

Wikiprint Book

Title: HICDEP

Subject: Format Specification

Version: 1.50

Date: 2011-09-21

Table of Contents

About HICDEP	5
HICDEP 1.50	6
General data format	6
Overview of data tables	6
Diagram	6
Main tables	6
Resistance tables	7
Mother-Child tables	7
Structure of data	8
From flat files towards a normalized structure	8
Technical considerations	8
Coding Conventions	9
Date codes	9
Day unknown	9
Month and day unknown	9
Year unknown	9
Specification of precision	9
ICD-10 codes	9
ATC codes	9
Other codes	10
Quality Assurance	10
tbIBAS - Basic clinical, background and demographic information	11
Core fields	11
Additional fields	11
QA Checks	11
MODE field	13
Coding Table	13
ORIGIN field	14
Coding Table	14
ETHNIC field	15
Coding Table	15
tbLTFU - Death and drop-out	16
Core fields	16
Additional fields	16
QA Checks	17
DROP_RS field	18
Coding Table	18
DEATH_R1 field	19
Coding Table	19
tbOVERLAP - Cross-cohort identification	20
Core fields	20
QA Checks	20
tbIVIS - Basic follow-up/visit related data	21
Core fields	21
Additional fields	21
QA Checks	21
tbIART - Antiretroviral treatment	23
Core fields	23
Additional fields	23
QA Checks	23
ART_ID field	25
Coding Table	25
ART_RS field	27
Coding Table	27
tbIMED - Other medication	28
Core fields	28
Additional fields	28
QA Checks	28
MED_ID field	29
Coding Table	29
tbIAE - Adverse Events	31
Core fields	31
Additional fields	31
QA Checks	31
AE_ID field	33
Coding Table	33
Coding Table: Pregnancy-related adverse events	33
Coding Table: Early childhood-related adverse events	33
Case Definitions	33
Case Definitions for AE_ID	34
AE_SPEC field	36
Coding Table	36

tbIDIS - Opportunistic infections	37
Core fields	37
Additional fields	37
QA Checks	37
DIS_ID field	38
Coding Table	38
Coding Table: Mother-to-child and paediatric specific	38
Case definitions	39
DIS_WD field	42
Coding Table	42
tbILAB_CD4 - Laboratory values	43
Core fields	43
Additional fields	43
QA Checks	43
tbILAB_RNA - Laboratory values	44
Core fields	44
Additional fields	44
QA Checks	44
RNA_T field	45
Coding Table	45
tbILAB_VIRO - Laboratory values - viro-/serology	46
Core fields	46
Additional fields	46
QA Checks	46
VS_ID field	48
Coding Table	48
VS_U field	49
Coding Table	49
VS_T field	50
Coding Table	50
tbILAB_BP - Laboratory values - Blood pressure	51
Core fields	51
QA Checks	51
BP_U field	52
Coding Table	52
tbILAB - Laboratory values	53
Core fields	53
Additional fields	53
QA Checks	53
LAB_ID field	55
Coding Table	55
LAB_U field	56
Coding Table	56
tbISAMPLES - Blood Samples	57
Core fields	57
Additional fields	57
QA Checks	57
tbILAB_RES - Resistance testing	59
Core fields	59
Additional fields	59
QA Checks	59
tbILAB_RES_LVL_1 - Nucleotide sequences (PRO, RT, GP41, GP120)	61
Core fields	61
Additional fields	61
QA Checks	61
tbILAB_RES_LVL_2 - Mutations	62
Core fields	62
QA Checks	62
tbILAB_RES_LVL_3 - Resistance test result	63
Core fields	63
Additional fields	63
QA Checks	63
Notes on the Pregnancy-Related Tables	64
Example	64
Important relations	64
tbIPREG - Pregnancy	65
Core fields	65
QA Checks	66
tbIPREG_OBS - Obstetrical problems during pregnancy	67
Core fields	67
QA Checks	67
tbIPREG_OUT - Pregnancy outcome	68
Core fields	68
QA Checks	68

tbIDELIVERY_MUM - Delivery information related to the mother	70
Core fields	70
QA Checks	70
tbIDELIVERY_CHILD - Delivery information related to the child(ren)	72
Core fields	72
QA Checks	72
tbINEWBORN - Newborn	74
Core fields	74
QA Checks	74
tbINEWBORN_ABNORM - Abnormalities	76
Core fields	76
QA Checks	76
Change log	77
Version 1.50	77
Version 1.30	77
Version 1.25	77
Version 1.21	77
Version 1.2	77
Version 1.1	77
Version 1.00	78
Version 0.50 and 0.90	78
Version 0.38	78
Considerations for data management	79
Uniqueness of PATIENT ID	79
Duplicate records	79
Considerations for using the format to create a database	80
Administrative fields	80
Further normalisation	80
Lookup tables	80
Performance	81

About HICDEP

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](#)). 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 [HICDEP website](#). HICDEP is a format under constant improvement (currently under the EuroCoord Project, Grant Agreement No. 260694) and additions are made almost every year. Please refer to the [ChangeLog](#) for the most current updates and always use the tables available on the HICDEP website for most current coding lists for ART and MED drugs.

Members of EuroCoord WP4, September 2011

HICDEP 1.50

This article describes the HICDEP version 1.50 which was released on the 13th of September 2011. It is the latest released version. For a more detailed version history, please refer to the [ChangeLog](#).

General data format

The table pages referenced [in the overview](#) describe the specific tables' structure in detail and present a list of suggested codes, both standard and human readable.

All codes apart from trivial no, yes or unknown codes are presented as lookup tables, the usage of these are described in the the article [Considerations for using the format to create 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.

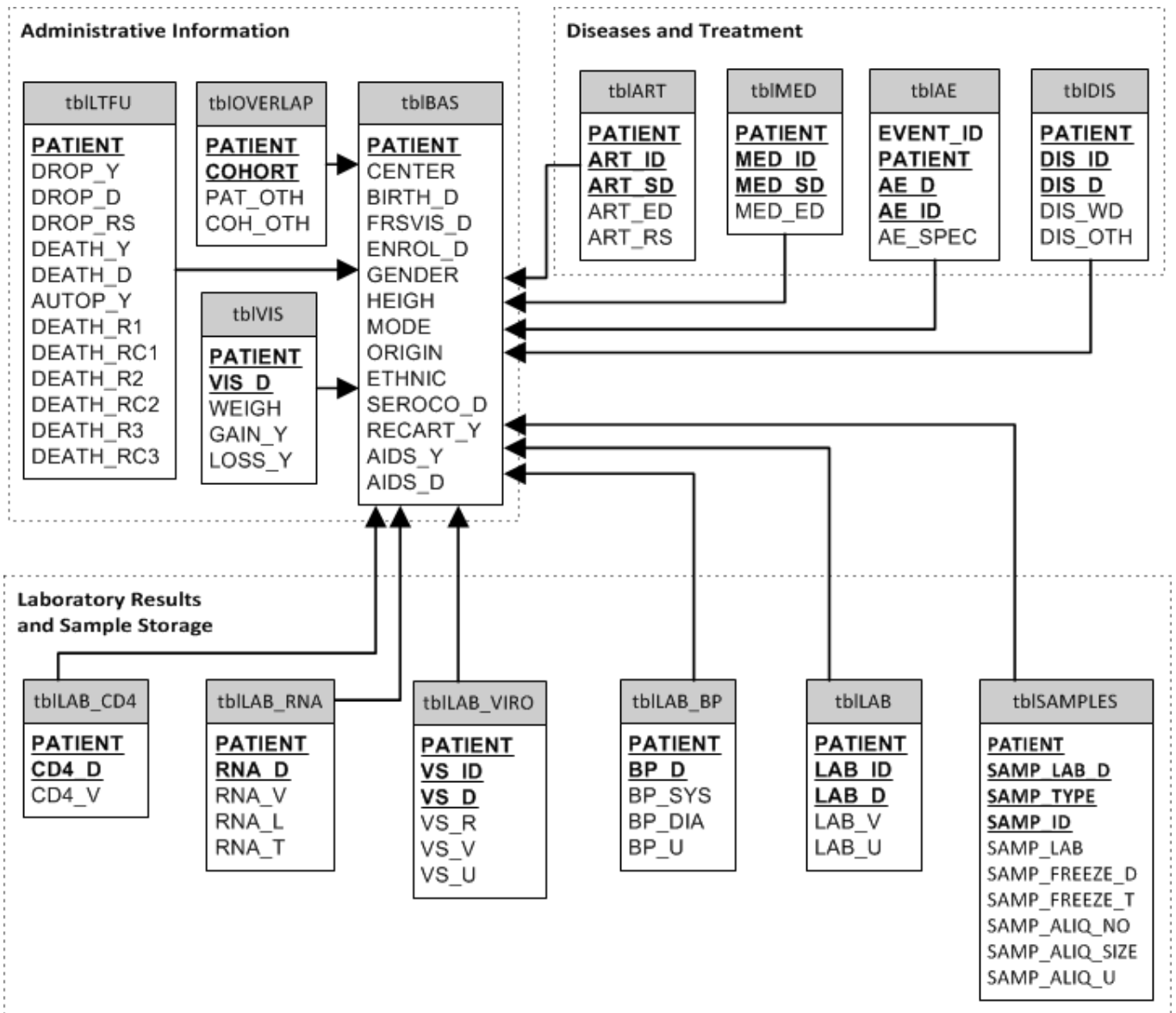
Issues regarding duplicates are discussed in [Considerations For Data Management](#).

Overview of data tables

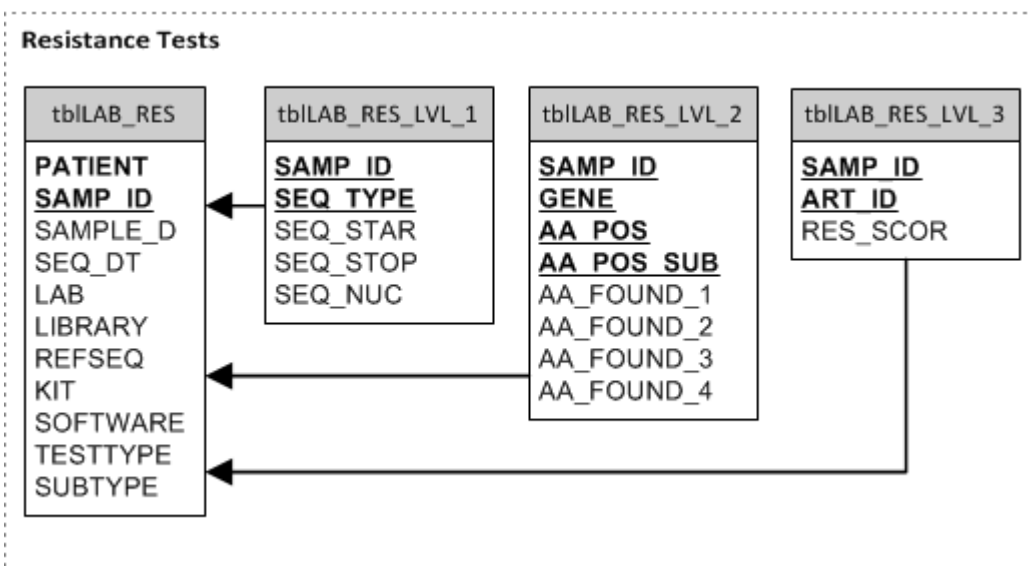
Table	Content
tblAE	holds type and date of adverse events including serious non-AIDS conditions
tblART	holds type of antiretroviral drug , start and stop dates and reason for stopping
tblBAS	holds basic information such as demographics, basic clinical information, date of AIDS diagnosis, death and drop-out information
tblDELIVERY_CHILD	holds delivery information related to the child
tblDELIVERY_MUM	holds delivery information related to the mother
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 .
tblNEWBORN	holds information related to newborns
tblNEWBORN_ABNORM	holds information related to abnormalities of newborns
tblOVERLAP	holds information on the patient's participation in other cohorts
tblPREG	holds general pregnancy -related information
tblPREG_OBS	holds information on obstetrical problems during pregnancy
tblPREG_OUT	describes the pregnancy outcome
tblSAMPLES	holds information on the storage of blood, urine and other biological samples
tblVIS	holds visit related information , weight, wasting.

