

# **ScanForm Data Collection Instrument Documentation**

**HIV Testing Services · ScanForm Documentation · Version 1.6**

2026-05-11

# Table of contents

<b>HTS 001 Client Intake Form</b>	<b>3</b>
What is this form? . . . . .	3
Form at a Glance . . . . .	3
Form Structure . . . . .	4
How ScanForm Processes This Form . . . . .	5
Documentation Sections . . . . .	6
Clinical Logic Summary . . . . .	6
About This Documentation . . . . .	7
<b>Variables</b>	<b>8</b>
Overview . . . . .	8
Page 1 Variables . . . . .	9
Pre-Test Counseling and Risk Assessment . . . . .	12
Page 2 Variables . . . . .	14
Summary Statistics . . . . .	20
OCR Models Reference . . . . .	21
<b>In-Phone Validation</b>	<b>22</b>
How In-Phone Validation Works . . . . .	22
Key Concepts . . . . .	23
Record Eligibility . . . . .	23
Validated Fields . . . . .	24
Fields Not Validated In-Phone . . . . .	26
Summary . . . . .	27
Important Limitation . . . . .	28
<b>Data Pipeline</b>	<b>29</b>
Pipeline Overview . . . . .	29
Pipeline Layers . . . . .	30
Data Flow Summary . . . . .	34
Notes for Data Managers . . . . .	36
<b>Data Quality Checks</b>	<b>37</b>
How Pipeline DQA Works . . . . .	37
Check Severity Levels . . . . .	38
Check Catalogue . . . . .	38
Expected Additional Checks . . . . .	42
Check Output Structure . . . . .	43
Summary . . . . .	44
Notes for Data Managers . . . . .	45

<b>Dashboard Indicators</b>	<b>46</b>
Overview	46
Indicator Index	46
Testing Volume	47
Testing Outcomes	48
Pre-Test Counseling Quality	49
Risk Assessment	50
TB Screening	51
STI Screening	52
Sex Partner Risk	53
Post-Test Counseling	54
Linkage to Services	55
Laboratory Results	56
Indicator Reference Table	57
Important Notes for Dashboard Developers	57

# HTS 001 Client Intake Form

HIV Testing Services · ScanForm Documentation · Version 1.6

## What is this form?

The **HTS 001 Client Intake Form** is the primary paper-based data collection instrument used in Nigeria’s HIV Testing Services programme. It is completed by a trained counsellor for every client who presents for an HTS session — capturing demographics, risk assessment, clinical screening, test results, and post-test counseling outcomes in a single encounter record.

The form is digitalised using **QED.ai ScanForm**: a field worker photographs the completed paper form using the ScanForm mobile app, and the image is processed on the server using Optical Character Recognition (OCR) to extract all data fields automatically. This document describes every aspect of the form, its data fields, and the automated systems that ensure data quality.

**This documentation is auto-generated** from the live XLSForm specification, banding configuration, in-phone validation rules, and dbt pipeline models. It updates automatically whenever any of these source files change, ensuring it always reflects the deployed form.

---

## Form at a Glance

Table 1: Form at a Glance

Property	Value
Form name	HTS 001 Client Intake Form
Form version	v1.6
Country	Nigeria
Programme	HIV Testing Services (HTS) – NG-HTS

<b>Form type</b>	ScanForm Photopack (two pages treated as one form)
<b>Paper size &amp; orientation</b>	A3 landscape – 420 x 297 mm
<b>Pages</b>	2 (Page 1: client demographics & pre-test counseling; Page 2: clinical screening & post-test)
<b>Records per form</b>	1 client per form
<b>Total variables</b>	~100 (including administrative and completion fields)
<b>OCR-banded fields</b>	~85 (all digit boxes, oval fields, mixed boxes, date fields)
<b>Handwriting fields (not OCR)</b>	7 (State, LGA, Facility Name, Comments, Completed By, Designation, Sign)
<b>Oval single-select fields</b>	~65 (all single-select bubble questions)
<b>Digit box fields</b>	~15 (ages, counts, scores, phone, codes)
<b>Date fields</b>	2 (Date of Visit; completion Date)
<b>In-phone validation active</b>	Yes – 72 fields with active checks
<b>Fields validated in-phone</b>	72 across both pages
<b>Pipeline DQA checks</b>	59 expected (14 integer conversion, 2 date, 25 multiple-selection, 18 custom)
<b>Dashboard indicators documented</b>	18 indicators across 7 thematic areas
<b>Form slug</b>	hts001-ng
<b>Form page identifiers</b>	HTS 001 P1 1.6 (Page 1) / HTS 001 P2 1.6 (Page 2)

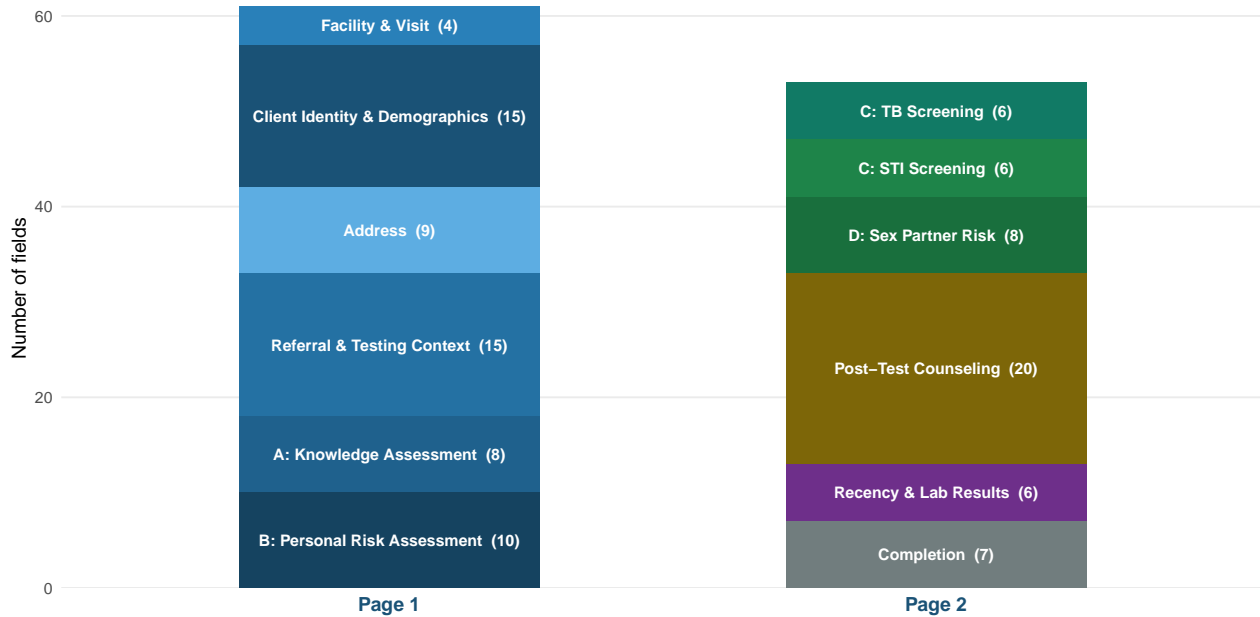
---

## Form Structure

The physical form is an A3 landscape sheet divided into two A4-sized pages, scanned together as a single photopack. Each page has its own QR code and Data Matrix barcode for unique identification.

## HTS 001 Client Intake --- Sections and Field Counts by Page

Each bar represents one physical page of the A3 photopack



## How ScanForm Processes This Form

The table below summarises the end-to-end journey of each HTS 001 record, from paper completion to dashboard display.

Table 2: End-to-end processing flow for HTS 001 records

Step	Actor	Action
1	Clinician	Completes paper form
2	Field worker	Photographs form with ScanForm app
3	Mobile app	Runs in-phone validation on 72 fields
4	Mobile app	Alerts worker if any check fails
5	Mobile app	Submits photo to server on pass
6	Server	OCR processing extracts all fields
7	Server pipeline	Pipeline DQA checks applied
8	Pipeline	Refinery models produce analysis-ready tables
9	Metabase	Dashboards updated with clean data

### Two layers of data quality protection

Every HTS 001 record passes through two independent quality gates before reaching dashboards:

1. **In-phone validation** — runs instantly on the mobile device when the photo is taken. Checks 72 fields for missing mandatory values and multi-selected single-select questions. If any check fails, the worker is alerted immediately and must correct the paper form before submitting.
2. **Pipeline DQA checks** — run server-side on the full dataset after OCR processing. Apply more sophisticated logic: integer and date parsing, value range checks, score consistency validation, and cross-field clinical rules. Failures are flagged as Warnings or Errors for data manager review.

---

## Documentation Sections

Table 3: Documentation sections overview

Section	Description
<b>Variables</b>	Complete inventory of every field on the form: variable names, field types, OCR models, conditional fields, and options for every single-select question. Organised by thematic section matching the physical form layout.
<b>Data Pipeline</b>	How raw OCR data flows from the ScanForm export API through dbt transformation layers (base, pre-clean, clean, checks, refinery) to produce analysis-ready tables. Describes key transformations, derived fields, and business logic at each stage.
<b>In-Phone Validation</b>	The 72 automated real-time checks applied by the ScanForm mobile app at the moment a photo is taken, before data is submitted. Documents which fields are validated, what check type applies, and the exact alert message shown to the field worker.
<b>DQA Checks</b>	Server-side data quality checks applied after OCR processing. Covers integer and date conversion checks, multiple-selection detection, and custom checks including score range validation, score consistency, and cross-field clinical logic. Explains Warning vs Error severity and recommended actions.
<b>Dashboard Indicators</b>	18 dashboard indicators derived from the refinery models: goal, numerator, denominator (with exact column names), disaggregations, visualization type, and data source. Covers testing volume, HIV positivity, TB/STI screening, post-test counseling, PrEP referral, and laboratory results.

---

## Clinical Logic Summary

Three clinical decision rules are printed directly on the physical form and enforced in the data pipeline.

### Rule 1 – HIV Testing Trigger

Test for HIV if Personal HIV Risk Assessment Score is  $\geq 1$  AND last HIV test was more than **3 months ago**.

*Fields involved:* risk\_assessment, time\_last\_hiv\_negative\_test\_result

### Rule 2 – TB Referral Trigger

If TB Screening Score is  $\geq 1$ , the client must be tested for Xpert MTB/RIF **or** referred to a TB service.

*Fields involved:* tb\_screening (sum of current\_cough, weight\_loss, fever, night\_sweats, lymphadenopathy)

### Rule 3 – PrEP Referral Trigger

If client tests **HIV negative** AND has a Sex Partner Risk Assessment Score  $\geq 1$  (Section D), refer client for **PrEP services**. Additionally, if HIV negative with Risk Score  $\geq 1$  or STI evidence, recommend **re-testing after 3 months**.

*Fields involved:* hiv\_test\_result, sex\_partner\_risk\_assessment, sti\_screening, prep\_referred

---

## About This Documentation

This documentation is automatically generated by **QED.ai** from the live form source files.

Table 4: Source files driving this documentation

Source file	What it drives
<b>HTS 001 Client Intake.xlsx</b> (XLSForm)	Variables page: field names, types, OCR models, conditional logic
<b>banding.json</b>	Field positions, preview geometry, OCR model assignments
<b>data validation.py</b>	In-Phone Validation page: check types, eligibility rules, alert messages
<b>*.sql (dbt models)</b>	Data Pipeline and DQA Checks pages: transformation logic, check definitions
<b>export config.json</b>	Column names and formats used in indicator definitions

The code generating this document is triggered automatically each time any of the above source files change, ensuring that documentation always accurately reflects the deployed form.

# Variables

HTS 001 Client Intake Form · NG-HTS · Nigeria · v1.6

## Overview

**Form:** NG-HTS 001 Client Intake | **Version:** v1.6 | **Country:** Nigeria | **Pages:** 2 (A3 landscape photopack)

This chapter documents every variable captured on the HTS 001 Client Intake ScanForm. Fields are grouped by thematic section matching the physical layout of the paper form.

### Field type legend:

Table 1: Field type legend

Field Type	Description
<b>Oval (single)</b>	Bubble oval – mark one option with an X; max one X per question
<b>Digit boxes</b>	Square digit boxes – one digit per box, 0-9
<b>Letter boxes</b>	Square letter boxes – one capital letter per box, A-Z
<b>Mixed boxes</b>	Combination of digit and letter boxes in a single field
<b>Date boxes</b>	Pre-structured date entry with separators; may contain pre-printed characters
<b>Handwriting</b>	Free-form handwriting zone – not OCR-processed by ScanForm

**Shading key:** Orange-shaded rows in section tables are **conditional fields** — printed on the paper form but should only be completed when the stated condition is met.

