| **VISIT** |
| Numeric:<br>0 = Baseline Visit<br>1 = First Follow Up Visit<br>2 = Second Follow Up Visit<br>etc. |

Often the above field is used for clinical trials databases where there is a need to associate the data directly with a given week's follow-up. Codes could then be the week number e.g. 4, 12, 24 etc or –1 for screening/randomisation and 0 for baseline visits.

In some cases it might be useful to have a separate field that defines the correct order of the periods. This becomes important where several dates are incomplete (unknown days, unknown months and possibly unknown years). The ordering by date would then not be correct.

One solution to this is use a PERI_ID field that numbers the periods from the 1st until Nth usage

**PERI_ID field**

| Period of usage (1st, 2nd, 3rd etc.) |
|---|
| **PERI_ID** |
| Numeric |

However this is an optional field that for most cohorts may not be needed. It also requires additional maintenance to keep it updated.

For databases that work with double data entry, such as most clinical databases, it becomes necessary to make each data entry unique and backwards traceable. For this to work a field like the above would have to be part of the primary key of each table that requires double data entry.

**ENTRY_ID field**

| Number of data entry |
|---|
| **ENTRY_ID** |
| Numeric<br>1 = first data entry<br>2 = second data entry<br>3 = comparison result of 1st and 2nd data entry<br>4 = final approved record including corrections |

With respect to performance, it might also be a good design in to have 3 copies of each table, one to hold the data while being entered and compared, one for the two data entries to be archived into once a final record has been approved and a table holding the final and approved values. This way it is avoided that queries will have to work on checking for ENTRY_ID = 4 and to select amongst a table holding 3 times the almost same data.

As part of an audit trail in a database a time stamp field could be added for each record to fix the exact time when the record last was inserted or updated. Along with the time stamp name of the user who entered or altered data can be recorded.

**T_STAMP and USER_LOG fields**

| Date and time of data entry | Username of user that last inserted or updated data |
|---|---|
| **T_STAMP** | **USER_LOG** |
| yyyy-mm-dd hh:mm:ss | Character |

Often it's necessary to keep a log of user action in a separate table. The above suggestion will only be valid for inserts and updates, and only be valid for the most recent action performed.

To record a complete audit trail a logging facility must be implemented. In most database management systems this is done using triggers on the tables. For any insert, update or delete actions performed on the data, the user, time stamp, old value and new value are recorded in the logging table.

The T_STAMP field could also include information about which time zone is relevant for data entry. Depending on database requirements this might in fact be mandatory if the FDA's 21 CRF part 11 on electronic records and signatures applies.

## Further normalisation

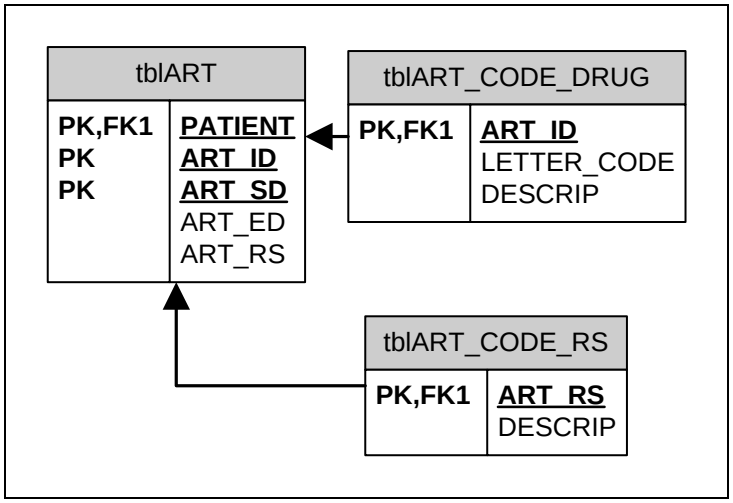
Depending on performance considerations it might be worth looking at how data are queried for data entry and data analysis. A smaller tblBAS table might increase performance: Since processing a smaller table is always faster than processing a larger table, one could put drop-out, death, birthday, date of aids diagnosis, etc. into separate tables and keep the core patient list in a separate master table

But if the database is used e.g. for BMI calculations directly on the running database, performance might be enhanced by keeping the patient list and the height together in the same table so that a query involves 2 tables (tblBAS and tblVIS) rather than perhaps 3 or more.

Another consideration is space. Although it may not be much of an issue, it will be possible to minimise the actual size of the database by putting fields that may be empty for most patients, like death information, into a separate table in a 0-1 to 1 relation to the master table.