Diagram

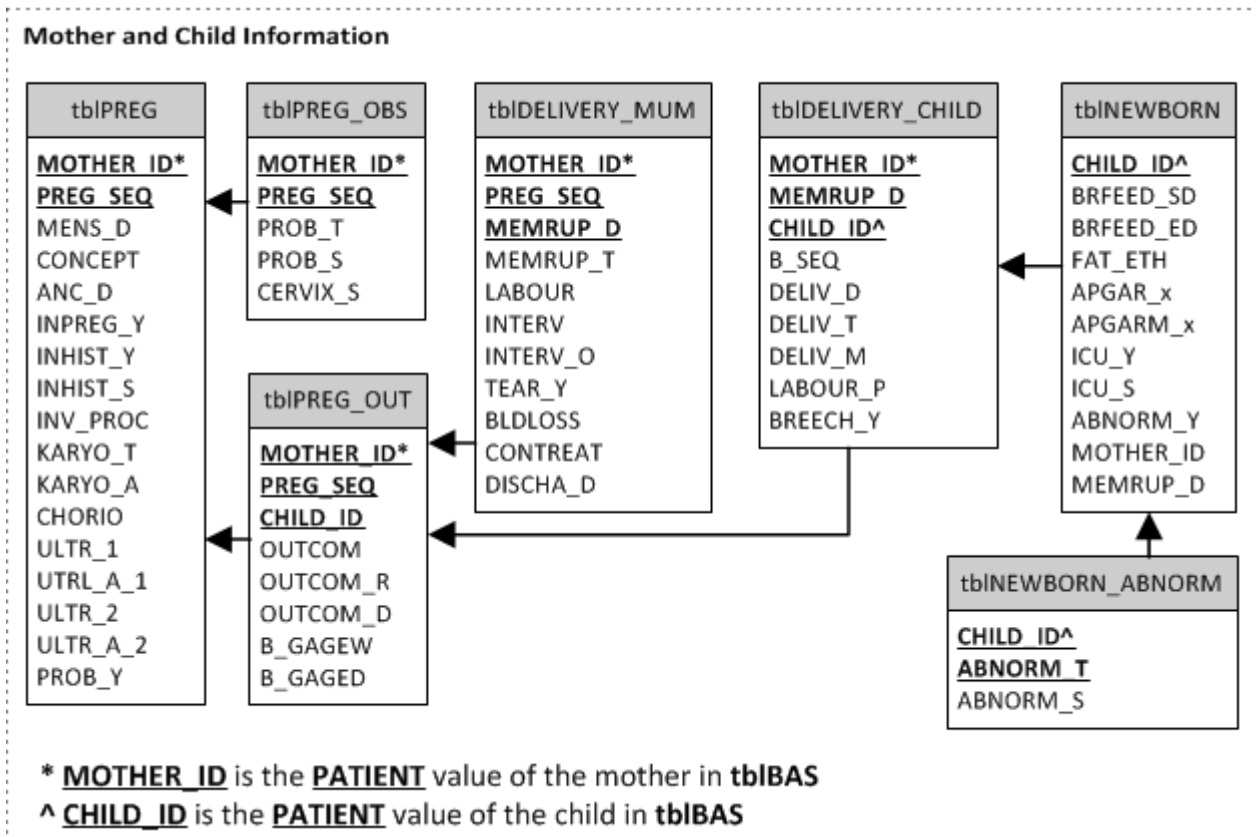
Main tables



Resistance tables



Mother-Child tables



Structure of data

From flat files towards a normalized 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 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:

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

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:

- Patient identifier
- Measurement type identifier

- Measured value
- Unit of value
- Date of determination

A record for usage of an antiretroviral drug would include:

- 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.

Coding Conventions

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 15th 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
U	Unknown

[ICD-10 codes](#)

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.

[ATC codes](#)