## Page 1 Variables

### Facility and Visit Information

Table 2: Facility and Visit – field overview

Variable	Label	Type	Re-quired	Condi-tional
facilityState	State	Handwriting	–	–
facilityLga	LGA	Handwriting	–	–
facilityName	Facility Name	Handwriting	–	–
dateVisit	Date of Visit	Date boxes	–	–

Table 3: Facility and Visit – field details

Variable	Format	OCR	Notes
facilityState	88 mm line	Not banded	Outside scannable area
facilityLga	88 mm line	Not banded	Outside scannable area
facilityName	120 mm line	Not banded	Outside scannable area
dateVisit	dd/mm/yyyy	date dd, date mm, two digits 20, date nearby yy	Century 20 pre-printed

### Client Identity and Demographics

Table 4: Client Identity – field overview

Variable	Label	Type	Re-quired	Condi-tional
firstName	Client First Name	Letter boxes	–	–
otherName	Middle Name	Letter boxes	–	–
surname	Surname	Letter boxes	–	–
age	Age (years)	Digit boxes	–	–
sex	Sex	Oval (single)	–	–
firstTimeVisit	First Time Visit	Oval (single)	–	–
employmentStatusId	Employment Status	Oval (single)	–	–
maritalStatus	Marital Status	Oval (single)	–	–
educationId	Education Level	Oval (single)	–	–
numWives	No. of Wives/Co-wives	Digit boxes	–	–
numChildren	No. of own children under 5 years	Digit boxes	–	–
clientCode	Client Code	Mixed boxes	–	–

recency	Recency Number	Digit boxes	-	-
phoneNumber	Client Telephone Number	Digit boxes	-	-

Table 5: Client Identity – field details

Variable	Format	OCR	Notes
firstName	17 boxes	string	
otherName	17 boxes	string	
surname	17 boxes	string	
age	3 boxes	int	
sex	Male / Female	select one or zero	
firstTimeVisit	Yes / No	select one or zero	
employmentStatusId	Employed / Unemployed / Student / Retired / Freelance	select one or zero	
maritalStatus	Married / Divorced / Widowed / Separated / Single	select one or zero	
educationId	None / Primary / Junior Secondary / Post Secondary / Quranic	select one or zero	
numWives	2 boxes	int	
numChildren	2 boxes	int	
clientCode	□□□-□□□-□□□-□	int x9, string x2	Batch-Bundle-Page- Check format
recency	4 boxes	int	
phoneNumber	11 boxes	int	

## Address

Table 6: Address – field overview

Variable	Label	Type	Re- quired	Condi- tional
line	Street Address / Description	Mixed boxes	-	-
nearestLandmark	Nearest Landmark/Compound	Mixed boxes	-	-
state	State of Residence	Letter boxes	-	-
lga	LGA of Residence	Letter boxes	-	-

Table 7: Address – field details

Variable	Format	OCR	Notes
line	72 mm	tokenwise string int	

nearestLandmark	36 mm	tokenwise string int
state	17 mm	tokenwise string int
lga	18 mm	tokenwise string int

**About tokenwise string int:** Address fields use this OCR model. It splits input into space-separated tokens and classifies each token as either all-letters or all-digits before applying the appropriate recognition engine. This prevents common misrecognitions such as reading the letter O as digit 0 in place names.

## Referral and Testing Context

Table 8: Referral and Testing – field overview

Variable	Label	Type	Re-quired	Condi-tional
referredFrom	Referred From	Oval (single)	–	–
referredFromOther	Referred From – specify	Letter boxes	–	Yes
testingSetting	Setting	Oval (single)	–	–
testingSettingOther	Setting – specify	Letter boxes	–	Yes
typeConsueling	Type of Session	Oval (single)	–	–
targetGroup	Target Group	Oval (single)	–	–
previouslyTested	Previously tested within last 3 months	Oval (single)	–	–
indexClient	Is client from an index client?	Oval (single)	–	–
relationWithIndexClient	Relation with index client	Oval (single)	–	Yes
indexClientCodeNew	Index Client Code (ScanForm)	Mixed boxes	–	Yes
indexClientCodeOld	Index Client Code (Legacy)	Digit boxes	–	Yes
pregnant	Client is Pregnant	Oval (single)	–	–
breastFeedingUnder6	Client breastfeeding under 6 months	Oval (single)	–	–
breastFeedingOver6	Client breastfeeding over 6 months	Oval (single)	–	–

Table 9: Referral and Testing – field details

Variable	Format	OCR	Notes
referredFrom	Self / TB / OPD / Blood Bank / STI / FP / Ward / Other	select one or zero	
referredFromOther	14 boxes	string	Fill when Referred From = Other
testingSetting	CT / TB / OPD / Outreach / Ward / Standalone HTS / Other	select one or zero	

testingSettingOther	12 boxes	string	Fill when Setting = Other
typeConsueling	Individual / Couple / Group / Previously Self-Tested	select one or zero	
targetGroup	(see form)	select one or zero	
previouslyTested	Yes / No	select one or zero	
indexClient	Yes / No	select one or zero	
relationWithIndexClient	Biological / Sexual / Social	select one or zero	Fill when Index Client = Yes
indexClientCodeNew	□□□-□□□-□□□-□	int x9, string x2	Fill when Index Client = Yes; one code only
indexClientCodeOld	□□□□-□□-□□	int	Fill when Index Client = Yes; one code only
pregnant	Yes / No	Not banded	Ensures linkage to PMTCT
breastFeedingUnder6	Yes / No	select one or zero	
breastFeedingOver6	Yes / No	select one or zero	

**Conditional fields – Index Client.** Three fields are conditionally required (shaded orange above):

- **relationWithIndexClient** – fill when Index Testing = **Yes**
- **indexClientCodeNew** – fill when Index Testing = **Yes**; complete only one code field
- **indexClientCodeOld** – fill when Index Testing = **Yes**; complete only one code field

## Pre-Test Counseling and Risk Assessment

All items in Sections A and B are marked No (0) or Yes (1) using a single oval bubble. Scores are the sum of responses and are entered in a single digit box.

### Section A – Knowledge Assessment

Table 10: Section A: Knowledge Assessment – field overview

Variable	Label	Type	Re-quired	Condi-tional
----------	-------	------	-----------	--------------

previousTestedHIVNegative	Previously tested HIV negative	Oval (single)	-	-
timeLastHIVNegativeTestResult	Time of last HIV Negative test Results	Oval (single)	-	-
clientInformHivTransRoutes	Client informed about HIV transmission routes	Oval (single)	-	-
clientInformRiskkHivTrans	Client informed about risk factors for HIV transmission	Oval (single)	-	-
clientInformPreventingHivTrans	Client informed on preventing HIV transmission methods	Oval (single)	-	-
clientInformPossibleTestResult	Client informed about possible test results	Oval (single)	-	-
informConsentHivTest	Informed consent for HIV testing given	Oval (single)	-	-
knowledgeAssessment	Knowledge Assessment Score (sum of 6)	Digit box	-	-

Table 11: Section A: Knowledge Assessment – field details

Variable	Format	OCR	Notes
previousTestedHIVNegative	No(0)/Yes(1)	select one or zero	
timeLastHIVNegativeTestResult	<1 month / 1-3 months / 4-6 months / >6 months	select one or zero	
clientInformHivTransRoutes	No(0)/Yes(1)	select one or zero	
clientInformRiskkHivTrans	No(0)/Yes(1)	select one or zero	
clientInformPreventingHivTrans	No(0)/Yes(1)	select one or zero	
clientInformPossibleTestResult	No(0)/Yes(1)	select one or zero	
informConsentHivTest	No(0)/Yes(1)	select one or zero	
knowledgeAssessment	1 box	int (0-6)	Auto-derived sum

**Clinical rule printed on form:** Test for HIV if Personal HIV Risk Assessment Score is  $\geq 1$  and last HIV test was more than 3 months ago.

## Section B – Personal HIV Risk Assessment (Last 3 months)

Table 12: Section B: Personal Risk Assessment – field overview

Variable	Label	Type	Re- quired	Condi- tional
everHadSexualIntercourse	Ever had sexual intercourse	Oval (single)	–	–
moreThanOneSexPartner- LastThreeMonths	More than 1 sex partner	Oval (single)	–	–
unprotectedVaginalSex	Unprotected Vaginal Sex	Oval (single)	–	–
uprotectedAnalSex	Unprotected Anal Sex	Oval (single)	–	–
bloodtransInlastThree- Months	Blood transfusion in last 3 months	Oval (single)	–	–
sexUnderInfluence	Sex under influence of drugs and alcohol	Oval (single)	–	–
stiLastThreeMonths	History of STI	Oval (single)	–	–
uprotectedSexWithCasu- alLastThreeMonths	Unprotected sex with casual partner	Oval (single)	–	–
uprotectedSexWithRegu- larPartnerLastThree- Months	Unprotected sex with regular partner	Oval (single)	–	–
riskAssessment	Personal HIV Risk Assessment Score (sum of 9)	Digit box	–	–

Table 13: Section B: Personal Risk Assessment – field details

Variable	Format	OCR	Notes
everHadSexualIntercourse	No(0)/Yes(1)	select one or zero	
moreThanOneSexPartner- LastThreeMonths	No(0)/Yes(1)	select one or zero	
unprotectedVaginalSex	No(0)/Yes(1)	select one or zero	
uprotectedAnalSex	No(0)/Yes(1)	select one or zero	
bloodtransInlastThree- Months	No(0)/Yes(1)	select one or zero	
sexUnderInfluence	No(0)/Yes(1)	select one or zero	
stiLastThreeMonths	No(0)/Yes(1)	select one or zero	
uprotectedSexWithCasu- alLastThreeMonths	No(0)/Yes(1)	select one or zero	
uprotectedSexWithRegu- larPartnerLastThree- Months	No(0)/Yes(1)	select one or zero	
riskAssessment	1 box	int (0-9)	Calculated sum

## Page 2 Variables

### Section C – Clinical TB Screening

Table 14: Section C: TB Screening – field overview

Variable	Label	Type	Re-quired	Condi-tional
currentCough	Current cough	Oval (single)	–	–
weightLoss	Weight loss	Oval (single)	–	–
fever	Fever	Oval (single)	–	–
nightSweats	Night sweats	Oval (single)	–	–
lymphadenopathy	Lymphadenopathy	Oval (single)	–	–
tbScreening	TB screening score (sum of 5)	Digit box	–	–

Table 15: Section C: TB Screening – field details

Variable	Format	OCR	Notes
currentCough	No(0)/Yes(1)	select one or zero	
weightLoss	No(0)/Yes(1)	select one or zero	
fever	No(0)/Yes(1)	select one or zero	
nightSweats	No(0)/Yes(1)	select one or zero	
lymphadenopathy	No(0)/Yes(1)	select one or zero	
tbScreening	1 box	int (0-5)	If $\geq 1$ refer for Xpert MTB RIF

**Clinical rule:** If TB screening score  $\geq 1$ , test for Xpert MTB RIF or refer to TB service.

## Section C – Syndromic STI Screening

Table 16: Section C: STI Screening – field overview

Variable	Label	Type	Re-quired	Condi-tional
urethralDischargeFemale	Female: Vaginal discharge or burning when urinating?	Oval (single)	–	–
lowerAbdominalPains	Female: Lower abdominal pains with/without vaginal discharge?	Oval (single)	–	–
urethralDischargeMale	Male: Urethral discharge or burning when urinating?	Oval (single)	–	–
complaintsOfScrotal	Male: Scrotal swelling and pain	Oval (single)	–	–
complaintsGenitalSore	Genital sore(s) or swollen inguinal lymph nodes?	Oval (single)	–	–
stiScreening	STI screening score (sum of 5)	Digit box	–	–

Table 17: Section C: STI Screening – field details

Variable	Format	OCR	Notes
urethralDischargeFemale	No(0)/Yes(1)	select one or zero	
lowerAbdominalPains	No(0)/Yes(1)	select one or zero	
urethralDischargeMale	No(0)/Yes(1)	select one or zero	
complaintsOfScrotal	No(0)/Yes(1)	select one or zero	
complaintsGenitalSore	No(0)/Yes(1)	select one or zero	
stiScreening	1 box	int (0-5)	If >=1 follow STI mgmt guidelines

