TasP

Antiretroviral Treatment as Prevention • ANRS 12249
Ukuphila kwami, ukuphila kwethu (my health for our health)

ANRS 12249 TasP trial

Definitions Indicators

version 3.0, 11th January 2017

Writing committee (alphabetical order): Collins, France, François, Frank, Joanna, Joseph, Marie-Louise, Rodolphe

Reviewed by Bruno, Laura, Kamal, Mélanie, Richard, Rosemary, Sylvie

Maintenance: Joseph

TasP

Version history

0.1	2014-04-17	First draft
0.8	2014-07-03	Version used for the DSMB held in July 2014
1.0	2014-09-24	Reorganisation of the document by Joseph Larmarange
1.1	2014-10-09	New section *HIV prevalence and incidence (DBS)* added New section *Retention in TasP clinics* added Definition of *Baseline TasP clinic visit* added (12.21) Correspondence between months and days added in *Preamble* Definition of *Assets Index* added
1.2	2014-10-22	New section *Estimated Cascade of HIV care*
1.3	2014-12-15	Section *Estimated Cascade of HIV care* updated New section *Computing ages* New section *Observed HIV status and observed HIV prevalence* New population *Ever diagnosed HIV positive* (10.7) New population *Individuals ever been in care* (12.22) Definitions regarding being in care have been updated to take into account data available from iDART Definitions of being on ART have been updated to be consistent with definitions of being in care
1.4	2015-01-08	Fixing several typos Precision added for *Seen in TasP clinics* (12.1) Precisions added for section 2 *Calendar rounds*.
1.5	2015-01-21	New section *Employment and active population*
2.0	2015-04-23	New concept *Data Collection Cycle* (cf. section 2) Definitions 6.2, 7.3, 7.4, 10.3 and 10.4 have been updated using data collection cycles. Section 20 and definitions 12.23, 15.9, 15.10 and 15.11 have been updated to take into account end of data collection.
2.1	2016-04-18	Definition of contact and rapid test updated taking into account the new trial status 50 (“rapid test done, but result unknown”)
3.0	2017-01-11	iDART is no longer used for defining care and ART. Introduction of ACCDB in the definition of care and ART. Also impacts the definition of *Newly diagnosed* Several definitions have been removed (in particular regarding ART coverage and treatment intensity) to be further explored by scientists.

The cascade of care section has also been removed from that document (and will be developed in a separate document).

Table of content

Preamble

This document aims at describing and **defining** the different concepts within the TasP ANRS 12249 trial, to make sure we use a common terminology/vocabulary when planning and conducting analyses.

This document also lists and defines a first set of key **populations** and **indicators**. Some of these indicators will become fixed variables generated automatically within the trial datasets. When appropriate, corresponding variables in analytical datasets are indicated.

Note that this is a **working document**, which will evolve and grow as TasP analyses progress. Therefore, some sections are more developed than others due to analysis conducted so far, while others still have to be developed more precisely.

Also note that this document is NOT an overall analysis plan of TasP data. It will be up to the different investigators leading specific analyses to define their study sub-population, the observation period (ever, per calendar round; exposure to x months of referral/care; right censoring taking into account trial/clinic exits etc.).

Technical notes (referring to the analytical datasets) are presented with a `specific font`.

Correspondence between months and days:

Months	Days	Months	Days	Months	Days	Months	Days
1	30	7	213	13	395	19	578
2	61	8	243	14	426	20	608
3	**91**	**9**	**274**	**15**	**456**	**21**	**639**
4	122	10	304	16	487	22	669
5	152	11	335	17	517	23	700
6	**183**	**12**	**365**	**18**	**548**	**24**	**730**

Months	Days	Months	Days	Months	Days	Months	Days
25	760	31	943	37	1125	43	1308
26	791	32	973	38	1156	44	1338
27	**821**	**33**	**1004**	**39**	**1186**	**45**	**1369**
28	852	34	1034	40	1217	46	1399
29	882	35	1065	41	1247	47	1430
30	**913**	**36**	**1095**	**42**	**1278**	**48**	**1460**

About numbering

Rules for sections and definitions numbering are formalized in version 1.0. The numbering should remain consistent in all versions higher than 1.0.

Main sections are numbered with one figure. Sub-sections should not be numbered. Sections have been numbered continuously in version 1.0. Sections order may evolve in later versions. However, sections should not be re-numbered (sections numbers should remain consistent across all versions higher than 1.0).

Definitions are numbered with two figures, i.e. a number like $x.y$ where x is the section number and y the sub-number of that definition within section x. As for sections, the definition numbers should remain consistent across all versions of that document higher than 1.0, even if the order of the text changes over time.

Some **definitions** refer to a **population** (a group of individuals with a common characteristics) or to an **indicator** (calculated from several populations, e.g. a proportion). However, to avoid any confusion all definitions should be unique (i.e. a population and an indicator cannot have the same definition number).

1. Cluster groups

The TasP trial has been implemented in several waves. The *cluster group* identifies when clusters have been opened.

Definition 1.1: *cluster groups*

» Group 1 clusters (n=4) are clusters opened in 2012
» Group 2 clusters (n=6) are clusters opened in 2013
» Group 3 clusters (n=12) are clusters opened in 2014

Cluster Id	Cluster Name	Cluster Group	Arm
1	Madwaleni	1	I
2	Shunqa	1	I
3	Embongolweni	1	C
4	Ntondweni	1	C
5	kwaGxaba	2	I
6	Makhambane	2	C
7	Cakula	3	C
8	kwaSqumbe	2	C
9	Bhekamandla	3	C
10	Egedeni	2	I
11	Mchakwini	2	C
12	Mfekayi	3	C
13	Makhwela	2	I
14	Esiphahleni	3	I
15	Nkundusi	3	C
16	Mazala	3	I
17	Chwebeni	3	C
18	Maqanda	3	I
19	Danyini	3	I
20	Esiqiwini	3	I
21	Odakaneni	3	C
22	Ingodini	3	I

TasP

2. Calendar rounds and data collection cycles

Definition 2.1: *calendar rounds*
A calendar round (CR) refers to the time period it took to implement one home-based HIV testing round in a cluster. The rank of the calendar round (i.e. first, second, etc.) is calculated per cluster. Calendar round are therefore in a **cluster perspective**.

The term *calendar rounds*, although useful for internal use, is maybe not appropriate for publications. An alternative expression could be ***home-based survey round***.

As all clusters didn't open at the same moment of time, calendar rounds are not synchronous between all clusters. An individual could migrate within the TasP and, therefore, could have been surveyed twice for a first calendar round, in two different clusters.

To avoid such problem, we could refer to data collection cycles.

Definition 2.2: *data collection cycles*
A data collection cycle (CC) refers to a time period where a cycle of home-based HIV testing was implemented. It's computed within a **global survey perspective**.

The table below gives the correspondence between data collection cycles and calendar rounds, per cluster group.

	Group 1 clusters	Group 2 clusters	Group 3 clusters
Collection cycle 1	CR 1		
Collection cycle 2	CR 2	CR 1	
Collection cycle 3	CR 3	CR 2	
Collection cycle 4	CR 4	CR 3	CR 1
Collection cycle 5	CR 5	CR 4	CR 2
Collection cycle 6	CR 6	CR 5	CR 3
Collection cycle 7	CR 7	CR 6	CR 4

Several datasets refer to a *CensusRoundId*. There is one *CensusRoundId* per cluster and per home-based survey rounds. The **Census Rounds** dataset provide, for each *CensusRoundId* the corresponding *ClusterId*, *CalendarRound* and *DataCollectionCycle*.

3. Logical rounds

SECTION KEPT FOR HISTORIC PURPOSE. THIS CONCEPT IS NOT USED ANYMORE IN ANALYTICAL DATASETS.

A logical round is an operational concept used in the early stages of the trial to organise the collection of trial forms and store trial status in the database. While calendar rounds are determined per cluster, logical rounds are determined per individual.

> **Definition 3.1:** *logical rounds*
> A logical round (LR) usually refers to the fact that a participant has experienced the trial and completed an individual questionnaire (IQ), but this is not systematic throughout the trial (there are still inconsistencies between LR and IQ).

Note: in theory, an individual remains in first logical round until he/she completes his/her first individual questionnaire. Then, he/she enters his/her second logical round until he/she completes his/her second individual questionnaire, and so on. At early stages in the trial, a logical round was considered as being completed if the individual provided a DBS or accepted a rapid test, even if he/she refused to complete the questionnaire. Rules changed in January 2013.

Individual *trial statuses* are collected per logical round. Each time a fieldworker visits a household, the trial status of an individual is updated by the fieldworker if needed. The first entered trial status and the last modified status per logical round are stored in the database, i.e. intermediate trial statuses for a same logical round are lost. In some occasions, the trial status of an individual could be modified in a TasP clinic by a nurse or an ART counsellor, in particular in case of exit or rapid test done in clinic. The trial status is composite indicator indicating if an individual has been contacted and, in case of contact, if a rapid test was done, the result of rapid test or self-reported HIV status.

4. Homesteads

THIS SECTION NEEDS ADDITIONAL WORK.

Homesteads refer to bundle structures. They could be:

- usable and occupied
- usable but unoccupied
- broken down
- destroyed or
- under construction.

This construction state is an operational status, used to follow field workers activities. At each survey round, the construction state is reset to missing. Fieldworkers are supposed to visit each homestead during each survey round and to update this construction state. However, as this state is reset at each survey round, it cannot be used for analytical purposes for phase 1 data, as we have not keept the history of this state.