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 (1st level, anatomical main group)
J05
  ANTIVIRALS FOR SYSTEMIC USE (2nd level, therapeutic subgroup)
```

J05A
 DIRECT ACTING ANTIVIRALS (3rd level, pharmacological subgroup)
 J05AE
 Protease inhibitors (4th level, chemical subgroup)
 J05AE02
 Indinavir (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 Combinations1
- 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 the [ATC Website](#) to learn about the structure of ATC. A fully updated database of ATC codes and DDD (Defined Daily Dosage) is available for querying.

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:

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:

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.

Quality Assurance

In order to verify the consistency and correctness of the data, QA checks are made before the data is used. The QA checks applying to a given table are listed at the bottom of its article. Additionally, a list of all QA checks, including checks which do not directly apply to the HICDEP tables themselves, is available [here](#).

tbIBAS - Basic clinical, background and demographic information

holds basic information such as demographics, basic clinical information, date of AIDS diagnosis, death and drop-out information

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
PATIENT	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
CENTER	character	Code for Clinic/Centre/Hospital where patient is seen.
BIRTH_D	yyyy-mm-dd	Birth date
FRSVIS_D	yyyy-mm-dd	First seen at clinic
ENROL_D	yyyy-mm-dd	Date of enrolment into the cohort
GENDER	numeric: <ul style="list-style-type: none"> 1 = Male 2 = Female 9 = Unknown 	Gender/sex
HEIGH	numeric (metric): 999 = Unknown	Height of patient at visit/most current
MODE	numeric. see coding table for valid codings.	Mode of infection
ORIGIN	character (1-3 letter/numeric codes). see coding table for valid codings.	Nationality or region of origin of patient
ETHNIC	numeric. see coding table for valid codings.	Ethnicity of patient
SEROCO_D	yyyy-mm-dd	Date of seroconversion
RECART_Y	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 9 = Unknown 	Has the patient received antiretroviral treatment?
AIDS_Y	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 9 = Unknown 	Has patient been given an AIDS diagnosis?
AIDS_D	yyyy-mm-dd	IF YES, date of AIDS diagnosis

Additional fields

For mode of infection and origin 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 and ORIGIN should be able to cover all possible values.

Field name	Format	Description
MODE_OTH	character	Mode of infection OTHER
ORI_OTH	character	Origin of patient OTHER

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES

AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tblBAS	WithinTable	BW001	AIDS date < SEROCO_D		YES
tblBAS	WithinTable	BW002	Duplicate patients		YES
tblBAS	WithinTable	BW003	First 3 chars of PATIENT don't form valid cohort code	CascadeOnly	YES
tblBAS	WithinTable	BW004	Missing patient		YES
tblBAS	WithinTable	BW005	Missing CENTER		YES
tblBAS	WithinTable	BW006	Missing BIRTH_D		YES
tblBAS	WithinTable	BW007	Missing FRVIS_D		YES
tblBAS	WithinTable	BW008	Missing ENROL_D		YES
tblBAS	WithinTable	BW009	Missing GENDER		YES
tblBAS	WithinTable	BW010	Missing HEIGH		YES
tblBAS	WithinTable	BW011	Missing MODE		YES
tblBAS	WithinTable	BW012	Missing MODE_OTH if MODE=90		YES
tblBAS	WithinTable	BW013	Missing Origin		YES
tblBAS	WithinTable	BW014	Missing ETHNIC		YES
tblBAS	WithinTable	BW015	Missing SEROCO_D		YES
tblBAS	WithinTable	BW016	Missing REART_Y		YES
tblBAS	WithinTable	BW017	Missing AIDS_Y		YES
tblBAS	WithinTable	BW018	Missing AIDS_D if AIDS_Y=1		YES
tblBAS	WithinTable	BW019	BIRTH_D out of range (15-85 yrs)	CascadeOnly	YES
tblBAS	WithinTable	BW020	BIRTH_D out of range (<18)	PENTA	YES
tblBAS	CrossTable	BC001	AIDS_Y=1 but no records in tblART		YES
tblBAS	CrossTable	BC002	AIDS_Y=0, but AIDS-defining records in tblDIS		YES
tblBAS	CrossTable	BC003	AIDS_Y=1 but no AIDS-defining records in tblDIS table		YES

MODE field

containing table

[tbIBAS](#)

explanation of variable

Mode of infection

format of data

numeric. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

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

ORIGIN field

containing table

[tblBAS](#)

explanation of variable

Nationality or region of origin of patient

format of data

character (1-3 letter/numeric codes). see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

Code	Region of 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-letter](#) or [3-letter](#) codes.

ETHNIC field

containing table

[tbIBAS](#)

explanation of variable

Ethnicity of patient

format of data

numeric. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

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

tbILTFU - Death and drop-out

holds data in death and drop-out

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
<u>PATIENT</u>	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
<u>DROP_Y</u>	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 	Has the patient DROPPED OUT?
<u>DROP_D</u>	yyyy-mm-dd	IF YES, Date of Last Visit
<u>DROP_RS</u>	numeric. see coding table for valid codings.	IF YES, Reason for DROP
<u>DEATH_Y</u>	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 	Has the patient died?
<u>DEATH_D</u>	yyyy-mm-dd	Date of Death
<u>AUTOP_Y</u>	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 9 = Unknown 	Was an autopsy Performed?
<u>DEATH_R1</u>	character. see coding table for valid codings.	Cause of death
<u>DEATH_RC1</u>	character with codes: <ul style="list-style-type: none"> 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
<u>DEATH_R2</u>	character. see coding table for valid codings.	Cause of death
<u>DEATH_RC2</u>	character with codes: <ul style="list-style-type: none"> 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
<u>DEATH_R3</u>	character. see coding table for valid codings.	Cause of death
<u>DEATH_RC3</u>	character with codes: <ul style="list-style-type: none"> I = Immediate cause U = Underlying cause/condition C = Contributing cause N = Not available 	Coding of causal relation of the code given in DEATH_R3 to the death

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 \(CoDe\)](#). You are 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.

Additional fields

Field name	Format	Description
<u>ICD10_1</u>	character	Cause of death as ICD-10 if available
<u>ICD10_2</u>	character	Cause of death as ICD-10 if available
<u>ICD10_3¹</u>	character	Cause of death as ICD-10 if available
<u>DEATH_OT</u>	character	Reason for death ? other - description
<u>L_ALIVE</u>	yyyy-mm-dd	Last date known to be alive

¹: List of *ICD10_#* in place 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.

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbLTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbLTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbLTFU	WithinTable	LFW001	DROP_Y and DEATH_Y both non-null		YES
tbLTFU	WithinTable	LFW002	DEATH_Y and DROP_RS both non-null		YES
tbLTFU	WithinTable	LFW003	Any of DEATH_Rx or DEATH_RCx non-null but DEATH_Y=0		YES
tbLTFU	WithinTable	LFW004	R2/RC2 non-null but R1/RC1 null		YES
tbLTFU	WithinTable	LFW005	R3/RC3 non-null but R2/RC2 null		YES
tbLTFU	WithinTable	LFW006	Duplicate patients		YES
tbLTFU	WithinTable	LFW007	Missing DROP_Y		YES
tbLTFU	WithinTable	LFW008	Missing DROP_D if DROP_Y=1		YES
tbLTFU	WithinTable	LFW009	Missing DROP_RS if DROP_Y=1		YES
tbLTFU	WithinTable	LFW010	Missing DEATH_Y		YES
tbLTFU	WithinTable	LFW011	Missing DEATH_D if DEATH_Y=1		YES
tbLTFU	WithinTable	LFW012	DEATH_D non null but DEATH_Y=0		YES
tbLTFU	CrossTable	LFC001	Patient not found in tbIBAS		YES
tbLTFU	CrossTable	LFC002	patient in tbIBAS hasn't got a record in tbLTFU		YES
tbLTFU	CrossTable	LFC003	tbIBAS has AIDS=0 but DEATHRx =8.1		YES

DROP_RS field

containing table

[tblLTFU](#)

explanation of variable

IF [DROP_Y](#) = 1, Reason for DROP

format of data

numeric. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

Code	Reason for Drop Out
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	Patient moved to another country
4	Patient moved and is followed by another centre
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

Note: If consent is withdrawn, all patient data except for the patient id and reason for drop out *may* have to be deleted.

DEATH_R1 field

containing table

[tblLTFU](#)

explanation of variable

Cause of death

format of data

character. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

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
50	Sudden infant death
50.1	Child abuse
90	Other
91	Suicide
92	Drug Overdose
93	Accident
99	Unknown, Fatal case with no information

tbOVERLAP - Cross-cohort identification

holds information on the patient's participation in other cohorts

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
<u>PATIENT</u>	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
<u>COHORT</u>	character	Code/name of the cohort
<u>PAT_OTH</u>	character	Unique patient identifier in other cohorts
<u>COH_OTH</u>	character	Name of the cohort

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).

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tblLTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tblLTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tblBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbOVERLAP	WithinTable	OW001	Invalid other cohort		YES
tbOVERLAP	CrossTable	OC001	PATIENT not found in tblBAS for that cohort		YES
tbOVERLAP	CrossTable	OC002	PAT_OTH not found in tblBAS for that overlapping cohort		YES

tbIVIS - Basic follow-up/visit related data

holds visit related information, weight, wasting.

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
PATIENT	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
VIS_D	yyyy-mm-dd	Date of patient visit
WEIGH	numeric (metric: kg): 999 = Unknown	Weight of patient at visit
GAIN_Y	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 9 = Unknown 	Is the patient gaining fat in the abdomen, neck, breast or other defined locations?
LOSS_Y	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 9 = Unknown 	Is the patient experiencing loss of fat from extremities, buttocks or face?

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.

Additional fields

The following fields meant to be used **for children and infants**.

The following fields are meant to be used **for infants**:

Field name	Format	Description
HEIGH_P	numeric	Height/length of patient at visit in percentiles
WEIGH_P	numeric	Weight of patient at visit in percentiles
HEADC	numeric	Head circumference measured in millimeters (mm)
HEADC_P	numeric	Head circumference in percentiles
BREASTF_Y	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 9 = Unknown 	Currently Breastfeeding?

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbIVIS	WithinTable	VW001	Duplicate records, same VIS_D		YES

tblVIS	WithinTable	VW002	Height decreasing over time		YES
tblVIS	WithinTable	VW003	Height out of acceptable range		YES
tblVIS	WithinTable	VW004	Weight out of acceptable range		YES
tblVIS	CrossTable	VC001	patient has no record in BAS table		YES
tblVIS	CrossTable	VC002	No weights within 3 mths of starting FPV/DRV	EPPICC	YES

tbIART - Antiretroviral treatment

holds type of antiretroviral drug, start and stop dates and reason for stopping

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
PATIENT	character (or numeric if possible)	identifies patient
ART_ID	character. see coding table for valid codings.	represents the antiretroviral treatment
ART_SD	yyyy-mm-dd	date of initiation of treatment
ART_ED	yyyy-mm-dd	date of stopping treatment
ART_RS	character. see coding table for valid codings.	reason for stopping treatment

