Study

Health and Demographic Surveillance System (HDSS) is an ongoing study conducted by the ALPHA_NETWORK.

Overview

The data that are used to create the ALPHA specifications come from health-and-demographic-surveillance-system (HDSS) studies being conducted in Eastern and Southern Africa where HIV is a major public health problem. In these settings, national vital registration and population-based health information are scarce. HDSS are an attempt to address this dearth of data.

Development of structured documentation for ALPHA specifications inevitably refers to the concepts that are core to HDSS. In addition, it also refers to objects, relationships and attributes from the HDSS reference data model.

Data Pipeline

ALPHA data management is in the process of migrating from a statistical software platform called Stata to a Business Intelligience (BI) platform called Pentaho.

Using Pentaho, raw data from each ALPHA site is moved through a series of steps. Together they form a data pipeline where the output from one step becomes the input for one or more subsequent steps. In what follows we provide an algorithm overview of each step along with its inputs and outputs.

01_Site_Specific_ETL_for_6.1 Step

Overview

Purpose

Creates staging tables from member centre specific data. The staging tables are then transformed further to create the ALPHA specification 6.1

Algorithm

[step: 1.1] :
This algorithm step references the following study concepts:

[step: 1.2] :
This algorithm step references the following study concepts:

Data Sources for the 01_Site_Specific_ETL_for_6.1 Step

Data Targets for the 01_Site_Specific_ETL_for_6.1 Step

02_Core_ETL_for_Raw_6.1 Step

Overview

Purpose

Creates ALPHA specification 6.1 in event format from staging tables created in the site specific ETL business process (01 Site Specific ETL for 6.1)

Algorithm

[step: 2.1] Generate anonymised IDs: Generate anonymised unique-identifiers
This algorithm step references the following study concepts: <u>unique-identifier</u>

[step: 2.2] Map original unique-identifiers to anonymised IDs: Create a mapping between original and anonymised IDs
This algorithm step references the following study concepts:

[step: 2.3] Store the mapping between original and anonymised IDs: Store the IDs mapping information where it can be accessed internally in the future
This algorithm step references the following study concepts:

[step: 2.4] Create Raw Spec 6.1 from staging data: Create Raw Spec 6.1 from staging data
This algorithm step references the following study concepts:

Data Sources for the 02_Core_ETL_for_Raw_6.1 Step

Data Targets for the 02_Core_ETL_for_Raw_6.1 Step

[1] HH_ID_Anonymise_Map

Note that this dataset has a companion dataset called [HH_ID_Anonymise_Map](#).

study_name [VARCHAR(15)]
hhold_id [VARCHAR(100)]
anonymised_hhold_id [INT]

[2] HH_ID_Anonymise_Map

Note that this dataset has a companion dataset called [HH_ID_Anonymise_Map](#).

No description is available.

[3] Indv_ID_Anonymise_Map

Note that this dataset has a companion dataset called [Indv_ID_Anonymise_Map](#).

study_name [VARCHAR(15)]
idno [VARCHAR(100)]
anonymised_idno [INT]

[4] Indv_ID_Anonymise_Map

Note that this dataset has a companion dataset called [Indv_ID_Anonymise_Map](#).

No description is available.

[5] Raw_61_Event_Format

Note that this dataset has a companion dataset called [Raw_61_Event_Format](#).

recnr [INT]
study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

[6] Raw_61_Event_Format

Note that this dataset has a companion dataset called [Raw_61_Event_Format](#).

No description is available.

03_Core_ETL_Raw_6.1_Dataset_Quality_Metrics Step

Overview

Calculate Data Quality Metrics

Purpose

Assesses the quality of the data in the raw specification created in business process, 02 Core ETL for Raw 6.1, on the basis of a set of quality metrics

Algorithm

[step: 3.1] Compile a list of Quality Metrics: Compile a list of quality metrics relevant to the data specification
This algorithm step references the following study concepts:

[step: 3.2] Create events consistency matrix: Create events consistency matrix showing the logical ordering of event sequences
This algorithm step references the following study concepts:

[step: 3.3] Compile residency starting events: Identify in the data, events that start a residency episode (birth, external-immigration, enumeration, becoming eligible for a study, found after being lost to follow-up)
This algorithm step references the following study concepts: [residency](#) [birth](#) [migration](#)

[step: 3.4] Compile residency ending events: Identify in the data, events that end a residency episode (external-outmigration, death, became ineligible for study, lost to follow-up, internal-outmigration, present in the study (right censored))
This algorithm step references the following study concepts: [residency](#) [death](#) [migration](#)

[step: 3.5] Compile legal and illegal start events: Review the identified start events and distinguish between legal and illegal ones
This algorithm step references the following study concepts:

[step: 3.6] Compile legal and illegal end events: Review the identified end events and distinguish between legal and illegal ones
This algorithm step references the following study concepts:

[step: 3.7] Compile legal and illegal transitions: Review all transitions between two events and distinguish between legal and illegal ones
This algorithm step references the following study concepts:

[step: 3.8] Compile illegal, missing or unknown sex: Compile illegal, missing or unknown sex

This algorithm step references the following study concepts:

[step: 3.9] Compile illegal, missing or runknown Date of Birth (DOB): Compile illegal, missing or unknown DOB
This algorithm step references the following study concepts:

[step: 3.10] Compile quality metrics: Calculate numbers of legal and illegal start events, end events, event transitions, sex values, out of range DOBs and missing sex and DOBs
This algorithm step references the following study concepts:

Data Sources for the 03_Core_ETL_Raw_6.1_Dataset_Quality_Metrics Step

[1] Raw_61_Event_Format

recnr [INT]
study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

Data Targets for the 03_Core_ETL_Raw_6.1_Dataset_Quality_Metrics Step

[1] DoBQualityMetrics

Note that this dataset has a companion dataset called QualityMetrics. DoB quality metrics added to the QualityMetrics table in Staging database.