## Lookup tables

In a running database the #_ID fields could be implemented as a foreign key to a linked lookup table containing all possible codes and their corresponding definitions in a text string.




**ART lookup table**

This setup not only enables integrity of the data, but also defines the domain[5] for the #_ID values and enables data to both become human readable and easily recoded[6].

An important note on lookup tables is that the performance on a large database can be slowed significantly by using character based keys to link lookup tables with the primary table as it is presented in this document. A work around is to use numeric value for the codes.

## *Performance*

As already outlined in the above section, there are also performance issues that may have to be considered.

[5] Domain is a term in the definition of the relational database model that defines a set of allowed values for a given set of fields (attributes), the mixing of different domains is not allowed in order to preserve the integrity of a relational and normalised model.

[6] Easily recoded permanently if the relation is specified as cascade on update or recoded dynamic by selecting a different column from the lookup table when querying the data through SQL

When using the suggested data types presented in this document for a database implementation, it may be worth looking at the actual data at hand when defining the final data types. The aim of this document is to present a format that will work between cohorts with rather different setups.

If it is at all possible in many cases there may be a large performance gain by using numeric instead of character fields. Character fields have been suggested here for, amongst others, the PATIENT field. If the PATIENT id is purely numeric it's worth using a numeric data type since it always faster for querying than a character field.

Whenever the field has to be character, make sure that only the needed amount of space is assigned for the field length; there is no need to assign 50 characters of memory if the field in fact only stores a 3-letter code.

# Appendix III – Case definitions

## *tblDIS - Opportunistic infections case definitions*

| Code | Severe Opportunistic Infections | Definitions |
|---|---|---|
| DEM | AIDS dementia complex | Definitive Disabling cognitive and/or motor dysfunction, or milestone loss in a child, and no other causes by CSF exam and brain imaging or by autopsy |
| | | Presumptive Same as above but no CSF and brain imaging performed |
| BCNE | Bacterial pneumonia, recurrent (>2 episodes within 1 year) | Definitive New X-ray evidence not present earlier and culture of pathogen that typically causes pneumonia (other than *P .carinii* or M. tuberculosis |
| | | Presumptive Acute radiological findings (new X-ray evidence not present earlier) and acute clinical findings |
| CANO | Candidiasis, oesophogeal | Definitive/autopsy Gross inspection by endoscopy/autopsy or by microscopy (histology) |
| | | Presumptive Recent onset retrosternal pain on swallowing and confirmed oral or pharyngeal candidiasis |
| CRCO | Cryptococcosis, extrapulm. | Definitive/autopsy Microscopy, culture of, or antigen detection in affected tissue |
| CRSP | Cryptosporidiosis (duration > 1 month) | Definitive/autopsy Microscopy. Duration of diarrhoea for more then 1 month |
| CMVR | Cytomegalovirus (CMV) chorioretinitis | Presumptive Loss of vision and characteristic appearance on serial ophtalmoscopy, progressing over serial months |
| CMVO | CMV – other location | Definitive/autopsy Microscopy (histology or cytology) |
| HERP | Herpes simplex virus ulcers (duration > 1 month) or pneumonitis/esophagitis | Definitive/autopsy Microscopy, culture of, or antigen detection in affected tissue |
| HIST | Histoplasmosis, extrapulm. | Definitive/autopsy Microscopy, culture of, or antigen detection in affected tissue |
| WAST | HIV Wasting Syndrome | Definitive Weight loss (over 10% of baseline) with no other cause, and 30 days or more of either diarrhoea or weakness with fever |
| ISDI | Isosporiasis diarrhoea (duration > 1 month) | Definitive/autopsy Microscopy (histology or cytology). Duration of diarrhoea for more than 1 month |
| LEIS | Leishmaniasis, visceral | Definitive/autopsy Histology or culture of *Leishmania* amastigotes in bone marrow or detection of amastigotes in tissue/fluid from affected<br>organ in a patient with symptoms and signs consistent with disseminated Leishmaniasis |
| MCDI | Microsporidosis diarrhoes (dur. > 1 month) | Definitive/autopsy Stool microscopy or rectal biopsy in patient with persistent diarrhoea |
| MC | Mycobact. avium complex (MAC) or Kanasii, extrapulm. | Definitive Culture |
| MCP | Mycobact. tuberculosis pulm. | Definitive Culture |
| MCX | Mycobact. tuberculosis extrapulm | Definitive Culture |
| MCPO | Mycobact. pulm. , other | Definitive Culture (indicate type) |
| | | Presumptive Acid fast bacteria (species not identified by culture) in sputum |

