

3/23/2016

Data Quality Report

Tuberculosis (TB) Registry database

1993-2015

Maintained by Manitoba Health that include all individuals who have been diagnosed with active TB disease, individuals who have contacted with the case and those with LTBI who receive treatment, demographic information of clients, Xray, drug dosage and test information.



UNIVERSITY
OF MANITOBA

Faculty of
Health Sciences

Say Hong
Manitoba Centre for Health Policy



Contents

Overview of Database.....	4
Data Model.....	5
Accuracy: Completeness and Correctness (VIMO Table).....	6
Client Table	6
Address Table.....	8
Case Table	9
Clinic Visit Table	11
Contact Table	12
Contact List Comment Table	13
Daily Drug Table	14
Direct Observed Therapy Drug Doses Table	16
Direct Observed Therapy Drug List Table	18
Direct Observed Drug Therapy Master Table	19
Drug Reaction Table.....	20
Drug Summary Table.....	21
Diagnosis Table	22
Hematology Result Table	23
HIV Test Table	24
Hospital Table	25
Immigrant Tracking Table	26
Inpatient Drug Table	27
Other Diseases Table.....	28
Other Test Table.....	29
Sources table.....	31
Study Table.....	32
Study Master Table	33
Surveillance Follow up Table.....	34

This report was based on methods and concepts described in the following document:

Azimaee M, Smith M, Lix L, Ostapyk T, Burchill C, Hong SP. *MCHP Data Quality Framework*. Winnipeg, Manitoba, Canada: Manitoba Centre for Health Policy, University of Manitoba; 2013. http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departamental_units/mchp/protocol/media/Data_Quality_Framework.pdf

Survey Table	36
Survey Comment Table	37
Survey Master Table	38
Symptom Table	39
Tuberculin Table.....	40
Visual test table.....	41
Xray Table.....	42
Referential Integrity Check	44
Linkability	46
Phin Type.....	46
Linkability for active case and contact individuals.....	46
Agreement	46
Internal Validity: Trend Analysis	47

Summary

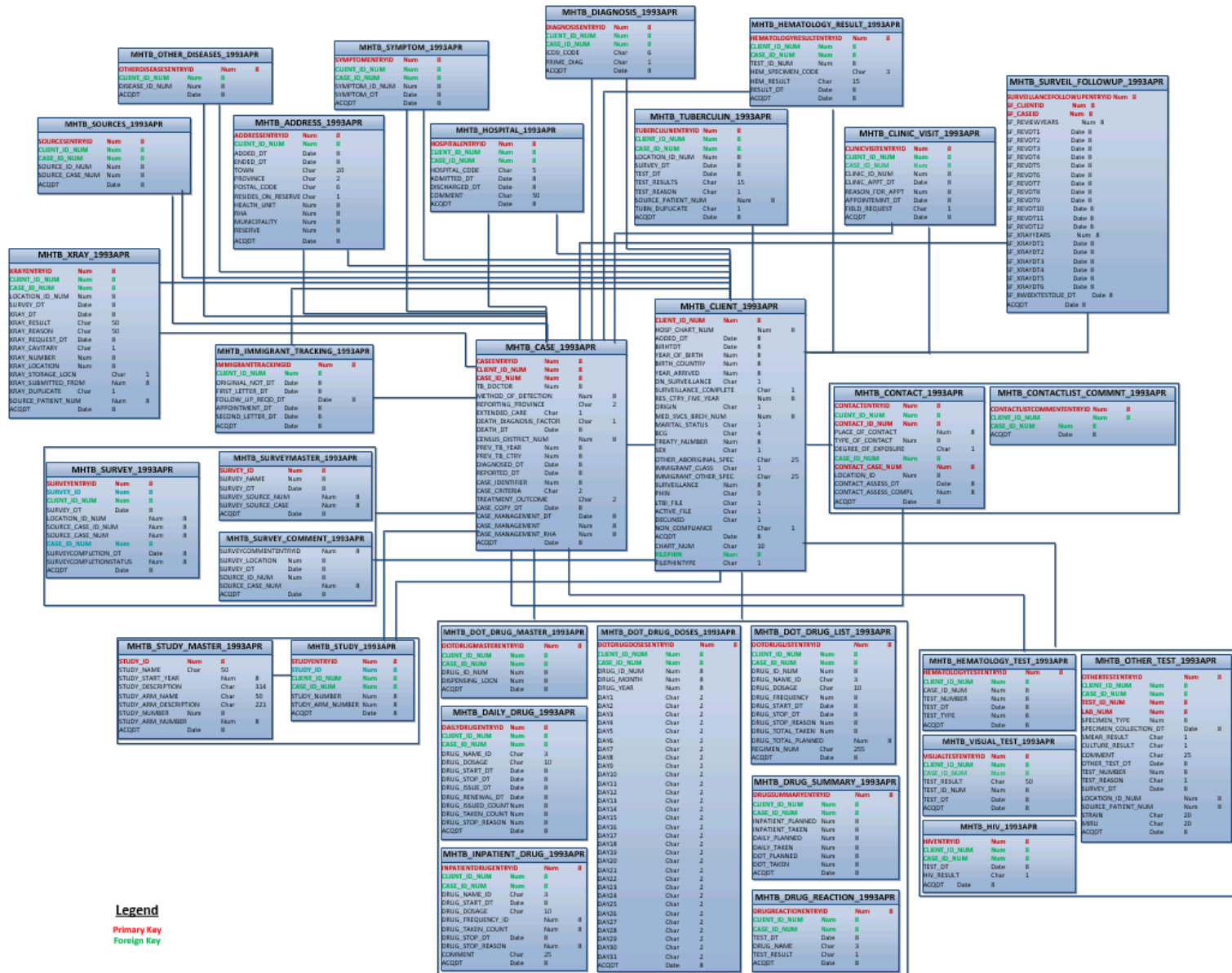
- Linkable records are low at about 50%. However, the client table contains both the active case as well as contact individuals. Percentage of linkable records for the active case is at about 72% and 62% for the contact individuals.

Overview of Database

Domain	SPDS Table	Dataset Label	Number of Records	Number of Fields
HEALTH	MHTB_ADDRESS_1993APR	MH Tuberculosis Registry - address table: 1993 onward	76074	13
HEALTH	MHTB_CASE_1993APR	MH Tuberculosis Registry - Case table: 1993 onward	4128	22
HEALTH	MHTB_CLIENT_1993APR	MH Tuberculosis Registry - Client: 1993 onward	72422	28
HEALTH	MHTB_CLINIC_VISIT_1993APR	MH Tuberculosis Registry - clinic visit table: 1993 onward	846	9
HEALTH	MHTB_CONTACTLIST_COMMNT_1993APR	MH Tuberculosis - ContactList comment: 1993 onward	427	4
HEALTH	MHTB_CONTACT_1993APR	MH Tuberculosis Registry - Contact table: 1993 onward	47880	12
HEALTH	MHTB_DAILY_DRUG_1993APR	MH Tuberculosis Registry - daily drug table: 1993 onward	8468	13
HEALTH	MHTB_DIAGNOSIS_1993APR	MH Tuberculosis Registry - diagnosis table: 1993 onward	19990	6
HEALTH	MHTB_DOT_DRUG_DOSES_1993APR	MH Tuberculosis Registry - Direct Observed Therapy Drug Doses: 1993 onward	49047	38
HEALTH	MHTB_DOT_DRUG_LIST_1993APR	MH Tuberculosis Registry - Direct Observed Therapy Drug List: 1993 onward	30498	14
HEALTH	MHTB_DOT_DRUG_MASTER_1993APR	MH Tuberculosis Registry - Direct Observed Therapy Drug Master: 1993 onward	9003	6
HEALTH	MHTB_DRUG_REACTION_1993APR	MH Tuberculosis Registry - Drug Reaction: 1993 onward	13009	7
HEALTH	MHTB_DRUG_SUMMARY_1993APR	MH Tuberculosis Registry - Drug Summary: 1993 onward	8364	10
HEALTH	MHTB_HEMATOLOGY_RESULT_1993APR	MH Tuberculosis Registry - Hematology Result: 1993 onward	36	8
HEALTH	MHTB_HIV_1993APR	MH Tuberculosis Registry - HIV: 1993 onward	2894	6
HEALTH	MHTB_HOSPITAL_1993APR	MH Tuberculosis Registry - Hospital: 1993 onward	2428	8
HEALTH	MHTB_IMMIGRANT_TRACKING_1993APR	MH Tuberculosis Registry - Immigrant Tracking: 1993 onward	4759	8
HEALTH	MHTB_INPATIENT_DRUG_1993APR	MH Tuberculosis Registry - Inpatient Drug: 1993 onward	12816	12
HEALTH	MHTB_OTHER_DISEASES_1993APR	MH Tuberculosis Registry - Other Diseases: 1993 onward	941	4
HEALTH	MHTB_OTHER_TEST_1993APR	MH Tuberculosis Registry - Other Test: 1993 onward	14879	19
HEALTH	MHTB_SOURCES_1993APR	MH Tuberculosis Registry - Sources: 1993 onward	53460	6
HEALTH	MHTB_STUDY_1993APR	MH Tuberculosis Registry - Study: 1993 onward	73	7
HEALTH	MHTB_STUDY_MASTER_1993APR	MH Tuberculosis Registry - Study Master: 1993 onward	6	9
HEALTH	MHTB_SURVEIL_FOLLOWUP_1993APR	MH Tuberculosis Registry - Surveillance Follow up: 1993 onward	173	25
HEALTH	MHTB_SURVEY_1993APR	MH Tuberculosis Registry - Survey: 1993 onward	37743	11
HEALTH	MHTB_SURVEY_COMMENT_1993APR	MH Tuberculosis Registry - Survey Comment: 1993 onward	44	6
HEALTH	MHTB_SURVEY_MASTER_1993APR	MH Tuberculosis Registry - Survey Master: 1993 onward	1026	6
HEALTH	MHTB_SYMPTOM_1993APR	MH Tuberculosis Registry - Symptom: 1993 onward	11814	6
HEALTH	MHTB_TUBERCULIN_1993APR	MH Tuberculosis Registry - Tuberculin: 1993 onward	71905	11
HEALTH	MHTB_VISUAL_TEST_1993APR	MH Tuberculosis Registry - Visual Test: 1993 onward	485	7
HEALTH	MHTB_XRAY_1993APR	MH Tuberculosis Registry - Xray: 1993 onward	77634	17