The concept of *homestead entry* refers to the number of homesteads usable, occupied and visited that accepted entry of the fieldworkers and be calculated overall, for the whole trial duration (homestead ever entered), or for each calendar round. The corresponding indicator would be number of homesteads entered / number of homestead visited.

WARNING: homestead and household level definitions and indicators (such as homestead contact, homestead entry, household contact, household participation) are not very precise at this stage and need more work.

5. Households

THIS SECTION NEEDS ADDITIONAL WORK.

Usually there is one household per homestead but occasionally there could be more than one household in each homestead

Household participation

The concept of *household participation* refers to the households agreeing to participate in the trial, i.e. the key informant (head of household or not) provides the list of household members (and responds to the household assets form). The corresponding indicator would be number of households registered / number of households contacted.

WARNING: we don't have yet an operational definition of contacted households and registered households. These definitions need more work.

Households' assets index

The asset index score is developed to proxy for household wealth. The asset index is built as the weighted sum of indicators or dummy variables for whether a household possessed certain assets. The category-weights are given by the normalized scores on the first factorial axis coming out of a multiple correspondence analysis (MCA). The higher the asset index score, the higher the implied socio-economic status of that household.

The assets score is computed for each HHI form completed by a household. Therefore, it could change over time.

Definition 5.1: *Household's assets index score*
Normalized score of house characteristics and assets owned by a household (see methodological notes). An higher means better living conditions.

The assets index score is provided in *AssetsScore* variable in **HHIs** dataset.

Methodological note:

- *HHIs whose most variables were missing have been excluded.*
- *Missing values ('missing', 'refused' and 'don't know' answers) have been imputed. If the household completed several questionnaires, missing values were replaced with the closest value for that variable. Otherwise, they were replaced with the modal value of that variable as observed in the same cluster during the same calendar round.*
- *Variables which have so-called rare categories, less than 5%, were excluded (HasTelephone; HasBedNet; HasMotorcycle; HasKombiLorryTractor).*

- » *DrinkWaterSource, ToiletType and MainCookingFuel have been recoded before computing MCA (see R code in appendix).*
- » *MCA was computed using only HHIs collected during calendar round 1 (to avoid over-representation of clusters and households surveyed several times). HHIs collected during calendar round 2 or higher were projected on the first axe according to the results of the MCA.*
- » *The assets score corresponds to the coordinates on the first axe. They have been normalized according to mean and standard deviation observed during the first round.*

Several methods are used in literature to create categories from the assets index score, the more common being to use quintiles or three groups (40%-40%-20%). According to the distribution of the index score (first two quintiles being similar), we decided to use the 40%-40%-20% approach.

Definition 5.2: *Household's assets index categories*
According to his household's assets index score for a specific HHI, an household could be classified as:

- » low socio-economic status
- » middle socio-economic status
- » high socio-economic status

Note: percentile cut-off values were calculated taking into account only HHIs completed during the first calendar round, to avoid over-representation of clusters and households surveyed several times.

The assets index score is provided in *AssetsScoreCat* variable in **HHIs** dataset.

Assets index score and categories should be interpreted with cautious as its interpretation is still under debate. A consensus leads to interprete it as more reflective of longer-run household wealth or living standards, failing to take into account of short-run or temporary interruptions or shocks on the household (Filmer & Pritchett 2001).

Households' composition

TO BE DEVELOPPED

6. Individual registration and eligibility

As per protocol, an individual is eligible for the trial if:

- 16 years or older;
- member of a household (as defined by key informant at the time of the household visit) and
- resident (defined as usually sleeping at least 4 days per week in this household).

For one individual, the eligibility status can change over time. For example, a participant in-migrating to the survey area at the beginning of the second calendar round, will have been non-eligible during the first calendar round and be eligible for the first time during the second calendar round.

During home-based visits, all identified eligible individuals are registered in the notebooks. However, individuals residing in households never contactable or in households who always refused to participate have not been registered. Therefore, the total number of registered individuals slightly underestimates the total number of the eligible resident population.

Population 6.1: *registered individuals*
Number of eligible individuals registered at any point during a home-based survey round.

Note: An individual is registered during any given calendar round if he/she is eligible. All individuals who have been eligible at least once are registered and all registered individuals have been eligible at least once.

Population 6.2: *eligible individuals per round X*
An individual is considered as eligible for a given home-based survey round if:

- he/she was registered during that round or in a previous round AND
- he/she was not exited for that specific round (see section 17).

Note: due to possible internal migration, this indicator is computed per data collection cycle.

```
Variables CC1Eligible, CC2Eligible, etc. in Individuals dataset could be
used to determine population 6.2
```

7. Individual contact

Population 7.1: *Ever contacted individuals*
Individuals who have ever been contacted (seen and talked to) by a fieldworker at any given calendar round.

Contacts are calculated from trial statuses. More precisely, a contact is defined as a trial status equal to 5 'refused all', 20 'rapid test refused, referred to clinic', 25 'no rapid test, known positive', 30 'no rapid test, known positive and on ART', 35 'rapid test done, positive result', 36 'rapid test done, indeterminate result', 37 'rapid test done, invalid result', 40 'rapid test done, negative result', 45 'rapid test refused, follow-up next survey-round' or 50 'rapid test done, but result unknown' (cf. variable *Contact* in **Trial Statuses** dataset).
Population 7.1 corresponds to individuals with *EverContacted==1* in **Individuals** dataset, i.e. to individuals with at least one trial status considered as a contact. The variable *NbContacts* indicates the total number of contacts for each registered individual.

Indicator 7.2: *Overall contact rate*
Number of individuals who have ever been contacted (seen and talked to) by fieldworker (population 7.1) / number of individuals (population 6.1)

Contact rates may be calculated for the whole trial duration (individuals ever contacted, indicator 7.2), or for each calendar round (ind. 7.4).

Population 7.3: *Individuals contacted per round*
Individuals who have ever been contacted during a given home-based survey round (data collection cycle).

Note: as in few occasions, some individuals have been contacted more than once in a given round, the computation of this population should be done as 'ever contacted' during that round. Due to the fact that some individual could migrate within the area, this indicator is computer per data collection cycle rather than calendar rounds.

Population 7.3 can be identified using variables *CC1EverContacted*, *CC2EverContacted*, *CC3EverContacted*, etc. in **Individuals** dataset.

Indicator 7.4: *Contact rate per round*
Number of individuals who have been contacted during a home-based survey round (population 7.3) / number of individuals eligible for that survey round (population 6.2).

8. Individual participation

There has been debate as to whether we should consider participation as having completed an IQ only, or as having completed any component of the trial: an IQ with or without providing a DBS with or without accepting a rapid test. And whether we should consider the fact that an IQ (or DBS) has been documented or the fact that a CZ1 (and/or CZ2) was signed and collected (there are some inconsistencies between consent forms collection/capture and IQ/DBS/RT collection).

Most cascade indicators are based on contacted individuals, and do not refer to the concept of participation. More specifically, in the case of a population trial as TasP, our targeted population remains the overall resident population (i.e. even included non-contacted individuals).

For a question of simplicity and to follow the latest definition of logical rounds, participation is defined according to completion of home-based individual questionnaires. Therefore, participants are the individuals for whom we have personal details. For certain analyses, an additional requirement of DBS and or of rapid test could be appropriate.

Population 8.1: *Participants*
Individuals having completed at least one individual question (IQ).

```
Participation can be computed using IQ1Completed, IQ2Completed and
IQ3Completed in Individuals dataset.
```

Indicator 8.2: *Overall participation rate*
Number of participants (population 8.1) / number of individuals ever contacted (population 7.1)

Participation rates could also be calculated per calendar round.

Population 8.3: *Participants at calendar round X*
Individuals having completed at least one individual question (IQ) during calendar round X.

Note: There are some cases where an individual completed more than one IQ during a given calendar round. We will consider participation as having "completed at least one IQ".

Indicator 8.4: *Participation rate at calendar round X*
Number of participants at calendar round X (population 8.3) / number of individuals ever contacted during calendar round X (population 7.3)

18. HIV prevalence and incidence (DBS)

DRAFT TO BE DISCUSSED

This section describes how to compute HIV prevalence and HIV incidence from DBS (Dry Blood Spots) results. In TasP, we collect many king of data regarding HIV status (DBSs, rapid HIV tests, self-reported status…). However, the protocol clearly indicates that HIV prevalence and HIV incidence should be computed only from DBSs collected during the home-based survey.

DBSs results are stored in variable *DBSResult* in **DBS Results** dataset.
This variable could be:
» Positive
» Negative
» Indeterminate (as reported by the lab) or
» Unknown, which means that a DBS was collected on the field but that no result was provided back by the lab. It could occurs if (i) the lab didn't have enough time to perform the test; (ii) the sample was lost; or (iii) the sample was not good enough to perform a test.

First DBS HIV prevalence

This indicator compute an HIV prevalence based on the *first valid DBS* available for each individual (i.e. the first DBS with a positive or a negative result, indeterminate and unknown being excluded). It's therefore computed among the registered participants having at least one valid DBS.

Indicator 18.1: *first DBS HIV prevalence*
Number of individuals whose first valid DBS is positive divided by the number of individuals having at least one valid DBS.