CentreId [VARCHAR(64)]
MetricTable [VARCHAR(64)]
QMetric [VARCHAR(64)]
Illegal [INT]
Legal [INT]
Total [INT]
Metric [NUMERIC(5:2)]
RunDate [DATETIME)]

[2] DoBValues

Note that this dataset has a companion dataset called DoBValues.

MetricTable [VARCHAR(128)]
DoBStatus [VARCHAR(25)]
Individuals [INT]
RunDate [DATETIME]

[3] DoBValues

Note that this dataset has a companion dataset called DoBValues.

MetricTable [VARCHAR(64)]
DoBStatus [VARCHAR(64)]
Individuals [INT]
RunDate [DATETIME)]

[4] EndingEvents

Note that this dataset has a companion dataset called EndingEvents.

MetricTable [VARCHAR(128)]
EventCode [VARCHAR(3)]
Individuals [INT]
RunDate [DATETIME]

[5] EndingEvents

Note that this dataset has a companion dataset called EndingEvents.

MetricTable [VARCHAR(64)]
EventCode [VARCHAR(64)]
Individuals [INT]
RunDate [DATETIME)]

[6] EndingQualityMetrics

Note that this dataset has a companion dataset called QualityMetrics. End events quality metrics added to the QualityMetrics table in Staging database.

CentreId [VARCHAR(64)]
MetricTable [VARCHAR(64)]
QMetric [VARCHAR(64)]
Illegal [INT]
Legal [INT]
Total [INT]
Metric [NUMERIC(5:2)]
RunDate [DATETIME)]

[7] EventCrossTab

Note that this dataset has a companion dataset called [TransitionsCrosstab](#).

MetricTable [VARCHAR(128)]
EventCode [VARCHAR(3)]
NONE [INT]
BTH [INT]
DTH [INT]
ENT [INT]
ENU [INT]
EXT [INT]
IMG [INT]
OBE [INT]
OBS [INT]
OMG [INT]
RunDate [DATETIME]

[8] IllegalTransitions

Note that this dataset has a companion dataset called [IllegalTransitions](#).

MetricTable [VARCHAR(128)]
EventCode [VARCHAR(3)]
NextEventCode [VARCHAR(3)]
numRecs [INT]
maxDate [DATETIME]
RunDate [DATETIME]

[9] IllegalTransitions

Note that this dataset has a companion dataset called [IllegalTransitions](#).

MetricTable [VARCHAR(64)]
EventCode [VARCHAR(64)]
NextEventCode [VARCHAR(64)]
numRecs [INT]
maxDate [DATETIME]
RunDate [DATETIME)]

[10] QualityMetrics

Note that this dataset has a companion dataset called [StartQualityMetrics](#).
Note that this dataset has a companion dataset called [TransitionQualityMetrics](#).
Note that this dataset has a companion dataset called [SexValueQualityMetrics](#).
Note that this dataset has a companion dataset called [EndingQualityMetrics](#).
Note that this dataset has a companion dataset called [DoBQualityMetrics](#).

CentreId [VARCHAR(15)]

MetricTable [VARCHAR(128)]
QMetric [VARCHAR(50)]
Illegal [INT]
Legal [INT]
Total [INT]
Metric [FLOAT(53)]
RunDate [DATETIME]

[11] SexValueQualityMetrics

Note that this dataset has a companion dataset called [QualityMetrics](). Sex values quality metrics added to the QualityMetrics table in Staging database.

CentreId [VARCHAR(64)]
MetricTable [VARCHAR(64)]
QMetric [VARCHAR(64)]
Illegal [INT]
Legal [INT]
Total [INT]
Metric [NUMERIC(5:2)]
RunDate [DATETIME)]

[12] SexValues

Note that this dataset has a companion dataset called [SexValues]().

MetricTable [VARCHAR(128)]
Sex [VARCHAR(3)]
Individuals [INT]
RunDate [DATETIME]

[13] SexValues

Note that this dataset has a companion dataset called [SexValues]().

MetricTable [VARCHAR(64)]
Sex [VARCHAR(64)]
Individuals [INT]
RunDate [DATETIME)]

[14] StartQualityMetrics

Note that this dataset has a companion dataset called [QualityMetrics](). Start events quality metrics added to the QualityMetrics table in Staging database.