Data Model

Tuberculosis (TB) Registry

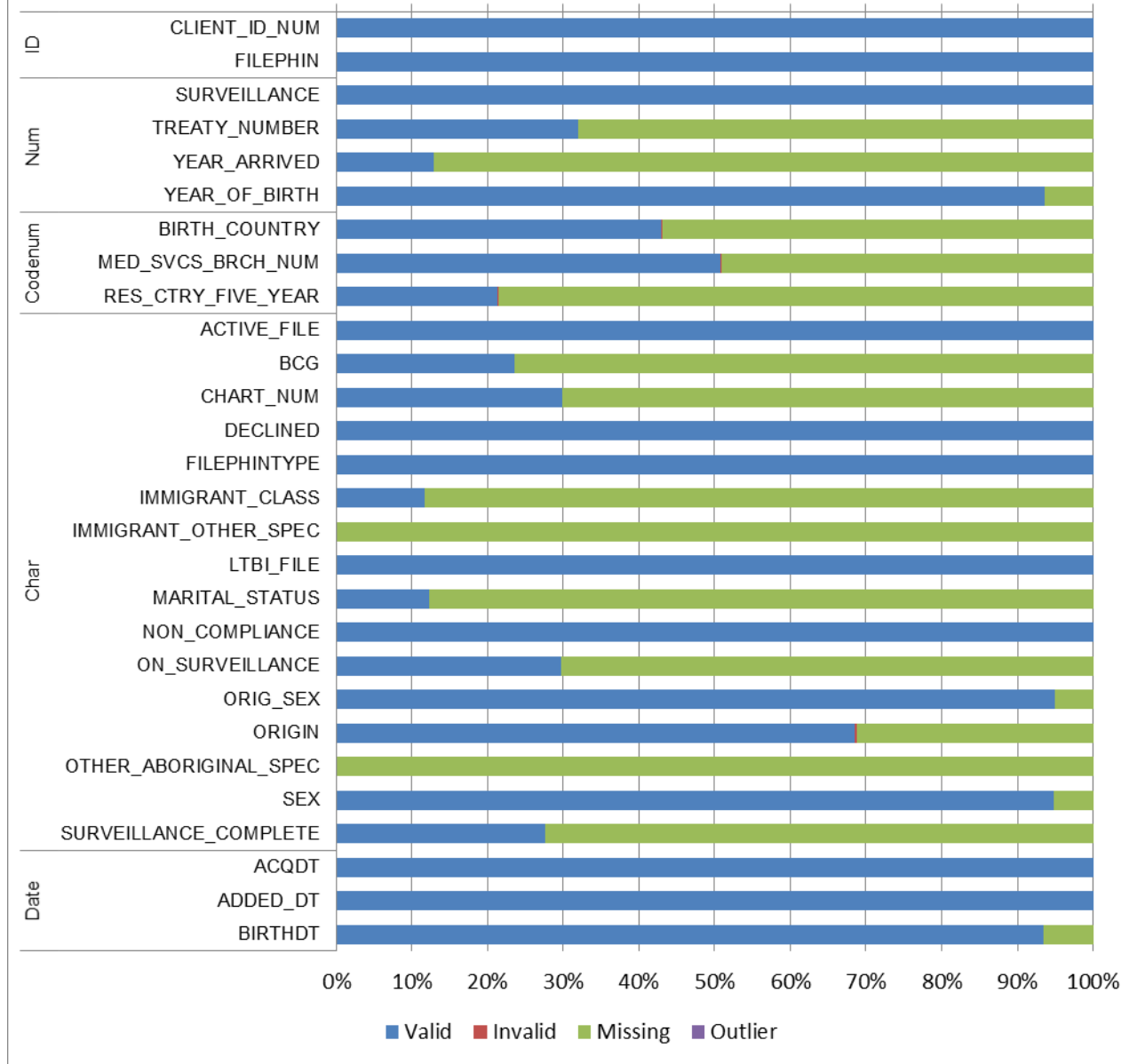


Accuracy: Completeness and Correctness (VIMO Table)

Client Table

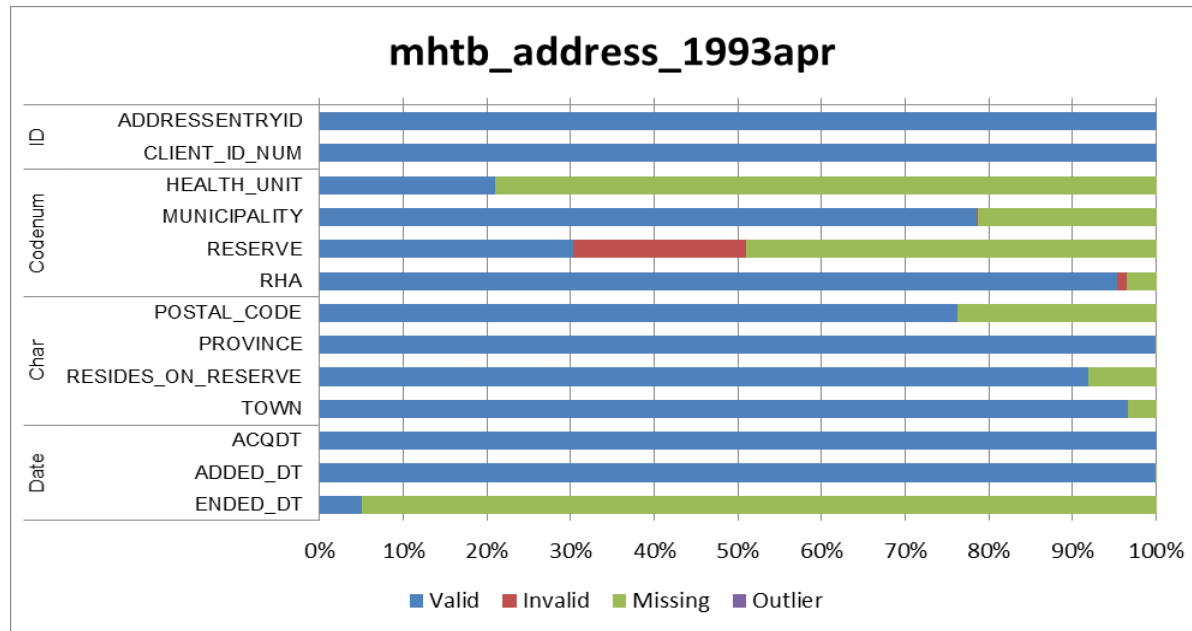
Dataset Label: MH Tuberculosis Registry - Client: 1993 onward			Records: 72422				Legend (Potential Data Quality Problems) :						
Dataset Name: mhtb_client_1993apr			Period: 1993-2015				None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
SUPPRESSED = Variables being suppressed in data file = No variance or 100% missing value			Legend for comment column										
			Blank = variables have not been tested (no formats have been specified for the variables)										
			✓ = Variables have been tested against the associated formats and no invalid values found										
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment	
ID	CLIENT_ID_NUM	Individual client ID number	100.00		.00								
	FILEPHIN	MH SCrambled PHIN	100.00		.00								
Num	SURVEILLANCE	surveillance	100.00		.00		0.00	1.00	.06	.00	.24		
	TREATY_NUMBER	Treaty number for status clients	32.00		68.00		SUPPRESSED						
	YEAR_ARRIVED	If foreign born, year client arrived in Canada	12.85		87.15		0.00	22003.00	2007.69	2009.35	299.10		
	YEAR_OF_BIRTH	Client year of birth	93.55		6.45		0.00	20144.00	1970.53	1973.01	102.44		
Top 10 Observed Values													
Codenum	BIRTH_COUNTRY	Country in which client was born	42.89	.21	56.91		909, 608, 356, 999, 231, 156, 404, 704, 408, 736					230, 282, 287, 345, ... (151 Invalid Obs. in total)	
	MED_SVCS_BRCH_NUM	If status, medical services branch number	50.78	.18	49.03		0, 420301, 420401, 410301, 420201, 420501, 420701, 410601,					85682, 111111, 310109, 420602, 545102 (132 Invalid Obs. in total)	
	RES_CTRY_FIVE_YEAR	Country in which client resided five years ago	21.32	.13	78.54		909, 608, 704, 9999, 356, 705, 156, 230, 344, 616					230, 282, 287, 345, ... (97 Invalid Obs. in total)	
Char	ACTIVE_FILE	Active TB file	100.00		.00		0, 1						
	BCG	Year of BCG, or may be a comment such as "scar" indicating BCG even if year not known	23.50		76.50		none, scar, 1995, 1996, 1994, 1997, 1991, 1989, 1990, 1998						
	CHART_NUM	Hospital chart number	29.79		70.21		SUPPRESSED						
	DECLINED	Client declined treatment	100.00		.00		0, 1						
	FILEPHINTYPE	MCHP Method to determine FILEPHIN	100.00		.00		4, 0						
	IMMIGRANT_CLASS	Status of immigrant - refugee, landed immigrant etc.	11.73		88.27		L, N, R, U, O						✓
	IMMIGRANT_OTHER_SPEC	Other notes: student for visitor	.02		99.98		Student, I, student, 1987, Kosovo, Temporary Resident, Visitor,						
	LTBI_FILE	Latent TB file	100.00		.00		0, 1						
	MARITAL_STATUS	Marital status	12.20	.03	87.77		M, S, U, D, W, P, E						E (19 Invalid Obs. in total)
	NON_COMPLIANCE	Client not complying with treatment	100.00		.00		0, 1						
	ON_SURVEILLANCE	Is client on surveillance?	29.77		70.23		U, Y, N, T						
	ORIG_SEX	Original sex values	94.90		5.10		F, M, U, m, G, f						
	ORIGIN	origin - Canadian born, foreign born, Inuit etc.	68.57	.16	31.27		S, N, F, C, I, U, M						U (117 Invalid Obs. in total)
	OTHER_ABORIGINAL_SPEC	Note whether member of band outside of Manitoba.	.01		99.99		1395, Ft. McPherson Band, Nicickhousemenecaning ON, Sask.						
SEX	Sex (MCHP)	94.88		5.12		2, 1							
SURVEILLANCE_COMPLETE	Is surveillance complete?	27.63		72.37		U, N, Y, T							
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11					
	ADDED_DT	Date client data added to database	100.00		.00		1989-12-31	2016-01-11					
	BIRTHDT	Client date of birth	93.46		6.54		1886-10-16	9731-11-01					

