STI Data Validation Exercise before UCSR Entry

Lily Alexander

March 26, 2018

Abstract

Summary. This is a document to validate and check variables before they are integrated into the UCSR.

1 High-level codebook for all variables

disease Disease type: string (str3) unique values: 1 missing "": 0/38 tabulation: Freq. Value "HIV" 38 collapsed _____ type: string (str3), but longest is str2 missing "": 0/38 unique values: 1 tabulation: Freq. Value 38 "No" lead_author Lead Author type: string (str12) missing "": 0/38 unique values: 2

28 "Marseille, E" 10 "Rodrigues, R"

tabulation: Freq. Value

warning: variable has embedded blanks

ref_year Reference Year _____ type: numeric (int) range: [2011,2014] units: 1 unique values: 2 missing .: 0/38 tabulation: Freq. Value 28 2011 10 2014 Output unit output_unit2 type: string (str29) missing "": 0/38 unique values: 3 tabulation: Freq. Value 14 "cost per person served" 14 "cost per person-quarter of CM" 10 "cost per person-year" warning: variable has embedded blanks title type: string (str130) unique values: 2 missing "": 0/38 tabulation: Freq. Value "Case management to improve adherence for HIV-infected patients receiving antiretroviral therapy in Ethiopia: a micro-costing study" "Mobile phones to support adherence to antiretroviral therapy: what would it cost the Indian National AIDS Control Program?" warning: variable has embedded blanks journal_etc Journal

type: string (str34)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

28 "Cost Eff Resour Alloc. 2011; 9: 18" 10 "J Int AIDS Soc. 2014; 17(1): 19036"

warning: variable has embedded blanks

url URL

type: string (str56)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

28 "<https://www.ncbi.nlm.nih.gov/pmc/artic

les/PMC3264532/>."

10 "<https://www.ncbi.nlm.nih.gov/pmc/artic

les/PMC4154142/>."

iso_name Country

type: string (str8)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

28 "Ethiopia" 10 "India"

location Location

type: string (str12)

unique values: 15 missing "": 0/38

"Axum" examples:

> "Dessie" "Gondar" "Karnataka"

warning: variable has embedded blanks

type: numeric (byte) range: [1,2] units: 1 unique values: 2 missing .: 0/38 tabulation: Freq. Value 28 1 10 2 pop_density Urbanicity ----type: string (str5) unique values: 2 missing "": 0/38 tabulation: Freq. Value 6 "rural" 32 "urban" ______ ownership Ownership ______ type: string (str17) unique values: 2 missing "": 0/38 tabulation: Freq. Value 28 "international NGO" 10 "mixed" warning: variable has embedded blanks platform Platform type: string (str15), but longest is str14 unique values: 1 missing "": 0/38 tabulation: Freq. Value 38 "Fixed facility" warning: variable has embedded blanks Platform specificity type: string (str68), but longest is str37

unique values: 2 missing "": 0/38

tabulation: Freq. Value

10 "Hospital - Tertiary (teaching)"

28 "Unspecified health care facility type"

warning: variable has embedded blanks

id_class Intervention Category

type: string (str18)

unique values: 1 missing "": 0/38

tabulation: Freq. Value

38 "Treatment and Care"

warning: variable has embedded blanks

id_int Intervention

type: string (str23)

unique values: 1 missing "": 0/38

tabulation: Freq. Value

38 "Retention and Adherence"

warning: variable has embedded blanks

type: string (str298)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

"The case management services consist of adherence counseling and support; health education; peer support; and referral of clients to community-based organizations equipped to address specific barriers to

adherence such as malnutrition,

substance abuse or material needs for

clothin, rent, and food."

10 "The mHealth intervention deployed in

the HIVIND trial had two components; the first was an automated IVR call and the second was a passive, neutral, picture message (SMS). Every patient in the intervention arm of the trial received each component of the mHealth intervention once a week."

warning: variable has embedded blanks

clinical_monitoring

Clinical monitoring

type: string (str2), but longest is str1

unique values: 1 missing "": 0/38

tabulation: Freq. Value

38 "."

demand_generation

Demand generation

type: string (str2), but longest is str1

unique values: 1 missing "": 0/38

tabulation: Freq. Value 38 "."

counseling_content C

Counseling content

type: string (str52)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

10 "adherence counseling"

28 "adherence counseling, peer-support;

health education"

warning: variable has embedded blanks

type: string (str78)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

28 "Case Manager (NR), Adherence Supporters (NR), Case Management Supervisors (NR)"

10 "Counselors (NR)"

warning: variable has embedded blanks

supportive_care Supportive care

type: string (str67)

unique values: 2 missing "": 0/38

tabulation: Freq. Value 10 "."

