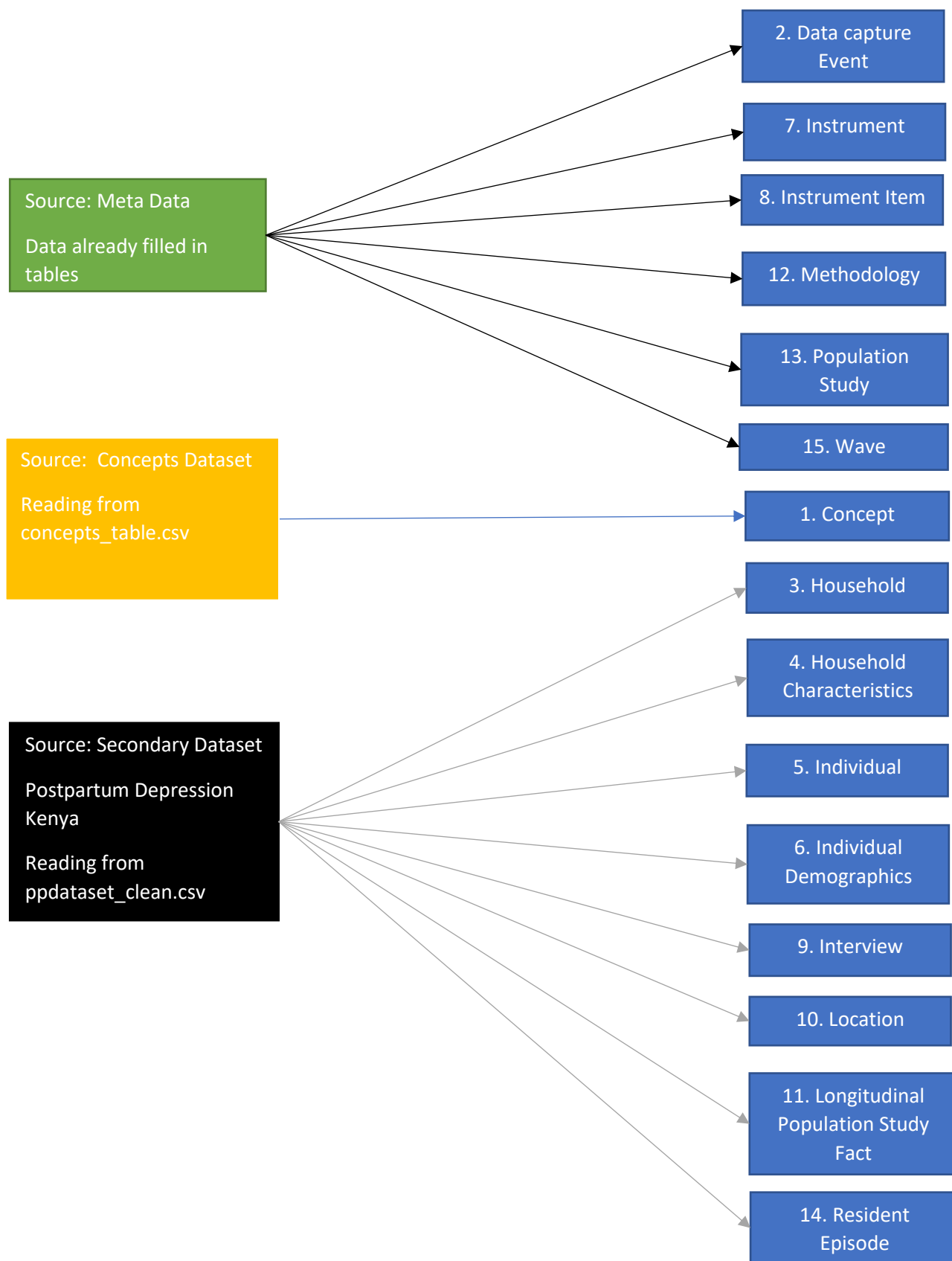


Source Data Mapping Approach to Mental Health Staging

Contents

1. Source: MetaData	3
Table name: Data capture Event	3
Table name: Instrument	3
Table name: Instrument Item	3
Table name: Methodology	4
Table name: Population Study	4
Table name: Wave	5
2. Source: Concept Dataset	6
Table name: Concept	6
3. Source: Secondary Dataset	7
Table name: Household	7
Table name: Household Characteristics	7
Table name: Individual	8
Table name: Individual Demographics	9
Table name: Interview	11
Table name: Location	11
Table name: Longitudinal Population Study Fact	12
Table name: Resident Episode	17



1. Source: MetaData

Metadata contained in the tables is from 12 population longitudinal studies (11 secondary and 1 primary) on depression, anxiety and psychosis.

