

IeDE Harmonist Data Toolkit Report

Report date: 2020-08-17

Example Harmonist Toolkit Report

Dataset Summary

Total number of patients in dataset: 28,089

Table 1: Table Summary

Table	Records	Patients	Age at Enrollment					
			0-4	5-9	10-14	15-19	20-24	Adults 25+
tblBAS	28089	28089	0	0	0	974	3388	23727
tblLTFU	28089	28089	0	0	0	974	3388	23727
tblVIS	1084237	26820	0	0	0	966	3316	22538
tblLAB_CD4	298041	27304	0	0	0	948	3317	23039
tblLAB_RNA	159234	15524	0	0	0	474	2014	13036
tblART	140435	25840	0	0	0	892	3016	21932
tblDIS	3750	3215	0	0	0	100	351	2764

Table 2: SITE in Dataset

SITE	Patients	tblLTFU	tblVIS	tblLAB_CD4	tblLAB_RNA	tblART	tblDIS
Hogwarts	594	594	537	557	461	536	36
Hufflepuff	11763	11763	11744	11447	239	11530	1933
Muggleton	5248	5248	5240	5145	4898	4397	650
Potterburg	3208	3208	3208	3168	3116	2723	69
Ravenclaw	1079	1079	0	983	983	904	0
Slytherin	458	458	413	429	308	426	38
Snapetown	4099	4099	4099	3971	3912	3800	345
Wizardville	1640	1640	1579	1604	1607	1524	144

Table 3: Summary of Uploaded IeDEA Tables

Table	Records	IeDEA DES Variables	Extra Variables
tblBAS	28089	PATIENT, PROGRAM, BIRTH_D, BIRTH_D_A, ENROL_D, ENROL_D_A, MODE, NAIVE_Y, PROPH_Y, RECART_Y, RECART_D, RECART_D_A, AIDS_Y, AIDS_D, AIDS_D_A, GENDER, HAART_D, HAART_D_A, CENTER	YEAR_ENROLLED, SITE
tblLTFU	28089	PATIENT, DROP_Y, DROP_D, DROP_D_A, DROP_RS, DEATH_Y, DEATH_D, DEATH_D_A, L_ALIVE_D, L_ALIVE_D_A	CENTER, SITE
tblVIS	1084237	PATIENT, CENTER, VIS_D, VIS_D_A, CDC_STAGE, WHO_STAGE, PREG_Y	GENDER_ID, SITE
tblLAB_CD4	298041	PATIENT, CD4_D, CD4_D_A, CD4_V, CD4_U	CENTER, SITE
tblLAB_RNA	159234	PATIENT, RNA_D, RNA_D_A, RNA_V	CENTER, SITE

Table 3: Summary of Uploaded IeDEA Tables (*continued*)

Table	Records	IeDEA DES Variables	Extra Variables
tblART	140435	PATIENT, ART_ID, ART_SD, ART_SD_A, ART_ED, ART_ED_A	CENTER, SITE
tblDIS	3750	PATIENT, DIS_ID, DIS_D, DIS_D_A, DIS_ED, DIS_ED_A, DIS_OUTCOME	CENTER, SITE

Table 4: Summary statistics from uploaded tables

Value	Count	Percent
Deceased		
No	25773	91.8
Yes	2316	8.2
Treatment Naive at Enrollment		
No	2940	10.5
Yes	25149	89.5
Receive Antiretroviral Therapy (ART)		
Yes	25840	92.0
Missing	2249	8.0