mhtb_client_1993apr



Address Table

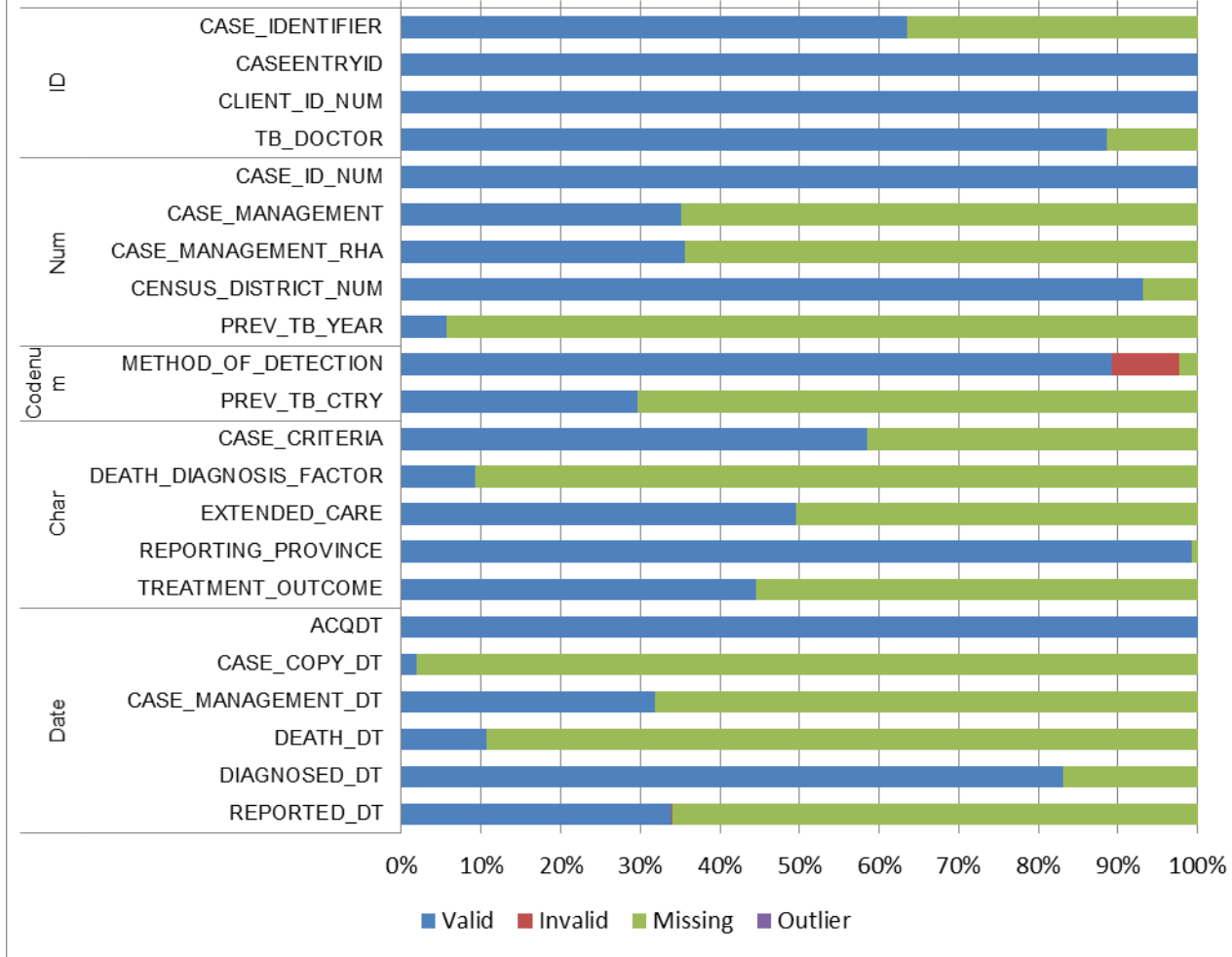
Dataset Label: MH Tuberculosis Registry - address table: 1993 onward				Records: 76074		Legend (Potential Data Quality Problems) :						
Dataset Name: mhtb_address_1993apr = No variance or 100% missing value				Period: 1993-2015		None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
Legend for comment column												
Blank = variables have not been tested (no formats have been specified for the variables)												
✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	ADDRESSEENTRYID	Automatically generated number for each unique address entry	100.00		.00							
	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
Top 10 Observed Values												
Codenum	HEALTH_UNIT	Number which corresponds to health unit	21.03		78.97		25, 24, 22, 23, 7, 20, 19, 1, 5, 4					✓
	MUNICIPALITY	Number which corresponds to municipality	78.47	.24	21.29		453, 452, 450, 382, 353, 357, 107, 379, 363, 129					0, 19, 22, 29, ... (184 Invalid Obs. in total)
	RESERVE	Number which corresponds to reserve name	30.28	20.67	49.05		0, 30, 17, 18, 8, 34, 44, 48, 23, 47					0 (15724 Invalid Obs. in total)
	RHA	Number which corresponds to regional health authority	95.31	1.20	3.49		10, 80, 70, 20, 40, 30, 45, 60, 15, 1					0, 1 (912 Invalid Obs. in total)
Char	POSTAL_CODE	6 digit postal code	76.21		23.79		R0B1A0, R0B0T0, R0B0M0, R0B0J0, R0B1C0, R0B1K0, R0B					
	PROVINCE	name of province	99.97		.03		MB, ON, NU, AB, SK, BC, QC, NS, NT, MN					
	RESIDES_ON_RESERVE	Yes/no	91.85		8.15		N, Y, U, n					
	TOWN	Name of city/community/town	96.56		3.44		Winnipeg, Nelson House, Island Lake, God's Lake Narrows, T					
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	ADDED_DT	Date address added	99.95		.05		1957-03-28	2016-01-11				
	ENDED_DT	Date address ended	4.96		95.04		1968-12-31	2016-01-10				