CentreId [VARCHAR(64)]
MetricTable [VARCHAR(64)]
QMetric [VARCHAR(64)]
Illegal [INT]

Legal [INT]
Total [INT]
Metric [NUMERIC(5:2)]
RunDate [DATETIME)]

[15] StartingEvents

Note that this dataset has a companion dataset called StartingEvents.

MetricTable [VARCHAR(128)]
EventCode [VARCHAR(3)]
Individuals [INT]
RunDate [DATETIME]

[16] StartingEvents

Note that this dataset has a companion dataset called StartingEvents.

MetricTable [VARCHAR(64)]
EventCode [VARCHAR(64)]
Individuals [INT]
RunDate [DATETIME)]

[17] TransitionQualityMetrics

Note that this dataset has a companion dataset called QualityMetrics. Transition events quality metrics added to the QualityMetrics table in Staging database.

CentreId [VARCHAR(64)]
MetricTable [VARCHAR(64)]
QMetric [VARCHAR(64)]
Illegal [INT]
Legal [INT]
Total [INT]
Metric [NUMERIC(5:2)]
RunDate [DATETIME)]

[18] TransitionsCrosstab

Note that this dataset has a companion dataset called EventCrossTab.

MetricTable [VARCHAR(64)]
EventCode [VARCHAR(64)]
None [INT]
BTH [INT]
DTH [INT]
ENT [INT]
ENU [INT]
EXT [INT]

IMG [INT]
OBE [INT]
OBS [INT]
OMG [INT]
RunDate [DATETIME)]

04_Core_ETL_to_Clean_6.1_Dataset Step

Overview

Clean the raw micro data set to obtain final micro data set

Purpose

Applies cleaning procedures to correct some inconsistencies identified in the quality assessment business process (03 Core ETL Raw 6.1 Dataset Quality Metrics) This business process does not clean all the errors identified, those requiring the attention of the ALPHA member centre are compiled in preparation for sending to the member centre

Algorithm

[step: 4.1] Check if first event is a legal first: Check if the first event to be ever recorded for each individual is enumeration, birth or external-immigration
This algorithm step references the following study concepts: [birth](#) [migration](#)

[step: 4.2] Change internal-immigration (ENT) to external-immigration (IMG) for first events: If first event is an internal-immigration change it to an external immigration
This algorithm step references the following study concepts: [migration](#) [internal-immigration](#)

[step: 4.3] Classify all other first events as illegal: Classify all first events other than enumeration, birth or external-immigration as illegal first events
This algorithm step references the following study concepts: [birth](#) [migration](#) [external-immigration](#)

[step: 4.4] Check if its a legal first event: Check if the marked as first event is a birth, an enumeration or an immigration from outside DSA
This algorithm step references the following study concepts: [birth](#) [migration](#)

[step: 4.5] Drop individuals with illegal start events:
This algorithm step references the following study concepts:

[step: 4.6] Check if last event is a legal last: Check if last events are external-outmigration, death, present in study site
This algorithm step references the following study concepts: [death](#) [migration](#)

[step: 4.7] Change internal-outmigration (EXT) to external-outmigration (OMG) for last event: If last event is an internal-outmigration change it to an external outmigration
This algorithm step references the following study concepts: [migration](#) [internal-outmigration](#)

[step: 4.8] Classify all other last events as illegal: Classify all last events other than external-outmigration,

death, present in study site as illegal last events
This algorithm step references the following study concepts: [death](#) [migration](#)

[step: 4.9] Drop individuals with illegal end events:
This algorithm step references the following study concepts:

[step: 4.10] Identify consecutive events and their dates: Identify current and next event and their dates
This algorithm step references the following study concepts:

[step: 4.11] Identify the event following each birth event: Check if a birth event is followed by a birth, an enumeration, external-immigration or internal-immigration
This algorithm step references the following study concepts: [birth](#) [migration](#) [external-immigration](#)

[step: 4.12] Identify the event following each death event: Check if a death event is followed by an event other than a NULL
This algorithm step references the following study concepts: [death](#)

[step: 4.13] Compile event pairs violating consistency matrix transitions: Review all other transitions in the data and record violations of consistency matrix
This algorithm step references the following study concepts:

[step: 4.14] Drop individuals with illegal transitions:
This algorithm step references the following study concepts:

[step: 4.15] Drop individuals with unknown sex or DOB:
This algorithm step references the following study concepts:

Data Sources for the 04_Core_ETL_to_Clean_6.1_Dataset Step

[1] Raw_61_Event_Format

recnr [INT]
study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

Data Targets for the 04_Core_ETL_to_Clean_6.1_Dataset Step

[1] Cleaned_61_Event_Format

recnr [INT]
study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [BIGINT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

[2] DeletedIndividualEvents

idno [VARCHAR(32)]
sex [INT]
dob [DATETIME]
eventnr [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
obs_date [DATETIME]
hhold_id [VARCHAR(32)]
Cause [VARCHAR(128)]

[3] IllegalEndEvents

recnr [INT]
study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
EventCount [INT]
event [INT]
EventCode [CHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]

obs_round [VARCHAR(2)]

[4] IllegalStartEvents

recnr [INT]
study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [CHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

[5] MicroDataCleanedSex

recnr [INT]
study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [BIGINT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