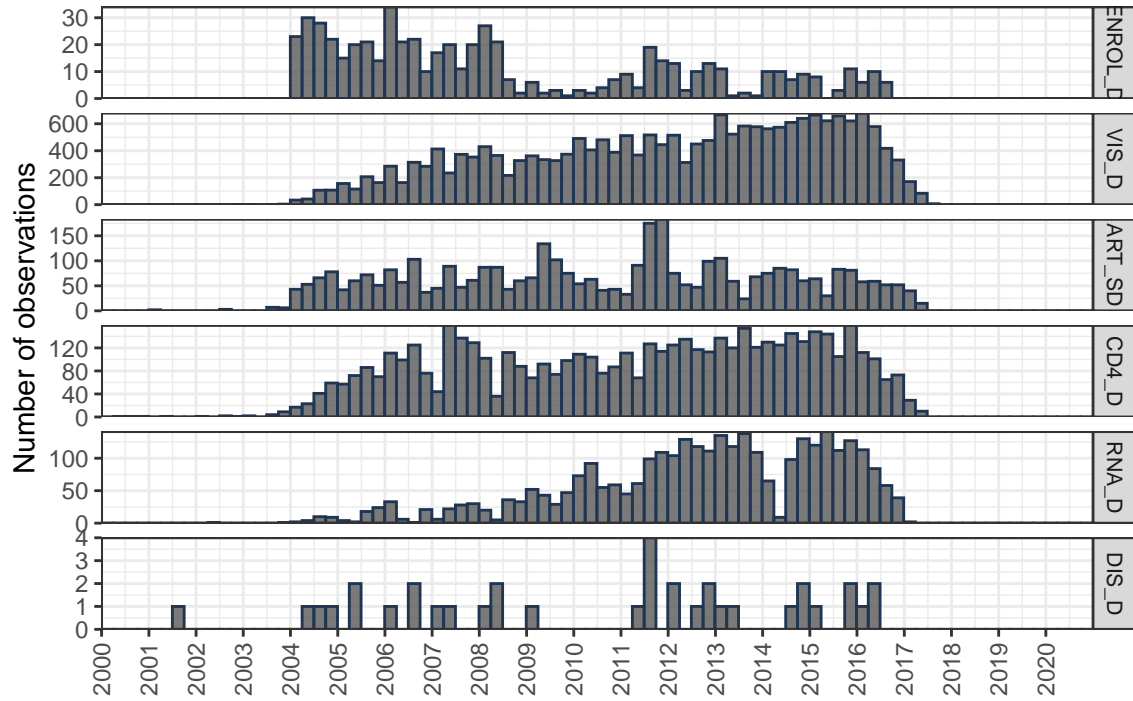
Table 5: Number of observations per year

Variable	< 2012	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
Enrolled	17245	2527	2859	2391	1439	1332	295	0	0	0	28088
Visits	527073	117341	124501	118630	120098	71094	5500	0	0	0	1084237
Deaths	1313	232	234	170	216	141	10	0	0	0	2316
Transfers Out	212	19	48	94	145	139	14	0	0	0	671
Viral Load	74750	15740	16776	16726	17731	15799	1708	0	0	0	159230
CD4	152935	30781	30465	33865	26465	20152	3378	0	0	0	298041