Case Table

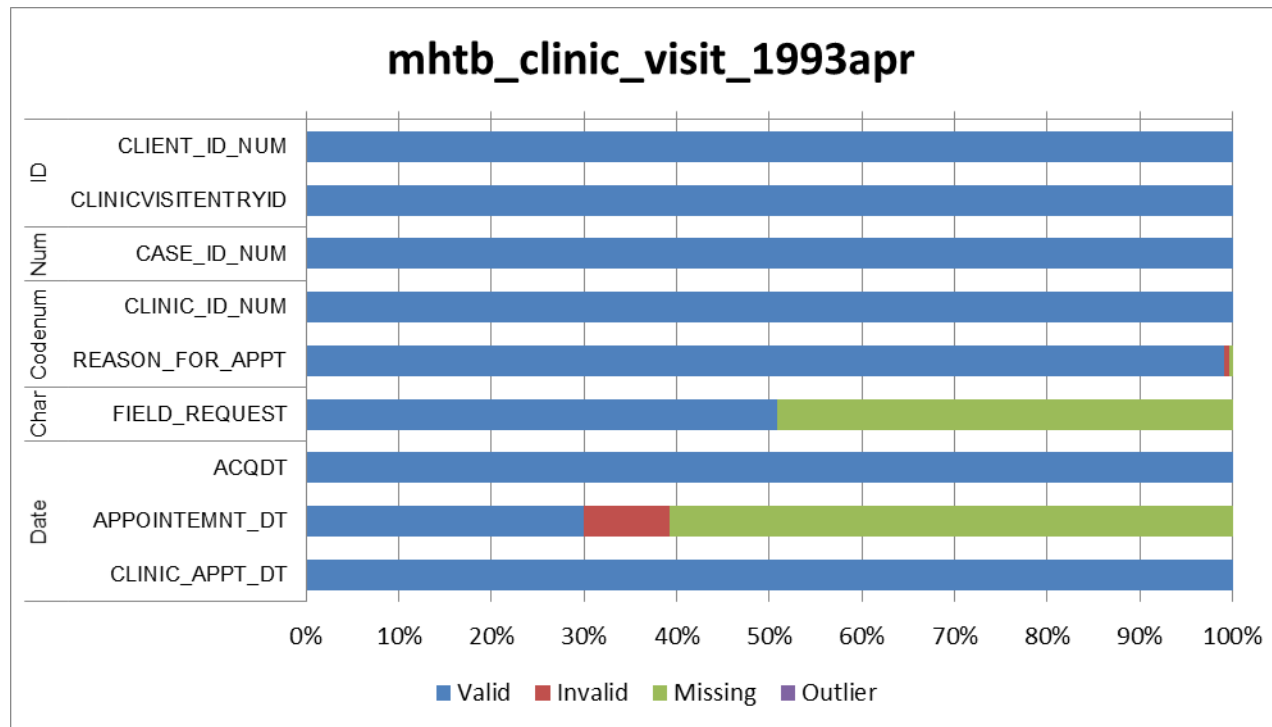
Dataset Label: MH Tuberculosis Registry - Case table: 1993 onward			Records: 4128			Legend (Potential Data Quality Problems) :						
Dataset Name: mhtb_case_1993apr = No variance or 100% missing value			Period: 1993-2015			None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables) ✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CASE_IDENTIFIER	7 digit number that combines year of diagnosis and case number in that year	63.57		36.43							
	CASEENTRYID	Automatically generated number for each unique address entry	100.00		.00							
	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	TB_DOCTOR	Unique identifier that links to physician name and location of practice	88.57		11.43							
Num	CASE_ID_NUM	Specific case #1 = first case, #2 = second case etc.	100.00		.00		0.00	4.00	1.05	1.00	.31	
	CASE_MANAGEMENT	How case has been managed	35.20		64.80		0.00	3.00	1.91	2.00	.55	
	CASE_MANAGEMENT_RHA	Number corresponding to the RHA that is responsible for managing the case	35.54		64.46		0.00	99.00	17.44	10.00	23.37	
	CENSUS_DISTRICT_NUM	Census district number	93.12		6.88		0.00	62.00	14.89	11.00	6.14	
	PREV_TB_YEAR	Year in which client previously diagnosed with TB if this is not case number 1	5.64		94.36		35.00	2014.00	1284.43	1652.56	924.52	
Top 10 Observed Values												
Codenum	METHOD_OF_DETECTION	Method by which this case was detected. Options include 'symptoms', 'screening' etc.	89.17	8.53	2.30		12, 2, 99, 30, 13, 0, 14, 3, 7, 57					-8, -7, -5, -4, ... (352 Invalid Obs. in total)
	PREV_TB_CTRY	Country in which client was previously diagnosed with TB	29.60		70.40		909, 231, 961, 962					✓
Char	CASE_CRITERIA	Positive culture (1), no positive culture (2), unknown (9)	58.43		41.57		1, 2, 9					✓
	DEATH_DIAGNOSIS_FACTOR	If client is deceased, three options as to whether TB was contributing factor	9.21		90.79		2, 3, 1					✓
	EXTENDED_CARE	extended_care	49.56		50.44		N, Y					
	REPORTING_PROVINCE	Province from which the TB case originated	99.20		.80		46, 61, 35, 62, 47, 48, 59, 12, 45					
Date	TREATMENT_OUTCOME	Options are 1-9. example death during treatment, treatment complete, absconded etc.	44.57		55.43		2, 1, 3, 7, 4, 6, 8					
	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	CASE_COPY_DT	date_case_copy	1.82		98.18		2007-01-12	2015-12-18				
	CASE_MANAGEMENT_DT	date_case_management	31.90		68.10		2007-01-12	2016-01-08				
	DEATH_DT	Date of death if clients deceased	10.73		89.27		1985-01-07	2015-12-12				
	DIAGNOSED_DT	Date current case was diagnosed	83.07		16.93		1937-11-01	2016-01-05				
	REPORTED_DT	Date current case was reported to TB unit	33.94	.02	66.04		1984-10-15	2213-12-16				2213-12-16 (1 Invalid Obs. in total)

mhtb_case_1993apr



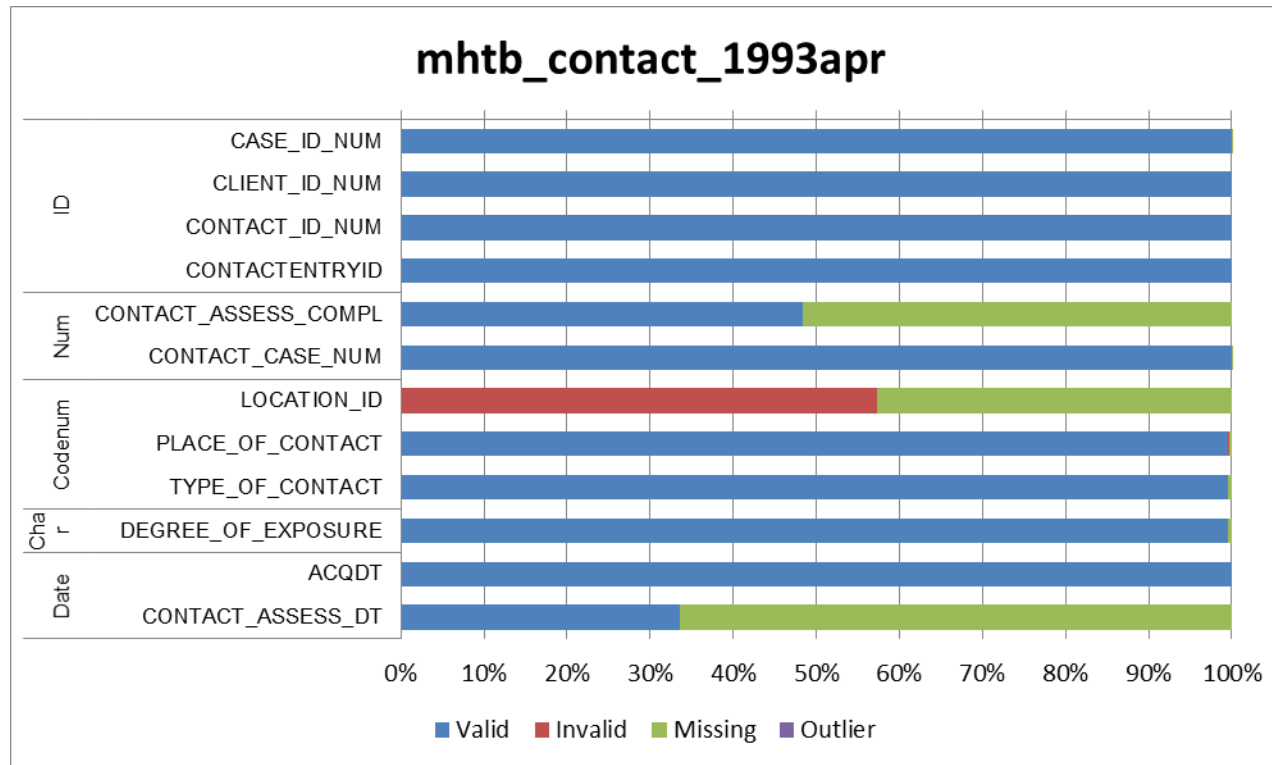
Clinic Visit Table

Dataset Label: MH Tuberculosis Registry - clinic visit table: 1993 onward				Records: 846	Legend (Potential Data Quality Problems) :							
Dataset Name: mhtb_clinic_visit_1993apr = No variance or 100% missing value				Period: 1993-2015	None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A				
Legend for comment column												
Blank = variables have not been tested (no formats have been specified for the variables)												
✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	CLINICVISITENTRYID	ClinicVisitEntryID	100.00		.00							
Num	CASE_ID_NUM	Specific case number. 1 = first case, 2 = second case etc.	100.00		.00		0.00	10.00	.43	.00	.70	
Top 10 Observed Values												
Codenum	CLINIC_ID_NUM	Specific clinic identifier where client was treated	100.00		.00		1, 3, 2, 4					✓
	REASON_FOR_APPT	Reason for appointment	99.05	.59	.35		6, 1, 2, 7, 4, 9, 10, 5, 14, 15					0 (5 Invalid Obs. in total)
Char	FIELD_REQUEST	Field request (yes/no)	50.83		49.17		N, Y					
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	APPOINTEMNT_DT	Time of appointment	30.02	9.22	60.76		1899-12-30	2000-10-05				78 invalid obs. out of [1900-01-01, 2016-03-11] range
	CLINIC_APPT_DT	Appointment date	100.00		.00		1990-08-19	2009-09-24				