[6] MicroDataEndCleaned

recnr [INT]
study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]

event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
"DROP" [VARCHAR(32)]

[7] MicroDataStartCleaned

recnr [INT]
study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
"DROP" [VARCHAR(32)]

[8] MicroDataTransitionsCleaned

recnr [INT]
study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
EventCount [INT]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

[9] tmpMicroData

recnr [INT]

study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [CHAR(3)]
EventCount [INT]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

05_Core_ETL_Clean_6.1_Dataset_Quality_Metrics Step

Overview

Calculate Data Quality Metrics

Purpose

Reruns data quality metrics first executed in Step 3 after events are cleaned in Step 4

Algorithm

[step: 5.1] **Compile a list of Quality Metrics:** Compile a list of quality metrics relevant to the data specification
This algorithm step references the following study concepts:

[step: 5.2] **Create events consistency matrix:** Create events consistency matrix showing the logical ordering of event sequences
This algorithm step references the following study concepts:

[step: 5.3] **Compile residency starting events:** Identify in the data events that start a residency episode (birth, external-immigration, enumeration, becoming eligible for a study, found after being lost to follow-up)
This algorithm step references the following study concepts: residency birth migration

[step: 5.4] **Compile residency ending events:** Identify in the data events that end a residency episode (external outmigration, death, became ineligible for study, lost to follow-up, internal-outmigration, present in the study (right censored))
This algorithm step references the following study concepts: residency death migration

[step: 5.5] **Compile legal and illegal start events:** Review the identified start events and distinguish between legal and illegal ones
This algorithm step references the following study concepts:

[step: 5.6] Compile legal and illegal end events: Review the identified end events and distinguish between legal and illegal ones
This algorithm step references the following study concepts:

[step: 5.7] Compile legal and illegal transitions: Review all transitions between two events and distinguish between legal and illegal ones
This algorithm step references the following study concepts:

[step: 5.8] Compile illegal, missing or unknown sex: Compile illegal, missing or unknown sex
This algorithm step references the following study concepts:

[step: 5.9] Compile illegal, missing or runknown dob: Compile illegal, missing or unknown dob
This algorithm step references the following study concepts:

[step: 5.10] Compile quality metrics: Calculate numbers of legal and illegal start events, end events, event transitions, sex values, out of range DOBs and missing sex and DOBs
This algorithm step references the following study concepts:

Data Sources for the 05_Core_ETL_Clean_6.1_Dataset_Quality_Metrics Step

[1] MicroDataCleanedSex

recnr [INT]
study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [BIGINT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

Data Targets for the 05_Core_ETL_Clean_6.1_Dataset_Quality_Metrics Step

[1] DoBQualityMetrics

CentreId [VARCHAR(64)]
MetricTable [VARCHAR(64)]
QMetric [VARCHAR(64)]
Illegal [INT]

Legal [INT]
Total [INT]
Metric [NUMERIC(5:2)]
RunDate [DATETIME)]

[2] DoBValues

MetricTable [VARCHAR(128)]
DoBStatus [VARCHAR(25)]
Individuals [INT]
RunDate [DATETIME]

[3] EndingEvents

MetricTable [VARCHAR(128)]
EventCode [VARCHAR(3)]
Individuals [INT]
RunDate [DATETIME]

[4] EndingQualityMetrics

CentreId [VARCHAR(64)]
MetricTable [VARCHAR(64)]
QMetric [VARCHAR(64)]
Illegal [INT]
Legal [INT]
Total [INT]
Metric [NUMERIC(5:2)]
RunDate [DATETIME)]

[5] IllegalTransitions

MetricTable [VARCHAR(128)]
EventCode [VARCHAR(3)]
NextEventCode [VARCHAR(3)]
numRecs [INT]
maxDate [DATETIME]
RunDate [DATETIME]

[6] SexValueQualityMetrics

CentreId [VARCHAR(64)]
MetricTable [VARCHAR(64)]
QMetric [VARCHAR(64)]
Illegal [INT]
Legal [INT]
Total [INT]
Metric [NUMERIC(5:2)]
RunDate [DATETIME)]

[7] SexValues

MetricTable [VARCHAR(128)]
Sex [VARCHAR(3)]
Individuals [INT]
RunDate [DATETIME]

[8] StartQualityMetrics

CentreId [VARCHAR(64)]
MetricTable [VARCHAR(64)]
QMetric [VARCHAR(64)]
Illegal [INT]
Legal [INT]
Total [INT]
Metric [NUMERIC(5:2)]
RunDate [DATETIME)]

[9] StartingEvents

MetricTable [VARCHAR(128)]
EventCode [VARCHAR(3)]
Individuals [INT]
RunDate [DATETIME]

[10] TransitionQualityMetrics

CentreId [VARCHAR(64)]
MetricTable [VARCHAR(64)]
QMetric [VARCHAR(64)]
Illegal [INT]
Legal [INT]
Total [INT]
Metric [NUMERIC(5:2)]
RunDate [DATETIME)]

[11] TransitionsCrosstab

MetricTable [VARCHAR(64)]
EventCode [VARCHAR(64)]
None [INT]
BTH [INT]
DTH [INT]
ENT [INT]
ENU [INT]
EXT [INT]
IMG [INT]
OBE [INT]
OBS [INT]