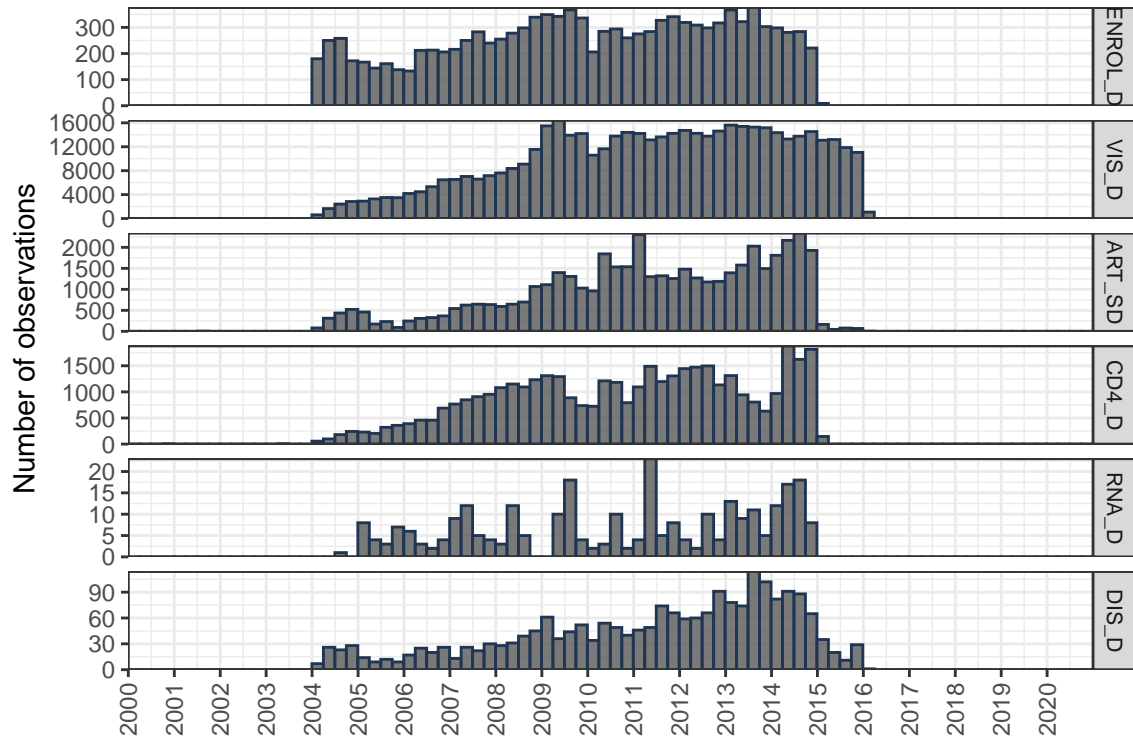
Histograms of important dates by SITE

SITE: Hogwarts

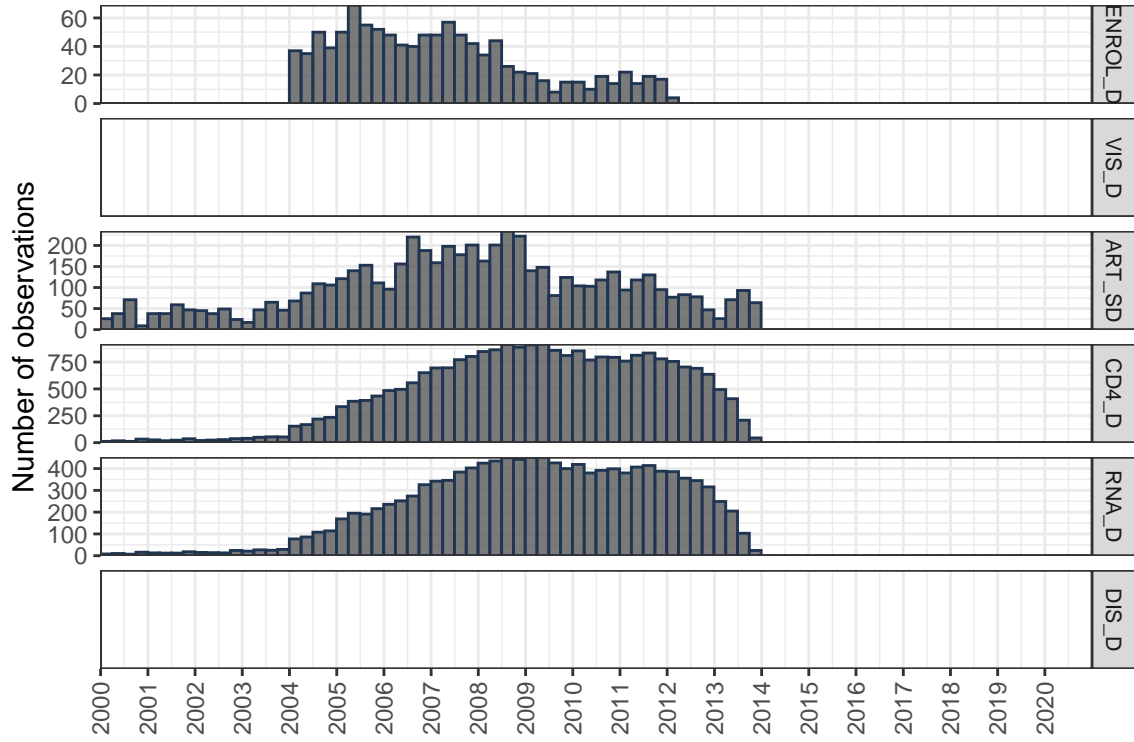
(Note: 23 observations before year 2000 and 1 observation after year 2020 are hidden)