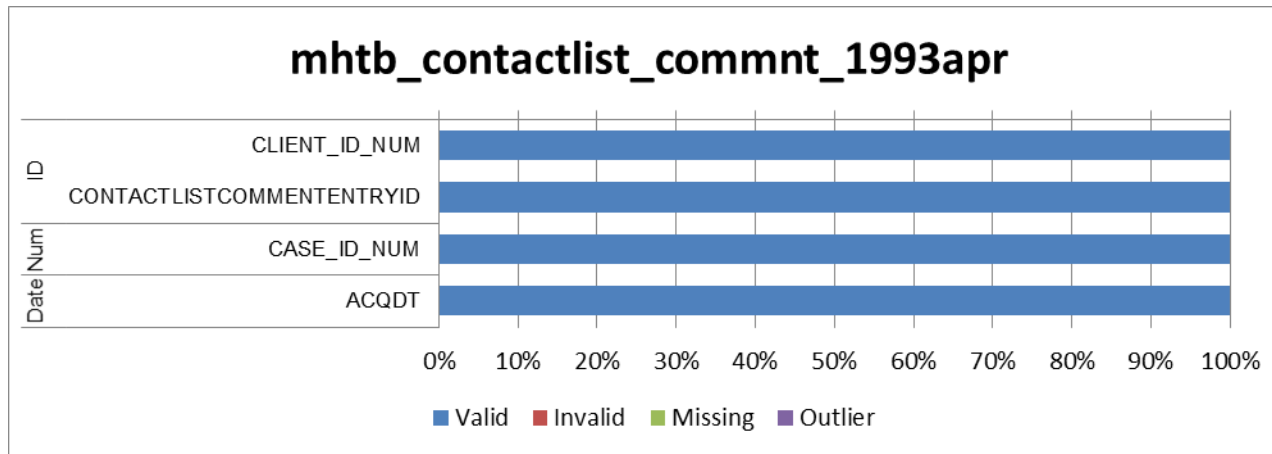
Contact Table

Dataset Label: MH Tuberculosis Registry - Contact table: 1993 onward		Records: 47880	Legend (Potential Data Quality Problems):									
Dataset Name: mhtb_contact_1993apr = No variance or 100% missing value		Period: 1993-2015	None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A						
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CASE_ID_NUM	This is the specific case number of the 'Case' (the client ID number)	100.00		.00							
	CLIENT_ID_NUM	Individual client ID number	100.00		.00							
	CONTACT_ID_NUM	Individual contact ID number	100.00		.00							
	CONTACTENTRYID	Automatically generated number for each unique contact entry	100.00		.00							
Num	CONTACT_ASSESS_COMPL	Whether contact assessment is complete and if further follow up/surveillance is required	48.36		51.64		0.00	3.00	2.00	3.00	1.37	
	CONTACT_CASE_NUM	This is the specific case number of the contact, if the contact also becomes a case	99.98		.02		0.00	5.00	.02	.00	.14	
Top 10 Observed Values												
Codenum	LOCATION_ID	Specific location where contact occurred	.00	57.25	42.74		0, 94					0 (27413 Invalid Obs. in total)
	PLACE_OF_CONTACT	Category of place of contact such as school, daycare, hospital etc.	99.52	.17	.31		1, 999, 44, 42, 3, 2, 4, 9, 43, 5					6 (83 Invalid Obs. in total)
	TYPE_OF_CONTACT	Type of contact	99.50	.01	.48		999, 15, 69, 6, 16, 14, 8, 7, 11, 10					37, 49, 50 (6 Invalid Obs. in total)
Char	DEGREE_OF_EXPOSURE	Degree of exposure to case	99.51		.49		4, 1, 2, 3					
	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
Date	CONTACT_ASSESS_DT	Date contact assessment done	33.52		66.48		2001-01-02	2016-01-07				



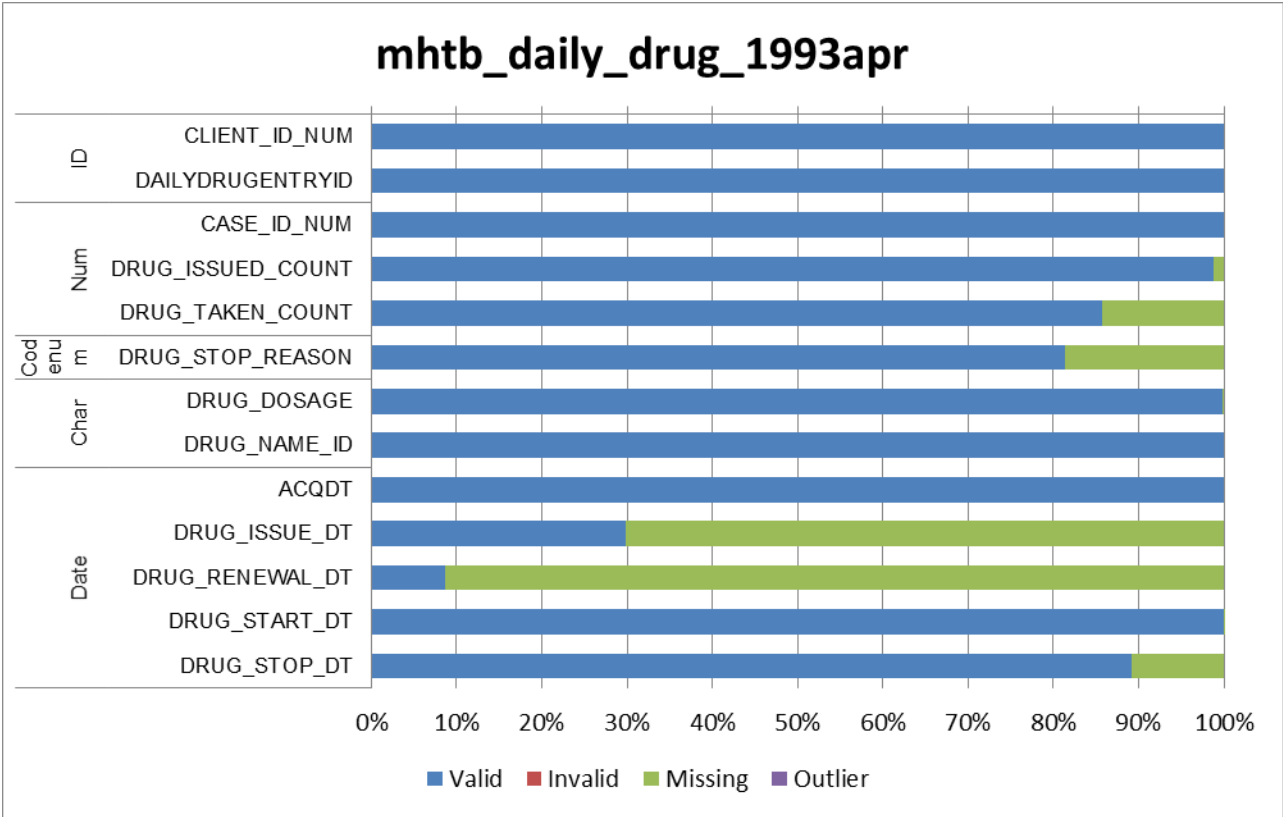
Contact List Comment Table

Dataset Label: MH Tuberculosis - Contact List comment: 1993 onward				Records: 427		Legend (Potential Data Quality Problems) :						
Dataset Name: mhtb_contactlist_commnt_1993apr = No variance or 100% missing value				Period: 1993-2015		None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
						Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)						
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Individual client ID number	100.00		.00							
	CONTACTLISTCOMMENTENTRYID	ContactListCommentEntryID	100.00		.00							
Num	CASE_ID_NUM	This is the specific case number of the "case" (the client ID number)	100.00		.00		0.00	10.00	1.04	1.00	.56	
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				