Indicator 18.1 could be computed using *FirstValidDBSResult* in **Individuals** dataset.

HIV incidence (DBS)

Only individuals with at least two DBSs and who's first DBS was negative are taken into account in this analysis. To estimate date of sero-conversion, a mid-approach is used, i.e. the date of sero-conversion is considered to be the middle point between the last negative DBS and the first positive DBS.

Definition 18.2: *DBS HIV Incidence*
Number of incident cases divided by person-years.

Incident cases are the number of individuals with a first DBS negative and at least one subsequent positive DBS. Person-years are equal to:

» Time between the first and the last DBS for individual who didn't convert (always negative)

» Time between the first DBS and the midpoint between last negative DBS and first positive DBS for seroconverters.

Note: a Poisson regression model with random effect should be used to take into account the cluster effect.

22. Observed HIV status and observed HIV prevalence

Individuals could be observed as being HIV positive within the TasP trial through DBS or HIV ascertainment. While DBS prevalence and incidence are used for the main outcome of the trial, an observed status is useful for process indicators.

Population 22.1: *Individuals whose HIV status has been observed*
The HIV status of an individual has been observed within the trial if that person provided at least one DBS and/or if this person was HIV ascertained at least once.

```
Population 22.1 is equal to individuals where variable HIVInfected is not
missing in Individuals dataset.
```

Population 22.2: *Individuals observed HIV positive*
An individual has been observed as HIV positive if he had any positive DBS and/or if he was ever ascertained HIV positive.

```
Population 22.2 is equal to HIVInfected==1 in Individuals dataset.
```

9. HIV testing and repeat testing

Note: mobile testing was implemented only in the last data collection cycle (in 2016).

Population 9.1: *Individuals ever tested*
Individuals who accepted at least once a rapid HIV test within TasP (at home or in a TasP mobile unit, or occasionally in TasP clinics).

Rapid tests are calculated from trial statuses. More precisely, having accepted a rapid test is defined as a trial status equal to 35 'rapid test done, positive result', 36 'rapid test done, indeterminate result', 37 'rapid test done, invalid result', 40 'rapid test done, negative result' or 50 'rapid test done, but result unknown' (cf. variable *RapidTested* in **Trial Statuses** dataset).
Population 9.1 corresponds to individuals with *EverRapidTested==1* in **Individuals** dataset, i.e. to individuals with at least one trial status considered as a rapid test. The variable *NbRapidTests* indicates the total number of tests for each registered individual.

Uptake

There has been discussion regarding if uptake of HIV testing should include or not in the denominator individuals self-reporting being HIV positive. Indeed, an HIV test is offered to ALL individuals contacted, regardless of their known HIV status. Some people aware of their HIV status will report their status to the fieldworkers, and "known HIV+" will be one of the reasons for refusing the test; but also some aware of their HIV status might not disclose and accept the test.

Therefore, two indicators are proposed:

- *crude uptake of testing*, calculated among all ever contacted individuals, a **process indicator** (a more appropriate process indicator being *HIV ascertainment uptake*, see section 10 page 23);
- *uptake of testing among those not reporting being HIV positive*, excluding individuals reporting being HIV positive, an **indicator of acceptability** of rapid test.

Indicator 9.2: *Crude uptake of testing*
Individuals ever tested (population 9.1) / individuals ever contacted (population 7.1)

Note: Participants who refuse to be tested at home are informed they can be tested in TasP clinics. In that specific case, the trial status corresponding to the test is associated with a clinic visit instead of a homestead visit. Therefore, it is also possible to calculate an ***uptake of home testing*** *(excluding tests performed in TasP clinics) and the proportion of those who have been tested in the clinic among all those who refused the test at home.*

Defining the denominator for the uptake of testing among those not reporting being HIV positive is more complex as individuals could have accepted a rapid test or refused for other reason at first contact before self-reporting being HIV positive in a subsequent contact. In

fact, only contacts before the first time the person self-reported to be HIV positive should be taken into account.

> **Indicator 9.3:** *Uptake of testing among those not reporting to be positive*
> Number of individuals who accepted at least once a rapid HIV test among those who never reported to be HIV positive or in a previous contact before self-reporting being HIV positive / Number of individuals ever contacted and who never self-reported to be HIV positive OR individuals contacted at least once before self-reporting to be HIV positive.

```
The first time someone self-reports being HIV positive could be seen as
right censor. Indicator 9.3 could be computed from a sub-sample of the
Trial Statuses dataset, by identifying for each individual if he/she ever
self-reported being HIV positive and, if yes, by removing all trial
statuses occurring after the first –self-report as HIV positive
(including that trial status).
```

Coverage

> **Indicator 9.4:** *Coverage of testing*
> Individuals ever rapid tested (population 9.4) / Registered individuals (population 6.1)

Note: Indicators 9.2, 9.3 and 9.4 could also be for each calendar round separately (to identify uptake trends over time) or per contact.

Repeat testing (among HIV negative)

This concept refers to the acceptability of a second/subsequent rapid test among individuals previously tested as HIV negative.

> **Indicator 9.5:** *Uptake of repeat testing*
> Number of individuals who ever accepted a second rapid HIV test / number of individuals who previously tested negative in the trial and who were subsequently contacted at least one more time in future calendar rounds

Note: repeat testing could also be calculated per contact (i.e. proportion of individuals accepting a rapid test at second contact among individuals tested negative at first contact).

Note 2: acceptability of rapid test at second contact among individuals who refused rapid test at first contact could also be computed to evaluate the capacity of repeat home-based testing to reach individuals who were previously refusers

Note 3: we could also think of estimating the rate of repeat testing among all those who previously tested negative in the trial (regardless of whether they were contacted a second time i.e. were offered repeat testing), to estimate coverage of repeat testing (rather than uptake); these individuals would need to have been eligible at least once after the first rapid test (i.e. someone from one of the additional clusters who entered the trial in calendar round 2 and was tested in that calendar round would not be eligible for repeat testing as there have been only 2 calendar rounds in these clusters so far).

10. HIV ascertainment uptake and coverage

HIV ascertainment is a **process indicator** and refers to the capacity of trial to ascertain the HIV status of participant, through rapid test or self-reporting as HIV positive. There is an asymmetry in the sense that we accept self-report as HIV+ (these individuals being referred to HIV care) but not self-report as HIV- (these individuals are encouraged to accept rapid test).

Population 10.1: *HIV status ever ascertained*
Number of individuals whose HIV status has been ascertained (i.e. had a valid rapid test result or self-reported to be HIV+) at least once

HIV ascertainment is computed from trial statuses. More precisely, being HIV ascertained is defined as a trial status equal to 25 'no rapid test, known positive', 30 'no rapid test, known positive and on ART', 35 'rapid test done, positive result' or 40 'rapid test done, negative result' (cf. variable *HIVAscertained* in **Trial Statuses** dataset).
Population 10.1 corresponds to individuals with *EverAscertained==1* in **Individuals** dataset, i.e. to individuals with at least one trial status considered as a HIV ascertainment. The variable *NbAscertainments* indicates the total number of ascertainments for each registered individual.

Note: by definition, indeterminate and invalid test results are excluded from HIV ascertainment. All rapid tests and self-reports are taken into account, whether the HIV ascertainment took place at home, in a mobile testing unit or in a TasP clinic.

Uptake

Indicator 10.2: *HIV ascertainment uptake*
Number of individuals whose HIV status has been ascertained (i.e. had a valid rapid test or self-reported to be HIV+) at least once (population 10.2) / Number of individuals ever contacted. (population 7.1)

This indicator may be calculated overall (HIV status ascertained at least once) or per calendar round or per eligibility or per contact.

Population 10.3: *HIV status ascertained per round X*
Number of individuals whose HIV status has ever been ascertained (i.e. had a valid rapid test result or self-reported to be HIV+) during a given home-based survey round (data collection cycle).

Population 10.3 can be identified using variables *CC1EverAscertained*, *CC2EverAscertained*, *CC3EverAscertained*, etc. in **Individuals** dataset.

Indicator 10.4: *HIV ascertainment uptake per round*
Number of individuals whose HIV status has ever been ascertained (i.e. had a valid rapid test or self-reported to be HIV+) during a given survey round (population 10.4) / Number of individuals ever contacted during the same survey round (population 7.3)

Note: due to internal migration, this indicator is computed per data collection cycle.

Newly diagnosed

Population 10.5: *Newly diagnosed among HIV ascertained*
We have no formal way to identify truly newly diagnosed individuals, but we can use a proxy, that is to estimate that an individual was newly diagnosed within the trial if:

» Had a rapid test AND
» Did not report in IQ that he/she was already positive AND
» Was not linked to ARTemis before date of first referral (see section 11 for definition of referral and section 12 for definition of care)
» Was not linked to ACCDB before date of first referral (see section 11 for definition of referral and section 12 for definition of care)

Population 10.5 corresponds to individuals where *NewlyDiagnosedAtReferral*==1 in **Individuals** dataset.

To compute this variable, we used the following conditions:
» *FirstReferralTS* equal to 35 (rapid tested)
» *CareAtReferral* equal to 3 (never been in care at referral)
» Considering all IQs completed before *DateFirstReferredToClinic* or within 7 days after *DateFirstReferredToClinic*, none of them had *KnowsPosMyself* equal to 1.