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 date and thus these fields are optional.

Field name	Format	Description
ART_DO	numeric	Dosage (mg or mL)
ART_FR	numeric: <ul style="list-style-type: none"> 1 = 1 daily dose/qd 2 = 2 daily doses/bid 3 = 3 daily doses/tid 4... = code gives number of daily doses 	Frequency

It may also be necessary to record the start and end time:

Field name	Format	Description
ART_ST?	hh:mm	Start hour and minute of the day
ART_ET?	hh:mm	Stop hour and minute of the day

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbIART	WithinTable	AW001	ART_RS not 0, but end date NULL		YES
tbIART	WithinTable	AW002	ART_RS=0 but end-date non NULL		YES
tbIART	WithinTable	AW003	ART_RS=98 yet ART_OTH is null		YES
tbIART	WithinTable	AW004	Duplicate records for same cohort, patient, art_id and art_sd		YES
tbIART	WithinTable	AW005	Dose out of range for those dose units and frequency	PaediatricOnly	YES

tblART	WithinTable	AW006	Missing art_fr	PaediatricOnly	YES
tblART	WithinTable	AW007	Missing art_do	PaediatricOnly	YES
tblART	WithinTable	AW008	Missing patient		YES
tblART	WithinTable	AW009	Missing art_id		YES
tblART	WithinTable	AW010	Missing art_sd		YES
tblART	WithinTable	AW011	Overlapping periods of same drug		YES
tblART	WithinTable	AW012	Double reporting - records reported for both combination drugs and their components		YES
tblART	WithinTable	AW013	Periods of overlap of contra-indicated drugs		YES
tblART	WithinTable	AW014	Restart of same drug without a stop		YES
tblART	WithinTable	AW015	ART_SD greater than or equal to ART_ED		YES
tblART	CrossTable	AC001	Patient has no record in table BAS		YES
tblART	CrossTable	AC002	Records exist in tblART yet RECART_Y=0 in tblBAS		YES

ART_ID field

containing table

[tblART](#)

explanation of variable

Code representing the antiretroviral treatment

format of data

character. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

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 - Soft 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.

See the [ATC Index](#) for the individual codes.

[Download this table as CSV file](#)

Code (Extended ATC Codes)	Anti-Retroviral Drugs
J05A	ART unspecified
J05A-BEV	Beviramat
J05A-PBT	Participant in Blinded Trial
J05AE	PI unspecified
J05AE-MOZ	Mozenavir (DMP-450)
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)
J05AF	NRTI unspecified
J05AF-ALO	Alovudine
J05AF-AMD	Amdoxovir (DADP)
J05AF-FOZ	Fozivudine tidoxi
J05AF-LDN	Lodensine (trialdrug)
J05AF-RVT	Reverset
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 (ViiREAD)
J05AF08	Adefovir (PREVEON)
J05AF09	Emtricitabine
J05AF10	Entecavir
J05AF11	Telbivudine
J05AG	NNRTI unspecified
J05AG-CPV	Capravirine
J05AG-DPC083	DPC 083
J05AG-DPC961	DPC 961
J05AG-EMV	Emivirine (MKC442)
J05AG-ETV	Etravirine (TMC 125)
J05AG-LOV	Loviride
J05AG-RPV	Rilpivirine (TMC-278)
J05AG01	Nevirapine (VIRAMUN)
J05AG02	Delavirdine (U-90152) (RESCRIPTOR)
J05AG03	Efavirenz (DMP-266) (STOCRIN, SUSTIVA)
J05AR01	Combivir (Zidovudine/Lamivudine)
J05AR02	Kivexa (Lamivudine/Abacavir)
J05AR03	Truvada (Tenofovir/Emtricitabine)
J05AR04	Trizivir (Zidovudine/Lamivudine/Abacavir)
J05AR05	Douvir-N (Zidovudine/Lamivudine/Nevirapine)
J05AR06	Atripla (Emtricitabine/Tenofovir/Efavirenz)
J05AX-EVG	Elvitegravir (Gilead)
J05AX-VIC	Vicriviroc (Schering)
J05AX07	Enfuvirtide (Fuzeon, T-20)
J05AX08	Raltegravir (Merck)
J05AX09	Maraviroc (Pfizer)
L01XX05	Hydroxyurea/Hydroxycarbamid (Litalir)

ART_RS field

containing table

[tblART](#)

explanation of variable

Reason for stopping treatment

format of data

character. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

Code	Reason for 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	Hyperlactataemia/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 too 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.

tbIMED - Other medication

holds type, start and stop dates for other HIV related medicines.

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
<u>PATIENT</u>	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
<u>MED_ID</u>	character. see coding table for valid codings.	Code representing the treatment
<u>MED_SD</u>	yyyy-mm-dd	Date of Initiation of Treatment
<u>MED_ED</u>	yyyy-mm-dd	Date of stopping treatment

Additional fields

Please see [tblART - Antiretroviral treatment](#).

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tblLTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tblLTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tblBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbIMED	WithinTable	MW001	Duplicate records		YES
tbIMED	WithinTable	MW002	MED_ONG=0 and MED_ED null		YES
tbIMED	WithinTable	MW003	MED_ONG=1 and MED_ED non-null		YES
tbIMED	WithinTable	MW004	Missing MED_ID		YES
tbIMED	WithinTable	MW005	Missing MED_SD		YES
tbIMED	WithinTable	MW006	Mising MED_ED		YES
tbIMED	WithinTable	MW007	Overlapping periods of same drug		YES
tbIMED	WithinTable	MW008	MED_ED < MED_SD		YES
tbIMED	CrossTable	MC001	Patient has no records in tblBAS		YES

MED_ID field

containing table

[tblMED](#)

explanation of variable

Code representing the treatment

format of data

character. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

See also the [notes on the extended ATC-Codes](#) and the [ATC Index](#).

TODO

what about the last for entries in the following table?

↓ [Download this table as CSV file](#)

Codes (Extended ATC-Codes)	Other HIV-related drugs
A10A	Insulin or derivatives hereof
A10B	Oral antidiabetic agents
A14A	Anabolic steroids/appetite stimulants
B01AC	Anti-platelets
C-HYP	Other anti-hypertensive agents [C02, C03, C04, C07, C08]
C09	ACE inhibitors
C10	Lipid lowering agents
HICDEP_code	HICDEP_description
J01AA08	Minocycline (MINOCIN)
J01EA01	Trimethoprim (MONOTRIM, NOFIL)
J01EC02	Sulfadiazine
J01EE	Cotrimoxazole - Comb. of sulfonamides and trimethoprim (BACTRIM, EUSAPRIM, NOFIL)
J01EE01	Sulfamethoxazole and trimethoprim (Bactrim)
J01EE03	Sulfamethole and trimethoprim - Cosoltrime (MADERAN)
J01FA09	Clarithromycine (KLACID)
J01FA10	Azithomycine (ZITHROMAX)
J01FF01	Clindamycine (DALACIN)
J01GB06	Amikacine (AMIKINE)
J01MA02	Ciprofloxacin (CIPROXINE, CILOXAN)
J01MA12	Levofloxacin (TAVANIC)
J01MA14	Moxifloxacin
J01RA02	Cosoltrime (MADERAN)
J02AA01	Amphotericin B (FUNGIZON)
J02AB	Imidazoles (DAKTARIN, NIZORAL, PEVARYL ?)
J02AB02	Ketoconazole
J02AC01	Fluconazole (DIFLUCAN)
J02AC02	Itraconazole (SPORANOX)
J02AC03	Voriconazole
J02AC04	Posaconazole
J02AX01	Flucytosine
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 (ZIVORAX)
J05AB04	Ribavirin
J05AB06	Ganciclovir (CYMEVENE)
J05AB09	Famciclovir
J05AB11	Valaciclovir (VALTEX)
J05AB12	Cidofovir (VISTIDE)
J05AB15	Valganciclovir
J05AD01	Foscarnet (FOSCAVIR)
L01AA01	Cyclophosphamide (ENDOXAN)
L01AD02	CCNU (LOMUSTINE)
L01AX04	Dacabazine (DTIC - Dome)
L01BA01	Methotrexate
L01CA01	Vinblastin (VELBE)
L01CA02	Oncovin (VINCRISTINE)
L01CB01	Etoposide (VEPESIDE, EXITOP 100)
L01DB01	Doxorubicine, Adriamycine (DOXIL, CAELYX, ADRIBLASTIN)
L01DC01	Bleomycine
L01XB01	Procarbazine (NATULAN)
L03AA02	G-CSF/Filgrastim (NEUPOGEN)
L03AB	Interferons
L03AB-AL2	Peginterferon alfa-2a/alfa-2b (PEGINTRON, PEGASYS)
L03AB10	Peginterferon alfa-2b (PEGINTRON)
L03AB11	Peginterferon alfa-2a (PEGASYS)
L03AC-IL2	Interleukin 2 (PROLEUKIN)
P01AX06	Atovaquone (WELLVONE, MEPRONE)
P01BA03	Primaquine
P01BD01	Pyrimethamine (DARAPRIM)
P01BD51	Pyrimethamine/Sulfadoxine (FANSIDAR)
P01CX01	Pentamidine aerosol (PENTACARNET)
V03AF03	Folate of calcium (LEUCOVORINE)
	Antibiotics
	Corticosteroids
	Tocolysis
	Cerclage

tblAE - Adverse Events

holds type and date of adverse events including serious non-AIDS conditions

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
EVENT_ID	numeric	Unique Event Identifier (foreign key to the different event tables)
PATIENT	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
AE_D	yyyy-mm-dd	date of event
AE_ID	character. see coding table for valid codings.	identifies type of event
AE_SPEC	character. see coding table for valid codings.	further specification
SRCDOC_Y	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 	whether the source documentation is available
SRCDOC_D	yyyy-mm-dd	date for source documentation verification
VERIFY_Y	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 	Has the monitor verified the source documentation?
VERIFY_D	yyyy-mm-dd	date for monitor verification
APPROV_Y	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 	final verification/approval
APPROV_D	yyyy-mm-dd	final verification date
APPROV_S	character	signature for final verification

Additional fields

Field name	Format	Description
AE_Y	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 9 = Unknown 	has the patient had an event?
AE_NAME	character	full name of the event
AE_DESCRIP	character	full description of the event
AE_R_Y	numeric: <ul style="list-style-type: none"> 0 = not related 1 = definitive 2 = remote/unlikely 3 = possible 4 = probable 	relation to treatment

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tblLTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tblLTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tblBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES

AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES

AE_ID field

containing table

[tblAE](#)

explanation of variable

Code to identify event

format of data

character. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

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

Coding Table: Pregnancy-related adverse events

[Download this table as CSV file](#)

Code	Adverse Event
ANEM	Anemia (<10g/l)
FEV	Fever (> 38 0C and > 1 day)
URITINF	Urinary tract infection
HEMATOMA	Hematoma
ENDOM	Endometritis
WOUINF	Wound infection
PERIT	Peritonitis
PNEU	Pneumonia
SEPSIS	Sepsis (fever and pos blood culture)
THROMB	Thromboembolism
DIC	Disseminated intravascular coagulation (DIC)
SUBI	Subileus/ ileus
HEMOR	Hemorrhage
PSY	Psychosis
PREECL	Preeclampsia/eclampsia
ANEMBL	Severe anemia requiring blood transfusion

Coding Table: Early childhood-related adverse events

[Download this table as CSV file](#)

DEVD	Developmental delay
SEIZ	Seizures
ONEU	Other neurological symptoms (use with AE_DESCRIP)
OABN	Other abnormal findings (use with AE_DESCRIP)

Case Definitions

see [separate article](#).

Case Definitions for [AE_ID](#)

HICDEP Code	ICD-10 codes	Adverse Event	Definition
AMI	I21.9	Acute myocardial infarction	Definitive myocardial infarction (MI) i) definitive electrocardiogram (ECG), ii) symptoms together with probable ECG and abnormal enzymes, iii) typical symptoms, abnormal enzymes and ischaemic/non-codable/not available ECG, or iv) fatal cases with naked-eye appearance of fresh MI and/or recent coronary occlusion found at necropsy. Please see the MONICA manual for further criteria.
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. 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. Please see the MONICA manual and the DAD MOOP (Manual of Operations) for further criteria
DIA	E14 (also E10 ? insulin dependent and E11 non-insulin-dependent)	Unspecified diabetes mellitus	The diagnostic criteria is: fasting blood glucose > 7 mmol/l Please see the ADA (the American Diabetes Association) criteria for classification.
ICP - BYP	n/a	Coronary artery by-pass grafting	Procedure
ICP - END	n/a	Carotid endarterectomy	Procedure
ICP - ANG	n/a	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 B. A kidney transplant, documented in a clinical note Confirmed: A or B Probable: Not applicable
AVN		Avascular necrosis in the femoral 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