Daily Drug Table

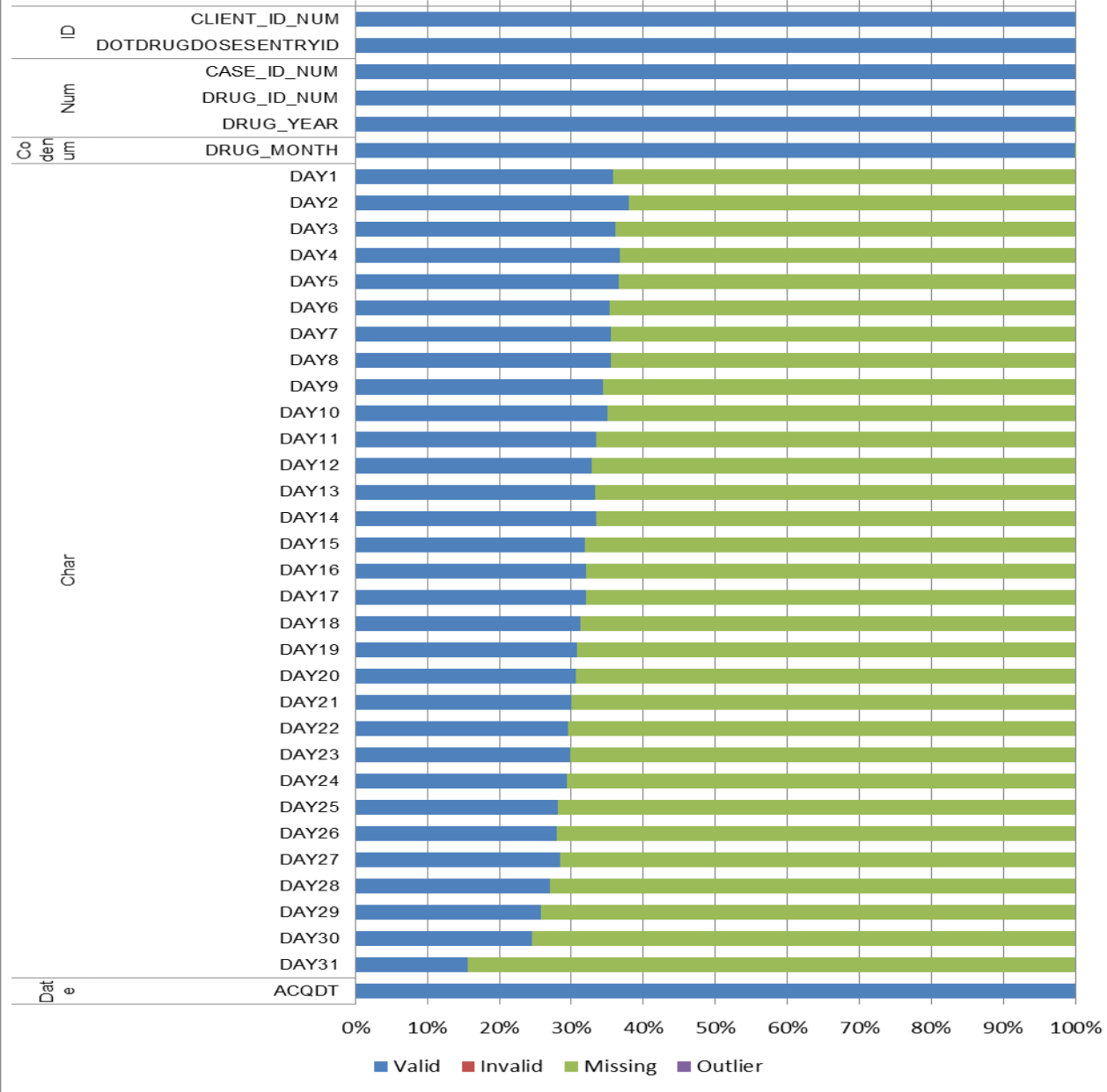
Dataset Label: MH Tuberculosis Registry - daily drug table: 1993 onward			Records: 8468				Legend (Potential Data Quality Problems) :					
Dataset Name: mhtb_daily_drug_1993apr = No variance or 100% missing value			Period: 1993-2015				None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A		
Legend for comment column												
Blank = variables have not been tested (no formats have been specified for the variables)												
✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	DAILYDRUGENTRYID	DailyDrugEntryID	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	3.00	.16	.00	.39	
	DRUG_ISSUED_COUNT	Number of pills issued	98.83		1.17		1.00	2970.00	196.57	197.46	104.05	
	DRUG_TAKEN_COUNT	Number of pills taken	85.69		14.31		0.00	2880.00	177.49	188.19	112.13	
Top 10 Observed Values												
Char	DRUG_STOP_REASON	Reason drug stopped	81.33		18.67		1, 4, 64, 53, 41, 82, 8, 6, 54, 172					✓
	DRUG_DOSAGE	Drug dosage	99.92		.08		300 mg, 25 mg, 600 mg, 50 mg, 1200 mg, 1500 mg, 500 mg, 800 mg					
Date	DRUG_NAME_ID	Drug name	100.00		.00		INH, B6, RMP, EMB, PZA, CM, CIP, RIF, AZT, BIX					✓
	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	DRUG_ISSUE_DT	Date drug issued	29.78		70.22		1985-03-01	2016-01-04				
	DRUG_RENEWAL_DT	Date drug renewed	8.63		91.37		1989-09-05	2016-03-01				
	DRUG_START_DT	Date drug started	99.93		.06		1954-10-03	2213-01-31				2213-01-31 (1 Invalid Obs. in total)
	DRUG_STOP_DT	Date drug ended	89.14	.01	10.85		1955-08-12	2215-05-15				2215-05-15 (1 Invalid Obs. in total)



Direct Observed Therapy Drug Doses Table

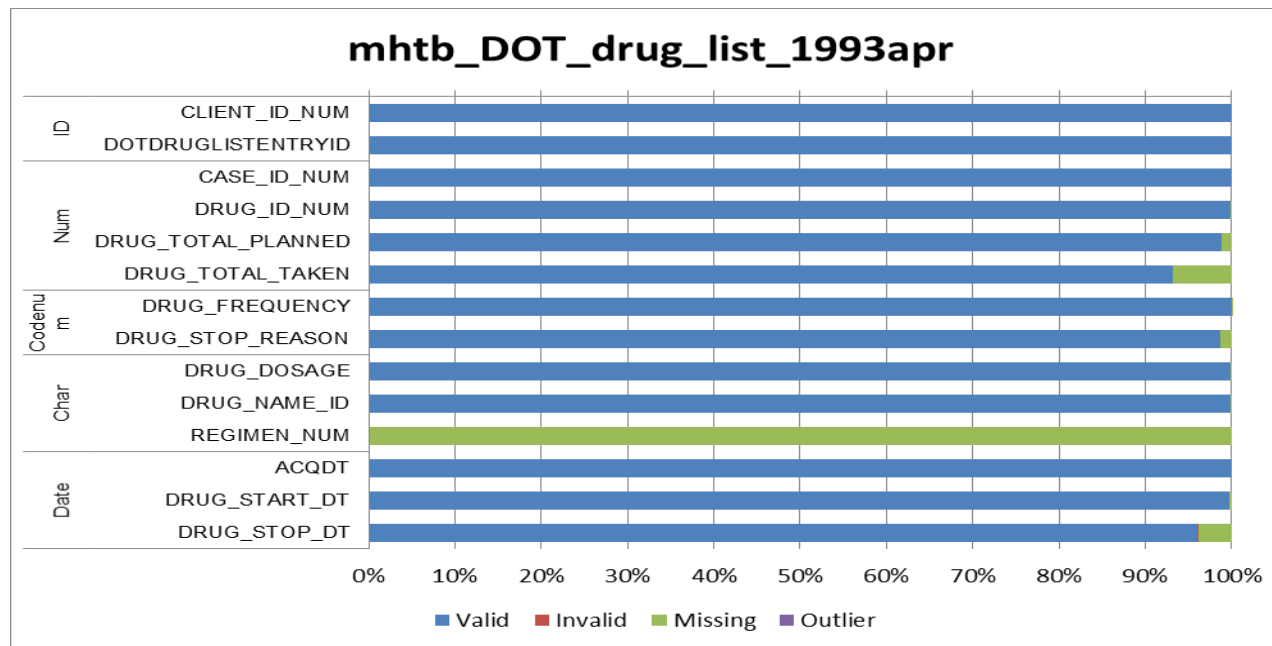
Dataset Label: MH Tuberculosis Registry - Direct Observed Therapy Drug Doses: 1				Records: 49047		Legend (Potential Data Quality Problems) :						
Dataset Name: mhdtb_DOT_drug_doses_1993apr				Period: 1993-2015		None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
= No variance or 100% missing value				Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)								
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	DOTDRUGDOSEENTRYID	DOTDrugDosesEntryID	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	7.00	.42	.33	.56	
	DRUG_ID_NUM	ID number of drug, link to drug name and dosage	100.00		.00		0.00	11.00	1.02	1.00	.18	
	DRUG_YEAR	Year in which drug dispensed	99.99		.01		1992.00	2016.00	2006.84	2007.33	4.11	
Top 10 Observed Values												
Codenum	DRUG_MONTH	Month in which drug dispensed	99.99	.00	.01		9, 10, 11, 3, 8, 7, 1, 2, 5, 4				2009 (1 Invalid Obs. in total)	
Char	DAY1	Day as to where drug was dispensed	35.76		64.24		-, NS, HC, XX, PH, X, RC, H6, P, S					
	DAY2	Day as to where drug was dispensed	38.01		61.99		NS, -, HC, XX, PH, X, RC, H6, P, S					
	DAY3	Day as to where drug was dispensed	36.09		63.91		NS, -, HC, XX, PH, X, RC, H6, S, P					
	DAY4	Day as to where drug was dispensed	36.77		63.23		NS, -, HC, XX, PH, X, RC, S, H6, P					
	DAY5	Day as to where drug was dispensed	36.53		63.47		NS, -, HC, XX, PH, X, RC, H6, P, S					
	DAY6	Day as to where drug was dispensed	35.28		64.72		NS, -, HC, XX, PH, X, RC, H6, S, DU					
	DAY7	Day as to where drug was dispensed	35.54		64.46		NS, -, HC, XX, PH, X, RC, H6, S, P					
	DAY8	Day as to where drug was dispensed	35.47		64.53		NS, -, HC, XX, PH, X, RC, H6, S, P					
	DAY9	Day as to where drug was dispensed	34.36		65.64		NS, -, HC, XX, PH, X, RC, H6, S, P					
	DAY10	Day as to where drug was dispensed	34.95		65.05		NS, -, HC, XX, PH, X, RC, H6, S, P					
	DAY11	Day as to where drug was dispensed	33.52		66.48		NS, -, HC, XX, PH, X, RC, H6, S, P					
	DAY12	Day as to where drug was dispensed	32.80		67.20		NS, -, HC, XX, PH, X, RC, H6, S, P					
	DAY13	Day as to where drug was dispensed	33.32		66.68		NS, -, HC, XX, PH, X, RC, H6, S, P					
	DAY14	Day as to where drug was dispensed	33.41		66.59		NS, -, HC, XX, PH, X, RC, H6, P, S					
	DAY15	Day as to where drug was dispensed	31.95		68.05		NS, -, HC, XX, PH, X, RC, S, H6, P					
	DAY16	Day as to where drug was dispensed	32.09		67.91		NS, -, HC, XX, PH, X, RC, H6, S, DU					
	DAY17	Day as to where drug was dispensed	32.10		67.90		NS, -, HC, XX, PH, X, RC, H6, S, P					
	DAY18	Day as to where drug was dispensed	31.28		68.72		NS, HC, -, XX, PH, X, RC, S, P, H6					
	DAY19	Day as to where drug was dispensed	30.84		69.16		NS, HC, -, XX, PH, X, RC, H6, S, P					
	DAY20	Day as to where drug was dispensed	30.65		69.35		NS, HC, -, XX, PH, X, RC, P, S, H6					
	DAY21	Day as to where drug was dispensed	30.00		70.00		NS, HC, -, XX, PH, X, RC, P, H6, S					
	DAY22	Day as to where drug was dispensed	29.60		70.40		NS, HC, -, XX, PH, X, RC, P, H6, S					
	DAY23	Day as to where drug was dispensed	29.87		70.13		NS, HC, -, XX, PH, X, RC, H6, P, S					
	DAY24	Day as to where drug was dispensed	29.39		70.61		NS, HC, -, XX, PH, X, RC, P, S, H6					
	DAY25	Day as to where drug was dispensed	28.08		71.92		NS, HC, -, XX, PH, X, RC, P, H6, S					
	DAY26	Day as to where drug was dispensed	27.93		72.07		NS, HC, XX, -, PH, X, RC, P, H6, S					
	DAY27	Day as to where drug was dispensed	28.47		71.53		NS, HC, XX, -, PH, X, RC, P, H6, S					
	DAY28	Day as to where drug was dispensed	27.03		72.97		NS, HC, XX, PH, -, X, RC, S, P, H6					
	DAY29	Day as to where drug was dispensed	25.74		74.26		NS, HC, XX, PH, X, -, RC, H6, P, S					
	DAY30	Day as to where drug was dispensed	24.59		75.41		NS, HC, XX, PH, X, RC, -, P, H6, S					
	DAY31	Day as to where drug was dispensed	15.62		84.38		NS, HC, XX, PH, X, RC, -, H6, P, S					
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				