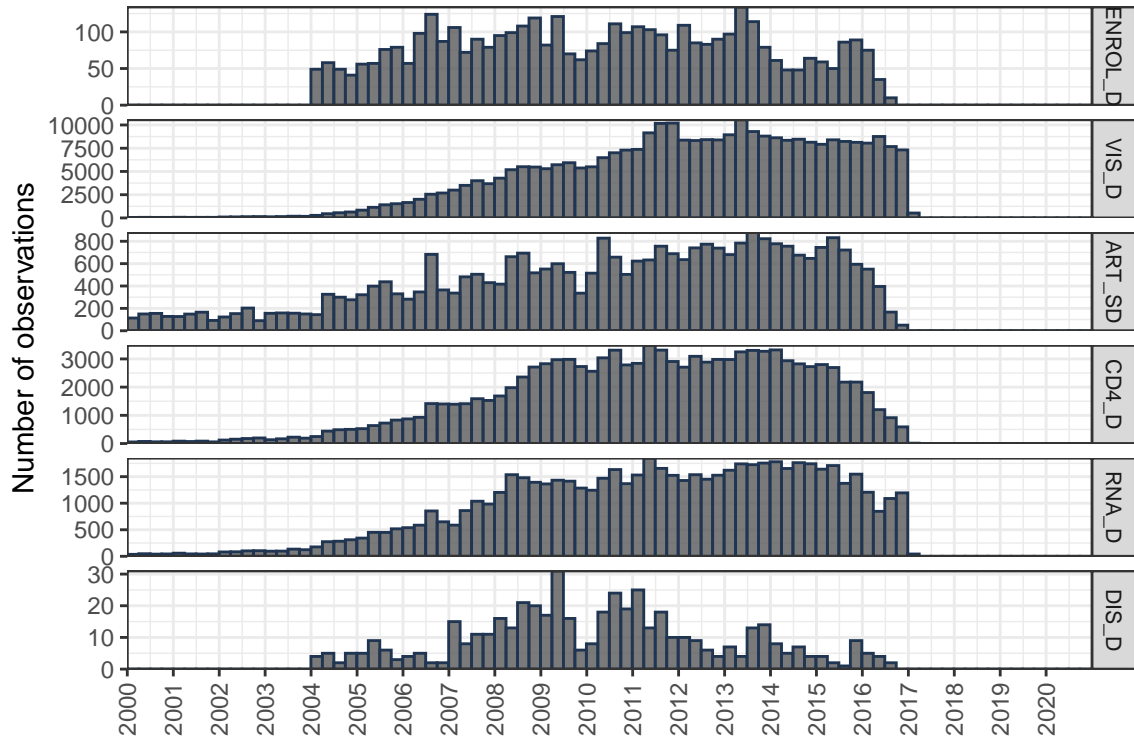
SITE: Hufflepuff



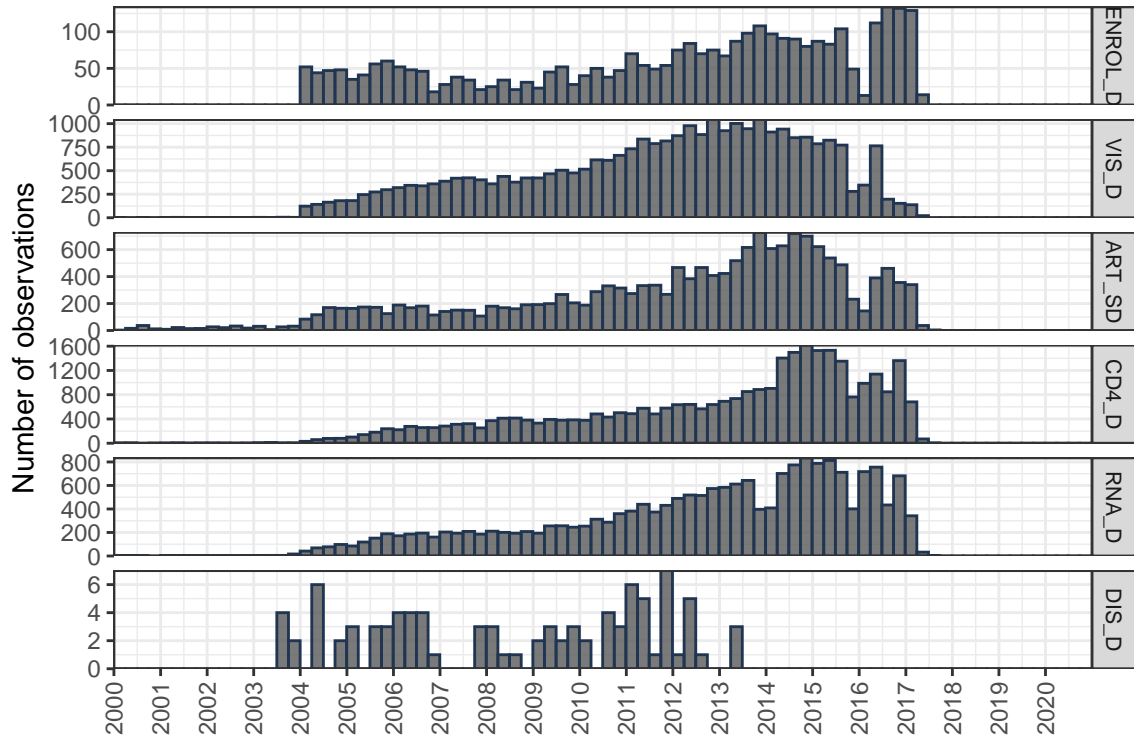
SITE: Ravenclaw (Note: 566 observations before year 2000 are hidden)