Coverage

Indicator 10.6: *HIV ascertainment coverage*
Number of individuals whose HIV status has ever been ascertained (population 10.1) / number of registered individuals registered (population 6.1)

Note: this may be calculated overall (HIV status ascertained at least once) or per calendar round or per eligibility.

11. Referral to HIV care (to TasP clinics)

An individual is referred to TasP clinics (for HIV care) if he/she was tested positive by a rapid test or if he/she reported to fieldworkers as being HIV positive (i.e. if HIV status was ascertained and positive).

Population 11.1: *Referred to TasP clinics*
Number of individuals ever ascertained HIV positive.

An individual is ascertained HIV positive if he/she has a trial status equal to 25 'no RT, known positive', 30 'no RT, known positive and on ART' or 35 'RT done, positive result'.
Population 11.1 corresponds to individuals where *EverReferredToTasPClinic==1* in **Individuals** dataset.
The date of the first referral is documented in *DateFirstReferredToClinic*. *NbReferrals* indicates the total number of times the person was referred to TasP clinics.

Note: this doesn't include individuals invited to go to TasP clinic for a rapid test due to an invalid or indeterminate result. Referral to HIV care is equal to the concept of ***positive HIV ascertainment*** *(cf. section 10 for a definition of HIV ascertainment).*

12. Engagement with care

About definition of being in care in DoH clinics

The ARTemis database collects data about the Hlabisa HIV care programme implemented in DoH (Department of Health) clinics within the sub-district. TasP participants have been linked with ARTemis database based on names, South African ID, date of birth and cell phone numbers. However, it is possible that some TasP participants cannot be found in ARTemis although they are in care in DoH.

ARTemis collects only data in the DoH clinics in Hlabisa. Therefore, we have no information regarding care received by participants in other places (e.g. private doctor, traditional healer, outside Hlabisa …).

Since January 2013, the database is not updated[1] anymore, except results of lab tests (viral loads and CD4 counts) obtained from the national laboratory (NHLS). So we can't use clinic visit dates to identify individuals in care in DoH. Therefore, the definition of care in DoH clinics is based on dates of CD4 count and viral loads as a proxy. A 13 months threshold is used, following Lessells *JAIDS* 2011[2].

In 2016, a new database named ACCDB (Africa Centre Clinical Database) has been developed, merging historical data from ARTemis with process data from Tier.net, the current system implemented in DoH clinics. A difference with ARTemis is that ACCDB contains only data for ART patients (patients in pre-ART care are not included in that dataset). However, ACCDB contains the date of clinical visits and ART prescriptions. TasP participants have been linked to ACCDB using names, South African ID, data of birth and cell phone numbers.

In December 2016, lab tests results were not yet available through ACCDB. However, such data should be added to ACCDB in 2017.

[1] The situation should change in 2015 as there is a project to restart full data collection in ARTemis. At that time, it should be possible to have a better way to identify individuals being on ART outside TasP clinics.

[2] Richard J Lessells et al., "Retention in HIV Care for Individuals Not Yet Eligible for Antiretroviral Therapy: Rural KwaZulu-Natal, South Africa," *JAIDS Journal of Acquired Immune Deficiency Syndromes* 56, no. 3 (March 2011): e79–86, doi:10.1097/QAI.0b013e3182075ae2.

To be consistent with the definitions of care in TasP clinics (less than 3 months late to next appointment), considering that all patients in ACCDB are supposed to be on ART and to have monthly visits, a window of 4 months (3+1) is used to define being in care in ACCDB.

Definition 12.1: *being in care in DoH clinics*
An individual is considered to be in care in a DoH clinic of Hlabisa

» if he/she is linked to ARTemis database and if/she has a CD4 count and/or a viral load (VL) recorded in ARTemis within the previous 13 months (395 days) OR

» if he/she is linked to ACCDB database and if he/she had a clinic visit within 4 months (122 days).

LinkedToArtemis in **Individuals** dataset indicates individuals who were successfully linked with the ARTemis database and *LinkedToACCDB* the individuals linked to ACCDB.
ARTemis data are stored in **ARTemis Patients** and **ARTemis Lab Results**[3] datasets. ACCDB data are store in **ACCDB Encounters** (visits) and **ACCDB Encounter Therapies** (ART prescriptions).

Care status in DoH at referral

All individuals ever referred to TasP clinics (population 11.1) could be divided in three groups, according to their care status in DoH at time of referral (i.e. at the date of the first referral to TasP clinics):

» currently in care in DoH (population 12.2);

» previously in care in DoH but lost to follow-up (population 12.3);

» never been in care in DoH (population 12.4).

Population 12.2: *Currently in care in DoH at referral*
Trial participant ever referred to care (population 11.1), linked to ARTemis and who has at least one CD4 and/or VL in ARTemis within 13 months (395 days) prior to the date of first referral OR linked to ACCDB with at least one clinic visit in the previous 4 months (122 days).

[3] Venous blood samples collected in TasP clinics and sent to DoH lab for CD4 count (quality control) are not recorded in **ARTemis Lab Results** dataset.

Population 12.2 corresponds to individuals with *CareAtReferral==1* in **Individuals** dataset, more precisely individuals following one of the following conditions:

» *EverReferredToTasPClinic*==1 AND *DaysARTemisLTBR*<=395

OR

» *EverReferredToTasPClinic==1 AND DaysACCDBLVBR<=122*

Population 12.3: *Previously in care in DoH but lost to follow-up at referral*
Trial participant ever referred to care (population 11.1), linked to ARTemis, with at least CD4 / VL in ARTemis before referral or one clinic visit in ACCDB before referral, but considered as actively in care (i.e. not part of population 12.2).

Population 12.3 corresponds to individuals with *CareAtReferral==2* in **Individuals** dataset, more precisely individuals where *EverReferredToTasPClinic==1* AND (*DaysARTemisLTBR* > 0 OR DaysACCDBLVBR > 0) AND not being already considered in care (pop. 12.2).

Population 12.4: *Never been in care in DoH at referral*
Trial participant ever referred to care (population 11.1), with no lab test result in ARTemis or clinic visit in ACCDB prior to referral.

Population 12.4 corresponds to individuals with *CareAtReferral==3* in **Individuals** dataset.

Note: by definition population 12.2 + population 12.3 + population 12.4 = population 11.1.

About definition of baseline clinic visit in TasP clinics

We have several sources of data that could be used to define baseline clinic visit:

- clinic visits recorded in the database;
- the type of clinic visits (i.e. baseline, W2, W4, non-protocol, etc.);
- having completed a CBC (baseline form in clinics);
- date of the first completed CHE (clinic history and examination form) completed by the nurse each time a patient come to clinic.

Using clinics visits dataset is not appropriate as in few occasions a clinic visit could be created even if the patient didn't physically come in the clinic (for example in case of exit) or if the patient came for rapid test in the clinic. The type of clinic visit isn't the best option has there are inconsistencies. In particular, the counsellor could make a mistake when select the type of visit on the computer.

The best option is to consider that the baseline clinic visit is the visit where the CBC form has been completed by the nurse, as during the baseline visit the nurse is supposed to complete both CBC and CHE. However, two elements need to be taken into account:

» very few individuals completed 2 CBC forms;
» very few individuals don't have a CBC form but have CHE forms.

Therefore, baseline clinic visit could be defined as the first clinic visit where a CBC and/or a CHE has been completed. Such definition is consistent with the way that retention in TasP clinics is computed (see section 19).

Definition 12.21: *Baseline TasP clinic visit*
The baseline TasP clinic visit for a patient is defined as the first TasP clinic visit where a CBC and/or a CHE form was completed.

```
Date of the baseline clinic visit is available in variable
BaselineClinicVisitDate in Individuals dataset.
```

Seen in TasP/DoH clinics after referral

Population 12.5: *Seen in TasP clinic after referral*
An individual has been seen in TasP clinic after referral if he/she was referred to care and he/she had a baseline clinic visit (definition 12.21) in TasP clinics.

```
Population 12.5 corresponds to individuals where SeenInTasPClinic==1 in
Individuals dataset, i.e. a BaselineClinicVisitDate is available and
EverReferredToTasPClinic==1.
```

Population 12.6: *Seen in DoH clinic after referral*
An individual has been seen in DoH clinic after referral if he/she had a Viral Load or a CD4 count recorded in ARTemis after the date of first referral or if he/she has a clinic visit in ACCDB after the date of the first referral.

```
Population 12.6 is equal to individuals where DateARTemisFTAR (date of
first test after referral) is populated (i.e. not missing) OR where
DateACCDBFVAR (date of first visit after referral) is populated (i.e. not
missing) in Individuals dataset.
```

Population 12.7: *Seen both in TasP and DoH clinic after referral*
An individual has been seen in both TasP and DoH clinics after referral if he/she had a baseline clinic visit in TasP clinics AND [if he/she had a Viral Load or a CD4 count recorded in ARTemis OR a clinic visit in ACCDB after the date of first referral].

Entry into care after referral

This indicator looks at the event *entry in care* over time among individuals not in care at time of referral (population 12.3 and 12.4).

Population 12.8: *Ever entered into care after referral*
Number of HIV+ individuals referred to care and not in care at time of referral (population 12.3 and 12.4) who visited a TasP clinic after referral (population 12.5) or have a CD4 and/or a VL recorded in ARTemis after referral (population 12.6)

Population 12.8 is equal to individuals where *EverEnteredCareAR*=1 in **Individuals** dataset.

Entry into care is a time-dependent phenomenon. Therefore, it is important to take into account time of observations since the first referral.