OMG [INT]
RunDate [DATETIME)]

06_CORE_ETL_to_Anonymise_Dataset Step

Overview

Purpose

Randomises the individual and household identifiers to anonymise the data

Algorithm

[step: 6.1] Bring together original unique-identifiers and anonymised IDs: Bring together original and anonymised IDs in the cleaned spec 6.1>
This algorithm step references the following study concepts:

[step: 6.2] Remove original IDs from the cleaned spec 6.1: Create cleaned spec 6.1 with only anonymised IDs
This algorithm step references the following study concepts:

[step: 6.3] Preserve ID Mappings: Preserve an internal mapping of original IDs to the anonymised IDs
This algorithm step references the following study concepts:

Data Sources for the 06_CORE_ETL_to_Anonymise_Dataset Step

[1] Cleaned_61_Event_Format

recnr [INT]
study_name [VARCHAR(15)]
idno [VARCHAR(32)]
hhold_id [VARCHAR(32)]
hhold_id_extra [VARCHAR(32)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [BIGINT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

[2] HH_ID_Anonymise_Map

study_name [VARCHAR(15)]
hhold_id [VARCHAR(100)]
anonymised_hhold_id [INT]

[3] Indv_ID_Anonymise_Map

study_name [VARCHAR(15)]
idno [VARCHAR(100)]
anonymised_idno [INT]

Data Targets for the 06_CORE_ETL_to_Anonymise_Dataset Step

[1] Anonymised_61_Event_Format

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [BIGINT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

07_CORE_Consolidate_Start_and_End_Events Step

Overview

Purpose

Quantifies proportion of records that are duplicates in terms of unique-identifier, event and event date, cleans the duplicates and drops individuals with single events

Algorithm

[step: 7.1] Dates of first events: Identify the date for each individual's first event to be ever recorded
This algorithm step references the following study concepts:

[step: 7.2] Dates of last events: Identify the date for each individual's last event to be ever recorded
This algorithm step references the following study concepts:

[step: 7.3] Identify and compile duplicates: Identify duplicates in terms of ID, event and event date
This algorithm step references the following study concepts:

[step: 7.4] Remove duplicates: Remove duplicate record in terms of ID, event and event date identified in 7.3
This algorithm step references the following study concepts:

[step: 7.5] Identify individuals with single events: Identify individuals with total events amounting to 1 or less
This algorithm step references the following study concepts: individuals

[step: 7.6] Drop individuals with single events: Drop individuals with total events amounting to 1 or less
This algorithm step references the following study concepts: individuals

[step: 7.7] Adjust events numbering: Adjust events numbering to account for dropped events
This algorithm step references the following study concepts:

[step: 7.8] Quantify proportion of records that are duplicates: Quantify percentage of events that are duplicates as a quality metric
This algorithm step references the following study concepts:

Data Sources for the 07_CORE_Consolidate_Start_and_End_Events Step

[1] Anonymised_61_Event_Format

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [BIGINT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

Data Targets for the 07_CORE_Consolidate_Start_and_End_Events Step

[1] EndEvents

EventCode [VARCHAR(64)]
EventYear [INT]
EventMonth [INT]
Count [INT)]

[2] QualityMetrics

CentreId [Unknown]
MetricTable [Unknown]
QMetric [Unknown]
Illegal [Unknown]
Legal [Unknown]
Total [Unknown]
Metric [Unknown]
RunDate [Unknown]

[3] S01

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [BIGINT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]

[4] S01_Dropped

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]

eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

[5] S01_Duplicates

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [INT]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

[6] S01_Temp2

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [BIGINT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]

[7] S01_Temp

recnr [INT]
study_name [VARCHAR(15)]

idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

[8] StartEvents

EventCode [VARCHAR(64)]
EventYear [INT]
EventMonth [INT]
Count [INT])

[9] Summary

study_name [VARCHAR(25)]
Records [INT]
LastDate [DATETIME]
FirstDate [DATETIME])

08_CORE_Verify_Temporal_Integrity Step

Overview

Verify Temporal Integrity

Purpose

Assesses the ordering, in time, of dates for consecutive/ successive events, compiles those with illogical timing, quantifies their proportion and drops individuals with wrongly timed successive events

Algorithm

[step: 8.1] Identify an individual's dates of consecutive events: For each individual in the data, determine the dates of consecutive events event date and next event date
This algorithm step references the following study concepts:

[step: 8.2] Use a future date if next event date is NULL: Replace NULL next event dates with a date in the future

This algorithm step references the following study concepts:

[step: 8.3] Assess temporal integrity of consecutive events dates: Check if event date is less than next event date and record violations
This algorithm step references the following study concepts:

[step: 8.4] Quantify proportion event dates violating temporal integrity: Quantify proportion event dates violating temporal integrity as a quality metric
This algorithm step references the following study concepts:

[step: 8.5] Drop individuals with temporal integrity violations: Drop individuals with temporal integrity violations
This algorithm step references the following study concepts: individuals

Data Sources for the 08_CORE_Verify_Temporal_Integrity Step

[1] S01

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [BIGINT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]