<p>CLD</p>		<p>Chronic liver disease ?severe clinical manifestations</p>	<p>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 (i) bleeding from gastric or esophageal varices (ii) hepatic encephalopathy stage III or IV (iii) hepatorenal syndrome A. 2 liver transplantation documented in a clinical note B. Pathology report or fibro-scan report documenting severe liver fibrosis or cirrhosis (Metavir F3 or F4 or fibroscan liver stiffness >= 8 kPa) Confirmed: A1 and B; or A2 Probable: A1</p>
<p>NADM</p>		<p>Non AIDS defining cancers</p>	<p>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 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 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 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. Confirmed: A or B Probable: C Possible: D * 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.</p>

AE_SPEC field

containing table

[tblAE](#)

explanation of variable

Code to further specify the event identified by [AE_ID](#).

format of data

character. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

Code (AE_ID)	Code (AE_SPEC)	Description
AMI	DAMI	Definitive Myocardial infarction
AMI	PAMI	Possible Myocardial infarction
ICP	ANG	Invasive Cardiovascular Procedures: Coronary angioplasty/stenting
ICP	BYP	Invasive Cardiovascular Procedures: Coronary artery by-pass grafting
ICP	END	Invasive Cardiovascular Procedures: Carotic endarterectomy
NADM	ALL	Leukemia: Acute lymphoid
NADM	AML	Leukemia: Acute myeloid
NADM	ANUS	Anus cancer
NADM	BLAD	Bladder cancer
NADM	BRCA	Breast cancer
NADM	CERV	Cervical dysplasia/carcinoma in situ
NADM	CLL	Leukemia: Chronic lymphoid
NADM	CML	Leukemia: Chronic myeloid
NADM	COLO	Colon cancer
NADM	COTC	Connective tissue cancer
NADM	HDL	Hodgkin lymphoma
NADM	KIDN	Kidney cancer
NADM	LEUK	Leukemia: unspecified
NADM	LIPC	Lip cancer
NADM	LIVR	Liver cancer
NADM	LUNG	Lung cancer
NADM	MALM	Malignant melanoma
NADM	MEAC	Metastasis: of adenocarcinoma
NADM	MEOC	Metastasis: of other cancertype
NADM	MESC	Metastasis: of squamuos cell carcinoma
NADM	META	Metastasis: unspecified
NADM	MULM	Multiple myeloma
NADM	PENC	Penile cancer
NADM	PROS	Prostate cancer
NADM	RECT	Rectum cancer
NADM	STOM	Stomach cancer
NADM	TESE	Testicular seminoma
NADM	UTER	Uterus cancer
STR	SHAE	Stroke: Haemorrhagia
STR	SINF	Stroke: Infarction
STR	SUNK	Stroke: Unknown

tbIDIS - Opportunistic infections

holds type and date of CDC-C diseases.

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
<u>PATIENT</u>	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
<u>DIS_ID</u>	character. see coding table for valid codings.	Code to identify event
<u>DIS_D</u>	yyyy-mm-dd	Date of event
<u>DIS_WD</u>	numeric. see coding table for valid codings.	Means of diagnosis
<u>DIS_OTH</u> ¹	character	Other location, only to be filled out if code alone is not sufficient

¹ DIS_OTH might be part of the record's unique identification

Additional fields

Please see [tblAE - Adverse Events](#) for specification on optional fields.

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tblLTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tblLTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tblBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbIDIS	WithinTable	DW001	Duplicate records for same DIS_ID and same DIS_D		YES
tbIDIS	WithinTable	DW002	Miscoded DIS_WD as codes on table definition		YES
tbIDIS	WithinTable	DW003	Miscoded DIS_ID - as in code list attached to table definition		YES
tbIDIS	WithinTable	DW004	DIS_D missing		YES
tbIDIS	WithinTable	DW005	DIS_ID missing		YES
tbIDIS	WithinTable	DW006	Same event recorded twice - 2 records, same DIS_ID, DIS_D within 6 months		YES
tbIDIS	CrossTable	DC001	Patient has no record in BAS		YES
tbIDIS	CrossTable	DC002	AIDS-defining records, yet AIDS=0 in tblBAS		YES
tbIDIS	CrossTable	DC003	First AIDS-defining DIS_D not equal to AIDS_D in tblBAS		YES

DIS_ID field

containing table

[tblDIS](#)

explanation of variable

Code to identify event

format of data

character. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

Code	Severe Opportunistic Infection/Malignancies
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
ISDI	Isosporiasis diarrhoea (duration > 1 month)
LEIS	Leishmaniasis, visceral
MCDI	Microsporidiosis diarrhoea (dur. > 1 month)
MC	Mycobact. avium complex (MAC) or Kanasii, extrapulm.
MCP	Mycobact. tuberculosis pulm.
MCX	Mycobact. tuberculosis extrapulm
MCPO	Mycobact. pulm., other
MCXO	Mycobact. extrapulm., other
PCP	Pneumocystis carinii pneumonia (PCP)
LEU	Progressive multifocal leucoencephalopathy
SAM	Salmonella bacteriaemia (non-typhoid) (recurrent)
TOX	Toxoplasmosis, brain
FBLS	Focal Brain lesion
KS	Kaposi Sarcoma
HG	Hodgkins Lymphoma
NHG	Non-Hodgkin Lymphoma -not specified
NHGB	Non-Hodgkin Lymphoma ? Burkitt (Classical or Atypical)
NHGI	Non-Hodgkin Lymphoma ? Diffuse large B-cell lymphoma (Immunoblastic or Centroblastic)
NHGU	Non-Hodgkin Lymphoma - Unknown/other histology
NHGP	Non-Hodgkin Lymphoma - Primary Brain Lymphoma
CRVC	Cervical Cancer

Coding Table: Mother-to-child and paediatric specific

[Download this table as CSV file](#)

Stage	Code	CDC disease description
A	CA-LYM	Lymphadenopathy
A	CA-HEY	Hepatomegaly
A	CA-SPL	Splenomegaly

A	CA-DER	Dermatitis
A	CA-PAR	Parotitis
A	CA-URI	Recurrent or persist. UR infection, sinusitis, or otitis media
B	CB-ANE	Anemia
B	CB-BMP	Bacterial meningitis, pneumonia, or sepsis
B	CB-CAN	Candidiasis oropharyngeal for >2 months (age>6)
B	CB-CMY	Cardiomyopathy
B	CB-CMN	CMV onset before 1 month
B	CB-DIA	Diarrhea (recurrent or chronic)
B	CB-FEV	Fever (lasting >1 month)
B	CB-HEP	Hepatitis
B	CB-HSS	Herpes simplex stomatitis (>2 episodes in 1 year)
B	CB-HCV	HSV bronchitis, pneumonitis, esophagitis (<1 month)
B	CB-HZO	Herpes zoster, multidermatomal or relapse
B	CB-LEI	Leiomyosarcoma
B	CB-LYM	Lymphoid interstitial pneumonia or pulmonary lymphoid hyperplasia complex
B	CB-NEP	Nephropathy
B	CB-NOC	Nocardiosis
B	CB-TON	Toxoplasmosis (start before 1 month)
B	CB-VAR	Varicella, disseminated
C	CC-EPD	Pneumocystis disease, extrapulmonary
C	CC-TOD	Toxoplasmosis disseminated
C	CC-COM	Cryptococcal meningitis
C	CC-GEN	M. genavense disease
C	CC-LOB	Lymphoma, primary, cerebral
C	CC-ICC	Carcinoma, cervical, invasive
C	CC-ILE	Intracerebral lesions, indetermined

Case definitions

[Download this table as CSV file](#)

Code	Severe Opportunistic Infection/Malignancies	Definitive/Autopsy or presumptive?	Definition
DEM	AIDS dementia complex	D	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
DEM	AIDS dementia complex	P	Same as above but no CSF and brain imaging performed
BCNE	Bacterial pneumonia, recurrent (>2 episodes within 1 year)	D	New X-ray evidence not present earlier and culture of pathogen that typically causes pneumonia (other than P .carinii or M. tuberculosis)
BCNE	Bacterial pneumonia, recurrent (>2 episodes within 1 year)	P	Acute radiological findings (new X-ray evidence not present earlier) and acute clinical findings
CANO	Candidiasis, oesophageal, bronchi, trachea, or lungs	D/A	Gross inspection by endoscopy/autopsy or by microscopy (histology)
CANO	Candidiasis, oesophageal, bronchi, trachea, or lungs	P	Recent onset retrosternal pain on swallowing and confirmed oral or pharyngeal candidiasis
CRCO	Cryptococcosis, extrapulm.	D/A	Microscopy, culture of, or antigen detection in affected tissue
CRSP	Cryptosporidiosis (duration > 1 month)	D/A	Microscopy. Duration of diarrhoea for more than 1 month

CMVR	Cytomegalovirus (CMV) chorioretinitis	P	Loss of vision and characteristic appearance on serial ophthalmoscopy, progressing over serial months
CMVO	CMV ? other location	D/A	Microscopy (histology or cytology)
HERP	Herpes simplex virus ulcers (duration > 1 month) or pneumonitis/esophagitis	D	Microscopy, culture of, or antigen detection in affected tissue
HIST	Histoplasmosis, extrapulm.	D/A	Microscopy, culture of, or antigen detection in affected tissue
WAST	HIV Wasting Syndrome	D	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)	D/A	Microscopy (histology or cytology). Duration of diarrhoea for more than 1 month
LEIS	Leishmaniasis, visceral	D/A	Histology or culture of Leishmania amastigotes in bone marrow or detection of amastigotes in tissue/fluid from affected organ in a patient with symptoms and signs consistent with disseminated Leishmaniasis
MCDI	Microsporidiosis diarrhoea (dur. > 1 month)	D/A	Stool microscopy or rectal biopsy in patient with persistent diarrhoea
MC	Mycobact. avium complex (MAC) or Kanasii, extrapulm.	D	Culture
MCP	Mycobact. tuberculosis pulm.	D	Culture
MCX	Mycobact. tuberculosis extrapulm	D	Culture
MCPO	Mycobact. pulm., other	D	Culture (indicate type)
MCPO	Mycobact. pulm., other	P	Acid fast bacteria (species not identified by culture) in sputum
MCXO	Mycobact. extrapulm., other	D	Culture (indicate type)
MCXO	Mycobact. extrapulm., other	P	Acid fast bacteria (species not identified by culture) on microscopy of normally sterile body fluid/tissue
PCP	Pneumocystis carinii pneumonia (PCP)	D	Microscopy (histology or cytology)
PCP	Pneumocystis carinii pneumonia (PCP)	P	Recent onset of dyspnoea on exertion or dry cough, and diffuse bilateral infiltrates on chest X-ray and pO ₂ <70 mmHg and no evidence of bacterial pneumonia
LEU	Progressive multifocal leucoencephalopathy	D/A	Microscopy (histology or cytology)
LEU	Progressive multifocal leucoencephalopathy	P	Progressive deterioration in neurological function and CT/MR scan evidence
SAM	Salmonella bacteraemia (non-typhoid) (recurrent)	D	Culture
TOX	Toxoplasmosis, brain	D	Microscopy (histology/cytology)
TOX	Toxoplasmosis, brain	P	Recent onset focal neurological abnormalities or reduced level of consciousness, and mass effect lesion on scan, and specific therapy response
FBLS	Focal Brain lesion	?	TODO:: To be updated ASAP
KS	Kaposi Sarcoma	D/A	Histology
KS	Kaposi Sarcoma	P	Characteristic erythematous/violaceous plaque-like lesion on skin or mucous membranes
HG	Hodgkins Lymphoma	?	TODO:: To be updated ASAP
NHG	Non-Hodgkin Lymphoma -not specified	?	TODO:: To be updated ASAP
NHGB	Non-Hodgkin Lymphoma ? Burkitt (Classical or Atypical)	D	Histology

NHGI	Non-Hodgkin Lymphoma ? Diffuse large B-cell lymphoma (Immunoblastic or Centroblastic)	D	Histology
NHGU	Non-Hodgkin Lymphoma - Unknown/other histology	?	TODO:: To be updated ASAP
NHGP	Non-Hodgkin Lymphoma - Primary Brain Lymphoma	D	TODO:: To be updated ASAP
NHGP	Non-Hodgkin Lymphoma - Primary Brain Lymphoma	P	Recent onset of focal neurological symptoms and signs or reduced level of consciousness, CT/MR 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	D/A	Histology

DIS_WD field

containing table

[tblDIS](#)

explanation of variable

Means of diagnosis

format of data

numeric. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

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

tbLAB_CD4 - Laboratory values

holds date and value of CD4 measurements.

Note: If needed, a CD8 table (tbLAB_CD8) could be formed from the same structure.

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
PATIENT	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
CD4_D	yyyy-mm-dd	Date of measurement
CD4_V	numeric (per microliter): -1 = undetectable or detection limit as negative value	Value of CD4 measurement

Additional fields

[CD4_V](#) 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:

Field name	Format	Description
CD4_U	numeric with codes (or full string): <ul style="list-style-type: none"> 1 = cells/μl 2 = % 	Unit of measurement

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbLAB_CD4	WithinTable	CW001	CD4 value out of range		YES
tbLAB_CD4	WithinTable	CW002	Duplicate records for same date		YES
tbLAB_CD4	WithinTable	CW003	Missing CD4_D		YES
tbLAB_CD4	WithinTable	CW004	Missing CD4_V		YES
tbLAB_CD4	WithinTable	CW005	Missing CD4_U		YES
tbLAB_CD4	WithinTable	CW006	Miscoded CD4_U as coding list on table definition		YES
tbLAB_CD4	WithinTable	CW007	CD4_U=2 (percentage) and CD4_V>100		YES
tbLAB_CD4	WithinTable	CW008	CD4_U=(1 or 3) and CD4_V>3000		YES