> 28 "nutrition support; substance abuse; material needs (clothing, rent)"

warning: variable has embedded blanks

visits Visit type & number

type: string (str24)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

10 "monthly follow-up visits"

28 "visits (NR)"

warning: variable has embedded blanks

referrals Referrals

type: string (str2), but longest is str1

unique values: 1 missing "": 0/38

tabulation: Freq. Value 38 "."

method

type: string (str1)

unique values: 1 missing "": 0/38

tabulation: Freq. Value

38 "."

id_tech Technology

type: string (str39)

unique values: 4 missing "": 0/38

tabulation: Freq. Value 28 "."

3 "SMS reminders" 4 "SMS+IVR"

3 "interactive voice response (IVR)

system"

warning: variable has embedded blanks

treatment_phase Treatment phase

type: string (str2), but longest is str1

unique values: 1 missing "": 0/38

tabulation: Freq. Value 38 "."

arv_regimen ARV Regimen

type: string (str2), but longest is str1

unique values: 1 missing "": 0/38

tabulation: Freq. Value 38 "."

treatment Treatment

type: string (str2), but longest is str1

unique values: 1 missing "": 0/38

tabulation: Freq. Value 38 "."

screening_diagnoses Screening and diagnosis

type: string (str3), but longest is str1

unique values: 1 missing "": 0/38

tabulation: Freq. Value 38 "."

community_awareness

Community awareness

type: string (str2), but longest is str1

unique values: 1 missing "": 0/38

tabulation: Freq. Value 38 "."

health_system Health system

type: string (str39)

unique values: 4 missing "": 0/38

tabulation: Freq. Value 28 "."

3 "SMS reminders"

4 "SMS+IVR"

3 "interactive voice response (IVR)

system"

warning: variable has embedded blanks

id_activities Costed activities

type: string (str63)

unique values: 4 missing "": 0/38

tabulation: Freq. Value

3 "adherence counseling, IVR"
3 "adherence counseling, SMS"
4 "adherence counseling, SMS+IVR"

28 "adherence counseling; health education;

peer support; referrals"

warning: variable has embedded blanks

start_month Start Month

type: numeric (byte)

range: [10,10] units: 1 unique values: 1 missing .: 10/38

tabulation: Freq. Value

28 10 10 .

Start Year start_year

type: numeric (int)

range: [2008,2008] units: 1 unique values: 1 missing .: 10/38

tabulation: Freq. Value

28 2008 10

end_month End Month

type: numeric (byte)

range: [9,9] units: 1 unique values: 1 missing .: 10/38

tabulation: Freq. Value

28 9 10 .

end_year End Year

type: numeric (int)

range: [2009,2009] units: 1 unique values: 1 missing .: 10/38

tabulation: Freq. Value

28 2009

10 .

period_portrayed Total Months

type: numeric (byte)

range: [12,12] units: 1 unique values: 1 missing .: 0/38

tabulation: Freq. Value

38 12

year_intro Year introduced at study site

type: numeric (int)

range: [2006,2006] units: 1 unique values: 1 missing .: 10/38

tabulation: Freq. Value 28 2006

10 .

coverage Coverage

type: numeric (byte)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

id_pop_dem_std Target group (demographic)

type: string (str18)

unique values: 1 missing "": 0/38

tabulation: Freq. Value

38 "General Population"

warning: variable has embedded blanks

id_pop_clin_std Target group (clinical)

type: string (str12)

unique values: 1 missing "": 0/38

tabulation: Freq. Value

38 "HIV positive"

warning: variable has embedded blanks

pop_age Average Age

type: string (str30)

unique values: 2 missing "": 0/38

tabulation: Freq. Value 10 "."