**Clinical rule:** If STI screening score  $\geq 1$ , follow syndromic STI management guidelines or refer.

### Section D – Sex Partner Risk (Last 3 months)

*Preamble on form: “Have you had sex with a partner who is HIV positive and falls in any of the categories below?”*

Table 18: Section D: Sex Partner Risk – field overview

Variable	Label	Type	Re-quired	Condi-tional
sexPartnerHivPositive	Had sex with a partner who is HIV positive?	Oval (single)	–	–
newDiagnosedHivlast-ThreeMonths	Partner newly diagnosed HIV, started treatment <3-6 months ago?	Oval (single)	–	–
currentlyArvForPmtct	Partner pregnant and receiving ARV for PMTCT?	Oval (single)	–	–
partnerAdolescent	Partner adolescent 10-19 yrs, HIV infected, on ARV or NOT?	Oval (single)	–	–
knowHivPositiveOnArv	Partner on ARV with unsuppressed VL?	Oval (single)	–	–
knowHivPositiveAfter-LostToFollowUp	Partner returned to treatment after Lost to Follow Up?	Oval (single)	–	–
sexPartnerUnpro- tectedAnal	Unprotected anal sex?	Oval (single)	–	–
sexPartnerRiskAssessment	Sex Partner Risk Assessment Score (sum of 7)	Digit box	–	–

Table 19: Section D: Sex Partner Risk – field details

Variable	Format	OCR	Notes
sexPartnerHivPositive	No(0)/Yes(1)	select one or zero	

newDiagnosedHivlast- ThreeMonths	No(0)/Yes(1)	select one or zero	
currentlyArvForPmtct	No(0)/Yes(1)	select one or zero	
partnerAdolescent	No(0)/Yes(1)	select one or zero	
knowHivPositiveOnArv	No(0)/Yes(1)	select one or zero	
knowHivPositiveAfter- LostToFollowUp	No(0)/Yes(1)	select one or zero	
sexPartnerUnpro- tectedAnal	No(0)/Yes(1)	select one or zero	
sexPartnerRiskAssessment	1 box	int (0-7)	Calculated sum

## Post-Test Counseling

Table 20: Post-Test Counseling – field overview

Variable	Label	Type	Required	Conditional
hivTestResult	HIV Test Result	Oval (single)	–	–
hivTestBefore	Tested for HIV before within this year?	Oval (single)	–	–
hivRequestResult	HIV Request and Result form signed by tester(s)	Oval (single)	–	–
hivRequestResultCt	HIV Request and Result form filled with CT Intake Form	Oval (single)	–	–
clientReceivedHivTestResult	Client received HIV test result	Oval (single)	–	–
postTestCounseling	Post test counseling done	Oval (single)	–	–
riskReduction	Risk reduction plan developed	Oval (single)	–	–
postTestDisclosure	Post test disclosure plan developed	Oval (single)	–	–
bringPartnerHivtesting	Will bring partner(s) for HIV testing	Oval (single)	–	–
childrenHivtesting	Will bring own children under 5 years for HIV testing	Oval (single)	–	–
informationFp	Provided with information on FP and dual contraception	Oval (single)	–	–
partnerFpThanCondom	Client/Partner use FP methods (other than condom)	Oval (single)	–	–
partnerFpUseCondom	Client/Partner use condoms as FP method	Oval (single)	–	–
correctCondomUse	Correct condom use demonstrated	Oval (single)	–	–
condomProvidedToClient	Condoms provided to client	Oval (single)	–	–
condomProvidedToClient- Count	How many condoms provided to client	Digit boxes	–	–
lubricantProvidedToClient	Lubricants provided to client	Oval (single)	–	–
lubricantProvidedToClient- Count	How many lubricants provided to client	Digit boxes	–	–
referredToServices	Client referred to other services	Oval (single)	–	–
discordantCouple	Discordant couple?	Oval (single)	–	–
prepReferred	Referred for PrEP?	Oval (single)	–	–

Table 21: Post-Test Counseling – field details

Variable	Format	OCR	Notes
hivTestResult	Negative / Positive	select one or zero	
hivTestBefore	Not prev. tested / Prev. negative / Prev. positive in care / Prev. positive not in care	select one or zero	
hivRequestResult	No(0)/Yes(1)	select one or zero	
hivRequestResultCt	No(0)/Yes(1)	select one or zero	
clientReceivedHivTestResult	No(0)/Yes(1)	select one or zero	
postTestCounseling	No(0)/Yes(1)	select one or zero	
riskReduction	No(0)/Yes(1)	select one or zero	
postTestDisclosure	No(0)/Yes(1)	select one or zero	
bringPartnerHivtesting	No(0)/Yes(1)	select one or zero	
childrenHivtesting	No(0)/Yes(1)	select one or zero	
informationFp	No(0)/Yes(1)	select one or zero	
partnerFpThanCondom	No(0)/Yes(1)	select one or zero	
partnerFpUseCondom	No(0)/Yes(1)	select one or zero	
correctCondomUse	No(0)/Yes(1)	select one or zero	
condomProvidedToClient	No(0)/Yes(1)	select one or zero	
condomProvidedToClient-Count	2 boxes	int	
lubricantProvidedToClient	No(0)/Yes(1)	select one or zero	
lubricantProvidedToClient-Count	2 boxes	int	
referredToServices	No(0)/Yes(1)	select one or zero	
discordantCouple	No(0)/Yes(1)	select one or zero	
prepReferred	No(0)/Yes(1)	select one or zero	

### Clinical rules printed on form:

- If client tests HIV negative and has a Risk Assessment Score  $\geq 1$  or evidence of an STI syndrome, recommend re-testing after 3 months.
- If HIV negative and score  $\geq 1$  in Section D, refer client for PrEP services.

### Recency, Syphilis, CD4 and Hepatitis Testing

Table 22: Recency and Lab – field overview

Variable	Label	Type	Re-quired	Condi-tional
recencyTest	Recency test with RTRI	Oval (single)	–	Yes
syphilisTestResult	Syphilis Test Result	Oval (single)	–	–
cd4SemiQuantitative	CD4 Test Result – Semi-Quantitative	Oval (single)	–	Yes

cd4FlowCytometry	CD4 Test Result – Flow Cytometry (cells/m3)	Digit boxes	–	Yes
hepatitisBTest	Hepatitis B Virus Test Result	Oval (single)	–	–
hepatitisCTest	Hepatitis C Virus Test Result	Oval (single)	–	–

Table 23: Recency and Lab – field details

Variable	Format	OCR	Notes
recencyTest	Recent / Long Term / Invalid	select one or zero	HIV-positive clients only
syphilisTestResult	Non-Reactive / Reactive	select one or zero	
cd4SemiQuantitative	<=200 / >200	select one or zero	HIV-positive clients only
cd4FlowCytometry	4 boxes	int	HIV-positive clients only
hepatitisBTest	Negative / Positive	select one or zero	
hepatitisCTest	Negative / Positive	select one or zero	

**Conditional fields:** recencyTest, cd4SemiQuantitative, and cd4FlowCytometry are printed on the form but should only be completed for **HIV-positive clients**.

## Form Completion

Table 24: Completion – field overview

Variable	Label	Type	Re-quired	Condi-tional
comment	Comments	Handwriting	–	–
completedBy	Completed by	Handwriting	–	–
designation	Designation	Handwriting	–	–
sign	Sign	Handwriting	–	–
endDate	Date (completion)	Date boxes	–	–
photoTaken	Photo taken	Oval (single)	–	–
discardPage	Discard both pages	Oval (single)	–	–

Table 25: Completion – field details

Variable	Format	OCR	Notes
comment	170 mm line	Not banded	Outside scannable area
completedBy	40 mm line	Not banded	Outside scannable area

designation	40 mm line	Not banded	Outside scannable area
sign	40 mm line	Not banded	Outside scannable area
endDate	dd/mm/yyyy	date dd, date mm, two digits 20, date nearby yy	Century 20 pre-printed
photoTaken	(check)	select one or zero	Administrative
discardPage	(check)	select one or zero	Discards full photopack entry

## Summary Statistics

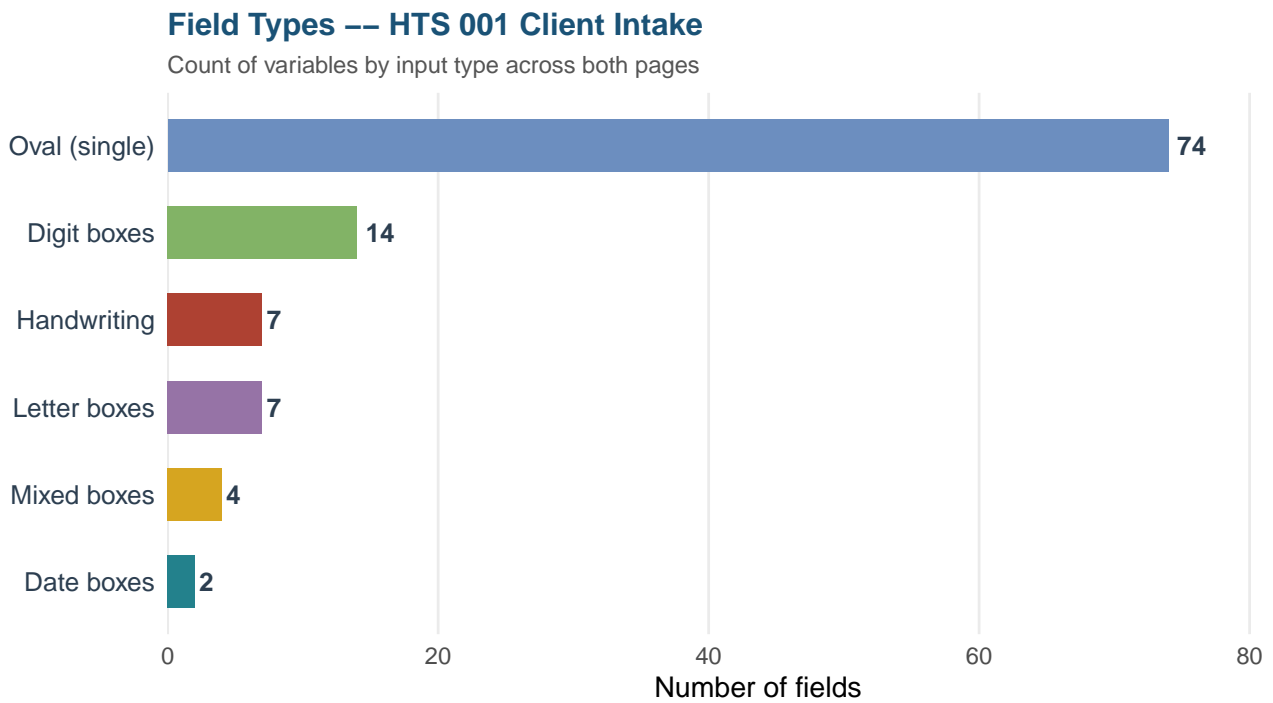


Figure 1: Distribution of field types across all form variables

Table 26: Form variable summary

Metric	Count
<b>Total fields</b>	<b>108</b>
<b>OCR-banded fields</b>	<b>100</b>
Handwriting (not banded)	7
Oval (single-select)	74
Digit boxes	14

Letter boxes	7
Mixed boxes	4
Date boxes	2
Conditional fields	8
Required fields	0

---

## OCR Models Reference

Table 27: OCR models used in this form

OCR Model	Characters Detected	Typical Use in this Form
<b>select one or zero</b>	Oval marked (X) or empty – enforces max one X per question group	All single-select oval questions
<b>int</b>	Digits 0-9	Ages, counts, phone numbers, codes, scores
<b>string</b>	Letters A-Z	Names, text identifiers, specify fields
<b>string int</b>	Letters A-Z and digits 0-9	Mixed alphanumeric fields
<b>tokenwise string int</b>	Mixed tokens – each space-separated token classified as all-letters or all-digits	Street addresses, landmarks, LGA and state fields
<b>date dd</b>	Integer 01-31 (day of month)	Day part of date fields
<b>date mm</b>	Integer 01-12 (month)	Month part of date fields
<b>date nearby yy</b>	2-digit year close to current year (+/- 1)	Year part of date fields
<b>two digits 20</b>	Always detects 20 – used for pre-printed century digits in year fields	Pre-printed 20 in year boxes
<b>int (score range)</b>	Integer sum of preceding Yes/No ovals in assessment sections	Knowledge, risk, TB, STI, sex partner score boxes
<b>Not banded</b>	Free handwriting – field is not OCR-processed	Clinician notes, signatures, facility names, free text

**Why select one or zero instead of single?** All single-choice oval questions use **select one or zero** rather than the basic **single** model. This model enforces a maximum of one filled oval per question group, automatically flagging entries where a respondent has accidentally crossed more than one bubble – improving downstream data quality without requiring manual verification.