| Code | Severe Opportunistic Infections | Definitions |
|---|---|---|
| MCXO | Mycobact. extrapulm. , other | Definitive Culture (indicate type) |
| | | Presumptive Acid fast bacteria (species not identified by culture) on microscopy of normally sterile body fluid/tissue |
| PCP | Pneumocystis carinii pneumonia (PCP) | Definitive Microscopy (histology or cytology) |
| | | Presumptive Recent onset of dyspnoea on exertion or dry cough, and diffuse bilateral infiltrates on chest X-ray and pO2 <70 mmHg and<br>no evidence of bacterial pneumonia |
| LEU | Progressive multifocal leucoencephalopathy | Definitive/autopsy Microscopy (histology or cytology) |
| | | Presumptive Progressive deterioration in neurological function and CT/MR scan evidence |
| SAM | Salmonella bacteriaemia (non-tyhpoid) (recurrent) | Definitive Culture |
| TOX | Toxoplasmosis, brain | Definitive Microscopy (histology/cytology) |
| | | Presumptive Recent onset focal neurological abnormalities or reduced level of consciousness, and mass effect lesion on scan, and specific<br>therapy response |
| FBLS | Focal Brain lesion | To be updated asap |
| **Code** | **Malignancies** | |
| KS | Kaposi Sarcoma | Definitive/autopsy Histology |
| | | Presumptive Characteristic erythematous/violaceous plaque-like lesion on skin or mucous membranes |
| HG | Hodgkins Lymphoma | To be updated asap |
| NHG | Non-Hodgkin Lymphoma -not specified | To be updated asap |
| NHGB | Non-Hodgkin Lymphoma – Burkitt (Classical or Atypical) | Definitive: Histology |
| NHGI | Non-Hodgkin Lymphoma – Diffuse large B-cell lymphoma (Immunoblastic or Centroblastic) | Definitive: Histology |
| NHGU | Non-Hodgkin Lymphoma - Unknown/other histology | To be updated asap |
| NHGP | Non-Hodgkin Lymphoma - Primary Brain Lymphoma | Definitive: To be updated asap |
| | | Presumptive: Recent onset of focal neurological symptoms and signs or reduced level of consciousness, CT/MR<br>scan evidence of a lesion or lesions having mass effect, no response to toxo therapy, no evidence of lymphoma outside the brain |
| CRVC | Cervical Cancer | Definitive/autopsy Histology |

## *tblAE - Adverse Events case definitions*

| HICDEP Code (letter codes) | Possible ICD-10 codes | Adverse Event | Definitions |
|---|---|---|---|
| AMI | I21.9 | Acute myocardial infarction | Definitive myocardial infarction (MI)<br>i) definitive electrocardiogram (ECG),<br>ii) symptoms together with probable ECG and abnormal enzymes,<br>iii) typical symptoms, abnormal enzymes and ischaemic/non-codable/not available ECG, or<br>iv) fatal cases with naked-eye appearance of fresh MI and/or recent coronary occlusion found at necropsy.<br>Please see the MONICA manual for further criteria:<br>http://www.ktl.fi/publications/monica/manual/index.htm |
| STR | I64.9 | Stroke, not specified as haemorrhage or infarction | Rapidly developed clinical signs of focal or global disturbance of cerebral function lasting more than 24 hours (unless interrupted by surgery or death), with no apparent cause other than a cardiovascular origin. Secondary stroke caused by trauma should be excluded.<br>The differentiation between infarction and haemorrhage should be based on results of cerebral scanning or necropsy. In case of uncertainty (results not interpretable, or test not performed), please indicate so on the event form.<br>Please see the MONICA manual for further criteria:<br>http://www.ktl.fi/publications/monica/manual/index.htm<br>And the DAD MOOP (Manual of Operations):<br>http://www.cphiv.dk/main.asp?submenu_id=26&topimg=12 |
| DIA | E14 (also E10 – insulin dependent and E11 non-insulin-dependent) | Unspecified diabetes mellitus | The diagnostic criteria is: fasting blood glucose > 7 mmol/l<br>Please see the ADA (the American Diabetes Association) criteria for classification. |
| ICP - BYP | na | Coronary artery by-pass grafting | Procedure |
| ICP - END | na | Carotic endarterectomy | Procedure |
| ICP - ANG | na | Coronary angioplasty/stenting | Procedure |