28 "85% aged 20-45; 15% aged 46-65"

warning: variable has embedded blanks

pop sex Gender

pop_sex Gende

type: string (str11)

unique values: 2 missing "": 0/38

tabulation: Freq. Value 10 "."

28 "58% females"

warning: variable has embedded blanks

pop_ses SE

type: string (str14)

unique values: 2 missing "": 0/38

tabulation: Freq. Value 10 "."

28 "75% unemployed"

warning: variable has embedded blanks

type: string (str14)

unique values: 2 missing "": 0/38

tabulation: Freq. Value 10 "."

28 "37% illiterate"

warning: variable has embedded blanks

hiv_prev HIV Prevalence

type: numeric (byte)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

tb_prev TB Prevalence

type: string (str2), but longest is str1

unique values: 1 missing "": 0/38

tabulation: Freq. Value 38 "."

tb_rx_resistance TB Drug Resistance

type: string (str1)

unique values: 1 missing "": 0/38

tabulation: Freq. Value 38 "."

costing_purpose Costing Purpose

type: string (str213)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

28 "To estimate unit costs of case

management services for ART patients in three provinces in Ethiopia, document the variation in costs, and explore opportunities for greater efficiency in

the provision of CM services"

10 "To study the costs of mobile phone

reminder strategies (mHealth

interventions) to support adherence in the context of India's National AIDS

Control Program (NACP)."

warning: variable has embedded blanks

timing Timing

type: string (str15)

unique values: 1 missing "": 0/38

tabulation: Freq. Value

38 "cross-sectional"

country_sampling Country Sampling

type: string (str3), but longest is str1

unique values: 1 missing "": 0/38

tabulation: Freq. Value

geo_sampling_incountry Geographic Area In Country Sampling

type: string (str9)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

10 "."

28 "purposive"

site_sampling Site Sampling

type: string (str1)

unique values: 1 missing "": 0/38

tabulation: Freq. Value 38 "."

px_sampling Patient Sampling

type: string (str14)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

28 "entire program"

10 "random"

warning: variable has embedded blanks

sample_size_derived Sample size formally derived

type: string (str3)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

28 "no"

10 "yes"

controls

type: string (str3)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

28 "no"

10 "yes"

econ_perspective_actual Perspective

type: string (str8)

unique values: 1 missing "": 0/38

tabulation: Freq. Value

38 "provider"

econ_costing Economic / Financial

type: string (str14)

missing "": 0/38 unique values: 2

tabulation: Freq. Value

28 "economic"

10 "financial only"

warning: variable has embedded blanks

Real World / Per Protocol real_world

type: string (str12)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

10 "per protocol" 28 "real world"

warning: variable has embedded blanks

asd_costs Above Service Costs Included

type: string (str19)

missing "": 0/38 unique values: 1

tabulation: Freq. Value

38 "some costs included"

warning: variable has embedded blanks

list_asd_costs Above Service Cost List

type: string (str177)

missing "": 0/38 unique values: 2

Value tabulation: Freq.

"development of intervention and web interface, recording intervention messages in different languages,

equipment (laptop), hardware
maintenance, program officer salary,

overhead"

28 "overhead & administrative costs at headquarters"

-

warning: variable has embedded blanks

omitted_costs Omitted Costs

type: string (str29)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

28 "Patient costs"

10 "Patient costs, Buliding/space"

warning: variable has embedded blanks

sensitivity_analysis Sensitivity Analysis

type: string (str7)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

10 "limited"
28 "none"

scale Economies of scale

type: string (str8)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

28 "analyzed" 10 "ignored"

research_costs Research Costs Included

type: string (str8)

unique values: 1 missing "": 0/38

tabulation: Freq. Value

38 "excluded"

unrelated_costs

Unrelated Costs Included

type: string (str2)

unique values: 1 missing "": 0/38

tabulation: Freq. Value 38 "no"

._____

overhead

Overhead Costs Included

type: string (str3)

unique values: 1 missing "": 0/38

tabulation: Freq. Value

38 "yes"

overhead_costs

Overhead Costs List

type: string (str108)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

10 "equipment (laptop, phone) for

demonstration, management personnel, "

28 "management personnel, curriculum development, office rental, travel

costs, telecommunications, IT,

utilities"

warning: variable has embedded and trailing blanks

Potential distortions pot_distortions _____

type: string (str4)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

10 "."