tbLAB_CD4	WithinTable	CW009	CD4 counts spike up or down suddenly (large change in less than a year)		YES
tbLAB_CD4	CrossTable	CC001	Patient has no record in BAS		YES

tbILAB_RNA - Laboratory values

holds date, value, detection limit and type of viral assay.

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
PATIENT	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
RNA_D	yyyy-mm-dd	Date of Measurement/Sample
RNA_V	numeric: -1 = undetectable or detection limit as negative value	HIV-RNA measurement value
RNA_L	numeric	Lower Limit of HIV-RNA Assay
RNA_T	numeric. see coding table for valid codings.	IF AVAILABLE, What type of VIRAL ASSAY was used for this measurement?

Additional fields

Field name	Format	Description
RNA_UL	numeric	IF AVAILABLE, Upper Limit of assay

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbILAB_RNA	WithinTable	RW001	Duplicate records for same RNA_D		YES
tbILAB_RNA	WithinTable	RW002	RNA_V=-1 and RNA_L missing		YES
tbILAB_RNA	WithinTable	RW003	RNA_V=+1 and RNA_U missing		YES
tbILAB_RNA	WithinTable	RW004	Missing RNA_D		YES
tbILAB_RNA	WithinTable	RW005	Missing RNA_V		YES
tbILAB_RNA	WithinTable	RW006	RNA_V > 10 million		YES
tbILAB_RNA	WithinTable	RW007	RNA_V < RNA_L		YES
tbILAB_RNA	WithinTable	RW008	RNA_V > RNA_U		YES
tbILAB_RNA	WithinTable	RW009	RNA_V < 0		YES
tbILAB_RNA	CrossTable	RC001	Patient has no record in BAS		YES

RNA_T field

containing table

[tblLAB_RNA](#)

explanation of variable

IF AVAILABLE, What type of VIRAL ASSAY was used for this measurement?

format of data

numeric. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

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
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

tbILAB_VIRO - Laboratory values - viro-/serology

holds test results for viro-/serological tests (hepatitis etc.)

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
PATIENT	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
VS_ID	character. see coding table	Viral test
VS_D	yyyy-mm-dd	Measurement date
VS_R	numeric: <ul style="list-style-type: none"> 1 = Positive 0 = Negative 9 = Unknown/borderline 	Measurement result
VS_V	numeric	Measurement value (HCV-RNA & HBV-DNA only) (copies/ml)
VS_U	character. see coding table for valid codings.	Measurement unit

Additional fields

Field name	Format	Description
VS_LL	numeric	IF AVAILABLE, Lower limit of assay
VS_UL	numeric	IF AVAILABLE, Upper limit of assay
VS_T	character. see coding table for valid codings.	IF AVAILABLE, type of ASSAY used for this measurement

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbILAB_VIRO	WithinTable	LVW001	Patient doesn't have a record in BAS		YES
tbILAB_VIRO	WithinTable	LVW002	Missing patient		YES
tbILAB_VIRO	WithinTable	LVW003	Missing VS_ID		YES
tbILAB_VIRO	WithinTable	LVW004	Missing VS_D		YES
tbILAB_VIRO	WithinTable	LVW005	Missing VS_R		YES
tbILAB_VIRO	WithinTable	LVW006	Missing VS_V		YES
tbILAB_VIRO	WithinTable	LVW007	Missing VS_U		YES
tbILAB_VIRO	WithinTable	LVW008	More or less than exactly 1 positive HIV test per patient		YES
tbILAB_VIRO	WithinTable	LVW009	More than 1 negative HIV test for a patient		YES
tbILAB_VIRO	WithinTable	LVW010	Date of negative test after date of positive test		YES

tblLAB_VIRO	WithinTable	LVW011	Duplicate records		YES
-------------	-------------	--------	-------------------	--	-----

VS_ID field

containing table

[tblLAB_VIRO](#)

explanation of variable

Viral test

format of data

character. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

Code	Viral Test
BVA	Bacterial vaginosis unspecified method
BVAC	Bacterial vaginosis - clinical
BVAG	Bacterial vaginosis - gram stain
CHLA	Chlamydia
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

VS_U field

containing table

[tblLAB_VIRO](#)

explanation of variable

Measurement unit

format of data

character. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

Code	Test measurement unit
1	copies/mL
2	IU/mL
3	Geq (millions of genome equivalent)

VS_T field

containing table

[tblLAB_VIRO](#)

explanation of variable

IF AVAILABLE, the type of ASSAY used for this measurement

format of data

character. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

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
6	Other

tblLAB_BP - Laboratory values - Blood pressure

holds date, diastolic and systolic values and unit of blood pressure measurements.

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
PATIENT	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
BP_D	yyyy-mm-dd	Date of Measurement/Sample
BP_SYS	numeric	Systolic Blood Pressure
BP_DIA	numeric	Diastolic Blood Pressure
BP_U	numeric. see coding table for valid codings.	Unit of measurement

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tblLTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tblLTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tblBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES

BP_U field

containing table

[tblLAB_BP](#)

explanation of variable

Unit of measurement

format of data

numeric. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

Code	Unit for blood pressure
1	mmHg
2	cmHg
3	Kpa

tbILAB - Laboratory values

holds type, date, value and unit of laboratory tests.

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
PATIENT	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
LAB_ID	character. see coding table for valid codings.	Code representing the measurement
LAB_D	yyyy-mm-dd	Date of measurement/sample
LAB_V	numeric: -1 = undetectable or detection limit as negative value	Value of measurement
LAB_U	numeric. see coding table for valid codings.	Unit of measurement

Additional fields

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

Field name	Format	Description
LAB_FA	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 9 = Unknown 	Was the blood sample taken while fasting?
LAB_ST	<ul style="list-style-type: none"> WB = Whole Blood P = Plasma S = Serum 	Specimen type

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).

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbILAB	WithinTable	LW001	Duplicate records		YES
tbILAB	WithinTable	LW002	Missing LAB_D		YES
tbILAB	WithinTable	LW003	Missing LAB_V		YES
tbILAB	WithinTable	LW004	Missing LAB_U		YES
tbILAB	WithinTable	LW005	Missing LAB_ID		YES
tbILAB	WithinTable	LW006	Missing LAB_LLN		YES
tbILAB	WithinTable	LW007	Missing LAB_ULN		YES
tbILAB	WithinTable	LW008	Missing LAB_FA		YES
tbILAB	WithinTable	LW009	Missing LAB_ST		YES
tbILAB	WithinTable	LW010	LAB_V out of range for unit LAB_U		YES

tblLAB	CrossTable	LC001	Patient has no record in BAS table		YES
--------	------------	-------	---------------------------------------	--	-----

LAB_ID field

containing table

[tblLAB](#)

explanation of variable

Code representing the measurement

format of data

character. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

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
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

LAB_U field

containing table

[tblLAB](#)

explanation of variable

Unit of measurement

format of data

numeric. see [coding table](#) for valid codings.

exists since HICDEP version

[1.30](#)

Coding Table

[Download this table as CSV file](#)

in case of measurement of	Unit Code	Unit String
Alanin-Aminotransferase	5	IU/L (u/L)
Alanin-Aminotransferase	11	µkat/L
Albumine	2	gm/L
Alk. Phosphatase	5	IU/L (u/L)
Amylase	5	IU/L (u/L)
Amylase	11	µkat/L
Creatinine	6	µmol/L
Glucose	1	mmol/L
Haemoglobin	1	mmol/L
Haemoglobin	2	gm/L
Haemoglobin	3	gm/dL
Lactate	1	mmol/L
Lactate	4	mg/dL
Lymphocyte count	8	1E+9/L
Lymphocyte count	9	1E+6/L
Lymphocyte count	10	cells/µL
Platelet count	8	1E+9/L
Platelet count	9	1E+6/L
Platelet count	10	cells/µL
Quick/INR	7	INR
Serum HDL	1	mmol/L
Serum HDL	2	gm/L
Serum HDL	3	gm/dL
Serum HDL	4	mg/dL
Serum HDL	5	IU/L (u/L)
Serum Triglyceride	1	mmol/L
Serum Triglyceride	2	gm/L
Serum Triglyceride	4	mg/dL
Total Bilirubin	6	µmol/L
Total Cholesterol	1	mmol/L
Total Cholesterol	2	gm/L
Total Cholesterol	3	gm/dL
Total Cholesterol	4	mg/dL
WBC count	8	1E+9/L
WBC count	9	1E+6/L
WBC count	10	cells/µL

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

tbISAMPLES - Blood Samples

This table contains information on the storage of blood, urine and other samples stored in a laboratory.

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
PATIENT	character (or numeric if possible)	patient cohort identifier
SAMP_LAB_D	YYYY-mm-dd	date when the sample was taken
SAMP_TYPE	character: <ul style="list-style-type: none"> BS = blood serum BP = blood plasma C = viable cells D = cell pellet (DNA) S = semen OTH:x = other sample type x (none of the above) TODO find further types of samples	type of the sample
SAMP_ID	character	identification symbol allowing the localization of the sample in the laboratory
SAMP_LAB	character	laboratory where the samples are stored
SAMP_FREEZE_D	YYYY-mm-dd	date when the sample was frozen
SAMP_FREEZE_T	hh:mm	time when the sample was frozen
SAMP_ALIQ_NO	numeric	number of aliquots available
SAMP_ALIQ_SIZE	numeric	size of the aliquot: <ul style="list-style-type: none"> in ml for serum, plasma and cell pellet aliquots in millions of cells for viable cell aliquots
SAMP_ALIQ_U	character: <ul style="list-style-type: none"> 0 = millions of cells 1 = ml 	unit of measurement for the SAMP_ALIQ_SIZE value

Additional fields

Field name	Format	Description
SAMP_LAB_T	hh:mm	time when the sample was taken
SAMP_TEMP	numeric	temperature of the storage unit containing the samples (in °C)
SAMP_DEFROST	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 9 = Unknown 	have the samples already been defrosted?

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES

AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES

tbILAB_RES - Resistance testing

holds background information on the resistance test, laboratory, library, kit, software and type of test

Note: This table is tightly linked to [tblLAB_RES_LVL_1](#), [tblLAB_RES_LVL_2](#) and [tblLAB_RES_LVL_3](#).

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.

These four tables are designed to capture several possible formats the clinics and cohorts might have recorded resistance test data in. Once this data is gathered it should like all other tables be quality assessed.

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
PATIENT	character (or numeric if possible)	Code to identify patient (Cohort Patient ID)
SAMP_ID	character (or numeric if possible)	The assigned sample ID
SAMPLE_D	yyyy-mm-dd	Date of the actual sample taken (NOT the test date)
SEQ_DT	yyyy-mm-dd hh:mm	Date and time when the sequencing was performed
LAB	character	Name of laboratory where the test was performed
LIBRARY	character	Library/algorithm used to identify resistance mutations
REFSEQ	character	Name/identifier of reference HIV strain used to find mutations
KIT	character	Vendor and version/name of the kit used for the test
SOFTWARE	character	Software and version used to determine resistance
TESTTYPE	numeric: <ul style="list-style-type: none"> 1 = Genotype 2 = Phenotype 9 = Other 	Type of test
SUBTYPE	character	Subtype of HIV-RNA

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* key¹.

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.

¹: However this raises the issue about several aliquots from the same day will look like duplicates in the tables.

Field name	Format	Description
SAMP_LAB	character (or numeric if possible)	The assigned sample ID at the lab where the resistance test is performed.
SAMP_INT	character (or numeric if possible)	The assigned sample ID from the centre.

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tblLTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tblLTFU		YES

AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tblBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tblLAB_RES	WithinTable	LRW001	Duplicate records for same patient on same date		YES