mhtb_DOT_drug_doses_1993apr



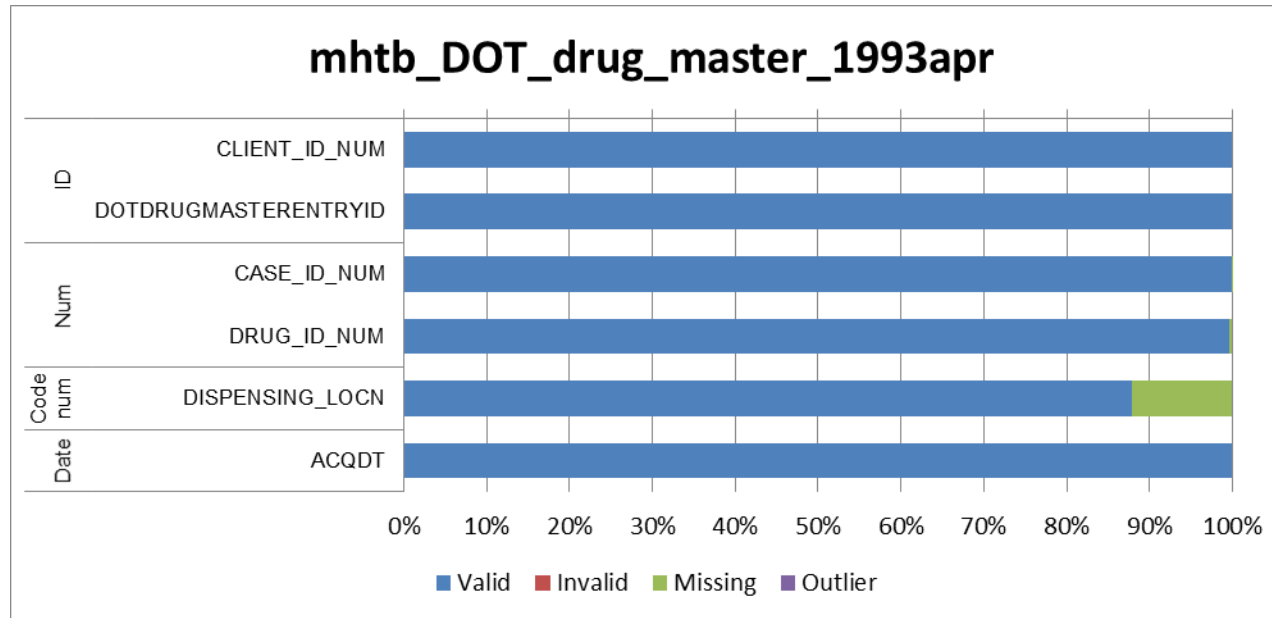
Direct Observed Therapy Drug List Table

Dataset Label: MH Tuberculosis Registry - Direct Observed Therapy Drug List: 1993 onward				Records: 30498				Legend (Potential Data Quality Problems):				
Dataset Name: mhtb_DOT_drug_list_1993apr				Period: 1993-2015				None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A	
= No variance or 100% missing value												
Legend for comment column												
Blank = variables have not been tested (no formats have been specified for the variables)												
✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	DOTDRUGLISTENTRYID	Auto generated identification number	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	61.00	.74	1.00	.63	
	DRUG_ID_NUM	ID number of drug	99.93		.07		0.00	21.00	1.59	1.00	.94	
	DRUG_TOTAL_PLANNED	Number of doses planned	98.85		1.15		0.00	800.00	37.07	25.78	34.72	
	DRUG_TOTAL_TAKEN	Number of doses taken	93.16		6.84		0.00	730.00	31.23	17.72	32.43	
Top 10 Observed Values												
Codenum	DRUG_FREQUENCY	Frequency by which drug is to be taken (once per week etc.)	99.95		.05		2, 5, 3, 20, 1, 9, 7, 11, 6, 10					✓
	DRUG_STOP_REASON	Reason drug stop	98.61		1.39		1, 8, 172, 76, 53, 70, 55, 82, 7, 64					✓
Char	DRUG_DOSAGE	Drug dosage in mgs	99.94		.06		900 mg, 600 mg, 25 mg, 50 mg, 300 mg, 2000 mg, 1500 mg, 1					
	DRUG_NAME_ID	Short form of drug name	99.93		.07		INH, RMP, B6, PZA, EMB, MOX, SM, RIF, Lev, RPT					✓
	REGIMEN_NUM	Regimen number	.00		100.00							
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	DRUG_START_DT	Date drug started	99.67	.03	.30		1985-09-10	2215-09-08				8 invalid obs. out of [1900-01-01, 2016-03-11] range
	DRUG_STOP_DT	Date drug ended	96.08	.04	3.88		1986-03-06	2215-10-09				12 invalid obs. out of [1900-01-01, 2016-03-11] range



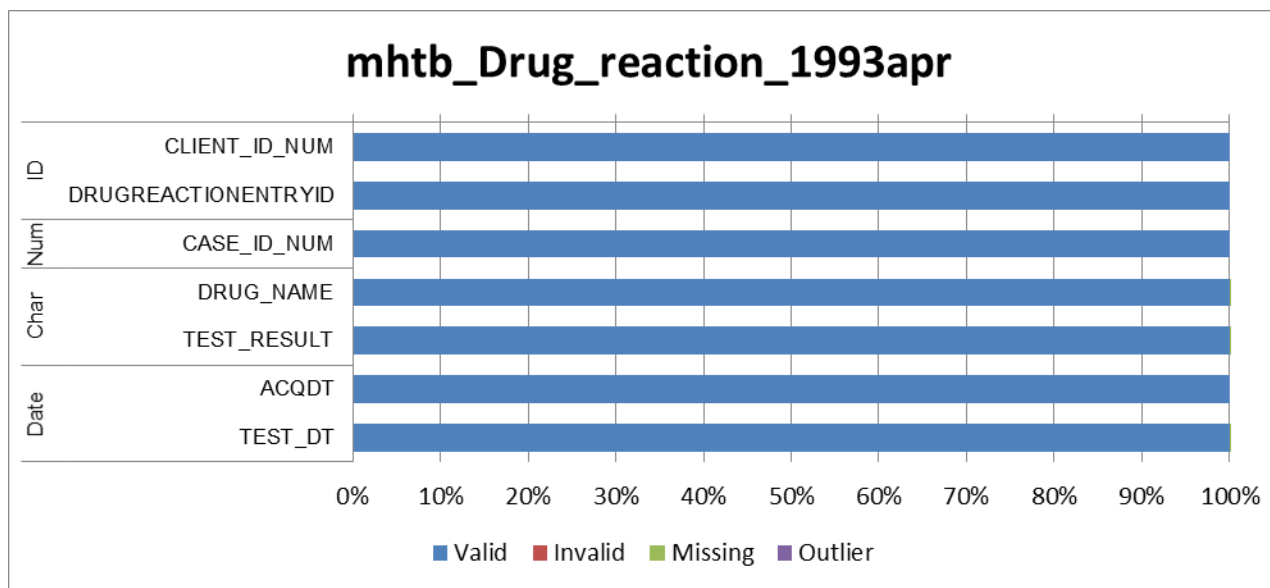
Direct Observed Drug Therapy Master Table