Metadata for subsequent data received from primary collection sites will be added to the tables.

Table name: Data capture Event

Logs events related to data capture during the study.

Destination Field	Description	Type	Comment
data_capture_id	Unique identifier for the data capture event	INT	Primary Key
wave_id	References the wave during which the event occurred	INT	
instrument_id	References the instrument used	INT	
completion_status		VARCHAR	
data_quality_indicator		VARCHAR	
mode_of_collection_description		VARCHAR	
mode_of_collection_type		VARCHAR	
data_capture_collector		VARCHAR	
data_source_description		VARCHAR	
data_source_type		VARCHAR	
data_capture_event_date		DATE	

Table name: Instrument

Contains information about the instruments used in data collection.

Destination Field	Description	Type	Comment
instrument_id	Unique identifier for the instrument	INT	Primary Key
name	Name the instrument	TEXT	
description	What the instrument measures	TEXT	
instrument_type_concept_id		INT	
version		FLOAT	
version_date		DATE	
language_concept_id		INT	

Table name: Instrument Item

Details items within each instrument.

Destination Field	Description	Type	Comment
instrument_item_id	Unique identifier for the instrument item	INT	Primary Key

instrument_id	References the instrument	INT	
name	Instrument item question number	VARCHAR	
description	Instrument item question description	TEXT	
instrument_item_type_concept_id		INT	
instrument_item_concept_vocabulary		VARCHAR	
instrument_item_concept_vocabulary_id		INT	
alternative_instrument_item_concept_vocabulary		VARCHAR	
alternative_instrument_item_concept_vocabulary_id		INT	
result_not_null_answer_list_concept_vocabulary		VARCHAR	

Table name: Methodology

Details the methodologies used in the study.

Destination Field	Description	Type	Comment
methodology_id	Unique identifier for the methodology	INT	Primary Key
data_collection_methodology_description		VARCHAR	
data_collection_methodology_type		VARCHAR	
time_method_description		VARCHAR	
time_method_type		VARCHAR	
sampling_procedure_description		VARCHAR	
sampling_procedure_type		VARCHAR	
data_collection_software_name		VARCHAR	
data_collection_software_version		VARCHAR	
data_collection_software_package_type		VARCHAR	
quality_statement_standard_name		VARCHAR	
quality_statement_standard_description		VARCHAR	
population_study_id	Unique identifier for the study	INT	

Table name: Population Study

Contains overarching information about the population study. Each methodology and wave of data collection is associated with a population study.

Destination Field	Description	Type	Comment
name	Unique identifier for the methodology	VARCHAR	
description		TEXT	
country		TEXT	
abstract		TEXT	
phenotype_description		TEXT	

outcome_phenotype_description		TEXT	
covariates_description		TEXT	
analyses_supported_text		TEXT	
version		FLOAT	
version_date		DATE	
citation_creators		TEXT	
citation_contributors		TEXT	
universe_spatial_coverage_text		TEXT	
population_study_id	Unique identifier for the study	INT	Primary Key
doi_registry		TEXT	
doi_value		TEXT	
url		TEXT	
citation_title		TEXT	
citation_publisher		TEXT	
citation_language_concept_id		FLOAT	
keywords		TEXT	
universe_spatial_coverage_concept_id		TEXT	
universe_spatial_temporal_coverage		TEXT	
analyses_supported_concept_id		FLOAT	

Table name: Wave

Details waves of data collection.