# In-Phone Validation

HTS 001 Client Intake Form · NG-HTS · Nigeria · v1.6

## How In-Phone Validation Works

**Form:** NG-HTS 001 Client Intake | **Version:** v1.6 | **Country:** Nigeria

This page documents the automated real-time validation rules applied by the ScanForm mobile app at the moment a field worker photographs this form. If any check fails, the app immediately alerts the worker and prompts them to correct the paper form and retake the photo — **before the data is submitted to the server**. No human verifier is involved in this process.

The table below summarises the in-phone validation workflow.

Table 1: In-phone validation workflow

Step	Actor	Action
1	Field worker	Photographs the completed paper form.
2	Mobile app	Lightweight OCR runs on device to detect filled/unfilled fields.
3	Mobile app	All configured checks are evaluated against the detected marks.
4	Mobile app	If all checks pass, data is submitted to the server.
5	Field worker	If any check fails, an alert is shown immediately on screen.
6	Field worker	Worker corrects the paper form and retakes the photo.

**Why validate on the phone?** Catching errors at the point of data capture — while the client and clinician are still present — is far cheaper than discovering them during data quality audits weeks later. In-phone validation closes the feedback loop instantly and prevents systematically missing or multi-selected fields from entering the database.

## Key Concepts

Table 2: In-phone validation – concept glossary

Concept	Meaning
<b>Record eligibility</b>	A patient row is only validated if a minimum number of its key fields contain data. Fully blank rows (unregistered slots) are silently skipped to avoid false alerts.
<b>boxals threshold for eligibility</b>	The minimum number of filled boxes in a field before that field counts as active for eligibility purposes. Example: a date field with threshold 2 is only considered active if at least 2 of its 8 boxes are filled.
<b>check boxes enough filled</b>	Required field check – fires if the field has zero filled boxes. Used for digit/letter box fields that must not be left blank.
<b>check checkboxes enough answers</b>	Minimum-one bubble check – fires if no bubble is crossed in the group. Used for mandatory single-select questions.
<b>check checkboxes not too many answers</b>	Maximum-one bubble check – fires if more than one bubble is crossed in the group. Used for all single-select questions to prevent accidental multi-selection.
<b>Exactly-one pair</b>	When both enough answers and not too many answers are applied to the same field, the combined effect is: exactly one bubble must be selected – no more, no less.
<b>At-most-one (solo)</b>	When only not too many answers is applied, the field is optional but cannot have more than one bubble selected.
<b>Discard criteria</b>	If the discard bubble is marked on the form, all validation checks for that record are suppressed – the row is intentionally void.
<b>PageNumberBlock(0)</b>	Alert message references the left-hand page (Page 1) of the two-page photopack.
<b>PageNumberBlock(1)</b>	Alert message references the right-hand page (Page 2) of the two-page photopack.

## Record Eligibility

Before any field-level check is evaluated, the app determines whether a given record (patient row) contains enough data to warrant validation. This prevents alerts on genuinely empty form slots.

### Eligibility rule for this form:

- A record is considered **active** when at least one key question field is filled (`checked_groups_threshold = 1`, with `checked_groups = None` meaning any registered field counts).
- If the **Discard both pages** bubble (`discardPage2`) is marked, the record is unconditionally skipped – no checks run.

- The minimum tries before a record can be skipped is **3** – the app will attempt to process the record at least three times before allowing it to be bypassed.

## Validated Fields

### Page 1 – Validated Fields

Table 3: Page 1 – validated fields (32 fields)

Variable	Field Label	Check type
dateVisit	Date of Visit	Required
referredFrom	Referred From	Exactly one
testingSetting	Setting	Exactly one
modality	Modality	Required
sex	Sex	At most one
firstTimeVisit	First Time Visit	At most one
maritalStatus	Marital Status	At most one
employmentStatusId	Employment Status	At most one
educationId	Education Level	At most one
typeConsueling	Type of Session	At most one
previouslyTested	Previously tested within the last 3 months	At most one
indexClient	Index Testing: Is client identified from an index client?	At most one
relationWithIndexClient	If yes – relation with index client	At most one
pregnant	Client is Pregnant	At most one
breastFeedingUnder6	Client breastfeeding < 6 months	At most one
breastFeedingOver6	Client breastfeeding > 6 months	At most one
previousTestedHIVNegative	Previously tested HIV negative	Exactly one
timeLastHIVNegativeTestResult	Time of last HIV Negative test Results	At most one
clientInformHivTransRoutes	Client informed about HIV transmission routes	At most one
clientInformRiskkHivTrans	Client informed about risk factors for HIV transmission	At most one
clientInformPreventingsHivTrans	Client informed on preventing HIV transmission methods	At most one
clientInformPossibleTestResult	Client informed about possible test results	At most one
informConsentHivTest	Informed consent for HIV testing given	At most one
everHadSexualIntercourse	Ever had sexual intercourse	Exactly one
moreThanOneSexPartnerLastThreeMonths	More than 1 sex partner	Exactly one
unprotectedVaginalSex	Unprotected Vaginal Sex	Exactly one
uprotectedAnalSex	Unprotected Anal Sex	Exactly one
bloodtransInlastThreeMonths	Blood transfusion in last 3 months	Exactly one
sexUnderInfluence	Sex under the influence of drugs and alcohol	Exactly one
stiLastThreeMonths	History of STI	Exactly one

Table 3: Page 1 – validated fields (32 fields) (continued)

Variable	Field Label	Check type
unprotectedSexWithRegularPartner-LastThreeMonths	Unprotected sex with regular partner in last 3 months	Exactly one
unprotectedSexWithCasualLastThreeMonths	Unprotected sex with casual partner in last 3 months	Exactly one

## Page 2 – Validated Fields

Table 4: Page 2 – validated fields (40 fields)

Variable	Field Label	Check type
currentCough	Current cough	At most one
weightLoss	Weight loss	At most one
fever	Fever	At most one
nightSweats	Night sweats	At most one
lymphadenopathy	Lymphadenopathy	At most one
urethralDischargeFemale	Female: Vaginal discharge or burning when urinating?	At most one
lowerAbdominalPains	Female: Lower abdominal pains with/without vaginal discharge?	At most one
urethralDischargeMale	Male: Urethral discharge or burning when urinating?	At most one
complaintsOfScrotal	Male: Scrotal swelling and pain	At most one
complaints genital	Genital sore(s) or swollen inguinal lymph nodes?	At most one
sexPartnerHivPositive	Had sex with a partner who is HIV positive?	At most one
newDiagnosedHivlastThreeMonths	Partner newly diagnosed HIV, started treatment <3-6 months ago?	At most one
currentlyArvForPmtct	Partner pregnant and receiving ARV for PMTCT?	At most one
partnerAdolescent	Partner adolescent 10-19 yrs, HIV infected, on ARV or NOT?	At most one
knowHivPositiveOnArv	Partner on ARV with unsuppressed VL?	At most one
knowHivPositiveAfterLostToFollowUp	Partner returned to treatment after Lost to Follow Up?	At most one
sexPartnerUnprotectedAnalSex	Unprotected anal sex (sex partner section)	At most one
hivTestResult	HIV Test Result	Exactly one
hivTestBefore	Have you been tested for HIV before within this year?	At most one
hivRequestResult	HIV Request and Result form signed by tester(s)	At most one
hivRequestResultCt	HIV Request and Result form filled with CT Intake Form	At most one
clientReceivedHivTestResult	Client received HIV test result	At most one
postTestCounseling	Post test counseling done	At most one
riskReduction	Risk reduction plan developed	At most one
postTestDisclosure	Post test disclosure plan developed	At most one
bringPartnerHivtesting	Will bring partner(s) for HIV testing	At most one
childrenHivtesting	Will bring own children <5 years for HIV testing	At most one
informationFp	Provided with information on FP and dual contraception	At most one

Table 4: Page 2 – validated fields (40 fields) (*continued*)

Variable	Field Label	Check type
partnerFpThanCondom	Client/Partner use FP methods (other than condom)	At most one
partnerFpUseCondom	Client/Partner use condoms as (one) FP method	At most one
correctCondomUse	Correct condom use demonstrated	At most one
condomProvidedToClient	Condoms provided to client	At most one
lubricantProvidedToClient	Lubricants provided to client	At most one
referredToServices	Client referred to other services	At most one
discordantCouple	Discordant couple?	At most one
recencyTest	Recency test with RTRI	At most one
cd4SemiQuantitative	CD4 Test Result - Semi-Quantitative	At most one
syphilisTestResult	Syphilis Test Result	At most one
hepatitisBTest	Hepatitis B Virus Test Result	At most one
hepatitisCTest	Hepatitis C Virus Test Result	At most one
providerId	Provider ID	Required

**Row shading key:** Green = Exactly one (mandatory single-select); Pink = Required (must not be blank); Unshaded = At most one (optional single-select).

## Fields Not Validated In-Phone

The fields listed below exist on the form and are OCR-processed, but are **not** subject to in-phone automated validation. This is a deliberate design choice – either the field is considered optional, its validation is handled downstream in the data pipeline, or adding an in-phone check was judged likely to produce excessive false alerts in field conditions.

Table 5: Fields present on form but not validated in-phone (19 fields)

Variable(s)	Field Label	Why not validated in-phone
photoTaken1 / photoTaken2	Photo taken (both pages)	Administrative flag – not a data quality concern
referredFromOther	Referred From – specify	Optional free-text; validated conditionally downstream
testingSettingOther	Setting – specify	Optional free-text; validated conditionally downstream
firstName / otherName / surname	Client names	Free-text letter boxes; completeness not enforced in-phone
age	Age (years)	Digit box; range checks handled in pipeline
numWives / numChildren	No. of wives / No. of children	Optional numeric fields
clientCode	Client Code	Pre-printed; format check handled in pipeline
recency	Recency Number	Optional identifier field

Table 5: Fields present on form but not validated in-phone (19 fields) (*continued*)

Variable(s)	Field Label	Why not validated in-phone
phoneNumber	Client Telephone Number	Optional contact detail
nearestLandmark / state / line / lga	Address fields	Free-text; completeness not enforced in-phone
indexClientCodeNew / indexClientCodeOld	Index Client Code (ScanForm / Legacy)	Conditional; only one should be filled; handled downstream
knowledgeAssessment	Knowledge Assessment Score	Derived sum; consistency check handled in pipeline
riskAssessment	Personal HIV Risk Assessment Score	Derived sum; consistency check handled in pipeline
tbScreening / stiScreening	TB / STI screening scores	Derived sums; consistency check handled in pipeline
sexPartnerRiskAssessment	Sex Partner Risk Assessment Score	Derived sum; consistency check handled in pipeline
condomProvidedToClientCount	How many condoms provided	Optional count; range check handled in pipeline
lubricantProvidedToClientCount	How many lubricants provided	Optional count; range check handled in pipeline
cd4FlowCytometry	CD4 Flow Cytometry (cells/m3)	Conditional (HIV+ only); range check handled in pipeline
date (completion)	Date signed by provider	Completion date; validated downstream