Data Targets for the 08_CORE_Verify_Temporal_Integrity Step

[1] QualityMetrics

CentreId [Unknown]
MetricTable [Unknown]
QMetric [Unknown]
Illegal [Unknown]
Legal [Unknown]
Total [Unknown]
Metric [Unknown]

RunDate [Unknown]

[2] S02

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]

[3] S02_DeletedTemporalViolations

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]

[4] S02_Violations

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]

residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]
NextDate [DATETIME]

09_CORE_Update_Event_Timing Step

Overview

Update Event timing

Purpose

Assesses and corrects migration event sequences. A movement out of the study area is defined as an external-outmigration (OMG) if the time between the external-outmigration and the subsequent external-immigration (IMG) is above a defined period of time (threshold) - e.g. six months.

Algorithm

[step: 9.1] Identify an individual's consecutive events and their dates: For each individual in the data, determine previous and current event and their corresponding dates
This algorithm step references the following study concepts:

[step: 9.2] Compile OMG-IMG pairs: From consecutive event pairs identified in 9.1, compile pairs where an external-outmigration is followed by an external-immigration
This algorithm step references the following study concepts: [migration](#) [external-outmigration](#)

[step: 9.3] Check if date difference between OMG and IMG is below threshold: Identify and compile OMG-IMG pairs with date differences below a recommended threshold
This algorithm step references the following study concepts:

[step: 9.4] Change IMG to ENT for OMG-IMG date difference below threshold: Change external-immigration to internal-immigration if the OMG-IMG date differences is below a recommended threshold
This algorithm step references the following study concepts: [migration](#) [external-immigration internal-immigration](#)

[step: 9.5] Quantify proportion OMG-IMG pairs with date differences below threshold: Quantify proportion OMG-IMG pairs with date differences below a recommended threshold as a quality metric
This algorithm step references the following study concepts:

[step: 9.6] Change OMG to EXT for OMG-IMG pair date difference below threshold: Change external-outmigration to internal-outmigration if the OMG-IMG date differences is below a recommended threshold

This algorithm step references the following study concepts: [migration](#) [internal-outmigration external-outmigration](#)

[step: 9.7] Compile EXT-ENT pairs: From consecutive event pairs identified in 9.1, compile pairs where an internal-outmigration is followed by an internal-immigration
This algorithm step references the following study concepts: [migration](#) [internal-outmigration](#)

[step: 9.8] Check if date difference between EXT and ENT is above threshold: Identify and compile EXT-ENT pairs with date differences above a recommended threshold
This algorithm step references the following study concepts:

[step: 9.9] Change ENT to IMG for EXT-ENT pair date difference above threshold: Change internal-immigration to external-immigration if the EXT-ENT pair date differences is above a recommended threshold
This algorithm step references the following study concepts: [migration](#) [external-immigration internal-immigration](#)

[step: 9.10] Quantify proportion EXT-ENT pairs with date differences above threshold: Quantify proportion EXT-ENT pairs with date differences above a recommended threshold as a quality metric
This algorithm step references the following study concepts:

[step: 9.11] Change EXT to OMG for EXT-ENT pair date difference above threshold: Change internal-outmigration to external-outmigration if the EXT-ENT pair date difference is above a recommended threshold
This algorithm step references the following study concepts: [migration](#) [internal-outmigration external-outmigration](#)

[step: 9.12] Assign times to event dates: Add recommended times to the event dates dependent on the event type to maintain temporal integrity and logical event sequences
This algorithm step references the following study concepts:

Data Sources for the 09_CORE_Update_Event_Timing Step

[1] S02

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]

Data Targets for the 09_CORE_Update_Event_Timing Step

[1] QualityMetrics

CentreId [Unknown]
MetricTable [Unknown]
QMetric [Unknown]
Illegal [Unknown]
Legal [Unknown]
Total [Unknown]
Metric [Unknown]
RunDate [Unknown]

[2] QualityMetrics

CentreId [Unknown]
MetricTable [Unknown]
QMetric [Unknown]
Illegal [Unknown]
Legal [Unknown]
Total [Unknown]
Metric [Unknown]
RunDate [Unknown]

[3] S03

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]

[4] S03_Recode_EXT

recnr [INT]

study_name [VARCHAR(15)]
idno [INT]
EventCount [INT]
eventnr [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
PrevEvent [VARCHAR(3)]
PrevDate [DATETIME]
PrevRecNr [INT]
DateDif [INT]

[5] S03_Recode_OMG

study_name [VARCHAR(15)]
recnr [INT]
idno [INT]
eventnr [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
PrevEvent [VARCHAR(3)]
PrevDate [DATETIME]
PrevRecNr [INT]
DateDif [INT]

[6] S03_Temp2

recnr [Unknown]
study_name [Unknown]
idno [Unknown]
hhold_id [Unknown]
hhold_id_extra [Unknown]
sex [Unknown]
dob [Unknown]
residence [Unknown]
eventnr [Unknown]
event [Unknown]
EventCode [Unknown]
event_date [Unknown]
type_of_date [Unknown]
obs_date [Unknown]
obs_round [Unknown]
EventCount [Unknown]