Destination Field	Description	Type	Comment
wave_id	Unique identifier for the wave	INT	Primary Key
name		VARCHAR	
description		TEXT	
instrument_model_type_concept_id		BIGINT	
start_date		DATE	
end_date		DATE	
kind_of_data_concept_id		INT	
authorizing_agency_concept_id		INT	
authorizing_statement		TEXT	
population_study_id	Unique identifier for the study	INT	

2. Source: Concept Dataset

Reading from concepts_table.csv

Table name: Concept

Stores metadata about various concepts used in the study. **No Primary Key Specified.**

Destination Field	Description	Type	Source Field	Logic	Comment
concept_vocabulary	Text identifier for the concept vocabulary i.e LOINC, SNOWMED, INSPIRE	TEXT	concept_vocabulary		
concept_code	Code representing the concept	TEXT	concept_code		
concept_text	Description or text of the concept.	TEXT	concept_text		
start_date	Validity period for the concept	DATE	start_date		
end_date	Validity period for the concept	DATE	end_date		
inspire_concept_id	Numeric identifiers for the concept	FLOAT	inspire_concept_id		
concept_id	Numeric identifiers for the concept	FLOAT	concept_id		Added concept id 3000000187 – Housewife 3000000188 – Other religion 3000000190 – Self-employed business 38003564 - Not Hispanic or Latino Vaginal delivery – 44784097 Caesarean Section - 4015701
score	Text score or rating for the concept i.e scores for questionnaire tools	TEXT	score		

3. Source: Secondary Dataset

Reading from ppdataset_clean.csv

Table name: Household

Contains information about households.

Destination Field	Description	Type	Source Field	Logic	Comment
household_id	Unique identifier for the household	INT		Auto-generated. Must continue from last ID	The unique key given to a unique household Primary Key
household_id_value	Text representation of the household ID	TEXT	ssid	Each unique individual assigned to be coming from a household	The purpose of this field is to link back to households.
location_id	References the location of the household	INT		Fetch id from location table using village_name	village_name for all households set to "Nairobi County" Foreign key: Yes
household_head_id	References the head of the household	INT		No household head column in source data	Set to 0

Table name: Household Characteristics

Records characteristics of households.

Destination Field	Description	Type	Source Field	Logic	Comment
household_characteristics_id	Unique identifier for household characteristics	BIGINT		Auto-generated. Must continue from last ID	The unique key given to a unique household characteristic Primary Key
household_id	References the household	INT		Fetch household_id from household table using household_id_value	The unique key given to a unique household Foreign key: Yes
wave_id	References the wave of data collection	INT		Check id from wave table	Population study id = 1 Set to:

					Baseline – wave = 1 Followup – wave = 2
household_characteristics_concept_id	Concept ID related to household characteristics	BIGINT	hsld	Check concept id from concept table	Map the concept_id of household size as follows Household Size of One -> 3000000212 Household Size of Two to Four -> 3000000213 Household Size of Five to Nine -> 3000000214 Household Size of Ten to Fourteen -> 3000000215 Household Size of Fifteen or More -> 3000000216
household_characteristics_concept_text	Text description of the concept	VARCHAR		Fetch household characteristics concept text from concept table using concept id	

Table name: Individual

Holds data about individuals in households.

Destination Field	Description	Type	Source Field	Logic	Comment
individual_id	Unique identifier for the individual	INT		Auto-generated. Must continue from last ID	The unique key given to a unique individual Primary Key

individual_id_value	Text representation of the individual ID	TEXT	ssid		The purpose of this field is to link back to persons in the source data.
household_id	References the household to which the individual belongs	INT		Fetch id from household table using household_id_value	Each unique individual assumed to be coming from a unique household Foreign key: Yes
gender_concept_id		INT		Study was on women – prenatal and postnatal	This field is meant to capture the biological sex of the individual. Map the sex values as follows Female -> 8532
first_wave_id		INT		First wave for this study is Baseline	population_study_id is 1. Set to 1 as per wave table
age_at_first_wave		INT	age		Age of individual at baseline.
year_of_birth		INT	ydob1	Extract the year from date of birth	
is_household_head		BOOLEAN		No household head column in source data	Set to “false”

Table name: Individual Demographics

Stores demographic data related to individuals.

Destination Field	Description	Type	Source Field	Logic	Comment
individual_demographics_id	Unique identifier for individual demographics	BIGINT		Auto-generated. Must continue from last ID	Primary Key
individual_id	References the individual	INT		Fetch individual_id from individual table using individual_id_value	The unique key given to a unique individual Foreign key: Yes
individual_concept_id	Concept IS related to individual demographics	BIGINT	rlg mrtl edc ocp1 mod	Check concept id from concept table	Map the concept_id as follows African Race - 38003600

					<p>Not Hispanic or Latino Ethnicity - 38003564</p> <p>Religion Protestant -4183092 Catholic - 4175384 Other religion - 3000000188</p> <p>Marital status Never Married - 45881671 Married - 45876756 Separated - 45884459</p> <p>highest level of education Tertiary (college/University) - 45876260 <=Primary - 45878725 Secondary - 45876261</p> <p>occupation Participant has a wage/salaried work - 45877708 Participant has self-employed business work -3000000190 Participant is a housewife by choice -3000000187 Participant is unable to find employment -45877709</p> <p>mode of delivery Vaginal delivery - 44784097 Caesarean Section - 4015701</p>
individual_concept_id_text	Text description of the concept	VARCHAR		Fetch individual demographics concept text from concept table using concept id	

Table name: Interview

Logs interviews conducted with individuals.