28 "none"

volunteer_time Valuing Volunteer Time ______ type: string (str3), but longest is str1 missing "": 0/38 unique values: 1 tabulation: Freq. Value 38 "." family_time Valuing Family Time type: string (str3), but longest is str1 missing "": 0/38 unique values: 1 tabulation: Freq. Value 38 "." px_costs_measured Patient-Incurred Costs Measured type: string (str12) unique values: 1 missing "": 0/38 tabulation: Freq. Value 38 "not included" warning: variable has embedded blanks Catastrophic Cost Calculated cat_cost ______ type: string (str4) unique values: 1 missing "": 0/38 tabulation: Freq. Value 38 "none" Reported Currency Year currency_yr

type: numeric (int)

range: [2009,2013] units: 1

unique values: 2 missing .: 0/38

tabulation: Freq. Value

28 2009 10 2013

iso_code Reported Currency

type: string (str3)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

10 "INR" 28 "USD"

currency_iso Currency of Data Collection

type: string (str3)

unique values: 2 missing "": 0/38

tabulation: Freq. Value

28 "ETB" 10 "INR"

type: string (str11)

unique values: 1 missing "": 0/38

tabulation: Freq. Value

38 "market only"

warning: variable has embedded blanks

current_x_rate Currency Exchange Rate

type: numeric (double)

range: [11.2,61.95] units: .01 unique values: 2 missing .: 0/38

tabulation: Freq. Value

28 11.2

discount_rate Discount Rate

type: numeric (byte)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

inflation Inflation rate

type: string (str2), but longest is str1

unique values: 1 missing "": 0/38

tabulation: Freq. Value 38 "."

mean_cost Mean Unit Cost

type: numeric (double)

range: [11.1416,1537.0665] units: 1.000e-08

unique values: 36 missing .: 0/38

mean: 175.994 std. dev: 396.569

percentiles: 10% 25% 50% 75% 90%

20.0549 28.9682 57.3792 100.523 160.439

si_personnel (SI)

type: numeric (double)

range: [355193.53,355193.53] units: 1

unique values: 1 missing .: 34/38

tabulation: Freq. Value

4 355193.53

34 .

si_per_service_delivery Personnel: Direct Service Delivery (SI) type: numeric (double) range: [39.600498,39.600498] units: 1 unique values: 1 missing .: 34/38 tabulation: Freq. Value 4 39.600498 34 si_per_support Personnel: Support (SI) type: numeric (double) units: 1 range: [355153.93,355153.93] unique values: 1 missing .: 34/38 tabulation: Freq. Value 4 355153.93 si_per_mixed_unspec Personnel: Mixed Unspecified (SI) type: numeric (float) units: . range: [.,.] unique values: 0 missing .: 38/38 tabulation: Freq. Value si_recurrent Recurrent Goods (SI) type: numeric (double) range: [60599.6,60680.885] units: .001 unique values: 3 missing .: 34/38 tabulation: Freq. Value 1 60599.6 1 60670.047 2 60680.885 34

si_rec_key_drugs Recurring: Supplies (key drugs) (SI)

type: numeric (float) range: [.,.] units: . unique values: 0 missing .: 38/38 tabulation: Freq. Value 38 . Recurring: Medical Supplies (excluding drugs) (SI) si_rec_med_int_supplies type: numeric (double) range: [10.838031,92.123263] units: 1.000e-06 unique values: 3 missing .: 34/38 tabulation: Freq. Value 1 10.838031 1 81.285233 2 92.123263 .----si_rec_nonmed_int_supplies Recurring: Non-medical Supplies (SI) type: numeric (float) range: [.,.] units: . unique values: 0 missing .: 38/38 tabulation: Freq. Value

si_rec_building_space Recurring: Building Space (SI)

type: numeric (float)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

Recurring: Other (SI) si_rec_other

type: numeric (double)

range: [60588.762,60588.762] units: 1 unique values: 1 missing .: 34/38

tabulation: Freq. Value

4 60588.762

34

type: numeric (double)

si_capital Capital (SI)

range: [28345.617,28345.617] units: 1 missing .: 34/38 unique values: 1

tabulation: Freq. Value

4 28345.617

34 .