The variable *DaysObservationAR* in **Individuals** dataset indicates the number of days of observation after first referral. This the difference between:
- date of exit and date of first referral if the individual exited the trial after referral;
- else, end date of data collection in clinics and date of first referral.

For phase 1, end date of data collection has been fixed to 31st May 2014 (datasets version 3.0 or higher).

Population 12.9: *Entered into care within X months*
Number of HIV+ individuals referred to care and not in care at time of referral (populations 12.3 and 12.4), observed at least X months after referral, who visited a TasP clinic after referral within X months or have a CD4 and/or a VL recorded within X months in ARTemis after referral.

Note: it is possible that some of these individuals are not actively in care anymore X months after referral. For such indicators, please refer to population 12.19 and indicator 12.20.

Indicator 12.10: *Entry into care within X months*
Proportion of HIV+ individuals who visited a TasP clinic within X months after referral or have a CD4 and/or a VL recorded in ARTemis within X months after referral (population 12.9) among individuals not in care at time of referral (i.e.' previously in care but lost to follow-up' population 12.3 or 'never been in care' population 12.4).

Note: entry into care can be analysed longitudinally (survival analysis) or within a specific period (i.e. within X months after referral).

Linkage to TasP clinics

These indicators are calculated over a period of time after referral and refers only to entry in TasP clinics.

Population 12.11: *Linked to TasP clinic within X months*
Number of HIV+ individuals who visited a TasP clinic within X months after referral and observed at least X months after referral

Variables *LinkedToTASP3M*, *LinkedToTASP6M*, *LinkedToTASP9M*, *LinkedToTASP12M*, *LinkedToTASP18M* and *LinkedToTASP24M* in **Individuals** dataset correspond to population 12.11, calculated respectively for 3 months (91 days), 6 months (183 days), 9 months (274 days), 12 months (365 days), 18 months (548 days) and 24 months (730 days)

Indicator 12.12: *Linkage to TasP clinic within X months*
Number of HIV+ individuals who visited a TasP clinic within X months after referral (population 12.11) / number of individuals ever referred to HIV care (population 11.11) AND observed at least X months after referral

Linkage to TasP clinic could be divided in two sub-groups, based on engagement with HIV care at time of referral:

- transfers in (patients in care at referral, population 12.2);
- novel entry into care (patients not in care at referral, populations 12.3 and 12.4).

For both concepts, indicators could be calculated among all individuals referred to care or only among individuals with the same care status at referral.

a. Transfers in

Population 12.13: *Transferred in TasP clinic within X months*
Number of HIV+ individuals who were "currently in care" (population 12.2) at time of referral and who linked to a TasP clinic within X months after referral (population 12.11) and observed at least X months.

Indicator 12.14: *Transfers in TasP clinic within X months among referred to care*
Transferred in TasP clinic within X months (population 12.13) / Ever referred to care (population 11.1) and observed at least X months

Indicator 12.15: *Transfers in TasP clinic within X months among in care at referral*
Transferred in TasP clinic within X months (population 12.13) / In care at referral (population 12.2) and observed at least X months

b. Novel entry into care in TasP clinic

Population 12.16: *Novel entered into care in TasP clinic within X months*
Number of HIV+ individuals who were "not in care" (populations 12.3 and 12.4) at time of referral and who linked to a TasP clinic within X months after referral (population 12.11) and observed at least X months.

Indicator 12.17: *Novel entry into care in TasP clinic within X months among referred to care*
Novel entered into care in TasP within X months (population 12.16) / Ever referred to care (population 11.1) and observed at least X months

Indicator 12.18: *Novel entry into care in TasP clinic within X months among not in care at referral*
Novel entered into care in TasP within X months (population 12.16) / Not in care at referral (population 12.3 + population 12.4) and observed at least X months

```
By definition, population 12.11 (linked to TasP clinic) is equal to
population 12.13 (transfers in) + population 12.16 (novel entry into
care). Also, indicator 12.12 = indicator 12.14 + indicator 12.17
```

Ever been in care

The purpose of this indicator is to identify all HIV positive individuals who have ever been in care at any time before the end of data collection.

Population 12.22: *Individuals ever been in care*
An observed HIV positive individuals (population 22.2) who was ever
» seen in TasP clinic (population 12.5) and/or
» linked to ARTemis database (i.e. was in care in DoH at any time).

```
Population 12.22 corresponds to individuals where EverInCare==1 in
Individuals dataset.
```

19. Retention in TasP clinics

A lot of retention into care indicators exist in literature. It is important to distinct retention within a programme, retention in a clinic and retention in care, i;e. programme perspective, clinic perspective or individual perspective. Also, there is a difference between retention in ART care for patients on ART and retention in HIV care (including pre-ART patients).

This section focuses on retention within TasP clinics among all HIV patients ever seen in TasP clinics. It doesn't take into account the care trajectory within DoH clinics (or other places).

Note: additional sections could be developed to deal with these other indicators.

About TasP clinic visits

Some clinic visits could be created in the database in some occasions where the patient was not seen physically in the clinic, for example when the clinic documents an exit (as an exit trial status needs to be associated with a visit in the database).

Therefore, analysis of retention in TasP clinics is based only on visits were a CBC (baseline form) and/or a CHE (Clinic History Examination) form is completed. A CHE is supposed to be completed by the nurse each time a patient is coming in a clinic, both for protocol and non-protocol visits.

```
Clinic visits were a CBC or a CHE has been completed can be identified in
the Visits dataset using CBCcompleted and CHEcompleted variables.
```

About clinic appointments

When a patient visits a clinic, the nurse/counsellor should schedule the next visit and capture it the database. However, in some occasions, the next appointment date is not available in the database.

In such cases, a *next appointment date* has been computed using the following rules based on the protocol:

- if the patient was on ART at the time of the visit, the computed appointment date is the visit date + one month (30 days);
- if the patient was not on ART at the time of the visit, the computed appointment date is the visit date + six months (183 days).

```
First, a variable OnART has been computed in the Visits dataset for each
visit where a CBC and/or a CHE has been computed
(CBCcompleted==1|CHEcompleted==1). As for now date of ART stop is not
known (if any ART interruption), OnART has been calculated as follow:
»  the individuals was already on ART at baseline
   (Individuals$TasPBaselineARTUse==1);
```

» the individuals initiated ART before the considered visit (**Individuals$***TasPBaselineARTUse*!=1 & Individuals$*TasPDateFirstARTPrescription*<=Visits$*VisitDate*).

Then, *NextAppointmentDateComputed* has been computed as follow:
» equal to *NextAppointmentDate* if *NextAppointmentDate* was defined;
» *VisitDate* + 30 days if *NextAppointmentDate* was missing and *OnART*==1;
» *VisitDate* + 183 days if *NextAppointmentDate* was missing and *OnART*==0.

Duration since enrolment in TasP clinics

Population 19.1: *Individuals enrolled in TasP clinics at least X months before end of data collection*
Individuals seen in TasP clinics (12.5) whose baseline clinic visit occurs at least X months before the end of data collection (i.e. the 1st of July 2016).

The number of days between TasP baseline clinic visit and the end date of data collection is provided in *DaysABCV* variable in **Individuals** dataset. Population 19.1 could be calculated from *DaysABCV*, e.g. individuals enrolled in TasP clinics at least 6 months (183 days) before the end of data collection are individuals where *DaysABCV*>=183.

Retention status in TasP clinic

This retention status in TasP clinic indicates, at a specific point of time, the status of a TasP patient regarding his/her care within TasP clinics. This status is based on the next appointment scheduled by the nurse/ARV counsellor, i.e. it indicates if the patient is waiting his/her next appointment or if he/she is late.

Furthermore, this status also takes into account if the patient was reported dead or if he/she was transferred out, i.e. he/she asked a letter for transferring his/her care to another service. A patient could be transferred out for several reasons, including the willingness to receive care in DoH clinics or migration outside the TasP area.

A patient could transfer out and then came back later in TasP clinics. Few patients have even been transferred out twice since the beginning of the trial. The following indicator takes into account the fact that a patient could reintegrate TasP clinics after having been transferred out. However, in that case, the status is still computed from the baseline clinic visit, i.e. his/her first visit in TasP clinics.

Finally, nurses and ARV counsellors could also complete an exit form in a patient was not seen in TasP for a very long time (lost to follow-up) and didn't ask to be transferred out. These exits forms have not been taken into account for the computation of the retention status in TasP clinic, considering that these patients will be classified as lost to follow up based on the date of their last appointment.

Definition 19.2: *Retention status in TasP clinic at MX after baseline clinic visit*
Each patient ever seen in TasP clinic and enrolled at least X months before the end of phase 1 (population 19.1) could be classified in one of the following category:

» *waiting for next appointment* if the appointment scheduled during his/her last visit before MX is supposed to occur after MX;
» *less than a month late*, i.e. at MX, the patient didn't attend his/her last schedule appointment and that appointment was planned within the month preceding MX;
» *one to three months late*, i.e. at MX, the patient didn't attend his/her last schedule appointment and that appointment was planned within 1 to 3 months before MX;
» *three to six months late*, i.e. at MX, the patient didn't attend his/her last schedule appointment and that appointment was planned within 3 to 6 months before MX;
» *six months or more late*, i.e. at MX, the patient didn't attend his/her last schedule appointment and that appointment was planned at least 6 months before MX;
» *transferred out*, i.e. at MX the patient was transferred out and didn't came back since the date of the transfer out;
» *dead*, i.e. at MX the patient was dead.