Destination Field	Description	Type	Source Field	Logic	Comment
interview_id	Unique identifier for the interview	INT		Auto-generated. Must continue from last ID	The unique key given to a unique interview date Primary Key
individual_id	References the individual interviewed	INT		Fetch individual_id from Individual table using individual_id_value	This column will be used to convert the wide to long. Foreign key: Yes
interview_date		DATE	dtstd, dtpnrf	dtstd as baseline study date dtpnrf as followup study date	
wave_id		INT		Check wave_id from Wave table	population_study_id = 1 Set to: Baseline – wave = 1 Followup – wave = 2
instrument_id		INT		Check instrument_id from the Instrument table EPDS instrument used	Set to 4

Table name: Location

Geographical information about locations.

Destination Field	Description	Type	Source Field	Logic	Comment
location_id	Unique identifier for the location	INT		Auto-generated. Must continue from last ID	The unique key given to a unique Location Primary Key: Yes
village_name		VARCHAR		Pregnant women recruited attending maternal and child health	Set to Nairobi County

				clinics in two public hospitals in Nairobi, Kenya	
place_kind		VARCHAR		Pregnant women residing in an urban, resource-poor setting	Set to Urban
latitude		FLOAT			The geocoded latitude of Nairobi county is - 1.3021282
longitude		FLOAT			The geocoded longitude of Nairobi county is 36.7203683

Table name: Longitudinal Population Study Fact

Records detailed facts and data points collected during the study.

Destination Field	Description	Type	Source Field	Logic	Comment
fact_id	Unique identifier for the fact	INT		Auto-generated. Must continue from last ID	Primary Key
individual_id	References to unique individual	INT		Fetch id from Individual table using individual_id_value	This column will be used to convert the wide to long. Foreign key: Yes
interview_id		INT		Fetch id from interview table using individual_id, wave_id and instrument_id	
resident_episode_id		INT		Fetch id from resident episode table using household_id and wave_id	
population_study_id	References to unique population study	INT		Check population_study_id from the Population study table	Set to 1
instrument_id		INT		Check instrument_id from the Instrument table EPDS instrument used	Set to 4
instrument_item_id		INT		Check instrument_item_id	Set id as below:

				from the Instrument Item table	lgh1, lgh2 -> 22 enjoy1, enjoy2 -> 23 blmd1, blmd2 -> 24 anxs1, anxs2 -> 25 scrd1, scrd2 -> 26 sad1, sad2 -> 29 thng1, thng2 -> 27 cryn1, cryn2 -> 30 slpn1, slpn2 -> 28 hrmg1, hrmg2 -> 31 scr1, scr2 -> 32
concept_id		INT	lgh1 enjoy1 blmd1 anxs1 scrd1 sad1 thng1 cryn1 slpn1 hrmg1 scr1 lgh2 enjoy2 blmd2 anxs2 scrd2 sad2 thng2 cryn2 slpn2 hrmg2 scr2		lgh1, lgh2 As much as I always could - 45878094 Not quite so much now - 45877849 Definitely not so much now - 45878095 Not at all - 45883172 enjoy1, enjoy2 As much as I ever did - 45885323 Rather less than I used to - 45879387 Definitely less than I used to - 45880856 Hardly at all – 45879388 blmd1, blmd2 No-never - 45880857 Yes-some of the time - 45880858 Not very often - 45880859 Yes-most of the time - 45885120 anxs1, anxs2 No-not at all - 45885119 Yes-sometimes - 45879997 Hardly ever - 45884598