## Summary

### In-Phone Validation Checks by Page and Type

HTS 001 Client Intake -- fields with active in-phone checks

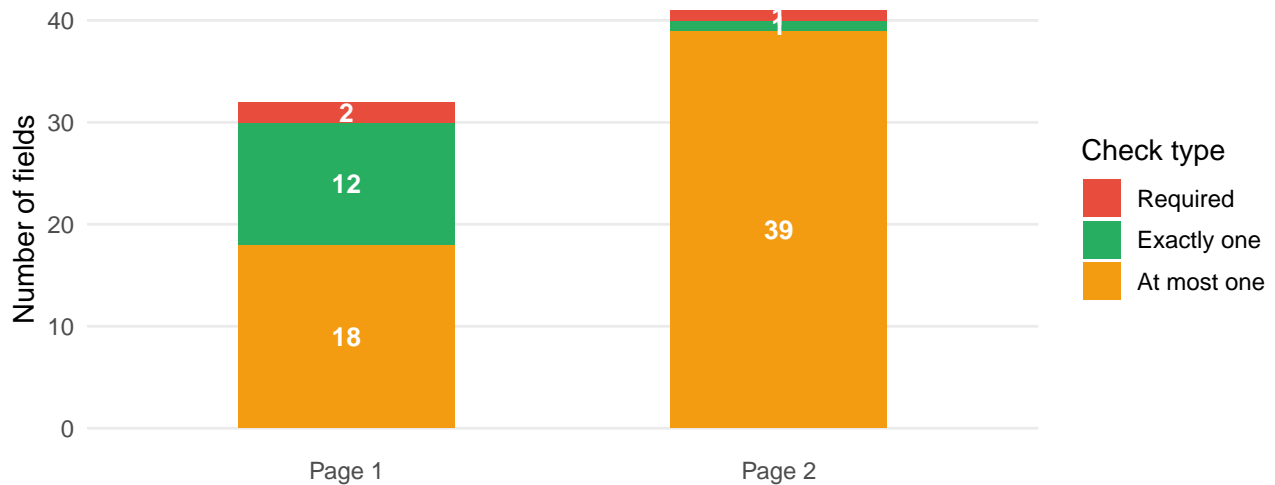


Figure 1: Distribution of in-phone check types across validated fields

Table 6: In-phone validation summary – HTS 001 Client Intake

Metric	Value
<b>Total fields on form (OCR-banded)</b>	<b>92</b>
<b>Fields with in-phone validation</b>	<b>73</b>
Fields NOT validated in-phone	19
Required checks (must not be blank)	3
Exactly one checks (mandatory single-select)	13
At most one checks (optional single-select)	57
Page 1 validated fields	32
Page 2 validated fields	41
Discard criteria fields	1 (discardPage2)
Minimum attempts before record can be skipped	3

---

## Important Limitation

In-phone validation uses a **lightweight on-device OCR** that determines only whether a box contains a mark or not — it does not read the actual character written. This means checks can confirm that *a* bubble was selected, but cannot verify *which* bubble was selected. Full character-level OCR and value range checks are applied after submission during server-side processing.

# Data Pipeline

HTS 001 Client Intake Form · NG-HTS · Nigeria · v1.6

## Pipeline Overview

**Form:** NG-HTS 001 Client Intake | **Version:** v1.6 | **Country:** Nigeria

This page describes how raw ScanForm scan data is transformed into analysis-ready tables through a series of dbt model layers. No dbt SQL models were provided for this form at time of documentation generation — the pipeline architecture below reflects the **standard ScanForm fast analytics pipeline** structure that applies to all forms on the platform.

**Note:** No dbt models were provided for this form. The sections below describe the expected pipeline structure and transformation logic inferred from the form design (XLS-Form, banding, and export configuration). This documentation will be updated automatically when dbt model files are added to the repository.

Table 1: Pipeline layers – HTS 001 Client Intake

Stage	Feeds into	Description
<b>Source: ScanForm Export API</b>	Base layer	Externally materialised export from ScanForm API; all fields as raw text.
<b>Base (raw export columns)</b>	Pre-clean layer	Thin wrapper selecting and renaming all export columns to snake case.
<b>Pre-clean (link pages, filter)</b>	Clean layer + Checks layer	Deduplication, discard filtering, empty record removal.
<b>Clean (cast, impute, reshape)</b>	Refinery clean	Type casting, date parsing, oval decoding, derived business logic flags.
<b>Checks (validation rules)</b>	Refinery checks	Rule-based validation and cross-field consistency checks.
<b>Refinery clean (mart-ready data)</b>	Metabase dashboards	Analysis-ready client intake dataset with date grouping and location joins.

<b>Refinery checks (mart-ready DQA)</b>	Metabase dashboards	Mart-ready DQA table with check counts and geographic breakdowns.
---	---------------------	---

## Pipeline Layers

The HTS 001 Client Intake data pipeline follows the **standard QED fast analytics architecture** — a sequence of dbt model layers, each with a well-defined responsibility. Data flows from the ScanForm export API through five transformation stages before reaching Metabase dashboards.

### Base Layer

The base model is a **thin wrapper** over the externally materialised ScanForm export. It selects all columns produced by the export configuration and applies minimal renaming to ensure consistent snake case column names throughout the pipeline.

#### What happens here:

- All columns arrive as raw text strings — no casting, no cleaning
- Underscores from incomplete or uncertain OCR are preserved (e.g. `_12_` rather than `12`)
- Both Page 1 and Page 2 columns are present in the same export row for this photopack form
- Administrative columns included: `page id`, `entry url`, `entry submission date`, `entry submitted by`, `row id`

Table 2: Key column groups arriving from the ScanForm export

Column group	Example raw values	Notes
<b>Date of Visit</b>	21/05/24, /05/24	Underscores indicate OCR uncertainty
<b>Client Code</b>	123-456-789-AB	Pre-printed; validated against expected format
<b>All oval fields</b>	0, 1, _	0 = No, 1 = Yes, underscore = uncertain
<b>Score fields</b>	3, _, 6	Integer sum of section responses
<b>Text box fields</b>	JOHN, JO_N	Underscores mark uncertain characters
<b>Handwriting fields</b>	(absent)	Handwriting zones are not OCR-processed

## Pre-Clean Layer

The pre-clean model is the **staging and filtering** layer. It applies record-level inclusion rules before any analytical transformations occur, ensuring that downstream clean and checks models operate on a consistent, well-scoped dataset.

### What happens here:

- **Duplicate exclusion:** Only entries with `newest entry status` IN ('NEWEST', 'DISABLED') are retained, removing superseded re-scans of the same page
- **Issue filtering:** Entries with scan processing failures are excluded
- **Discarded record exclusion:** Rows where `discard page = 1` are dropped
- **Empty record exclusion:** Rows where a minimum proportion of key fields are blank may be filtered
- **Column pruning:** Administrative columns not needed downstream are removed

This is the **last point** at which record inclusion logic is applied. Both the clean model and the checks model are built from the same pre-clean base, ensuring totals are consistent between the analytical dataset and the data quality report.

---

## Clean Layer

The clean model produces the **analysis-ready dataset** — one row per client intake encounter, with all fields correctly typed, reshaped, and labelled.

### Date parsing

Table 3: Date parsing transformations in the clean layer

Raw column	Transformation	Output column
date visit	Strip underscores, parse dd/mm/yyyy, cast to DATE	date visit clean
end date	Strip underscores, parse dd/mm/yyyy, cast to DATE	end date clean
Submission date	Used as fallback if date visit is invalid or missing	date fallback

## Numeric casting

All digit-box fields arrive as strings with possible leading/trailing underscores (e.g. `_2_`). The clean layer applies `numberFullyStripped`-style logic: removes surrounding underscores, casts to integer, and retains NULL for values that cannot be parsed (no imputation).

Affected fields: age, num wives, num children, recency, phone number, knowledge assessment, risk assessment, tb screening, sti screening, sex partner risk assessment, condom count, lubricant count, cd4 flow cytometry.

## Oval decoding

All `select one or zero` fields arrive as `"0"`, `"1"`, or `"_"`. The clean layer casts these to: `0` = integer 0 (No); `1` = integer 1 (Yes); `"_"` = NULL (uncertain / not verified).

## Derived and business logic fields

Table 4: Derived fields computed in the clean layer

Derived field	Logic	Purpose
<code>is index client</code>	<code>index client = 1</code>	Flag for index testing sub-population
<code>is hiv positive</code>	<code>hiv test result = 1</code>	Gate for recency, CD4, and hepatitis follow-up
<code>is first visit</code>	<code>first time visit = 1</code>	Distinguishes new vs returning clients
<code>knowledge score complete</code>	<code>knowledge assessment IS NOT NULL</code>	Whether the 6-item score was fully recorded
<code>risk score complete</code>	<code>risk assessment IS NOT NULL</code>	Whether the 9-item score was fully recorded
<code>refer for prep</code>	<code>prep referred = 1</code>	PrEP linkage indicator
<code>referred for tb</code>	<code>tb screening &gt;= 1</code>	TB referral flag per clinical rule
<code>referred for sti</code>	<code>sti screening &gt;= 1</code>	STI referral flag per clinical rule
<code>recommend retest</code>	<code>hiv test result = 0 AND (risk assessment &gt;= 1 OR sti screening &gt;= 1)</code>	3-month retest recommendation flag

## Conditional field handling

Table 5: Conditional fields and their validity flags

Field	Condition	Validity flag
<code>relation with index client</code>	<code>is index client = TRUE</code>	<code>relation valid</code>
<code>recency test</code>	<code>is hiv positive = TRUE</code>	<code>recency valid</code>

cd4 semi quantitative	is hiv positive = TRUE	cd4 valid
cd4 flow cytometry	is hiv positive = TRUE	cd4 valid

---

## Checks Layer

The checks model is a **data quality audit table** — a union of many individual check result rows, each describing one data quality issue found in one record. It is built from the pre-clean model (not the clean model) so that checks operate on the least-transformed data possible.

Table 6: Check categories expected in the checks layer

Check type	Fields checked	Description
<b>Date validity</b>	date visit, end date	Parsed date is within plausible range (not in future, not before 2010)
<b>Date completeness</b>	date visit	Day, month, or year component is missing (underscore)
<b>Score range</b>	knowledge assessment, risk assessment, tb screening, sti screening, sex partner risk assessment	Score exceeds the maximum possible sum for its section
<b>Score consistency</b>	All score fields vs constituent items	Score does not match count of 1 responses in constituent items
<b>Select-one violation</b>	All select one or zero fields	More than one oval marked in a single-select group
<b>Conditional completeness (index)</b>	relation with index client, index client code new, index client code old	Field is blank when index client = 1
<b>Conditional completeness (HIV+)</b>	recency test, cd4 semi quantitative	Field is blank when hiv test result = 1
<b>Cross-field logic</b>	breastfeeding under 6 + breastfeeding over 6	Both marked Yes simultaneously
<b>Cross-field logic</b>	hiv test result + recency test	Recency result recorded but HIV result is Negative
<b>Client code format</b>	client code	Does not match expected pattern ###-###-###-XX
<b>Phone number length</b>	phone number	Fewer than 11 digits recorded

**Check output structure:** Each row in the checks table contains: `page id`, `entry url`, `entry submission date`, `row id` (metadata to locate the record); a plain-language check description; check tags (field names involved, check category); and severity (Warning or Error).

## Refinery Layer

Refinery models are **mart-ready tables** placed in the `marts/` directory. They are the only models directly accessible in Metabase. They consume both the clean model and the checks model.

**refinery clean** — **analytical dataset** adds to the clean model:

- Stable date grouping columns for Metabase filtering: `visit month`, `visit year`, `visit quarter` (derived from `date visit clean`; falls back to `entry submission date` when visit date is invalid)
- Modality label decoded from the `testing setting` code using the legend printed on the physical form (1 = Emergency, 2 = Index, ... 14 = PMTCT Post ANC1)
- Location metadata joins (facility name, LGA, state) where a master facility list is available
- All derived flags from the clean layer forwarded unchanged

**refinery checks** — **DQA dashboard dataset** adds to the checks model:

- Same stable date grouping columns for trend analysis of data quality over time
- Check counts aggregated by `page id` to enable record-level DQA scores
- Location metadata for geographic breakdowns of data quality

---

## Data Flow Summary

Table 7: Pipeline layer summary – HTS 001 Client Intake

Layer	Input	Key operations	Row grain
<b>Base</b>	ScanForm materialised export	Column selection, snake case rename	1 row per ScanForm entry
<b>Pre-Clean</b>	Base model	Deduplication, discard filtering, empty record removal	1 row per valid entry
<b>Clean</b>	Pre-clean model	Type casting, date parsing, oval decoding, derived flags	1 row per client intake
<b>Checks</b>	Pre-clean model	Rule-based validation, cross-field consistency checks	1 row per data quality issue
<b>Refinery</b>	Clean model + Checks model	Date grouping, label decoding, location joins, aggregation	1 row per client intake (clean) / 1 row per issue (checks)

---

## Notes for Data Managers

**Interpreting underscore values.** In all base and pre-clean outputs, **underscores ( \_ ) represent OCR uncertainty** — a character that ScanForm could not read with confidence and that has not yet been verified by a human operator. Never treat an underscore as a zero or a space. In the clean layer these become NULL.

**Externally materialised exports cannot be updated in place.** The ScanForm export feeding this pipeline is **externally materialised**. Any change to the export configuration requires creating a new versioned export (e.g. `hts001 v2`), backfilling all historical data, updating dashboard references, and only then retiring the previous version. This process can take several days for high-volume forms.