| | | | |
|---|---|---|---|
| LAC | | Lactate acidosis | Elevated S-lactate > 2.5 mM (>22.3 mg/dL) AND plasma pH < 7.35 (alternatively: Bicarbonate/ $HCO_3^-$ <= 20 mM (<= 20 meq/L)) AND otherwise unexplained recent onset of at least one of the following: Abdominal distension, anorexia, abdominal pain, nausea, vomiting, diarrhea, increased liver function enzymes, jaundice, dyspnea, fever, neuropathy, generalized weakness, ascending neuromuscular weakness, myalgias, paresthesias, weight loss or hepatomegaly. |
| PAN | | Pancreatitis | Typical clinical history (i.e. severe abdominal pain), plus one or more of the following: elevated serum amylase > 1.5x ULN, elevated serum lipase, radiological findings. |
| ESRD | N18.0 (N18.8/9, N25.9, N26, N0.5, N04, N08) | End stage renal disease | A. Hemodialysis or peritoneal dialysis expected to last at least three months, documented in a clinical note<br>B. A kidney transplant, documented in a clinical note<br>Confirmed: A or B<br>Probable: Not applicable |
| AVN | | Avascular necrosis in the femural head | Diagnosed by the combination of clinical symptoms (pain, walking difficulties) and imaging findings (MRI, bone scintigraphy) |
| FRA | Several depending on location | Bone fracture | Diagnosed by X-ray |
| HEP | | Severe hepatic encephalopathy (stage III or IV) | Stage III: marked confusion, incoherent speech, asterixis, sleeping but arousable - Stage IV: coma |
| CLD | | Chronic liver disease –severe clinical manifestations | A. 1. Clinical symptoms of end-stage liver failure in patients with chronic liver disease, based on the diagnosis documented in a clinical note of either<br>(i) bleeding from gastric or esophageal varices<br>(ii) hepatic encephalopathy stage III or IV<br>(iii) hepatorenal syndrome<br>A. 2 liver transplantation documented in a clinical note<br>B. Pathology report or fibro-scan report documenting severe liver fibrosis or cirrhosis (Metavir F3 or F4 or fibroscan liver stiffness >= 8 kPa)<br>Confirmed: A1 and B; or A2<br>Probable: A1 |

| | | | |
|---|---|---|---|
| NADM | | Non AIDS defining cancers | A.Diagnosis of cancer (other than: AIDS defining (non-Hodgkin's lymphoma, Kaposi's sarcoma ), or invasive cervical cancer); and basal and squamous cell skin cancers) in a pathology report that established the diagnosis<br>B. Diagnosis of cancer (other than: AIDS defining (non-Hodgkin's lymphoma, Kaposi's sarcoma, or invasive cervical cancer); and basal and squamous cell skin cancers) in a hospital discharge summary or consultation note from the hospitalization or clinic visit during which the diagnosis was established<br>C. In the absence of A or B: Strong suspicion of cancer supported by (i) evidence from radiological or other imaging technique, (ii) or biochemical assay<br>D. In the absence of A, B or C: Strong suspicion of cancer by visual inspection (e.g. skin metastasis, suspected malignant melanoma, tissue growth resembling cancer visualized during endoscopy/anoscopy) not explained by other known conditions.<br>Confirmed: A or B<br>Probable: C<br>Possible: D<br>* The date of diagnosis is the month, day and year the tumor was first diagnosed by a recognized medical practitioner, whether clinically or microscopically confirmed. |

# Appendix IV – Change log

## *Version 1.30*

Updated version with these corrections/additions

tblART: Updated list of drugs

tblAE: Added EVEN_ID as unique identifier and link to detailed tables for each event (see document on D:A:D event tables) – this replaces the optional AE_NO field, AE_SPEC to further specify an event by coding. A series of basic verification fields have been added to allow for tracking of event status for source documentation availability, verification of documentation (through monitoring) and final approval of the event.
AE_R_Y - Relation to treatment: added more detailed codes.
Detailed table definitions for the D:A:D events are available at: http://www.cphiv.dk/HICDEP/Documents/tabid/159/Default.aspx

tblLAB: added several codes for various biomarker tests.

tblLAB_CD4: added CD4_U under optional fields to discriminate between CD4 % and CD4 cell count, so that the tblLAB_CD4 table can hold both types of measurements.

tblLAB_VIRO: added several codes for various virology and serology tests.