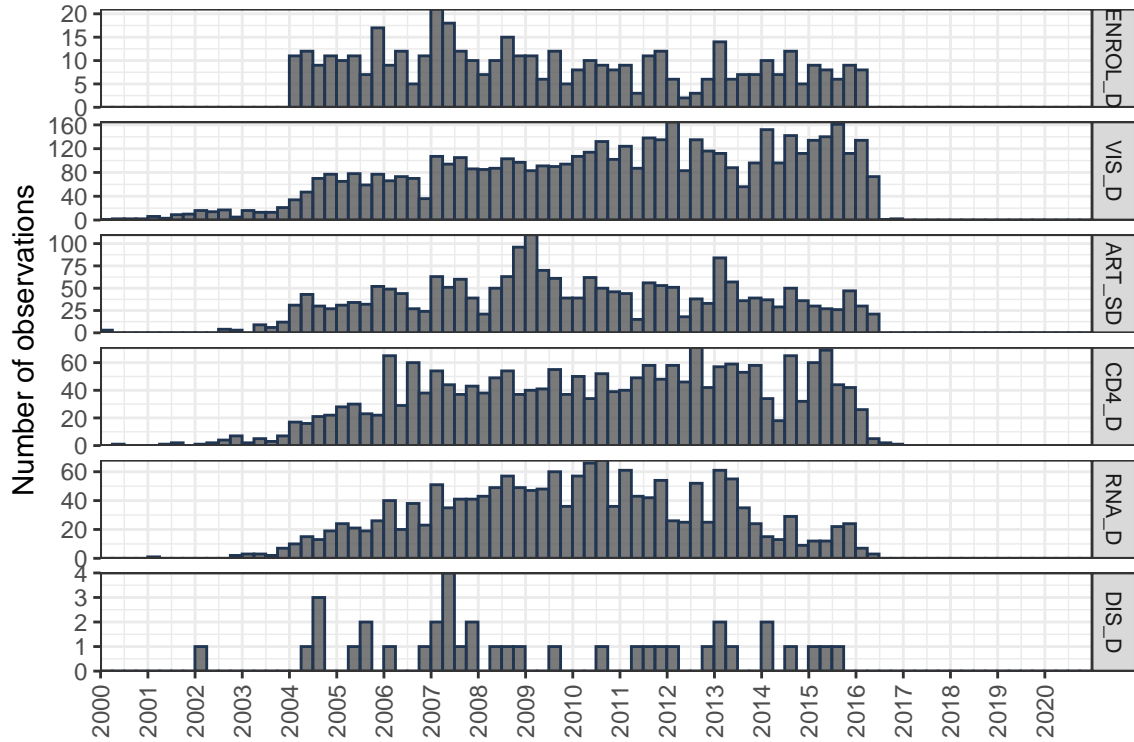
SITE: Snapetown (Note: 3681 observations before year 2000 are hidden)