tblLAB_RES	WithinTable	LRW002	Missing PATIENT		YES
tblLAB_RES	WithinTable	LRW003	Missing SAMP_ID		YES
tblLAB_RES	WithinTable	LRW004	Missing SAMPLE_D		YES
tblLAB_RES	WithinTable	LRW005	Missing SEQ_DT		YES
tblLAB_RES	WithinTable	LRW006	Missing LAB		YES
tblLAB_RES	WithinTable	LRW007	Missing LIBRARY		YES
tblLAB_RES	WithinTable	LRW008	Missing REFSEQ		YES
tblLAB_RES	WithinTable	LRW009	Missing KIT		YES
tblLAB_RES	WithinTable	LRW010	Missing SOFTWARE		YES
tblLAB_RES	WithinTable	LRW011	Missing TESTTYPE		YES
tblLAB_RES	WithinTable	LRW012	Missing SUBTYPE		YES
tblLAB_RES	WithinTable	LRW013	SEQ_DT has no time part		YES
tblLAB_RES	CrossTable	LRC001	PATIENT has no record in tblBAS		YES
tblLAB_RES	CrossTable	LRC002	This SAMP_ID has both LVL_1 and LVL_2 records		YES

tbILAB_RES_LVL_1 - Nucleotide sequences (PRO, RT, GP41, GP120)

holds nucleoside sequence for the PRO and RT sequences. No entry is made if the test was a phenotype test.

Note: This table is tightly linked to [tbILAB_RES](#).

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
SAMP_ID	character (or numeric if possible)	The assigned sample ID
SEQTYPE	character: <ul style="list-style-type: none"> PRO = PRO sequence RT = RT sequence GP41 = GP41 sequence GP120 = GP120 sequence 	Type of nucleotide sequence if available
SEQ_START	numeric	Start position for the sequence
SEQ_STOP	numeric	Stop position for the sequence
SEQ_NUC	character	Nucleotide sequence if available

Additional fields

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

Field name	Format	Description
SEQ_AA	character	Amino acid sequence if available (empty if test was phenotype)

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.

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES

tbILAB_RES_LVL_2 - Mutations

holds mutations and positions of PRO and RT sequences.

Note: This table is tightly linked to [tbILAB_RES](#).

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
SAMP_ID	character (or numeric if possible)	The assigned sample ID
GENE	character: <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • a = first • b = second • etc. 	Subposition used to code insertions
AA_FOUNDED_1	character. empty = Amino acid has been deleted.	Mutation (Amino acid) found in the sequence
AA_FOUNDED_2	character. empty = Amino acid has been deleted.	Mutation (Amino acid) found in the sequence (if more than 1)
AA_FOUNDED_3	character. empty = Amino acid has been deleted.	Mutation (Amino acid) found in the sequence (if more than 2)
AA_FOUNDED_4	character. empty = Amino acid has been deleted.	Mutation (Amino acid) found in the sequence (if more than 3)

[AA_FOUNDED_#](#) could be extended if mixtures with more than 4 amino acids are found.

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES

tbILAB_RES_LVL_3 - Resistance test result

holds resistance result in relation to antiretroviral drug.

Note: This table is tightly linked to [tbILAB_RES](#).

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
SAMP_ID	character (or numeric if possible)	The assigned sample ID
ART_ID	character. see coding table for valid codings.	Drug code of antiretroviral
RES_SCOR	character	Score of resistance or recommendation given from the test.

Additional fields

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

Field name	Format	Description
RES_CUT	character	Cut-off value for phenotype test result

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.

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES

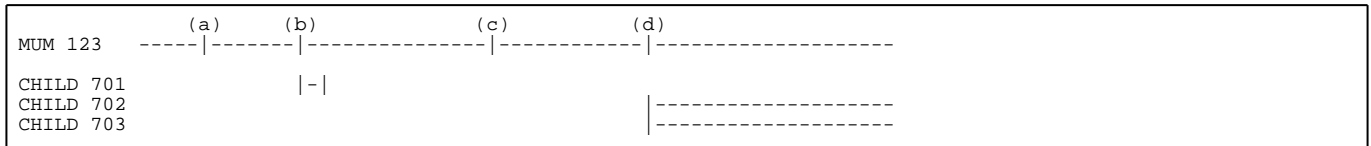
Notes on the Pregnancy-Related Tables

The correct and accurate modelling of a pregnancy and its outcomes is rather involved, and some assumptions had to be made. This page illustrates the model used in the pregnancy tables to which these assumptions lead.

Example

Consider the following example:

1. A female patient (123) gets pregnant at time (a).
2. At time (b) she has an abortion which causes the dead fetus of CHILD 701.
3. At time (c), she gets pregnant again.
4. The delivery of this pregnancy is at time (d), when the twins (CHILD 702 and CHILD 703) are born.



This story is described in the different pregnancy tables as follows:

There are two records in [tblPREG](#), one for each pregnancy:

MOTHER_ID	PREG_SEQ	MENS_D	...
123	1	(a)	...
123	2	(c)	...

There are three records in [tblPREG_OUT](#) describing the pregnancy outcome for each fetus:

MOTHER_ID	PREG_SEQ	CHILD_ID	...
123	1	701	...
123	2	702	...
123	2	703	...

There is one record in [tblDELIVERY_MUM](#):

MOTHER_ID	PREG_SEQ	MEMRUP_D	...
123	2	(d)	...

There are two records in [tblDELIVERY_CHILD](#):

MOTHER_ID	MEMRUP_D	CHILD_ID	...
123	(d)	702	...
123	(d)	703	...

There are two records in [tblNEWBORN](#):

CHILD_ID	...
702	...
703	...

Important relations

The following statements are true for a well-encoded data set:

1. For every record in [tblPREG](#), there are 1 or more records in [tblPREG_OUT](#) referring to it.
2. For every record in [tblPREG_OUT](#) that describes a delivery, there is exactly one record in [tblDELIVERY_CHILD](#) referring to it.
3. For every record in [tblDELIVERY_MUM](#), there are 1 or more records in [tblDELIVERY_CHILD](#) referring to it indirectly through [tblDELIVERY_MUM](#).
4. For every record in [tblNEWBORN](#) there is exactly one record in [tblDELIVERY_CHILD](#) referring to the same child.
5. For every record in [tblDELIVERY_MUM](#), there is some record in [tblPREG](#) to which it refers.

tbIPREG - Pregnancy

Please also read the [notes on pregnancy tables](#).

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
MOTHER_ID	Character (or numeric if possible)	Patient ID of mother of the child
PREG_SEQ	numeric	Sequence number of the pregnancy for the specified mother
MENS_D	yyyy-mm-dd	Date of last menstrual period (If date not known exactly please give approximated date)
CONCEPT	character: <ul style="list-style-type: none"> 1=Natural 2=Infertility treatment unspecified 2.1=IVF (In Vitro Fertilisation) 2.2=ICSI (IntraCytoplasmic Sperm Injection) 2.3=Ovulation induction 3=Artificial insemination 4=Self insemination 9=Unknown 	Conception
ANC_D	yyyy-mm-dd	Date of first antenatal care contact
INPREG_Y	character: <ul style="list-style-type: none"> 1=Yes 2=No: ectopic 3=No: missed abortion 4=No: death in utero (IUFT) 	At first gynaecological visit: intact intrauterine pregnancy?
INHIST_Y	numeric: <ul style="list-style-type: none"> 0=No 1=Yes 9=Unknown 	If no in INPREG_Y was a histological investigation of reason made.
INHIST_S	character	If yes, please specify reason
INV_PROC	<ul style="list-style-type: none"> 0=No 1=Yes, chorionic villus sampling (CVS) 2=Cordocentesis 3=Amniocentesis 9=Unknown 	Invasive procedure
KARYO_T	<ul style="list-style-type: none"> 0=Not done 1=Normal 2=Abnormal 9=Unknown 	Karyotype
KARYO_A	character	If abnormal, please specify
CHORIO	character: <ul style="list-style-type: none"> 1=Monochorionic 2=Dichorionic 3=Trichorionic etc. 9=Unknown 	For multiple pregnancies
ULTR_1	character: <ul style="list-style-type: none"> 0=No 1=Yes, normal 2=Yes, abnormal 9=Unknown 	Ultrasound 1. trimester
ULTR_A_1	character	If abnormal, please specify

<i>ULTR_2</i>	character: <ul style="list-style-type: none"> • 0=No • 1=Yes, normal • 2=Yes, abnormal • 9=Unknown 	Ultrasound 2. trimester
<i>ULTR_A_2</i>	character	If abnormal, please specify
<i>PROB_Y</i>	numeric: <ul style="list-style-type: none"> • 1 = Yes • 0 = No • 9 = Unknown 	did any obstetrical problems occur? (if yes, recorded in tblPREG_OBS)

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tblLTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tblLTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tblBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tblPREG	WithinTable	PW001	ANC_D		
tblPREG	WithinTable	PW002	INPREG_Y=2,3,4 but INHIST_Y null		
tblPREG	WithinTable	PW003	INPREG_Y=1 but INHIST_Y non null		
tblPREG	WithinTable	PW004	INHIST_Y=1 but INHIST_S null		
tblPREG	WithinTable	PW004	INHIST_Y=0 or 9 but INHIST_S non null		
tblPREG	WithinTable	PW005	KARYO_T=2 but KARYO_A null		
tblPREG	WithinTable	PW006	KARYO_T<>2 but KARYO_A non null		
tblPREG	WithinTable	PW007	ULTRA_x =2, but ULTR_A_x null		
tblPREG	WithinTable	PW008	ULTRA_x <>2, but ULTR_A_x non null		
tblPREG	CrossTable	PC001	PROB_Y = 1 but no records in tblPREG_OBS		
tblPREG	CrossTable	PC001	PROB_Y = 0 or 9 but records in tblPREG_OBS		

tbIPREG_OBS - Obstetrical problems during pregnancy

This table describes problems **during a pregnancy**. Abnormalities in newborns are recorded in [tbNEWBORN_ABNORM](#) instead.

Please also read the [notes on pregnancy tables](#).

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
MOTHER_ID	Character (or numeric if possible)	patient id of mother
PREG_SEQ	numeric	Sequence number of the pregnancy for the specified mother
PROB_T	character: <ul style="list-style-type: none"> 1 = Preterm contractions 2 = Shortened cervix 3 = Preterm rupture of membranes 4 = Antepartum bleeding 5 = Intrauterine growth retardation (IUGR) 6 = Preeclampsia/HELLP 7 = Hypertension 8 = Gestational diabetes (unspecified type) 8.1 = Gestational diabetes (Diet) 8.2 = Gestational diabetes (Insulin) 9 = Placental abruption 10 = Placenta praevia 99 = Other, specify in PROB_S 	type of obstetrical problem
PROB_S	character.	description of other (99) obstetrical problem
CERVIX_S	numeric (mm)	In case of shortened cervix, the length of the cervix in millimeters.

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbIPREG_OBS	CrossTable	POC001	MOTHER_ID+PREG_SEQ doesn't exist in tbIPREG		
tbIPREG_OBS	WithinTable	POW001	PROB_T=99 but PROB_S null		
tbIPREG_OBS	WithinTable	POW002	PROB_T<>99 but PROB_S non null		
tbIPREG_OBS	WithinTable	POW003	PROB_T=2 but CERVIX_S null		
tbIPREG_OBS	WithinTable	POW003	PROB_T<>2 but CERVIX_S non null		

tbIPREG_OUT - Pregnancy outcome

This table describes outcomes of pregnancies described in [tbIPREG](#).

Please also read the [notes on pregnancy tables](#).

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
MOTHER_ID	Character (or numeric if possible)	Patient ID of mother of the child
PREG_SEQ	numeric	Sequence number of the pregnancy for the specified mother
CHILD_ID	Character (or numeric if possible)	Patient ID of the child
OUTCOM	character: <ul style="list-style-type: none"> 1=Born alive, HIV negative 2=Born alive, HIV positive 3=Born alive, unknown HIV status 10=Stillborn 11=Spontaneous miscarriage 20=Termination: surgical 21=Termination: medication 	Pregnancy outcome
OUTCOM_R	<ul style="list-style-type: none"> 1=Fetus with malformation 2=Dead fetus 3=Unwanted pregnancy 8=Other 9=Unknown 	Reason for termination
OUTCOM_D	yyyy-mm-dd	Date of birth or termination of pregnancy
B_GAGEW	numeric	Gestational age in complete weeks at birth or termination
B_GAGED	numeric	Gestational age in days in addition to weeks at birth or termination

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tblLTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tblLTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tblBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbIPREG_OUT	CrossTable	PTC001	MOTHER_ID+PREG_SEQ doesn't exist in tbIPREG		
tbIPREG_OUT	CrossTable	PTC002	OUTCOM=1,2 or 3 and CHILD_ID doesn't exist in tbINEWBORN or tbIDELIVERYCHILD		
tbIPREG_OUT	CrossTable	PTC002	OUTCOM=10,11,20 or 21 and CHILD_ID exists in tbINEWBORN or tbIDELIVERYCHILD		
tbIPREG_OUT	WithinTable	PTW001	OUTCOME=20 or 21 and OUTCOM_R null		

tblPREG_OUT	WithinTable	PTW001	OUTCOME not 20 or 21 and OUTCOM_R non null		
-------------	-------------	--------	--	--	--

tbDELIVERY_MUM - Delivery information related to the mother

This table contains information about the delivery **specific to the mother**. Child-specific information is recorded in [tbDELIVERY_CHILD](#) instead.