Retention status is documented in *TasPRetentionStatusAtM1ABCV* to *TasPRetentionStatusAtM24ABCV* in the **Individuals** dataset.

Retention in TasP clinics rates (based on scheduled appointments)

Two retention rates are proposed:

» a *crude retention rate in TasP clinics*, where deaths and transfers out are taken into account in the denominator;
» a *net retention rate in TasP clinics*, where deaths are excluded from the denominator.

For both indicators, only individuals enrolled at least X months before the end of data collection are included.

An individual is considered as *lost to follow-up* if he/she is 3 months or more late, according to the TasP protocol.

Indicator 19.3: *Crude retention in TasP clinics rate at MX*
Individuals waiting for next appointment or less than 3 months late (see definition 19.2) divided by the number of individuals enrolled at least X months before the end of data collection (population 19.1)

Example for M6: number of individuals where *TasPRetentionStatusAtM6ABCV* is equal to 1, 2 or 3 divided by the number of individuals where *TasPRetentionStatusAtM6ABCV* is equal to 1, 2, 3, 4, 5, 6 or 7.

Indicator 19.4: *Net retention in TasP clinics rate at MX*
Individuals waiting for next appointment or less than 3 months late (see definition 19.2) divided by the number of individuals enrolled at least X months before the end of data collection (population 19.1) and not dead at MX (see definition 19.2).

Example for M6: number of individuals where *TasPRetentionStatusAtM6ABCV* is equal to 1, 2 or 3 divided by the number of individuals where *TasPRetentionStatusAtM6ABCV* is equal to 1, 2, 3, 4, 5 or 56

13. CD4 count

Baseline CD4

Definition 13.1: *Baseline CD4 count*
This refers to the first ever CD4 count measured in an individual.

For those who have never been in care: this would be equivalent to the first CD4 at presentation at the TasP clinic if they linked to TasP (see below).

For the known positives: baseline CD4 may be obtained for those linked to ARTemis database.

This data will not be known for individuals who are not in care in TasP and were not in care in ARTemis (those who transferred their care from outside the district).

At baseline visit

This refers to the first point-of care (PIMA) CD4 count measured in the TasP clinic in all HIV positive individuals. For the subset of individuals who have never had their CD4 count measured previously, this would correspond to their baseline CD4 count as indicated above

Indicator 13.2: *CD4 count at TasP baseline clinic visit*
PIMA CD4 measured in TasP clinic at baseline visit (individuals seen in TasP clinic, population 12.5).

Indicator 13.2 is documented in variable *TasPBaselineCD4* in **Individuals** dataset.
Note: a window of 7 days around baseline clinic visit date as also be taken into account.

At ART initiation

This refers to the PIMA CD4 count measured before the individual is initiated on ART in TasP clinics.

Indicator 13.3: *CD4 count at ART initiation in TasP clinics*
Last CD4 count measured in TasP clinic before date of ART initiation in TasP

Note: this indicator is computed only for patients initiating ART in TasP clinics (see section 14 for a definition of ART initiation).

Indicator 13.3 is documented in variable *TasPCD4AtARTInitiation* in **Individuals** dataset and is computed only for individuals where *TasPInitiatedART==1*.
Note: as for indicator 13.2, CD4 results from **CHEs** were added to **Lab Test results** to compute *TasPCD4AtARTInitiation*.

Note: there are on-going discussions about other ways of calculating this indicator. For example, it could the closest CD4 count between 4/8 weeks before ART initiation and ART initiation. CD4 count after ART initiation should be taken into account, even if closer to date of initiation than the previous CD4 count before initiation, because they can change very early (in the first week).

X months after baseline visit

Indicator 13.4: *CD4 count at X months after TasP baseline clinic visit*
This refers to the PIMA CD4 count measured X months after the baseline clinic visit, +/- 4-8 weeks, among individuals observed at least X months (+4/8 weeks).

Note: a consensus on the period (4 or 8 weeks) should be found.

Note: we could also calculate CD4 at M6 and M12 post-ART initiation.

14. ART initiation and uptake in TasP clinics

ART status among positive people at TasP baseline clinic visit

All individuals linked to TasP clinics will fall into 3 categories:

- » previously on ART but stopped ART;
- » ART naïve;
- » on ART at baseline.

Population 14.1: *Previously on ART but stopped ART at baseline clinic visit*
HIV+ individuals who were on ART in the past but not on ART anymore at TasP baseline clinic visit (ART defaulters).

Population 14.2: *ART-naïve at baseline clinic visit*
HIV+ individuals who are ART-naïve at TasP baseline clinic visit

Population 14.3: *On ART at baseline clinic visit*
HIV+ individuals who initiated ART in the past and still are under on ART at TasP baseline clinic visit.

ART status at baseline is documented in baseline clinic form (**CBCs** dataset) and in **Individuals** dataset, *TasPBaselineARTStatus* variable.

Note: by definition, population 14.1 + population 14.2 + population 14.3 (+ missing) = population 12.5 (seen in TasP clinic).

ART eligibility among patients not on ART at TasP baseline clinic visit

According to the protocol:

- » All individuals are eligible for ART in intervention arm
- » In control arm:
  - CD4 count <350 cells/mm3 irrespective of WHO clinical stage
  - WHO clinical stage 3 or 4 irrespective of CD4 count
  - those with a diagnosis of active tuberculosis (TB) irrespective of CD4 cell count
  - those who have a co-infection with hepatitis B virus (HBV) if the latter requires treatment
  - pregnant women

a. Eligibility according to the TasP protocol criteria

Population 14.4: *Eligible for ART according to TasP protocol at baseline clinic visit / at M X*
Number of patients, not on ART at baseline (populations 14.1 and 14.2), eligible for ART in TasP clinic according to the protocol criteria, at baseline clinic visit or at M X.

Note: eligibility at M X could be measured only among patients observed at least X months after TasP baseline clinic visit

```
Eligibility could be measured using CHEs dataset (Clinical History and
Examination forms).
```

Indicator 14.5: *Proportion eligible for ART according to TasP protocol at baseline clinic visit / at M X*
Eligible for ART according to TasP protocol at baseline clinic visit or at M X (population 14.5) / TasP patients not on ART at baseline clinic visit (population 14.1 + population 14.2) [and observed at least X months after baseline visit, if at M X].

b. Eligibility according to national guidelines

Population 14.6: *Eligible for ART according to national guidelines at baseline clinic visit / at M X*
Number of patients, not on ART at baseline (populations 14.1 and 14.2), eligible for ART according to national guidelines, at baseline clinic visit or at M X.

Note: eligibility at M X could be measured only among patients observed at least X months after TasP baseline clinic visit

```
Eligibility could be measured using CHEs dataset (Clinical History and
Examination forms).
```

Indicator 14.7: *Proportion eligible for ART according to national guidelines at baseline clinic visit / at M X*
Eligible for ART according to national guidelines at baseline clinic visit or at M X (population 14.7) / TasP patients not on ART at baseline clinic visit (population 14.1 + population 14.2) [and observed at least X months after baseline visit, if at M X].

Note: this indicator allows to identify patients in the intervention group who would not have been initiated on ART if they had been in the control group.

ART uptake after baseline visit among not on ART at baseline

Population 14.8: *Initiated ART in TasP clinic*
Individuals linked to TasP and not on ART at baseline (population 14.1 and 14.2) who ever initiated ART in TasP clinic.

Population 14.8 corresponds to individuals where *TasPInitiatedART==1* in **Individuals** dataset. This variable has been computed from **HIV Drug Treatment** dataset. An individual not on ART at baseline is considered to have initiated within TasP if he/she has at least one ART prescription recorded in this dataset. Date of ART initiation is documented in *DateTasPARTInitiation* (**Individuals** dataset), the earliest *DateStarted* found in **HIV Drug Treatments** for each individual initiating ART.

Indicator 14.9: *ART initiation uptake in TasP clinics after baseline visit*
This is the proportion of patients not on ART at baseline who initiated ART in TasP, i.e. population 14.8 / (population 14.1 + population 14.2).

Note: indicator 14.9 doesn't take into account time after baseline visit. See indicator 14.11 for a more precise indicator.

Population 14.10: *Initiated ART in TasP clinic within X months after baseline visit*
Number of individuals, not on ART at baseline, initiating ART in TasP within X months after baseline TasP clinic visit, and observed at least X months after baseline visit.

Indicator 14.11: *ART initiation in TasP clinics within X months after baseline visit*
Patients initiating ART within X months after baseline visit (population 14.10) / patients not on ART at baseline visit (regardless of ART eligibility status, i.e. population 14.1 + population 14.2) and observed at least X months after baseline visit.

ART uptake after eligibility

Population 14.12: *Initiated ART in TasP clinic within X months/weeks after eligibility*
Number of individuals, not on ART at baseline, initiating ART in TasP within X months/weeks after becoming eligible (see population 14.4 for a definition of eligibility in TasP clinics), and observed at least X months/weeks after eligibility.

Indicator 14.13: *ART initiation in TasP clinics within X months/weeks after eligibility*
Patients initiating ART within X months/weeks after eligibility (population 14.12) / patients not on ART at baseline visit (regardless of ART eligibility), becoming eligible for ART and observed at least X months/weeks after eligibility.