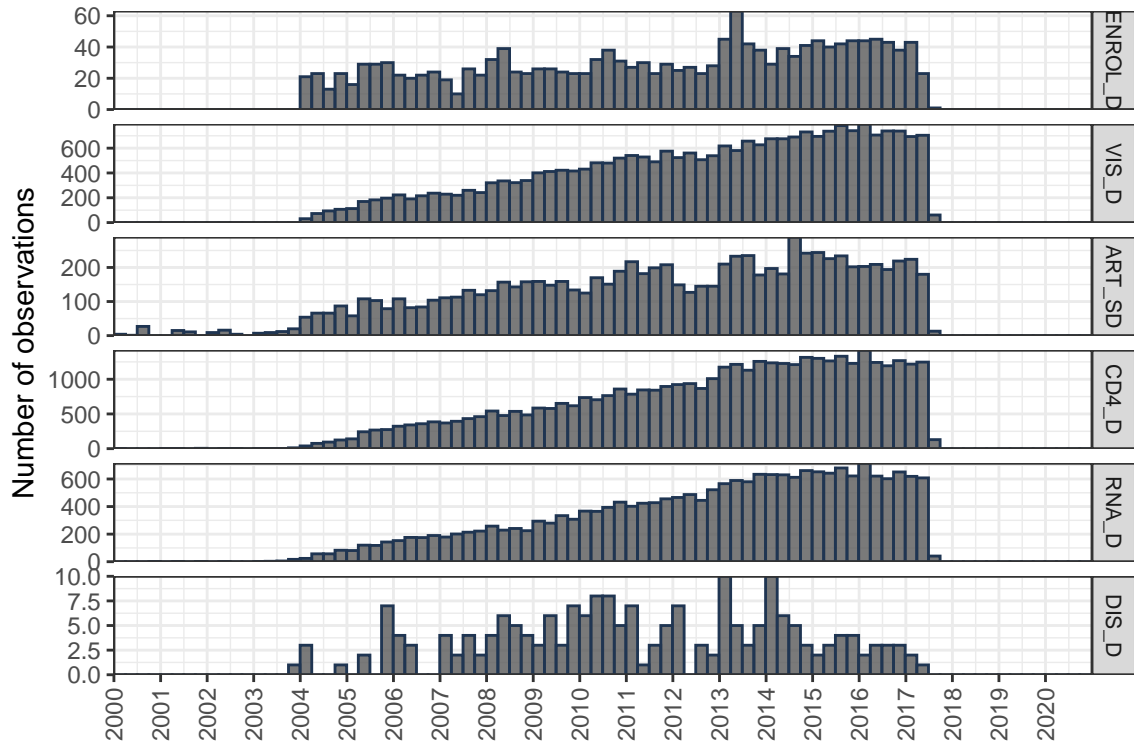
SITE: Potterburg (Note: 87 observations before year 2000 are hidden)



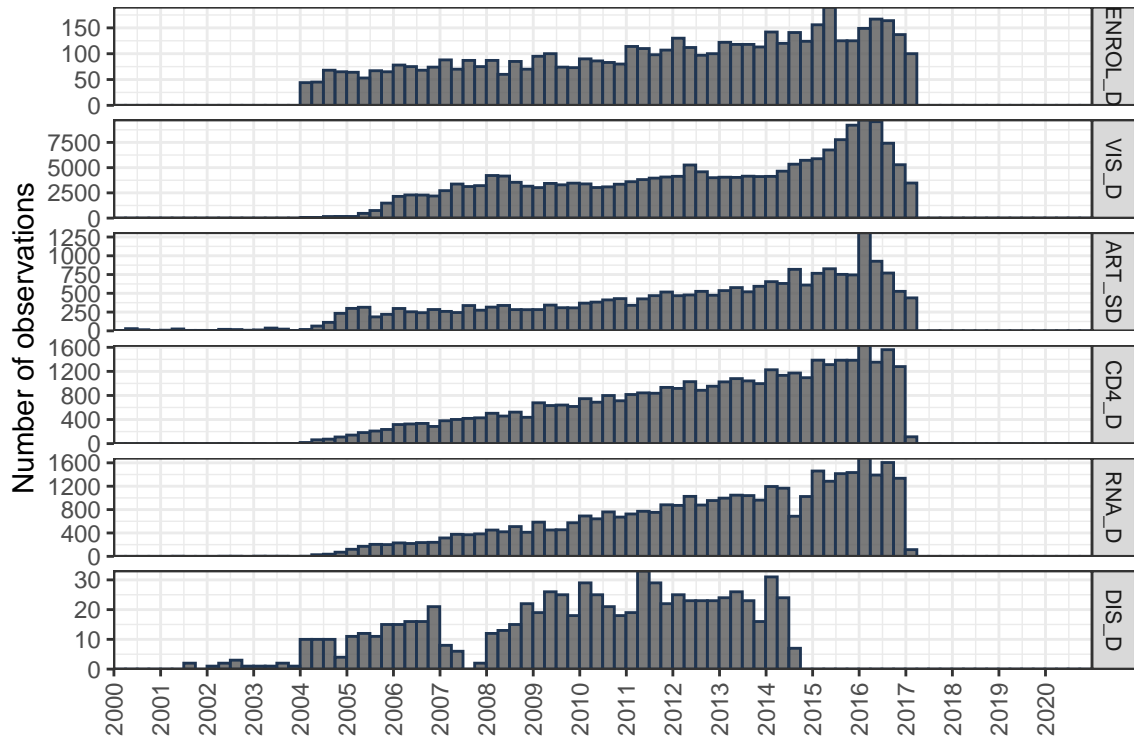
SITE: Slytherin (Note: 22 observations before year 2000 are hidden)