[7] S03_Temp2

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]

hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]

[8] S03_Temp

recnr [Unknown]
study_name [Unknown]
idno [Unknown]
hhold_id [Unknown]
hhold_id_extra [Unknown]
sex [Unknown]
dob [Unknown]
residence [Unknown]
eventnr [Unknown]
event [Unknown]
EventCode [Unknown]
event_date [Unknown]
type_of_date [Unknown]
obs_date [Unknown]
obs_round [Unknown]
EventCount [Unknown]

[9] S03_Temp

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]

10_CORE_Produce_Final_Core_MicroData_Files Step

Overview

Purpose

Produces the final dataset

Algorithm

[step: 10.1] Output events format data: Output data in the form of a calendar of events
This algorithm step references the following study concepts:

[step: 10.2] Identify residency episode start events: Identify events that start a residency episode in the events format data
This algorithm step references the following study concepts: residency

[step: 10.3] Identify residency episode end events: Identify events that end a residency episode in the events format data
This algorithm step references the following study concepts: residency

[step: 10.4] Bring together episode start and end events for each episode: Bring together in one record corresponding events that start and end each of the residency episodes in the data
This algorithm step references the following study concepts: residency episodes

[step: 10.5] Compile unmatching start and end events: Compile start and end events in one record but belonging to different individuals
This algorithm step references the following study concepts: individuals

[step: 10.6] Compile illegal start events: Compile illegal start events
This algorithm step references the following study concepts:

[step: 10.7] Compile illegal end events: Compile illegal end events
This algorithm step references the following study concepts:

[step: 10.8] Output episodes format data: Output data in the form of residency episodes
This algorithm step references the following study concepts: residency episodes

[step: 10.9] Generate MD5 checksum for events format data: Generate and store an MD5 fingerprint for the events format data
This algorithm step references the following study concepts:

[step: 10.10] Generate MD5 checksum for episodes format data: Generate and store an MD5 fingerprint for the residency episodes format data
This algorithm step references the following study concepts: residency episodes

Data Sources for the 10_CORE_Produce_Final_Core_MicroData_Files Step

[1] Final_61_Episode_Format

FileContent [VARCHAR(-1)]
FileSize [INT]

[2] Final_61_Event_Format

FileContent [VARCHAR(-1)]
FileSize [INT]

[3] S03

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]

Data Targets for the 10_CORE_Produce_Final_Core_MicroData_Files Step

[1] Final_61_Episode_Format

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
start_hhold_id [INT]
start_hhold_id_extra [DECIMAL(32:0)]
start_event [INT]

start_event_date [DATETIME]
start_type_of_date [INT]
start_obs_date [DATETIME]
start_obs_round [VARCHAR(2)]
end_hhold_id [INT]
end_hhold_id_extra [DECIMAL(32:0)]
end_event [INT]
end_event_date [DATETIME]
end_type_of_date [INT]
end_obs_date [DATETIME]
end_obs_round [VARCHAR(2)]

[2] Final_61_Event_Format

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [BIGINT]
event_count [BIGINT]
event [INT]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

[3] MD5-DataFingerPrint

11_CORE_Prepare_Data_Quality_Feedback Step

Overview

Prepare Site Feedback

Purpose

Compiles data quality assessment report to be shared with the member centre

Algorithm

[step: 11.1] Output duplicate records: Output duplicate records in terms of unique-identifier, event and event date

This algorithm step references the following study concepts: [unique-identifier](#)

[step: 11.2] Output core violations: Output core violations (illegal start, end events, illegal transitions, missing sex and missing DOBs
This algorithm step references the following study concepts:

[step: 11.3] Output dropped events: Output dropped events
This algorithm step references the following study concepts:

[step: 11.4] Output temporal violations: Output records with temporal violations
This algorithm step references the following study concepts:

[step: 11.5] Output offending migration events: Output offending migration events
This algorithm step references the following study concepts: [migration](#)

[step: 11.6] Produce quality metrics report: Produce a summary report of the quality metrics providing statistics to give the magnitude of the errors
This algorithm step references the following study concepts:

Data Sources for the 11_CORE_Prepare_Data_Quality_Feedback Step

[1] QualityMetrics

CentreId [VARCHAR(15)]
MetricTable [VARCHAR(128)]
QMetric [VARCHAR(50)]
Illegal [INT]
Legal [INT]
Total [INT]
Metric [FLOAT(53)]
RunDate [DATETIME]

[2] S01_Dropped

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]

obs_round [VARCHAR(2)]

[3] S01_Duplicates

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [INT]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]

[4] S02_DeletedTemporalViolations

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]
eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]

[5] S02_Violations

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
hhold_id [INT]
hhold_id_extra [DECIMAL(32:0)]
sex [INT]
dob [DATETIME]
residence [VARCHAR(5)]

eventnr [INT]
event [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
type_of_date [INT]
obs_date [DATETIME]
obs_round [VARCHAR(2)]
EventCount [INT]
NextDate [DATETIME]

[6] S03_Recode_EXT

recnr [INT]
study_name [VARCHAR(15)]
idno [INT]
EventCount [INT]
eventnr [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
PrevEvent [VARCHAR(3)]
PrevDate [DATETIME]
PrevRecNr [INT]
DateDif [INT]

[7] S03_Recode_OMG

study_name [VARCHAR(15)]
recnr [INT]
idno [INT]
eventnr [INT]
EventCode [VARCHAR(3)]
event_date [DATETIME]
PrevEvent [VARCHAR(3)]
PrevDate [DATETIME]
PrevRecNr [INT]
DateDif [INT]

Data Targets for the 11_CORE_Prepare_Data_Quality_Feedback Step

[1] DeletedTemporalViolations

No description is available.