Please also read the [notes on pregnancy tables](#).

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
MOTHER_ID	Character (or numeric if possible)	Patient ID of mother of the child
PREG_SEQ	numeric	Sequence number of the pregnancy for the specified mother
MEMRUP_D	yyyy-mm-dd	Date of rupture of membranes
MEMRUP_T	hh:mm	Time of rupture of membranes
LABOUR	character: <ul style="list-style-type: none"> 1=Spontaneous 2=Induced 3=No labour (elective C-section) 	Onset of labor
INTERV	character: <ul style="list-style-type: none"> 0=No 1=Fetal blood sampling FBS 2=Internal electrodes 3=1+2 90=Other 91=1+Other 92=2+Other 93=1+2+Other 	Interventions during delivery
INTERV_O	character	Interventions during delivery - other
TEAR_Y	numeric: <ul style="list-style-type: none"> 1=Yes 0=No 9=Unknown 	Episiotomy/tear
BLDLOSS	numeric (mL)	Estimated blood loss during delivery
CONTREAT	character: <ul style="list-style-type: none"> 0=No, treatment has been interrupted 1=Yes, at the foreseen intervals 2=Yes, but not at foreseen time points 9=Unknown 	Did the patient continue the usual antiretroviral therapy?
DISCHA_D	yyyy-mm-dd	Date of discharge from hospital

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES

tblDELIVERY_MUM	CrossTable	DMC001	MOTHER_ID doesn't exist in tblPREG		
tblDELIVERY_MUM	CrossTable	DMC002	MOTHER_ID doesn't exist in tblDELIVERY_CHILD with same MEMRUP_D		
tblDELIVERY_MUM	WithinTable	DMW001	PREG_SEQ invalid (PREG_SEQ>1 and PREG_SEQ=x, but PREG_SEQ=x-1 doesn't exist)		
tblDELIVERY_MUM	WithinTable	DMW002	DISCHA_D < MEMRUP_D		
tblDELIVERY_MUM	WithinTable	DMW004	INTERV=0-3 but INTERV_O non null		
tblDELIVERY_MUM	WithinTable	DMW003	INTERV=90,91, 92 or 93 but INTERV_O null		

tbDELIVERY_CHILD - Delivery information related to the child(ren)

This table holds information about a delivery **specific to a child**. Mother-specific data is kept in [tbDELIVERY_MUM](#).

Please also read the [notes on pregnancy tables](#).

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
MOTHER_ID	Character (or numeric if possible)	Patient ID of mother of the child
MEMRUP_D	yyyy-mm-dd	Date of rupture of membranes
CHILD_ID	Character (or numeric if possible)	Patient ID of the child
B_SEQ	numeric	If multiple births, indicate number (1=first born)
DELIV_D	yyyy-mm-dd	Date of delivery
DELIV_T	hh:mm	Time of delivery
DELIV_M	character: <ul style="list-style-type: none"> 1=Vaginally, spontaneous 2=Vaginally, forceps 3=Vaginally, vacuum 10= Cesarean section, primary/elective (before onset of labour and rupture of membrane) 11=Cesarean section, Secondary 	Mode of delivery
LABOUR_P	numeric	Duration of labour in hours (from cervical dilatation (>=3 cm) until delivery)
BREECH_Y	numeric: <ul style="list-style-type: none"> 1 = Yes 0 = No 9 = Unknown 	Was the child born from a breech presentation?

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbDELIVERY_CHILD	CrossTable	DCC001	MOTHER_ID doesn't exist in tbIPREG		
tbDELIVERY_CHILD	CrossTable	DCC002	CHILD_ID doesn't exist in tbIPREG_OUT		
tbDELIVERY_CHILD	WithinTable	DCW001	Duplicate records for same Mother_ID, CHILD_ID		
tbDELIVERY_CHILD	WithinTable	DCW002	B_SEQ >4		
tbDELIVERY_CHILD	WithinTable	DCW003	B_SEQ>1, but no record exists for same Mother, Child with B_SEQ=1		

tblDELIVERY_CHILD	WithinTable	DCW005	DELIV_D < MEMRUP_D		
tblDELIVERY_CHILD	WithinTable	DCW004	LABOUR_P >72		
tblDELIVERY_CHILD	WithinTable	DCW006	2 children, same mother, same MEMRUP_D, DELIV_D >DELIV_D , yet B_SEQ < B_SEQ		

tbNEWBORN - Newborn

Please also read the [notes on pregnancy tables](#).

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
CHILD_ID	Character (or numeric if possible)	Patient ID of the child
BRFEED_SD	yyyy-mm-dd	Breastfeeding, start date.
BRFEED_ED	yyyy-mm-dd	Breastfeeding, end date.
FAT_ETH	See Coding Table .	Ethnicity of father
APGAR_1	numeric	1 st APGAR score
APGARM_1	numeric	minute at which the 1 st APGAR test was performed
APGAR_2	numeric	2 nd APGAR score
APGARM_2	numeric	minute at which the 2 nd APGAR test was performed
APGAR_3	numeric	3 rd APGAR score
APGARM_3	numeric	minute at which the 3 rd APGAR test was performed
ICU_Y	numeric: <ul style="list-style-type: none"> • 1 = Yes • 0 = No • 9 = Unknown 	Referral to intensive/intermediate care unit?
ICU_S	character	if yes, specify reason
ICU_D	yyyy-mm-dd	Date of entry to intensive/intermediate care unit
ABNORM_Y	numeric: <ul style="list-style-type: none"> • 1 = Yes • 0 = No • 9 = Unknown 	did any abnormalities occur? (if yes, recorded in tbNEWBORN_ABNORM)

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbNEWBORN	CrossTable	NC001	CHILD_ID doesn't exist in tbIDELIVERY_CHILD		
tbNEWBORN	CrossTable	NC002	ABNORM_Y = 1, yet no records in tbNEWBORN_ABNORM		
tbNEWBORN	CrossTable	NC003	ABNORM_Y = 0 or 9, yet records in tbNEWBORN_ABNORM		
tbNEWBORN	WithinTable	NW001	BRFEED_SD>BRFEED_ED		

tblNEWBORN	WithinTable	NW002	APGARM_X's out of order (e.g APGARM_3 < APGARM_2)		
tblNEWBORN	WithinTable	NW003	ICU_Y=1, but ICU_S or ICS_D null		
tblNEWBORN	WithinTable	NW004	ICU_Y=0 or 9, but ICU_S or ICU_D non null		

tbNEWBORN_ABNORM - Abnormalities

Abnormalities in newborns are recorded here, one abnormality per row. The absence of a record is to be interpreted as "unknown whether the abnormality existed" since most cohorts only record positive events.

Please also read the [notes on pregnancy tables](#).

Core fields

Note: Fields marked **bold** form the unique identifier for a record of the table.

Field name	Format	Description
CHILD_ID	Character (or numeric if possible)	Patient ID of the child
ABNORM_T	character: <ul style="list-style-type: none"> • 1 = Birth defect(s) (detectable in physical examination including skin abnormalities) • 2 = Congenital infection(s) • 3 = Drug withdrawal syndrome • 4 . 1 = Neurological disorder(s): abnormal reflexes • 4 . 2 = Neurological disorder(s): abnormal motility • 4 . 3 = Neurological disorder(s): abnormal tonus • 90 = Other health problems 	type of abnormality
ABNORM_S	character	further specification of the abnormality

QA Checks

Table	Crosstable	Error Code	Description	Study specific	HICDEP?
AllTables	CrossTable	ATC001	any date in database after DEATH_D in tbILTFU		YES
AllTables	CrossTable	ATC002	any date in database after DROP_D in tbILTFU		YES
AllTables	CrossTable	ATC003	any date in database before BIRTH_D in tbIBAS		YES
AllTables	CrossTable	ATC004	any date in database in the future		YES
AllTables	CrossTable	ATC005	patients submitted previously who have been missed out		YES
AllTables	CrossTable	ATC006	Any fields not coded as coding lists on table definition		YES
tbNEWBORN_ABNORM	CrossTable	NAC001	CHILD_ID doesn't exist in NEWBORN		
tbNEWBORN_ABNORM	WithinTable	NAW001	ABNORM_T=90 but ABNORM_S null		

Change log

Version 1.50

- Added the pediatric tables: [tbIPREG](#), [tbIPREG_OBS](#), [tbIPREG_OUT](#), [tbIDELIVERY_MUM](#), [tbIDELIVERY_CHILD](#), [tbNEWBORN](#) and [tbNEWBORN_ABNORM](#).
- Added new table [tbISAMPLES](#) which holds information regarding sample storage.
- [tbIART](#): Updated list of drugs

Version 1.30

- [tbIART](#): Updated list of drugs
- [tbIAE](#):
 - added [EVENT_ID](#) as unique identifier and link to detailed tables for each event (see ¹) ? this replaces the optional [AE_NO](#) field.
 - added [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.
 - in [AE_R_Y](#) - Relation to treatment: added more detailed codes.
- [tbILAB](#): added several codes for various biomarker tests.
- [tbILAB_CD4](#): added [CD4_U](#) as optional fields to discriminate between CD4% and CD4 cell count, so that the [tbILAB_CD4](#) table can hold both types of measurements.
- [tbILAB_VIRO](#): added several codes for various virology and serology tests.
- [CaseDefinitions](#) updated with end stage renal disease, chronic liver disease and non-AIDS defining malignancies

¹:Detailed table definitions for the D:A:D events are available at <http://www.cphiv.dk/HICDEP/Documents/tabid/159/Default.aspx>

Version 1.25

- [tbIART](#): Updated list of drugs
- [tbIMED](#): Updated list of drugs
- [tbIDIS](#):
 - Changed wording for CANO to ?Candidiasis, oesophageal, bronchi, trachea, or lungs?
 - Added COCC - Coccidioidomycosis, disseminated or extrapulmonary
- [tbILAB](#): Added [LAB_ST](#) as additional field to code for type of specimen used for the measurement
- [tbILAB_CD4](#): Added [CD4_U](#) as additional field so the table can hold both percentage and absolute CD4 measurements
- [tbILAB_RNA](#): Added [RNA_UL](#) (upper limit of detection) to the list of additional fields.
- Added more viral assays to the list of [RNA_T_codes](#)
- [tbILAB_VIRO](#): Added unit field to [tbILAB_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.
- [tbILTFU](#): 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

- Added reasons for stopping treatment to table [tbIART_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

- added [CaseDefinitions](#)

Version 1.1

[tbIBAS](#):

- The table was split into *tblBAS* and *tblLTFU*. *tblLTFU* holds data on death and drop-out
 - Renamed *LOS_Y* to *LOSS_Y*
 - Renamed *GAL_Y* to *GAIN_Y*
- tblLAB_BLP*:
- Renamed table to *tblLAB_BP*
 - Renamed *BLP_D* to *BP_D*
 - Renamed *BLP_SYS* to *BP_SYS*
 - Renamed *BLP_DIA* to *BP_DIA*
 - Renamed *BLP_U* to *BP_U*

Version 1.00

tblBas:

- Renamed *BIRTHDAY* to *BIRTH_D*
- Renamed *FIRSTVIS* to *FRSVIS_D*
- Renamed *REC_ART* to *RECART_Y*

tblLAB:

- *LAB_U*: 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* 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_LVL1:

- Renamed *SEQ_ST* to *SEQ_STAR*
- *SEQ_STOP*: Added to the table to specify at which position in the sequence the sequencing was terminated

tblLAB_RES_LVL2:

- 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

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.

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 like in the following table to exist:

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

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

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

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

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
Visit number
Numeric: 0 = Baseline Visit 1 = First follow up visit 2 = Second follow up visit 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
Period of usage (1 st , 2 nd , 3 rd etc.)
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
Number of data entry
Numeric: 1 = first data entry 2 = second data entry 3 = comparison of 1 st and 2 nd data entry 4 = final approved record including corrections

With respect to performance, it might also be a good design 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	USER_LOG
Date and time of data entry	Username of user that last inserted or updated data
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 CFR part 11](#) on electronic records and signatures applies.

Further normalisation

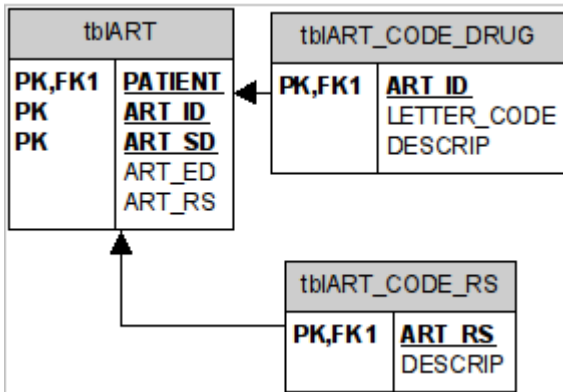
Depending on performance considerations it might be worth looking at how data are queried for data entry and data analysis. A smaller *tblBAS* table might increase performance: Since processing a smaller table is always faster than processing a larger table, one could put drop-out, death, birthday, date 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.



This setup not only enables integrity of the data, but also defines the domain¹ for the #_ID values and enables data to both become human readable and easily recoded².

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.

¹: 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.

²: 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

Performance

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

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.