					<p>Yes-very often - 45877850</p> <p>scrd1, scrd2 No-not at all - 45885119 No-not much - 45879389 Yes-sometimes - 45879997 Yes-quite a lot - 45885324</p> <p>sad1, sad2 No-not at all - 45885119 Not very often - 45880859 Yes-quite often - 45879391 Yes-most of the time - 45885120</p> <p>thng1, thng2 No-most of the time I have coped quite well - 45877851 No-I have been coping as well as ever - 45885325 Yes-sometimes I have not been coping as well as usual - 45881507 Yes-most of the time I have not been able to cope at all - 45879390</p> <p>cryn1, cryn2 No-never - 45880857 Only occasionally - 45885326 Yes-most of the time - 45885120 Yes-quite often – 45879391</p> <p>slpn1, slpn2</p>
--	--	--	--	--	---

					No-not at all - 45885119 No-Not very often - 45880859 Yes-sometimes - 45879997 Yes-most of the time - 45885120 hrmg1, hrmg2 Never - 45876662 Hardly ever - 45884598 Sometimes - 45882528 Yes-quite often - 45879391 scr1, scr2 1989297
value_type_concept_id		INT	lgh1 enjoy1 blmd1 anxs1 scrd1 sad1 thng1 cryn1 slpn1 hrmg1 scr1 lgh2 enjoy2 blmd2 anxs2 scrd2 sad2 thng2 cryn2 slpn2 hrmg2 scr2		lgh1, lgh2 -> 42870286 enjoy1, enjoy2 -> 42870287 blmd1, blmd2 -> 42870288 anxs1, anxs2 -> 42870289 scrd1, scrd2 -> 42870290 sad1, sad2 -> 42870293 thng1, thng2 -> 42870291 cryn1, cryn2 -> 42870294 slpn1, slpn2 -> 42870292 hrmg1, hrmg2 -> 42870295 scr1, scr2 -> 1989297
value_as_char		VARCHAR		Fetch value from concept table using concept_id and score	concept_text of response to questions.
value_as_num		FLOAT	lgh1 enjoy1 blmd1 anxs1 scrd1		lgh1 and lgh2 As much as I always could - 0 Not quite so much now - 1

			<p>sad1 thng1 cryn1 slpn1 hrmg1 scr1 lgh2 enjoy2 blmd2 anxs2 scrd2 sad2 thng2 cryn2 slpn2 hrmg2 scr2</p>	<p>Definitely not so much now - 2 Not at all - 3</p> <p>enjoy1, enjoy2 As much as I ever did - 0 Rather less than I used to - 1 Definitely less than I used to - 2 Hardly at all – 3</p> <p>blmd1, blmd2 No-never - 0 Yes-some of the time - 2 Not very often - 1 Yes-most of the time - 3</p> <p>anxs1, anx2 No-not at all - 0 Yes-sometimes - 2 Hardly ever - 1 Yes-very often - 3</p> <p>scrd1, scrd2 No-not at all - 0 No-not much - 1 Yes-sometimes - 2 Yes-quite a lot - 3</p> <p>sad1, sad2 No-not at all - 0 Not very often - 1 Yes-quite often - 2 Yes-most of the time – 3</p> <p>thng1, thng2 No-most of the time I have coped quite well - 1 No-I have been coping as well as ever - 0 Yes-sometimes I have not been coping as well as usual - 2</p>
--	--	--	--	--

					<p>Yes-most of the time I have not been able to cope at all - 3</p> <p>cryn1, cryn2 No-never - 0 Only occasionally - 1 Yes-most of the time - 3 Yes-quite often - 2</p> <p>slpn1, slpn2 No-not at all - 0 No-Not very often - 1 Yes-sometimes - 2 Yes-most of the time - 3</p> <p>hrmg1, hrmg2 Never - 0 Hardly ever - 1 Sometimes - 2 Yes-quite often – 3</p> <p>scr1, scr2 total scores</p>
value_as_concept		INT		Fetch value from concept table using concept_id	inspire_concept_id of EPDS tool questions and total score
is_indv_level		BOOLEAN			Set to NULL

Table name: Resident Episode

Tracks residence episodes of individuals.

Destination Field	Description	Type	Source Field	Logic	Comment
resident_episode_id	Unique identifier for the resident episode	INT		Auto-generated. Must continue from last ID	The unique key given to a unique resident episode Primary Key
location_id		INT		Fetch location_id from household table using household_id_value	
household_id		INT		Fetch household_id from household table	

				using household_id_value	
wave_id		INT		Check wave_id from Wave table	population_study_id = 1 Set to: Baseline – wave = 1 Followup – wave = 2