Note: the protocol stipulates within 2 weeks of enrolment. In literature, the period of 3 months is often used. This indicator could also be analysed longitudinally using KM and Cox models.

Note 2: This indicator could be calculated separately within intervention clusters stratified on national eligibility criteria.

ART status at a specific date

Definition 14.14: *ART status in TasP clinics at a specific date*

> An individual is considered to be actively on ART in a TasP clinic if he is actively in care (i.e. less than 3 months late to next appointment, cf. definition 19.2) in TasP clinic and if he received an ART prescription within 3 months (91 days).

15. ART coverage

ART coverage at the beginning of TasP

This indicator tries to estimate coverage of ART prior to the trial, i.e. at the implementation of the first calendar round, at population level. Individuals registered in calendar round 2 or later are excluded.

Population 15.1: *Individuals HIV+ at the beginning of the trial*
Number of individuals already HIV infected at the beginning of the trial (i.e. seroconverted individuals are excluded). HIV status at the beginning of the trial is:

» the result of the first valid DBS if the first valid DBS was performed during the first calendar round OR
» the result of the first HIV ascertainment if the first HIV ascertainment was performed during the first calendar round.

Population 15.1 corresponds to individuals where *HIVStatusBeforeTasP==1* in **Individuals** dataset. More precisely:

» if *DateFirstValidDBS>DateFirstAscertainment* & *FirstAscertainmentCR*==1, then *HIVStatusBeforeTasP=FirstAscertainment*
» if *DateFirstValidDBS<=DateFirstAscertainment* & *FirstDBSCR*==1, then *HIVStatusBeforeTasP=FirstValidDBSResult*
» if *DateFirstValidDBS*==MISSING & *DateFirstAscertainment*!=MISSING & *FirstAscertainmentCR*==1, then *HIVStatusBeforeTasP=FirstAscertainment* (no valid DBS)
» if *DateFirstValidDBS*!= MISSING & *DateFirstAscertainment*==MISSING & *FirstDBSCR*==1, then *HIVStatusBeforeTasP=FirstValidDBSResult* (never ascertained)

To identify individuals on ART, we used several data sources:

» ARTemis: this database collects information from DoH clinics within the sub-district of Hlabisa. However, since January 2013, the database is not updated[4], except results of lab tests (viral loads and CD4 counts). This database provides a date of ART initiation (if initiated before January 2013). There is no end date of ART (to identify individuals not on ART anymore). Viral Loads are still collected. An undetectable viral load will be used as a proxy of being on ART.
» iDART: this database from DoH pharmacists is not maintained anymore and cannot be used for phase 2.
» ACCDB: this database build on Tier.net contains ART prescriptions of patients in local DoH clinics. However, the completion of ACCDB is not yet equal to ARTemis. It is therefore necessary to combine both NHLS results from ARTemis and ACCDB to estimate ART coverage. In the future, it will maybe possible to use only ACCDB.

[4] The situation should change in 2015 as there is a project to restart full data collection in ARTemis. At that time, it should be possible to have a better way to identify individuals being on ART outside TasP clinics.

Population 15.2: *Individuals on ART at the beginning of the trial*
Number of individuals HIV infected at the beginning of the trial (population 15.2) and being on ART at that time. Being on ART at the beginning of the trial is defined as at least one of the following:

» being actively in care in DoH (cf. definition 12.1) and having at least one undetectable viral load within 13 months before the beginning of the trial OR
» being actively in care in DoH and having an ART prescription in ACCDB within 3 months.

Note: by definition, individuals cannot be on ART in TasP clinics at beginning of the trial.

```
Population 15.2 are individuals where OnARTAtBeginning is equal to 1
('Yes') in Individuals dataset. More precisely, this variable has been
computed as follow:
»  if InCareAtBeginning == 1 AND (DateHIVStatusBeforeTasP -
   DateARTemisLastUVLBHBST) <= 395 days
»  OR
»  if InCareAtBeginning == 1 AND (DateHIVStatusBeforeTasP -
   DateACCDBLastARTBHBST) <= 122 days
```

Indicator 15.3: *ART coverage at the beginning of TasP*
Proportion of HIV+ individuals on ART at the beginning of the trial among individuals HIV infected at the beginning of the trial (population 15.2 / population 15.1).

16. Viral load

THIS SECTION NEEDS ADDITIONAL WORK.

Undetectable Viral Load

This indicator could be calculated at different thresholds. We weed to be aware that over time the thresholds for detection has changed. Our assays have gone from <300 to <50 and now <40. For DoH it has been <25 for a while. So 400 is probably an OK threshold as the Quinn and Attia papers showed that there were no transmissions at viral load<400.

> **Definition 16.1:** *Virological suppression*
> An individual is considered as virally suppressed if he/she has a viral load strictly below 400 copies/mL.

Virological suppression/undetectable VL

This refers to individuals initiating ART within the trial who have viral load < 400 copies/mL.

First calculated among patients attending TasP clinics, then for those initiating outside the trial before transfer, the definition is the same but in some of them the ART initiation date will predate the trial.

Note: This indicator could be described at different VL thresholds, at different timepoints, over different period (first suppression; suppressed after x months on ART and over x number of months)

Note: Need to be aware that over time the thresholds for detection has changed. Our assays have gone from <300 to <50 and now <40. For DoH it has been <25 for a while. So 400 is probably an OK threshold as the Quinn and Attia papers showed that there were no transmissions at viral load<400.

Virological rebound

This refers to individuals on ART with previously suppressed viral load who ever developed a viral load >400 copies/mL on 2 consecutive occasions.

Discussion:

- *Rodolphe: Are you sure you will have always the same test or is it a threshold you define to take into account various thresholds+blips?*
- *Collins: This helps us to distinguish it from a viral blip; tricky because 400 cut-off is already slack. Usually cut-off is 50 and blip is 50-400 and immediately followed by another <50.*

Further complicated by the fact that SA guidelines have defined virological failure as viral load >1000 as below. Perhaps we should say .....400-1000 on 2 consecutive occasions.

Note: This indicator can also be calculated at different timepoints (6 months/24 months after virological suppression). Unless we want to answer specific questions like at what time point is rebound most common, otherwise virological rebound falls under virological failure as well, that is, it is a type of virological failure. The other type is for people who never achieved suppression which by definition has to have a time point to enable you make that decision.

Virological failure

This refers to individuals on ART with viral load>1000 copies/mL after 6 months on ART measured 2 months apart

17. Exits

THIS SECTION NEEDS ADDITIONAL WORK.

Trial exits

Trial exits or population exits refer to the proportion of individuals who exited the trial, among all trial participants.

- Outmigration: reported by head of household as no longer a resident member
- Death
- Not able anymore to provide an informed consent (for example: mentally disabled)

TasP clinic exits

TasP clinic exits refers to the proportion of patients who exited clinic care, among all trial participants enrolled in care.

- Death
- Transfer of care is when a participant indicates to a clinic staff that they are relocating and requests a transfer letter
- Loss to follow-up
    - not having been seen in TasP clinic ≥ 90 days for those on ART;
    - for those not on ART (not eligible or who choose not to start), loss to follow-up will be defined as > 9 months from last clinic visit and/or CD4 cell count.
- Decision/choice of the patient to exit clinic care without transferring care somewhere else (Is this classified as LTFU?)

Note: the current exit forms doesn't distinguish between individuals who exited only the clinic cohort but not the population cohort. (All of them appear with an Exit trial status). Some of Transfer Out and Loss to Follow-up corresponds to out-migration but not all of them. ➔ We will have to discuss how we want to manage that and do some specific recode.

NB: for phase 2, the database and field software will be changed to store properly residency and care episodes. We will also have two separate forms for Trial Exit and Clinic Exit.

```
In current datasets, exits are documented through trial statuses, with
the limitations explained before.
```

21. Computing ages

When collecting or updating a date of birth, fieldworkers could collect an incomplete date of birth, by indicating elements (year, month and/or day) not known. In such cases, a computed date of birth is computed by completing randomly the missing elements.

The collected date of birth is available in text variable *VPDateOfBirth* from **Individuals** dataset (U letter being used for unknown part of the date). The computed date of birth corresponds to *DateOfBirth*. In the case where date of birth is fully known, *VPDateOfBirth* and *DateOfBirth* are equal. The precision of the date of birth is indicated in *DateOfBirthPrecision*.

When year of birth is unknown, this computation generates an important bias as year is randomly selected between 1900 and 1999. Therefore, **ages should be computed only if at least year of birth is known**. When the year of birth is known but not the exact month and/or day, we can consider that the computed date of birth could be used for age calculation, as the age will be correct at more or less one year.

Mean age should be calculated using exact ages (i.e. an individual born the 1st of January 1980 is aged 34.25 years at the 1st of April 2014).

23. Employment and active population

Population 23.1 *Employed individuals*
Individuals are employed if they reported to be currently employed (full time or part time) at the time of the completion of an individual questionnaire.

Population 23.1 corresponds to the variable *Employed* in **IQs** dataset and has been computed from *CurrentlyInEmployment*. Due to the fact that some individuals reported to be employed when looking at *NonEmploymentTypeOtherSpec*, people with *NonEmploymentTypeRecoded*==8 are also considered to be employed.

Note: NonEmploymentTypeOtherSpec is only available at IQ1 and IQ3. By consequence, correction done using NonEmploymentTypeRecoded has been done only for IQ1 and IQ3.