**Score field consistency.** The five assessment scores (knowledge assessment, risk assessment, tb screening, sti screening, sex partner risk assessment) are **handwritten sums** entered by the clinician. The checks layer validates these against the individual item responses. Discrepancies are common and should be reviewed as part of routine data quality assurance — the pipeline retains the clinician-recorded score as the primary value and flags inconsistencies rather than overwriting them.

# Data Quality Checks

HTS 001 Client Intake Form · NG-HTS · Nigeria · v1.6

## How Pipeline DQA Works

**Form:** NG-HTS 001 Client Intake | **Version:** v1.6 | **Country:** Nigeria

This page documents the **server-side data quality checks** applied after OCR data has been submitted and processed through the dbt pipeline. These checks are complementary to in-phone validation: they operate on the full processed dataset, apply more sophisticated cross-field logic, and produce a structured audit trail used for data quality reporting and review.

**Note:** No dbt SQL check models were provided for this form. The content below describes the standard check architecture and the checks that would be expected for this form based on its field definitions, OCR models, and clinical logic. This page will be updated automatically when SQL model files are added to the repository.

The table below summarises the end-to-end pipeline flow from scanned data through to the Metabase DQA dashboard.

Table 1: Pipeline processing flow for HTS 001 records

Stage	Description
<b>Scanned &amp; submitted data</b>	Photo submitted to server after passing in-phone validation.
<b>Base layer (raw text)</b>	Raw OCR text extracted for all banded fields; no transformations applied.
<b>Pre-clean (filter, deduplicate)</b>	Records filtered for duplicates and discarded pages; base dataset established.
<b>Clean layer (cast, derive)</b>	Values cast to target types (integer, date); derived fields computed.
<b>Checks layer (one row per failure)</b>	Each failing check produces one row; passing records produce no rows.
<b>Refinery checks (mart-ready DQA)</b>	Check results materialised into mart-ready tables for dashboard consumption.
<b>Metabase DQA dashboard</b>	DQA flags surfaced in dashboard for monitoring by data managers.

## In-phone vs Pipeline Checks

Table 2: In-phone validation vs pipeline DQA checks

Dimension	In-phone validation	Pipeline DQA checks
<b>When</b>	At moment of photo capture	After server-side OCR processing
<b>What OCR sees</b>	Mark / no mark only (lightweight)	Full character-level OCR values
<b>Scope</b>	Single record, single photo	Entire dataset, all submissions
<b>Effect on worker</b>	Immediate alert; must correct before submitting	No immediate effect; flagged for data manager review
<b>Effect on data</b>	Prevents bad data entering the system	Flags or excludes records from clean dataset
<b>Logic available</b>	Field presence only	Value ranges, cross-field rules, date validity, score consistency

## Check Severity Levels

Table 3: Check severity levels

Severity	Meaning	Typical use cases
<b>Warning</b>	Record is flagged and appears in the DQA report, but is retained in the clean analytical dataset. The value is preserved as-is. A data manager may choose to investigate or correct it.	OCR uncertainty on non-critical fields; scores that are plausible but inconsistent with individual items; optional fields with unexpected values.
<b>Error</b>	Record is flagged AND excluded from the clean analytical dataset. It will not appear in counts, indicators, or dashboard aggregations until the issue is resolved.	Unparseable dates on mandatory date fields; critical identifiers that cannot be decoded; records where core mandatory fields are entirely missing.

## Check Catalogue

The checks below are organised by check macro type. Each section explains what the check does, which fields it applies to, what failure looks like, and what the recommended action is.

## Integer Conversion Checks

Macro: `batch check convert to int`

These checks verify that OCR text values in digit-box fields can be parsed as integers. A failure means the OCR returned a value containing non-digit characters – typically an underscore (uncertain OCR), a letter misread as a digit, or a blank – that cannot be converted to a number for analysis.

**Severity:** Warning – record retained in clean dataset; field value stored as NULL.

Table 4: Integer conversion check fields (14 fields)

Variable	Field Label	Boxes	Expected range
age	Age (years)	3	1-120
numWives	No. of Wives/Co-wives	2	0-99
numChildren	No. of own children <5 years	2	0-99
recency	Recency Number	4	Any 4-digit integer
phoneNumber	Client Telephone Number	11	11-digit Nigerian number
postalCode	Postal Code	7	Any integer
knowledgeAssessment	Knowledge Assessment Score	1	0-6
riskAssessment	Personal HIV Risk Assessment Score	1	0-9
tbScreening	TB Screening Score	1	0-5
stiScreening	STI Screening Score	1	0-5
sexPartnerRiskAssessment	Sex Partner Risk Assessment Score	1	0-7
condomProvidedToClient-Count	How many condoms provided	2	0-99
lubricantProvidedToClient-Count	How many lubricants provided	2	0-99
cd4FlowCytometry	CD4 Flow Cytometry (cells/m3)	4	0-9999

---

## Date Conversion Checks

Macro: `batch check convert to date`

These checks verify that date fields can be parsed into valid calendar dates in `dd/mm/yyyy` format. Failures include OCR uncertainty in any part of the date, invalid day/month combinations (e.g. 31 February), and dates outside a plausible range for this programme.

**Severity:** Warning for partial reads; Error for `dateVisit` when the date is entirely unreadable.

Table 5: Date conversion check fields

Variable	Label	Format	OCR models	Pre-printed	Severity
dateVisit	Date of Visit	dd/mm/yyyy	date dd, date mm, two digits 20, date nearby yy	Year century '20'	Error if fully blank; Warning if partial
endDate	Date (completion)	dd/mm/yyyy	date dd, date mm, two digits 20, date nearby yy	Year century '20'	Warning

**Note on the two digits 20 model:** Both date fields have the year century pre-printed as **20** on the physical form. The **two digits 20** OCR model always returns 20 for those two boxes regardless of what is written, so the century component is never uncertain. Only the decade and unit digits of the year (`date nearby yy` boxes) can produce uncertainty.

---

## Multiple-Selection Checks

Macro: `batch check melt columns`

For select-one questions, ScanForm exports each answer option as a separate boolean column. The melt check detects records where more than one option column is set to 1 for the same question – meaning the clinician accidentally marked multiple bubbles, or OCR misread shading as a filled bubble.

**Severity:** Warning – the record is retained but the field value is set to NULL in the clean dataset, since the intended answer is ambiguous.

**Note:** The `select one or zero` OCR model already enforces a maximum of one filled bubble during the OCR recognition phase. The melt check provides a second layer of defence at the pipeline level.

The 25 select-one fields subject to this check are split across the two pages of the photopack.

### Page 1 fields (18):

Table 6: Multiple-selection check fields – Page 1 (18 fields)

Variable	Field Label	Options (exported columns)
<code>referredFrom</code>	Referred From	self, tb, sti, fp, opd, ward, blood bank, other
<code>testingSetting</code>	Setting	ct, tb, sti, fp, opd, ward, outreach, standalone hts, other
<code>modality</code>	Modality	1-14 (numeric code)
<code>sex</code>	Sex	male, female
<code>firstTimeVisit</code>	First Time Visit	yes, no
<code>maritalStatus</code>	Marital Status	married, divorced, widowed, separated, single
<code>educationId</code>	Education Level	none, primary school, junior secondary, higher secondary, post secondary, quranic
<code>employmentStatusId</code>	Employment Status	employed, unemployed, retired, freelance, student
<code>typeConsueling</code>	Type of Session	individual, couple, group, previously self tested

Table 6: Multiple-selection check fields – Page 1 (18 fields) (*continued*)

Variable	Field Label	Options (exported columns)
previouslyTested	Previously tested within last 3 months	yes, no
indexClient	Index Testing: Is client from an index client?	yes, no
relationWithIndexClient	Relation with index client	biological, sexual, social
pregnant	Client is Pregnant	yes, no
breastFeedingUnder6	Client breastfeeding < 6 months	yes, no
breastFeedingOver6	Client breastfeeding > 6 months	yes, no
previousTestedHIVNegative	Previously tested HIV negative	0, 1
timeLastHIVNegativeTestResult	Time of last HIV Negative test Results	1mth, 13mth, 46mths, 6mths
everHadSexualIntercourse	Ever had sexual intercourse	0, 1

**Page 2 fields (7):**

Table 7: Multiple-selection check fields – Page 2 (7 fields)

Variable	Field Label	Options (exported columns)
hivTestResult	HIV Test Result	negative, positive
hivTestBefore	Have you been tested for HIV before this year?	not previously tested, previously negative, previously positive hiv care, previously positive not hiv care
recencyTest	Recency test with RTRI	recent, long term, negative, invalid
cd4SemiQuantitative	CD4 Test Result – Semi-Quantitative	lessThan200, over200
syphilisTestResult	Syphilis Test Result	non reactive, reactive
hepatitisBTest	Hepatitis B Virus Test Result	negative, positive
hepatitisCTest	Hepatitis C Virus Test Result	negative, positive

## Expected Additional Checks

Beyond the three standard macro-based check types above, the following **custom cross-field and range checks** are expected for this form based on its clinical logic. These would be implemented as individual CTEs in the checks SQL model.

Table 8: Expected custom cross-field and range checks (18 checks)

Check name	Rule	Severity
knowledge score range	Score must be between 0 and 6	Warning
risk score range	Score must be between 0 and 9	Warning
tb score range	Score must be between 0 and 5	Warning
sti score range	Score must be between 0 and 5	Warning
sex partner score range	Score must be between 0 and 7	Warning
knowledge score consistency	Recorded score must equal count of Yes responses across the 6 knowledge items	Warning
risk score consistency	Recorded score must equal count of Yes responses across the 9 risk items	Warning
tb score consistency	Recorded score must equal count of Yes responses across the 5 TB items	Warning
sti score consistency	Recorded score must equal count of Yes responses across the 5 STI items	Warning
sex partner score consistency	Recorded score must equal count of Yes responses across the 7 sex partner risk items	Warning
recency without positive hiv	recency test must be blank if hiv test result = Negative	Warning
cd4 without positive hiv	CD4 fields must be blank if hiv test result = Negative	Warning
breastfeeding contradiction	Both breastfeeding fields cannot simultaneously be Yes	Warning
index code both filled	indexClientCodeNew and indexClientCodeOld cannot both be filled	Warning
date visit in future	Parsed date of visit must not be after today's date	Error
date visit before programme	Parsed date of visit must not be before 2015-01-01	Warning
phone number length	Phone number must be exactly 11 digits when all boxes are filled	Warning
age range	Age must be between 1 and 120 when parseable	Warning

## Check Output Structure

Every failing check produces **one row** in the checks table. Passing records produce no rows. The checks table is a union of all individual CTE results.

Table 9: Checks table – output column structure

Column	Type	Description
<b>page id</b>	text	Unique page identifier from the ScanForm Data Matrix – used to locate the physical form.
<b>entry url</b>	text	Direct link to the ScanForm verification interface entry for this record.
<b>entry submission date</b>	timestamp	Timestamp when the photo was submitted to the server.

<b>row id</b>	text	Row identifier within the export (for line-listing forms with multiple records per page).
<b>check name</b>	text	Machine-readable check identifier (e.g. tb score range).
<b>check description</b>	text	Plain-language description of the issue, suitable for display to a data manager.
<b>severity</b>	text	Warning or Error – determines whether the record is excluded from the clean dataset.
<b>field name</b>	text	The specific variable that triggered this check.
<b>field value</b>	text	The raw OCR value (before casting) that failed the check.
<b>check tags</b>	text[]	Array of tags for filtering in the DQA dashboard (e.g. date, page1; score, tb).