SITE: Wizardville (Note: 74 observations before year 2000 are hidden)



SITE: Muggleton (Note: 78 observations before year 2000 are hidden)



Data Quality Summary

Note: Download Error Detail spreadsheet from the Toolkit for more information

Table 6: Summary of Errors

Description	Variable	Count	Percent
tbIBAS			
Missing Required Variable	SEX	28089	100
Invalid Code	HAART_D_A	2	0
Invalid Code	RECART_D_A	2	0
Future date	ENROL_D	1	0
Deprecated variable	GENDER	1	0
Deprecated variable	HAART_D	1	0
Deprecated variable	HAART_D_A	1	0
Invalid Code	MODE	1	0
tbLTFU			
Invalid Code	DROP_RS	4	0
Invalid Code	L_ALIVE_D_A	4	0
DEATH_D before RNA_D	DEATH_D	1	0
DEATH_D before VIS_D	DEATH_D	1	0
Y/N data in conflict with date	DROP_D	1	0
L_ALIVE_D before ENROL_D	L_ALIVE_D	1	0
L_ALIVE_D before RNA_D	L_ALIVE_D	1	0
L_ALIVE_D before VIS_D	L_ALIVE_D	1	0
tbIVIS			
Invalid Code	CDC_STAGE	1	0
tbILAB_CD4			
Value Above Expected Range	CD4_V	1	0
tbILAB_RNA			
Missing Required Variable	RNA_D	4	0
Duplicate Record	PATIENT	2	0
tbIART			
Duplicate Record	PATIENT	35	0
Missing Required Variable	ART_ID	34	0
Invalid Code	ART_SD_A	10	0
Invalid Code	ART_ED_A	3	0
ART_ED before ART_SD	ART_ED	1	0
Invalid Code	ART_ID	1	0
Missing Required Variable	ART_SD	1	0

Table 7: Summary of Warnings

Description	Variable	Count	Percent
tbIBAS			
BIRTH_D before 1920	BIRTH_D	3	0
Date before 1980	AIDS_D	1	0
Recently deprecated variable	CENTER	1	0
tbIVIS			
Date before 1980	VIS_D	1	0

Table 8: Summary of Invalid Codes

Variable	InvalidCode	Count	Percent
tblART			
ART_SD_A	Y<	10	0
ART_ED_A	Y<	3	0
ART_ID	J05AF05_J05AF03_J05AF07	1	0
tblBAS			
HAART_D_A	Y<	2	0
RECART_D_A	Y<	2	0
MODE	999	1	0
tblLTFU			
L_ALIVE_D_A	A	4	0
DROP_RS	99	3	0
DROP_RS	89	1	0
tblVIS			
CDC_STAGE	3	1	0

Table 9: Summary of Important Missing Variable Entries

table	variable	number	percent
tblART	ART_ID	34	0
tblART	ART_SD	1	0
tblBAS	RECART_Y	2249	8
tblBAS	SEX	28089	100
tblLAB_RNA	RNA_D	4	0

	tbILTFU	tbIVIS	tbILAB_CD4	tbILAB_RNA	tbIART	tbIDIS
Hogwarts	100	90	93	77	90	6
Hufflepuff	100	99	97	2	98	16
Muggleton	100	99	98	93	83	12
Potterburg	100	100	98	97	84	2
Ravenclaw	100	0	91	91	83	0
Slytherin	100	90	93	67	93	8
SnapeTown	100	100	96	95	92	8
Wizardville	100	96	97	98	92	8

Figure 1: Percent of Patients from tbIBAS Included

	ART_ID	ART_SD	MODE	RE CART_Y	RNA_D	SEX
Hogwarts	100	100	62	90	100	0
Hufflepuff	100	100	4	98	100	0
Muggleton	100	100	99	83	100	0
Potterburg	100	100	99	84	100	0
Ravenclaw	100	100	50	83	100	0
Slytherin	100	100	79	93	100	0
Snape town	100	100	86	92	100	0
Wizardville	99	100	93	92	100	0

Figure 2: Percent Variable Entries Complete (Not Missing/Unknown)