[2] DroppedEvents

No description is available.

[3] DuplicateEvents

No description is available.

[4] QualityMetrics

CentreId [VARCHAR(64)]
MetricTable [VARCHAR(64)]
QMetric [VARCHAR(64)]
Illegal [INT]
Legal [INT]
Total [INT]
Metric [NUMERIC(5:2)]
RunDate [DATETIME)]

[5] Recode_EXT

No description is available.

[6] S03_Recode_OMG

No description is available.

[7] Violationss

No description is available.

Study Concepts

[birth]
Pregnancies and their outcomes are recorded for all registered women in the HDSS regardless of the location at which the outcomes occur. An HDSS aims to record all outcomes including miscarriages, induced abortions, stillbirths and live births. All pregnancy outcomes are needed for fertility estimation All live births are registered as individual members of the HDSS

[death]
All deaths to registered members of the HDSS are recorded regardless of the place of occurrence of the death

[demographic-surveillance-area]
The demographic surveillance area (DSA) is an area with clearly and fairly permanent delineated boundaries, preferably recognizable on the ground (for example, rivers, roads, and clearly demarcated administrative boundaries).

[eligibility]
Every HDSS defines a population under surveillance. Individuals have places of residence and attachments to social units such as households. First the residential units, the social groups and the individuals need to be identified. A set of inclusion criteria to distinguish eligible from ineligible entities needs to be defined.

Residential and social units are eligible if they are situated in the DSA. Individuals are eligible if they are resident at eligible residential units or if they belong to eligible social units

[episodes]
Episodes are the logical complement to events. They are meaningful and identifiable segments of time started and ended by events.

[external-immigration]
This is when an individual or social unit migrates into the DSA from a location outside the DSA

[external-migration]
This involves residence changes between a residential unit in a DSA and one that's outside it.

[external-outmigration]
When an individual relocates from a residential unit in the DSA to a place outside the DSA

[hdss-core-events]
Three core events alter the size of the registered population in an HDSS. These are births, deaths and migration in consistency with the fundamental demographic equation: $Pt1 = Pt0 + Bt0, t1-Dt0, t1+It0, t1-Ot0, t1$. Pt1 –Population at time t1, Pt0 – Population at time t2, Bto, t1 – Births between time t0 and time t1, Dto, t1 – Deaths between time t0 and time t1, Ito, t1 – Immigrations between time t0 and time t1, Oto, t1 – Outmigration between time t0 and time t1

[health-and-demographic-surveillance-system]
A Health and Demographic Surveillance System(HDSS) is defined as a set of field and computing operations applied within a clearly demarcated geographic area to handle the longitudinal follow-up of well-defined entities or primary subjects (individuals, social units (e.g. households), and residential units (physical locations)) and their related demographic, socio-economic and health outcomes.

[individuals]
Individuals are the subject of primary interest in an HDSS. They are the people who are residents of a residential unit or members of a social unit

[initial-census]
Data collection done within the DSA to register and define the target population. During the initial census an extensible system of unique identifiers is assigned to the primary entities of the HDSS. The initial census also works as a basis for the development of the HDSS database

[internal-immigration]
Relocating into a residential unit within the DSA after exiting another also within the DSA

[internal-migration]
This involves residence changes from one residential unit to another in the same DSA. While internal migration does not alter the size of the registered population, it is essential to record internal migration to avoid double counting of individuals and to also correctly apportion exposure to social and physical environment. Decisions have to be made for deciding if migration has occurred based on the duration since the event..

[internal-outmigration]
Exiting a residential unit in the DSA in order to join another also within the DSA

[longitudinality]

The longitudinal measurement of demographic and health dynamics in the registered population by constantly updating a set of prescribed attributes for the primary subjects during rounds of follow up visits to the registered residents in the DSA

[membership]
The state of belonging to a social group. Membership associates individuals with social groups such as households

[migration]
The change of residence by a registered individual or social group (e.g., a household). There are two types of migration that occur among the registered population. These are internal and external migration.

[primary-hdss-entities]
Residential units, Social units and individuals comprise the main entities of interest in an HDSS.

[residency]
The state of being physically present in a given residential unit for a defined threshold of time. Residency is an essential pre-requisite for enumeration of individuals at risk of demographic events or disease exposure. Residency associates individuals with residential units

[residential-units]
Residential units are the places in the DSA where individuals live

[social-units]
these are groups to which individuals in the DSA belong

[unique-identifier]
A number or identification code used to uniquely identify the primary entities of an HDSS. Unique identifiers are assigned at baseline and they should be extensible to accommodate the addition of future primary entities