Appendix III – Case definitions updated with end stage renal disease, chronic liver disease and non-AIDS defining malignancies

## *Version 1.25*

Updated version with these corrections/additions

tblART: Updated list of drugs

tblMED: Updated list of drugs

tblDIS: Changed wording for CANO to 'Candidiasis, oesophageal, bronchi, trachea, or lungs'
Added COCC - Coccidioidomycosis, disseminated or extrapulmonary

tblLAB: Added LAB_ST as additional field to code for type of specimen used for the measurement

tblLAB_CD4: Added CD4_U as additional field so the table can hold both percentage and absolute CD4 measurements

tblLAB_RNA: Added RNA_UL (upper limit of detection) to the list of additional fields for tblLAB_RNA.
Added more viral assays to the list of RNA_T codes

tblLAB_VIRO: Added unit field to tblLAB_VIRO into the general format and VS_LL (lower limit of detection), VS_UL (upper limit of detection) and VS_T (type of test) and list of tests to the list of additional fields.

tblLTFU: Added DEATH_RC# to code for causal relation of the DEATH_R# code to the death in order to comply with CoDe and still maintain a format to be used for cohorts not using CoDe. ICD10_# fields have been moved to the list of additional fields.

## *Version 1.21*

Updated version with these corrections:

Added reasons for stopping treatment to table tblART_CODE_RS:

| Code | Coding for Reason of Stopping Treatment |
|---|---|
| 1.1 | Virological failure |
| 1.2 | Partial virological failure |
| 1.3 | Immunological failure – CD4 drop |
| 1.4 | Clinical progression |
| 90 | Side effects – any of the above but unspecified |
| 90.1 | Comorbidity |
| 92.1 | Simplified treatment available |

| | |
|---|---|
| 92.2 | Treatment to complex |
| 92.3 | Drug interaction |
| 93.1 | Structured Treatment Interruption (STI) – at high CD4 |
| 94.1 | Non-compliance |
| 96 | Pregnancy |
| 97 | Study treatment |

## *Version 1.2*

Updated version with these corrections:

Changed Appendix III – Change log to: Appendix IV – Change log and added Appendix III – Case definitions

## *Version 1.1*

Updated version with these corrections:

| Table | Field | Code | Changed to |
|---|---|---|---|
| tblBAS / tblLTFU | | | The table was split into tblBAS and tblLTFU. tblLTFU holds data on death and drop-out |
| tblBAS | LOS_Y | - | LOSS_Y |
| tblBAS | GAI_Y | - | GAIN_Y |
| tblLAB_BLP | - | - | Renamed the table to tblLAB_BP |
| tblLAB_BLP | BLP_D | - | BP_D |
| tblLAB_BLP | BLP_SYS | - | BP_SYS |
| tblLAB_BLP | BLP_DIA | - | BP_DIA |
| tblLAB_BLP | BLP_U | - | BP_U |

## Version 1.00

Updated version with these corrections:

| Table | Field | Code | Changed to |
|---|---|---|---|
| tblBAS | BIRTHDAY | - | BIRTH_D |
| tblBAS | FIRSTVIS | - | FRSVIS_D |
| tblBAS | REC_ART | - | RECART_Y |
| tblLAB | LAB_U | Numeric codes | This option has been dropped – please use the "unit codes/strings" as that is a safer way to code/represent the units – prefixing all "unit codes/strings" with a numeric value should however make analysis easier. |
| tblLAB_VIRO | - | - | New table added to capture mainly hepatitis measurements/tests. |
| tblLAB_RES | SEQ_DT | - | This was added to capture the time of sequencing in order to facilitate quality assurance of the data for contamination that might have happened during the sequencing. |
| tblLAB_RES<br>tblLAB_RES_LVL_1 - Nucleotide sequences | SEQ_ST | - | SEQ_STAR |
| tblLAB_RES<br>tblLAB_RES_LVL_1 - Nucleotide sequences | SEQ_STOP | - | Added to the table to specify at which position in the sequence the sequencing was terminated |
| tblLAB_RES<br>tblLAB_RES_LVL_2 Level 2 - Mutations | - | - | The table has been optimised for ease of analysis so that the mutation codes have been split into their components of amino acid position, sub position for insertions and 4 our more fields for mixtures of amino acids found in the sample |

## Version 0.50 and 0.90

First public versions that incorporated comments and corrections received from attendees at the 7th International Workshop on HIV Observational Databases, March 29th-30th 2003, Fiuggi, Italy and Stephen Hart.

## *Version 0.38*

Version presented at 7th International Workshop on HIV Observational Databases, March 29th-30th 2003, Fiuggi, Italy