## Summary

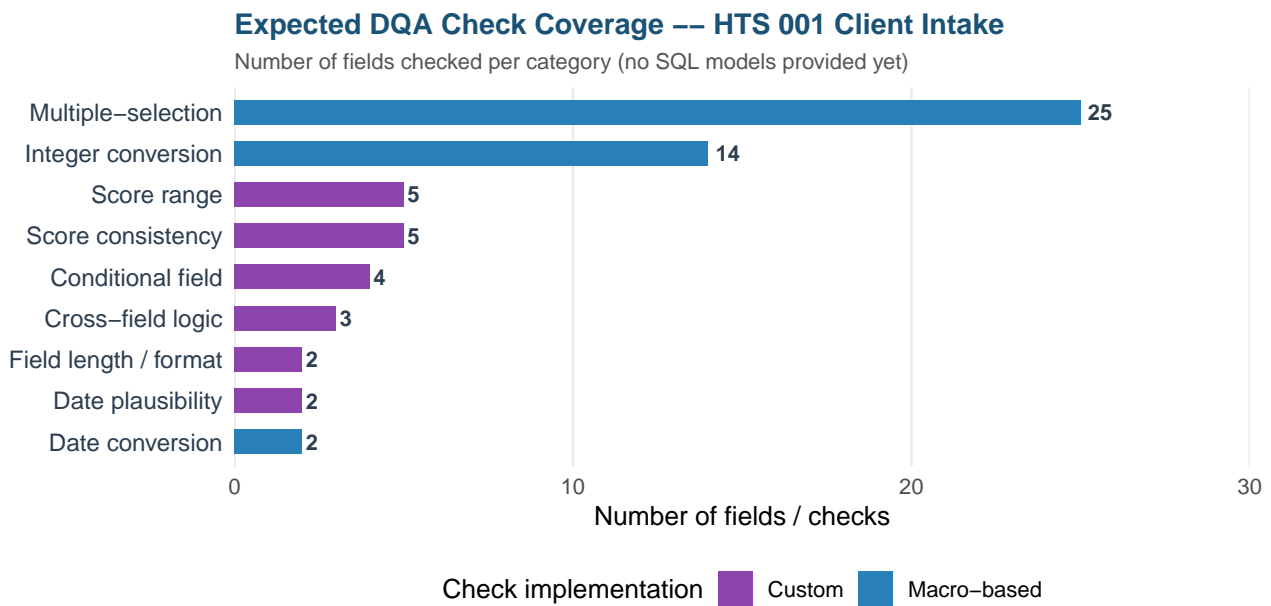


Figure 1: Expected pipeline check coverage by category

Table 10: Pipeline DQA check summary – HTS 001 Client Intake

Metric	Value
Macro-based integer conversion checks (fields)	14
Macro-based date conversion checks (fields)	2
Macro-based multiple-selection checks (questions)	25
Expected custom range checks	5
Expected custom score consistency checks	5

Expected custom cross-field / conditional checks	8
<b>Total expected checks</b>	<b>59</b>
Fields with Error-level severity	1 (dateVisit – fully unreadable)
Fields with Warning-level severity	58
<b>SQL models provided</b>	<b>0 – not yet implemented</b>

---

## Notes for Data Managers

**Score consistency checks require caution.** The five assessment scores (knowledge assessment, risk assessment, tb screening, sti screening, sex partner risk assessment) are handwritten sums entered by the clinician at the time of the consultation. It is common for these to differ from the automated sum of individual item responses, either due to arithmetic errors or because the clinician counted items differently. The pipeline **retains the clinician-recorded score** as the primary value and raises a Warning – it does not overwrite the recorded score with the computed sum.

### How to act on DQA flags:

- **Error records** are excluded from indicator calculations automatically. A data manager should review the physical form (use the entry URL link to open the scan in ScanForm) and either correct the OCR value through manual verification or mark the record as permanently invalid.
- **Warning records** remain in the clean dataset. Review is recommended for high-volume Warning categories (e.g. score inconsistencies) to assess whether a systematic training issue exists at a facility.

# Dashboard Indicators

HTS 001 Client Intake Form · NG-HTS · Nigeria · v1.6

## Overview

**Form:** NG-HTS 001 Client Intake | **Version:** v1.6 | **Country:** Nigeria

This page documents the dashboard indicators that can be derived from the HTS 001 Client Intake refinery models. Each indicator follows the QED indicator template format: goal, definitions, numerator, denominator, disaggregations, visualization type, and data source. Indicators are grouped by thematic area matching the sections of the physical form.

**Note:** No refinery SQL models were provided for this form. The indicators below are derived from the form's field definitions, clinical logic, and standard HTS programme indicator frameworks (Nigeria PEPFAR/NASCP). Column names used in numerator and denominator definitions reflect the expected clean dataset column names based on the XLSForm variable names. This page will be updated automatically when refinery model files are added to the repository.

---

## Indicator Index

Table 1: Indicator Index – HTS 001 Client Intake

#	Indicator Name	Thematic Area	Visualization
1	Total HTS clients tested	Testing volume	Count / time series
2	HIV positivity rate	Testing outcomes	Proportion / bar
3	First-time testers proportion	Testing outcomes	Proportion
4	Index testing yield	Testing outcomes	Proportion / bar
5	Knowledge Assessment completion rate	Pre-test counseling	Proportion

Table 1: Indicator Index – HTS 001 Client Intake (*continued*)

#	Indicator Name	Thematic Area	Visualization
6	Mean Knowledge Assessment score	Pre-test counseling	Mean / distribution
7	Personal Risk Assessment completion rate	Risk assessment	Proportion
8	High-risk clients (Personal Risk Score $\geq 1$ )	Risk assessment	Proportion / bar
9	TB screening completion rate	TB screening	Proportion
10	TB symptom burden (score $\geq 1$ )	TB screening	Proportion / bar
11	STI screening completion rate	STI screening	Proportion
12	STI symptom burden (score $\geq 1$ )	STI screening	Proportion / bar
13	Sex Partner Risk Assessment completion rate	Sex partner risk	Proportion
14	Post-test counseling completion	Post-test counseling	Proportion / bar
15	Condom provision rate	Post-test counseling	Proportion / bar
16	PrEP referral rate	Linkage	Proportion / bar
17	Syphilis test positivity rate	Laboratory	Proportion
18	CD4 testing coverage among HIV-positive clients	Laboratory	Proportion

## Testing Volume

### 1. Total HTS Clients Tested

**Goal:** Monitor testing throughput over time and across facilities to assess programme reach and identify underperforming sites.

Table 2: Indicator 1: Total HTS Clients Tested

Element	Detail
<b>Numerator</b>	Count of all records where discard page is not 1 and date visit clean IS NOT NULL – every submitted, non-discarded client intake record.
<b>Denominator</b>	N/A – this is a count indicator.
<b>Disaggregations</b>	Facility name; State; LGA; Visit month; Visit year; Testing setting; Type of session; Sex
<b>Definition of terms</b>	discard page = 1 means the clinician marked the discard bubble – the entire photopack is excluded. A non-null date visit clean confirms the visit date was successfully OCR-read and parsed.
<b>Visualization</b>	Line chart (time series) for overall trend; bar chart disaggregated by facility or state for site comparison.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## Testing Outcomes

### 2. HIV Positivity Rate

**Goal:** Identify facilities, settings, and population groups with elevated HIV positivity to guide targeted testing resource allocation.

Table 3: Indicator 2: HIV Positivity Rate

Element	Detail
<b>Numerator</b>	Count of records where hiv test result = 'positive'.
<b>Denominator</b>	Count of records where hiv test result IS NOT NULL (a test result was recorded – either positive or negative).
<b>Disaggregations</b>	Facility; State; LGA; Visit month; Testing setting; Modality; Sex; Age group (derived from age: <15, 15-24, 25-34, 35-49, 50+); First-time tester; Type of session
<b>Definition of terms</b>	HIV Test Result is recorded in hiv test result with values 'negative' or 'positive'. Records where the field is NULL are excluded from both numerator and denominator to avoid deflating the rate.
<b>Visualization</b>	Bar chart by facility with a reference line showing the national/state average positivity rate; stacked bar disaggregated by sex.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

### 3. First-Time Testers Proportion

**Goal:** Track programme reach among individuals who have never previously been tested for HIV – a key PEPFAR coverage indicator.

Table 4: Indicator 3: First-Time Testers Proportion

Element	Detail
<b>Numerator</b>	Count of records where first time visit = 'yes'.
<b>Denominator</b>	Count of records where first time visit IS NOT NULL.
<b>Disaggregations</b>	Facility; State; LGA; Visit month; Testing setting; Sex; Age group (from age)
<b>Definition of terms</b>	first time visit = 'yes' indicates the client has never previously received an HIV test. Records where the field is NULL are excluded from both counts.
<b>Visualization</b>	Proportion displayed as a single KPI number with trend sparkline; bar chart by facility for site comparison.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## 4. Index Testing Yield

**Goal:** Assess the effectiveness of index testing as a case-finding strategy – clients identified through an index case typically have higher HIV positivity.

Table 5: Indicator 4: Index Testing Yield

Element	Detail
<b>Numerator</b>	Count of records where index client = 'yes' AND hiv test result = 'positive'.
<b>Denominator</b>	Count of records where index client = 'yes' AND hiv test result IS NOT NULL.
<b>Disaggregations</b>	Facility; State; Relation with index client (biological, sexual, social); Visit month
<b>Definition of terms</b>	An index client is one identified through contact tracing of a known HIV-positive person (index client = 'yes'). The relation with the index case is recorded in relation with index client (biological, sexual, or social contact).
<b>Visualization</b>	Bar chart comparing positivity rate for index-tested vs non-index-tested clients; breakdown by relationship type.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## Pre-Test Counseling Quality

### 5. Knowledge Assessment Completion Rate

**Goal:** Monitor whether all six knowledge assessment items are being completed by counsellors – a proxy for counseling session quality.

Table 6: Indicator 5: Knowledge Assessment Completion Rate

Element	Detail
<b>Numerator</b>	Count of records where knowledge assessment IS NOT NULL (score was recorded).
<b>Denominator</b>	Count of all valid (non-discarded) records.
<b>Disaggregations</b>	Facility; State; Visit month; Testing setting
<b>Definition of terms</b>	knowledge assessment is a single digit box recording the clinician-computed sum of all 6 knowledge items (max = 6). A non-null value indicates the score was written and successfully OCR-read.
<b>Visualization</b>	Proportion as KPI with trend over time; bar chart by facility to identify sites with low completion.
<b>Frequency</b>	Daily.

## 6. Mean Knowledge Assessment Score

**Goal:** Track average pre-test HIV knowledge among clients to evaluate counseling effectiveness and identify populations needing enhanced education.

Table 7: Indicator 6: Mean Knowledge Assessment Score

Element	Detail
<b>Numerator</b>	Sum of knowledge assessment across all records where the score is not null and is between 0 and 6.
<b>Denominator</b>	Count of records where knowledge assessment IS NOT NULL AND knowledge assessment BETWEEN 0 AND 6.
<b>Disaggregations</b>	Facility; State; Visit month; Sex; Age group (from age); First-time tester
<b>Definition of terms</b>	The Knowledge Assessment Score is the clinician-recorded sum of the 6 binary items in Section A (max = 6). Scores flagged by the DQA pipeline as inconsistent with individual item responses are retained but annotated.
<b>Visualization</b>	Distribution histogram of scores; mean score as KPI; line chart of mean score over time by facility.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## Risk Assessment

### 7. Personal Risk Assessment Completion Rate

**Goal:** Monitor whether the 9-item personal risk assessment is consistently completed – a required step before HIV testing in the standard protocol.

Table 8: Indicator 7: Personal Risk Assessment Completion Rate

Element	Detail
<b>Numerator</b>	Count of records where risk assessment IS NOT NULL.
<b>Denominator</b>	Count of all valid (non-discarded) records.
<b>Disaggregations</b>	Facility; State; Visit month; Testing setting
<b>Definition of terms</b>	risk assessment is the clinician-recorded sum of the 9 personal HIV risk items in Section B (max = 9). A non-null value indicates the score was written on the form and successfully read by OCR.
<b>Visualization</b>	Proportion as KPI; bar chart by facility highlighting sites below a threshold (e.g. < 80% completion).

<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## 8. High-Risk Clients (Personal Risk Score $\geq 1$ )

**Goal:** Identify the proportion of clients presenting with at least one personal HIV risk factor – per protocol, these clients should be recommended for HIV testing and, if negative, re-testing after 3 months.

Table 9: Indicator 8: High-Risk Clients (Personal Risk Score  $\geq 1$ )

Element	Detail
<b>Numerator</b>	Count of records where risk assessment $\geq 1$ .
<b>Denominator</b>	Count of records where risk assessment IS NOT NULL AND risk assessment BETWEEN 0 AND 9.
<b>Disaggregations</b>	Facility; State; Visit month; Sex; Age group (from age); HIV Test Result
<b>Definition of terms</b>	A client is considered high-risk if their Personal HIV Risk Assessment Score (risk assessment) is 1 or more. The score is the sum of 9 binary risk behaviour items from Section B. Per the clinical rule printed on the form, clients with score $\geq 1$ whose last HIV test was more than 3 months ago should be offered testing.
<b>Visualization</b>	Proportion as KPI; stacked bar showing score distribution (0, 1-3, 4-6, 7-9) by facility.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## TB Screening

### 9. TB Screening Completion Rate

**Goal:** Monitor whether all HTS clients are being screened for TB symptoms – a mandatory co-infection screening step in the Nigerian HTS protocol.