Population 23.2 *Working-age population*
The working-age population includes all persons aged 15–64 years.

In IQs dataset, population 22.2 are individuals at IQ where *AgeAtIQ* >=15 and *AgeAtIQ* <65.

Population 23.3 *Active population*
The active population includes all persons aged 15–64 years who are employed or looking for a job.

Population 23.3 are individuals with *ActiveStatus*==1, i.e. with *AgeAtIQ* >=15 and *AgeAtIQ* <65 and (*ProfessionalStatus*==1 or *ProfessionalStatus*=3)in **IQs** dataset.

Note: individuals who are still studying, pregnant, sick or injured, retired or doing nothing (not looking for a job) or who are <15 or ≥65 are considered to be inactive.

Note 2: in phase 1, this population can be computed only at IQ1 and IQ3 (as ProfessionalStatus is not available at IQ2).

Indicator 23.4 *Employment-to- population ratio (labor absorption rate)*
Individuals who are employed (cf. population 23.1) among the working-age population divided by the working age population (population 23.2).

Note: in phase 1, this indicator can be computed only at IQ1 and IQ3.

Indicator 23.5 *Employment-to-active population ratio*
Proportion of employed individuals (cf. population 23.1) among the active population (population 23.3) divided by the active population (population 23.3).

Note: in phase 1, this indicator can be computed only at IQ1 and IQ3.

Index

Appendix

R code for computing household's assets index score

```
# Empty rows excluded
thhi <-
hhi[!(DrinkWaterSource>90&ToiletType>90&IsElectrified>5&MainCookingFuel>90&IsHomest
eadOwnerAMember>5),]

# Recoding variables

thhi[DrinkWaterSource>90,DrinkWaterSource:=NA]
thhi[DrinkWaterSource%in%3:8,DrinkWaterSource:=3]

thhi[ToiletType>90,ToiletType:=NA]
thhi[ToiletType==1,ToiletType:=2]
thhi[ToiletType==4,ToiletType:=5]

thhi[IsElectrified>5,IsElectrified:=NA]

thhi[MainCookingFuel>90,MainCookingFuel:=NA]
thhi[MainCookingFuel==3,MainCookingFuel:=1]
thhi[MainCookingFuel==7,MainCookingFuel:=1]
thhi[MainCookingFuel%in%5:6,MainCookingFuel:=4]

thhi[IsHomesteadOwnerAMember>5,IsHomesteadOwnerAMember:=NA]

# Completing missingness with closest HHI if available
# If not, with the mode within the same cluster and the same calendar round

m <- function(x){
  tmp <- table(x, exclude=NA)
  return(as.integer(names(tmp[tmp==max(tmp)][1])))
}

thhi2 <- thhi

for (i in thhi[is.na(DrinkWaterSource),DocumentId]) {
  hhid <- thhi[DocumentId==i,HouseholdId]
  if (thhi[HouseholdId==hhid&!is.na(DrinkWaterSource),.N]>0) {
    tmp <- thhi[HouseholdId==hhid&!is.na(DrinkWaterSource)]
    tmp[,Days:=abs(as.integer(VisitDate-thhi[DocumentId==i,VisitDate]))]
    setkey(tmp,Days) # Sorting by Days
    thhi[DocumentId==i,DrinkWaterSource:=tmp[,first(DrinkWaterSource)]]
  } else {
    cr <- thhi[DocumentId==i,CalendarRound]
    cl <- thhi[DocumentId==i,ClusterId]

thhi[DocumentId==i,DrinkWaterSource:=m(thhi2[ClusterId==cl&CalendarRound==cr,DrinkW
aterSource])]
    # NB: we compute the mode on data avaialble before data correction
  }
}

for (i in thhi[is.na(ToiletType),DocumentId]) {
  hhid <- thhi[DocumentId==i,HouseholdId]
  if (thhi[HouseholdId==hhid&!is.na(ToiletType),.N]>0) {
    tmp <- thhi[HouseholdId==hhid&!is.na(ToiletType)]
    tmp[,Days:=abs(as.integer(VisitDate-thhi[DocumentId==i,VisitDate]))]
    setkey(tmp,Days) # Sorting by Days
    thhi[DocumentId==i,ToiletType:=tmp[,first(ToiletType)]]
  } else {
    cr <- thhi[DocumentId==i,CalendarRound]
    cl <- thhi[DocumentId==i,ClusterId]
```

```
thhi[DocumentId==i,ToiletType:=m(thhi2[ClusterId==cl&CalendarRound==cr,ToiletType])
]
    # NB: we compute the mode on data avaialble before data correction
  }
}

for (i in thhi[is.na(IsElectrified),DocumentId]) {
  hhid <- thhi[DocumentId==i,HouseholdId]
  if (thhi[HouseholdId==hhid&!is.na(IsElectrified),.N]>0) {
    tmp <- thhi[HouseholdId==hhid&!is.na(IsElectrified)]
    tmp[,Days:=abs(as.integer(VisitDate-thhi[DocumentId==i,VisitDate]))]
    setkey(tmp,Days) # Sorting by Days
    thhi[DocumentId==i,IsElectrified:=tmp[,first(IsElectrified)]]
  } else {
    cr <- thhi[DocumentId==i,CalendarRound]
    cl <- thhi[DocumentId==i,ClusterId]

thhi[DocumentId==i,IsElectrified:=m(thhi2[ClusterId==cl&CalendarRound==cr,IsElectri
fied])]
    # NB: we compute the mode on data avaialble before data correction
  }
}

for (i in thhi[is.na(MainCookingFuel),DocumentId]) {
  hhid <- thhi[DocumentId==i,HouseholdId]
  if (thhi[HouseholdId==hhid&!is.na(MainCookingFuel),.N]>0) {
    tmp <- thhi[HouseholdId==hhid&!is.na(MainCookingFuel)]
    tmp[,Days:=abs(as.integer(VisitDate-thhi[DocumentId==i,VisitDate]))]
    setkey(tmp,Days) # Sorting by Days
    thhi[DocumentId==i,MainCookingFuel:=tmp[,first(MainCookingFuel)]]
  } else {
    cr <- thhi[DocumentId==i,CalendarRound]
    cl <- thhi[DocumentId==i,ClusterId]

thhi[DocumentId==i,MainCookingFuel:=m(thhi2[ClusterId==cl&CalendarRound==cr,MainCoo
kingFuel])]
    # NB: we compute the mode on data avaialble before data correction
  }
}

for (i in thhi[is.na(IsHomesteadOwnerAMember),DocumentId]) {
  hhid <- thhi[DocumentId==i,HouseholdId]
  if (thhi[HouseholdId==hhid&!is.na(IsHomesteadOwnerAMember),.N]>0) {
    tmp <- thhi[HouseholdId==hhid&!is.na(IsHomesteadOwnerAMember)]
    tmp[,Days:=abs(as.integer(VisitDate-thhi[DocumentId==i,VisitDate]))]
    setkey(tmp,Days) # Sorting by Days

thhi[DocumentId==i,IsHomesteadOwnerAMember:=tmp[,first(IsHomesteadOwnerAMember)]]
  } else {
    cr <- thhi[DocumentId==i,CalendarRound]
    cl <- thhi[DocumentId==i,ClusterId]

thhi[DocumentId==i,IsHomesteadOwnerAMember:=m(thhi2[ClusterId==cl&CalendarRound==cr
,IsHomesteadOwnerAMember])]
    # NB: we compute the mode on data available before data correction
  }
}

# Creating df for analysis
# Note: some variables not used

v <- c(
  "DrinkWaterSource",
  "ToiletType",
  "IsElectrified",
```

```
  "MainCookingFuel",
  "IsHomesteadOwnerAMember",
  "HasCellphone",
  #"HasTelephone",
  "HasRadioStereo",
  "HasTV",
  "HasVideoDVD",
  "HasSofa",
  "HasBed",
  #"HasBedNet",
  "HasTableChairs",
  "HasSewingMachine",
  "HasKitchenSink",
  "HasPrimusCooker",
  "HasElecHotplate",
  "HasElecStoveOven",
  "HasGasCooker",
  "HasElecKettle",
  "HasFridgeFreezer",
  "HasCarBatteryElec",
  "HasCarBakkie",
  #"HasMotorcycle",
  "HasBicycle",
  #"HasKombiLorryTractor",
  "HasBlockMaker",
  "HasWheelbarrow",
  "HasHoeSpadeFork",
  "HasCattle",
  "HasOtherLivestock"
  )

fhhi <- thhi[,v,with=F]

for (i in 1:ncol(fhhi))
  fhhi[[i]] <- factor(fhhi[[i]])

# MCA computed on CR1 data
acm <- mca(fhhi[thhi[,CalendarRound==1],])

# Predict coordinates for all HHIs
thhi$AssetsScore <- predict(acm, fhhi)[,1]

# Normalized on CR1 data
thhi$AssetsScore <- (thhi$AssetsScore - mean(acm$rs[,1]))/sd(acm$rs[,1])

# Score should be positive for rich people
if(mean(thhi[HasTV==2,AssetsScore])>mean(thhi[HasTV==1,AssetsScore]))
  thhi$AssetsScore <- thhi$AssetsScore*(-1)

# Cutting at 40-40-20
thhi$AssetsCat <- as.integer(cut(thhi$AssetsScore, breaks=c(-
10,quantile(thhi[CalendarRound==1,AssetsScore],probs=c(0.4,0.8)),10)))
```