Capital: Equipment (medical) (SI) si_cap_medical_equip

type: numeric (double)

range: [11671.725,11671.725] units: 1 unique values: 1 missing .: 34/38

tabulation: Freq. Value

4 11671.725

34 .

si_cap_equip_non_medical_inter Capital: Equipment (non-medical) (SI)

type: numeric (double)

units: 1 range: [16673.893,16673.893] unique values: 1 missing .: 34/38

tabulation: Freq. Value

4 16673.893

34

si_cap_build Capital: Building/space (SI)

type: numeric (float)

range: [.,.] units: . unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

si_cap_vehic Capital: Vehicles (SI)

type: numeric (float)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

si_cap_other Capital: Other (SI)

type: numeric (float)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

i_mixed Mixed (SI

type: numeric (float)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

a_primary_sd Primary Service Delivery (A)

type: numeric (double)

range: [39.6005,131.72376] units: .00001 unique values: 3 missing .: 34/38

tabulation: Freq. Value

1 39.600498

1 120.88573

2 131.72376

a_prisd_rc_counseling Primary SD: R&C Counseling (A)

type: numeric (double)

range: [39.600498,39.600498] units: 1 unique values: 1 missing .: 34/38

tabulation: Freq. Value

4 39.600498

34 .

a_prisd_unspecified Primary SD: Unspecified (A)

type: numeric (double)

range: [81.28523,92.12326] units: .00001 unique values: 2 missing .: 35/38

tabulation: Freq. Value

1 81.285233 2 92.123263

35 .

a_secondary_sd Secondary Service Delivery (A)

type: numeric (float)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

a_ancillary (A)

type: numeric (float)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

a_anc_demand_generation Ancillary: Demand Generation (A) _____ type: numeric (float) range: [.,.] units: . unique values: 0 missing .: 38/38 tabulation: Freq. Value 38 . Ancillary: Lab Services (A) a_anc_lab_services type: numeric (float) range: [.,.] units: . unique values: 0 missing .: 38/38 tabulation: Freq. Value a_anc_adhreten Ancillary: Adherence/Retention (A) type: numeric (float) range: [.,.] units: . unique values: 0 missing .: 38/38 tabulation: Freq. Value 38 a_anc_unspecified Ancillary: Unspecified (A) type: numeric (float) range: [.,.] units: . unique values: 0 missing .: 38/38 tabulation: Freq. Value 38 . Operational (A) a_operational

type: numeric (double)

range: [444088.31,444099.15] units: .01

unique values: 2 missing .: 34/38

tabulation: Freq. Value

3 444088.31 1 444099.15

34

type: numeric (double)

range: [18132.858,18132.858] units: 1

unique values: 1 missing .: 34/38

tabulation: Freq. Value

4 18132.858

34 .

type: numeric (float)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

type: numeric (double)

range: [322639.84,322639.84] units: 1

unique values: 1 missing .: 34/38

tabulation: Freq. Value

4 322639.84

34 .

type: numeric (double)

range: [32514.092,32514.092] units: 1

unique values: 1 missing .: 34/38

tabulation: Freq. Value

4 32514.092

34 .

Operational: Training (A) a_ope_training

type: numeric (float)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

Operational: Transportation (A) a_ope_transportation

type: numeric (float)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

Operational: Mass Education (A) a_ope_massed

type: numeric (float)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

Operational: HMIS and Record-Keeping (A) a_ope_hmis

type: numeric (float)

units: . range: [.,.]

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .

a_ope_technology_development Operational: Technology development (A) _____

type: numeric (double)

range: [11671.725,11671.725] units: 1

unique values: 1 missing .: 34/38

tabulation: Freq. Value

4 11671.725

34 .

a_ope_unspecified Operational: Unspecified (A)

type: numeric (double)

range: [40371.666,40371.666] units: 1

unique values: 1 missing .: 34/38

tabulation: Freq. Value

4 40371.666

34 .

a_mixed Mixed (A)

type: numeric (float)

range: [.,.] units: .

unique values: 0 missing .: 38/38

tabulation: Freq. Value

38 .