Table 10: Indicator 9: TB Screening Completion Rate

Element	Detail
<b>Numerator</b>	Count of records where tb screening IS NOT NULL.
<b>Denominator</b>	Count of all valid (non-discarded) records.
<b>Disaggregations</b>	Facility; State; LGA; Visit month

<b>Definition of terms</b>	tb screening is the clinician-recorded sum of the 5 TB symptom items in Section C (max = 5): current cough, weight loss, fever, night sweats, lymphadenopathy. A non-null value means the screening was completed and the score was written and read.
<b>Visualization</b>	Proportion as KPI with facility-level bar chart.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## 10. TB Symptom Burden (Score $\geq 1$ )

**Goal:** Identify the proportion of HTS clients presenting with at least one TB symptom and therefore requiring referral for Xpert MTB/RIF testing or TB services.

Table 11: Indicator 10: TB Symptom Burden (Score  $\geq 1$ )

Element	Detail
<b>Numerator</b>	Count of records where tb screening $\geq 1$ .
<b>Denominator</b>	Count of records where tb screening IS NOT NULL AND tb screening BETWEEN 0 AND 5.
<b>Disaggregations</b>	Facility; State; Visit month; HIV Test Result; Sex
<b>Definition of terms</b>	Per the clinical rule on the form: if TB screening score $\geq 1$ , the client should be tested for Xpert MTB RIF or referred to a TB service. The five constituent symptoms are: current cough, weight loss, fever, night sweats, lymphadenopathy – each scored 0 (No) or 1 (Yes).
<b>Visualization</b>	Proportion as KPI; bar chart by facility; cross-tabulation with HIV Test Result to show TB/HIV co-burden.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## STI Screening

### 11. STI Screening Completion Rate

**Goal:** Monitor whether syndromic STI screening is consistently completed for all HTS clients, as required by the Nigerian HTS protocol.

Table 12: Indicator 11: STI Screening Completion Rate

Element	Detail
<b>Numerator</b>	Count of records where sti screening IS NOT NULL.

<b>Denominator</b>	Count of all valid (non-discarded) records.
<b>Disaggregations</b>	Facility; State; Visit month; Sex
<b>Definition of terms</b>	sti screening is the clinician-recorded sum of the 5 STI symptom items in Section C (max = 5). A non-null value means the syndromic STI screening was completed.
<b>Visualization</b>	Proportion as KPI; bar chart by facility.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## 12. STI Symptom Burden (Score $\geq 1$ )

**Goal:** Identify the proportion of HTS clients presenting with at least one STI symptom and therefore requiring referral for syndromic STI management.

Table 13: Indicator 12: STI Symptom Burden (Score  $\geq 1$ )

Element	Detail
<b>Numerator</b>	Count of records where sti screening $\geq 1$ .
<b>Denominator</b>	Count of records where sti screening IS NOT NULL AND sti screening BETWEEN 0 AND 5.
<b>Disaggregations</b>	Facility; State; Visit month; Sex; HIV Test Result
<b>Definition of terms</b>	Per the clinical rule on the form: if STI screening score $\geq 1$ , the clinician should follow syndromic STI management guidelines or refer. The five items are sex-differentiated: vaginal discharge, lower abdominal pain (female); urethral discharge, scrotal swelling (male); and genital sore / inguinal lymphadenopathy (any sex).
<b>Visualization</b>	Proportion as KPI; bar chart by facility; breakdown by sex to identify sex-specific STI burden patterns.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## Sex Partner Risk

### 13. Sex Partner Risk Assessment Completion Rate

**Goal:** Monitor whether Section D (sex partner risk) is being completed – clients with a score  $\geq 1$  who test HIV negative should be referred for PrEP services per protocol.

Table 14: Indicator 13: Sex Partner Risk Assessment Completion Rate

Element	Detail
<b>Numerator</b>	Count of records where sex partner risk assessment IS NOT NULL.
<b>Denominator</b>	Count of all valid (non-discarded) records.
<b>Disaggregations</b>	Facility; State; Visit month
<b>Definition of terms</b>	sex partner risk assessment is the clinician-recorded sum of the 7 sex partner risk items in Section D (max = 7). Questions ask whether the client has had sex with an HIV-positive partner meeting specific clinical criteria (on ARV with unsuppressed VL, newly diagnosed, LTFU-returned, etc.).
<b>Visualization</b>	Proportion as KPI; bar chart by facility.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## Post-Test Counseling

### 14. Post-Test Counseling Completion

**Goal:** Verify that all tested clients receive post-test counseling, risk reduction planning, and disclosure planning – key quality-of-care indicators.

Table 15: Indicator 14: Post-Test Counseling Completion

Element	Detail
<b>Numerator</b>	Count of records where post test counseling = 1 AND risk reduction = 1 AND post test disclosure = 1.
<b>Denominator</b>	Count of records where hiv test result IS NOT NULL (a test was conducted).
<b>Disaggregations</b>	Facility; State; Visit month; HIV Test Result; Sex
<b>Definition of terms</b>	Full post-test counseling is defined as completion of all three mandatory post-test steps: (1) post test counseling = 1 (counseling done), (2) risk reduction = 1 (risk reduction plan developed), (3) post test disclosure = 1 (disclosure plan developed). Each field is a binary 0/1 oval question.
<b>Visualization</b>	Proportion as KPI broken down by HIV result; bar chart by facility.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

### 15. Condom Provision Rate

**Goal:** Monitor condom provision as a prevention service delivery indicator – all clients should be offered condoms at the conclusion of the HTS session.

Table 16: Indicator 15: Condom Provision Rate

Element	Detail
<b>Numerator</b>	Count of records where condom provided to client = 1.
<b>Denominator</b>	Count of records where hiv test result IS NOT NULL.
<b>Disaggregations</b>	Facility; State; Visit month; HIV Test Result; Sex
<b>Definition of terms</b>	condom provided to client is a binary 0/1 oval field. When 1, the number of condoms distributed is additionally captured in condom provided to client count (integer, 2-box digit field). Lubricant provision is captured separately in lubricant provided to client and lubricant provided to client count.
<b>Visualization</b>	Proportion as KPI; mean condom count per client as supplementary metric; bar chart by facility.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## Linkage to Services

### 16. PrEP Referral Rate

**Goal:** Track the proportion of HIV-negative high-risk clients referred for PrEP – a key prevention cascade indicator under the Nigerian PEPFAR programme.

Table 17: Indicator 16: PrEP Referral Rate

Element	Detail
<b>Numerator</b>	Count of records where prep referred = 1.
<b>Denominator</b>	Count of records where hiv test result = 'negative' AND (risk assessment $\geq 1$ OR sex partner risk assessment $\geq 1$ ).
<b>Disaggregations</b>	Facility; State; Visit month; Sex; Age group (from age)
<b>Definition of terms</b>	Per the clinical rule printed on the form: HIV-negative clients with a Personal HIV Risk Assessment Score $\geq 1$ or a Sex Partner Risk Assessment Score $\geq 1$ in Section D should be referred for PrEP services. prep referred = 1 indicates a referral was made and recorded.
<b>Visualization</b>	Proportion as KPI; bar chart by facility to identify sites with low PrEP referral uptake; trend line over time.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

## Laboratory Results

### 17. Syphilis Test Positivity Rate

**Goal:** Monitor syphilis co-infection burden among HTS clients – particularly important for PMTCT programmes and key populations.

Table 18: Indicator 17: Syphilis Test Positivity Rate

Element	Detail
<b>Numerator</b>	Count of records where syphilis test result = 'reactive'.
<b>Denominator</b>	Count of records where syphilis test result IS NOT NULL (either 'reactive' or 'non reactive' was recorded).
<b>Disaggregations</b>	Facility; State; Visit month; Sex; HIV Test Result; Age group (from age)
<b>Definition of terms</b>	syphilis test result is a single-select oval field with options 'non reactive' and 'reactive'. Records where the field is NULL are excluded from both numerator and denominator. Cross-tabulation with hiv test result reveals HIV/syphilis co-infection.
<b>Visualization</b>	Proportion as KPI; bar chart by facility; cross-tabulation with HIV result as stacked bar.
<b>Frequency</b>	Daily.
<b>Data source</b>	refinery clean

### 18. CD4 Testing Coverage Among HIV-Positive Clients

**Goal:** Monitor the proportion of newly identified HIV-positive clients who receive CD4 testing – an indicator of ART eligibility assessment completeness.

Table 19: Indicator 18: CD4 Testing Coverage Among HIV-Positive Clients

Element	Detail
<b>Numerator</b>	Count of records where cd4 semi quantitative IS NOT NULL OR cd4 flow cytometry IS NOT NULL.
<b>Denominator</b>	Count of records where hiv test result = 'positive'.
<b>Disaggregations</b>	Facility; State; Visit month; CD4 result category: low ( $\leq 200$ cells/m <sup>3</sup> ) vs high ( $> 200$ cells/m <sup>3</sup> )
<b>Definition of terms</b>	CD4 testing is only applicable for HIV-positive clients. Two methods are captured: (1) cd4 semi quantitative – a single-select oval with options 'lessThan200' ( $\leq 200$ cells/m <sup>3</sup> ) or 'over200' ( $> 200$ cells/m <sup>3</sup> ); (2) cd4 flow cytometry – a 4-digit box for the exact count. A client is considered tested if either method has a non-null value.
<b>Visualization</b>	Proportion as KPI; bar chart of CD4 category distribution ( $\leq 200$ vs $> 200$ ) among positive clients by facility.
<b>Frequency</b>	Daily.

## Indicator Reference Table

Table 20: All Indicators – Quick Reference

#	Indicator	Key column(s)	Type	Refinery model
1	Total HTS clients tested	discard page, date visit clean	Count	refinery clean
2	HIV positivity rate	hiv test result	Proportion	refinery clean
3	First-time testers proportion	first time visit	Proportion	refinery clean
4	Index testing yield	index client, hiv test result, relation with index client	Proportion	refinery clean
5	Knowledge Assessment completion rate	knowledge assessment	Proportion	refinery clean
6	Mean Knowledge Assessment score	knowledge assessment	Mean	refinery clean
7	Personal Risk Assessment completion rate	risk assessment	Proportion	refinery clean
8	High-risk clients (Risk Score >= 1)	risk assessment	Proportion	refinery clean
9	TB screening completion rate	tb screening	Proportion	refinery clean
10	TB symptom burden (score >= 1)	tb screening	Proportion	refinery clean
11	STI screening completion rate	sti screening	Proportion	refinery clean
12	STI symptom burden (score >= 1)	sti screening	Proportion	refinery clean
13	Sex Partner Risk Assessment completion rate	sex partner risk assessment	Proportion	refinery clean
14	Post-test counseling completion	post test counseling, risk reduction, post test disclosure	Proportion	refinery clean
15	Condom provision rate	condom provided to client, condom provided to client count	Proportion	refinery clean
16	PrEP referral rate	prep referred, hiv test result, risk assessment	Proportion	refinery clean
17	Syphilis test positivity rate	syphilis test result	Proportion	refinery clean
18	CD4 testing coverage (HIV+)	cd4 semi quantitative, cd4 flow cytometry	Proportion	refinery clean

## Important Notes for Dashboard Developers

**NULL handling:** All proportion indicators should exclude records where the relevant field is NULL from both numerator and denominator unless the indicator explicitly measures completeness (in which case NULL counts as incomplete and the denominator is all records).

**Discard exclusion:** All indicators should filter out records where `discard page = 1` before any calculation.

**Score range filtering:** Indicators based on assessment scores (knowledge assessment, risk assessment, tb screening, sti screening, sex partner risk assessment) should additionally filter to records where the score is within its valid range (e.g. tb screening BETWEEN 0 AND 5) to exclude DQA-flagged outliers.

**Date grouping:** Use `visit month` and `visit year` from the refinery model for time series – these are stable derived columns that fall back to submission date when `date visit clean` is invalid.

**Conditional fields:** recency test, cd4 semi quantitative, and cd4 flow cytometry are only applicable for HIV-positive clients. Always gate these indicators with `hiv test result = 'positive'` in the denominator.

---

HTS 001 Client Intake Form • Version 1.6 • Nigeria • Documentation generated by QED.ai ScanForm