Dataset Label: MH Tuberculosis Registry - Direct Observed Therapy Drug Master: Records: 9003				Legend (Potential Data Quality Problems) :								
Dataset Name: mhtb_DOT_drug_master_1993apr = No variance or 100% missing value				Period: 1993-2015								
				None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A					
				Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)								
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	DOTDRUGMASTERENTRYID	Auto generated identification number	100.00		.00							
Num	CASE_ID_NUM	Specific case number	99.99		.01		0.00	4.00	.41	.34	.55	
	DRUG_ID_NUM	ID number of drug	99.62		.38		1.00	9.00	1.17	1.00	.53	
Top 10 Observed Values												
Code num	DISPENSING_LOCN	Where drug was dispensed	87.83	.01	12.16		56, 9, 8, 4, 51, 3, 53, 13, 15, 18					38 (1 Invalid Obs. in total)
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				



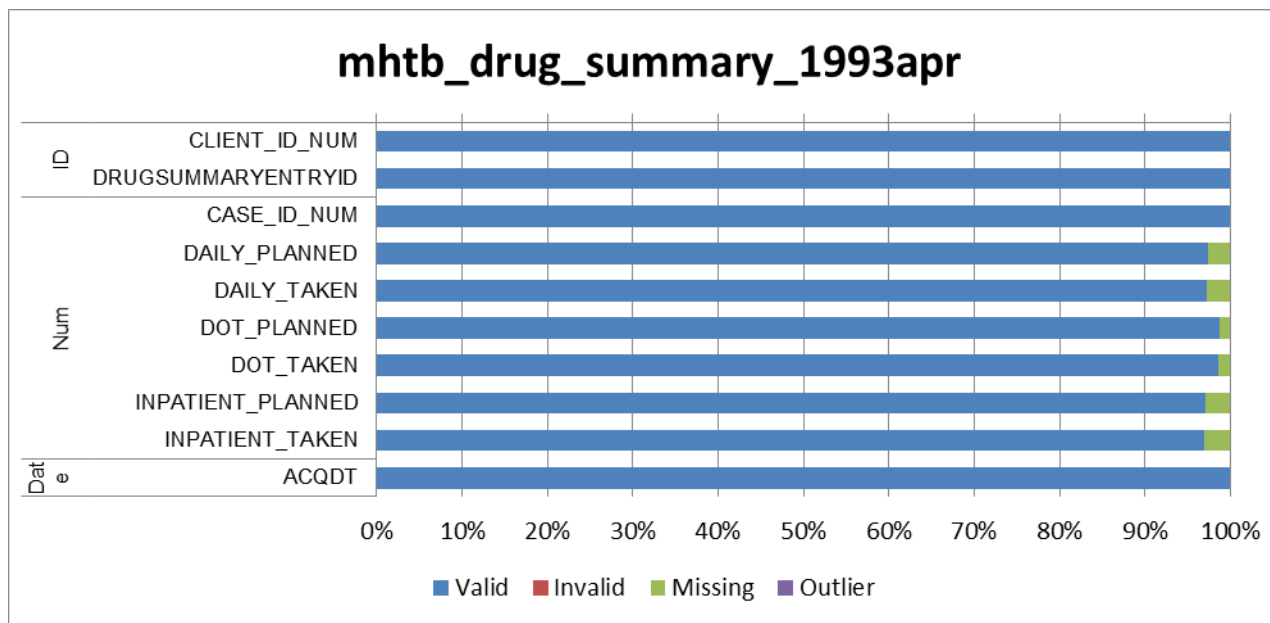
Drug Reaction Table

Dataset Label: MH Tuberculosis Registry - Drug Reaction: 1993 onward		Records: 13009		Legend (Potential Data Quality Problems) :								
Dataset Name: mhtb_Drug_reaction_1993apr = No variance or 100% missing value		Period: 1993-2015		None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A					
Legend for comment column												
Blank = variables have not been tested (no formats have been specified for the variables)												
✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	DRUGREACTIONENTRYID	Auto generated identification number	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	10.00	1.06	1.00	.33	
Top 10 Observed Values												
Char	DRUG_NAME	Name of drug to which reaction was tested	99.98	.01	.02		EMB, INH, RMP, PZA, SM, OFL, MOX, LIN, RIF, AMK					EM (1 Invalid Obs. in total)
	TEST_RESULT	Outcome of test (S - sensitive, R - resistant, U - Unknown)	99.98		.02		S, R, N					✓
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	TEST_DT	Date of drug reaction test	99.95	.03	.02		1881-08-27	2215-06-15				1881-08-27, 1882-02-10, 2044-08-18, 2215-06-15 (4 Invalid Obs. in total)



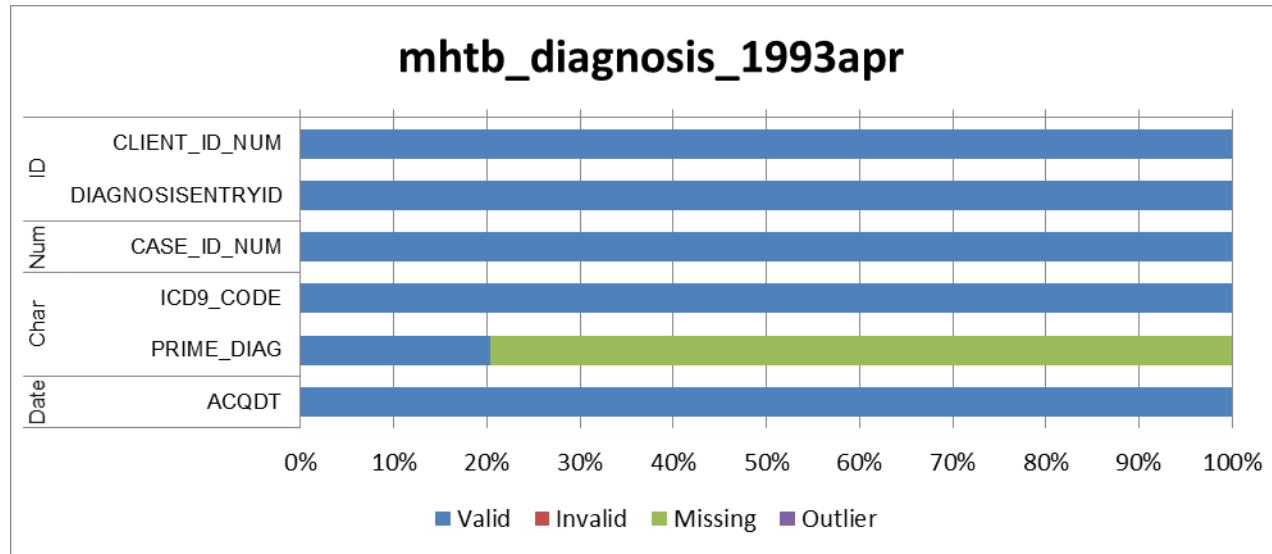
Drug Summary Table

Dataset Label: MH Tuberculosis Registry - Drug Summary: 1993 onward			Records: 8364			Legend (Potential Data Quality Problems) :						
Dataset Name: mhtb_drug_summary_1993apr = No variance or 100% missing value			Period: 1993-2015			None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	DRUGSUMMARYENTRYID	Auto generated identification number	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	9.00	.30	.00	.51	
	DAILY_PLANNED	Number of doses planned for daily regimen	97.36		2.64		0.00	852.00	60.00	.03	102.34	
	DAILY_TAKEN	Actual number of daily regimen doses taken	97.20		2.80		0.00	852.00	57.58	.02	99.82	
	DOT_PLANNED	Number of doses planned for Direct Observed Therapy	98.76		1.24		0.00	730.00	58.29	70.65	47.96	
	DOT_TAKEN	Actual number of doses taken on Direct Observed Therapy	98.65		1.35		0.00	730.00	52.17	59.64	47.75	
	INPATIENT_PLANNED	number of doses planned for inpatient	97.06		2.94		0.00	393.00	1.47	.00	9.81	
	INPATIENT_TAKEN	Actual number of doses taken as inpatient	96.96		3.04		0.00	393.00	1.47	.00	9.79	
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				



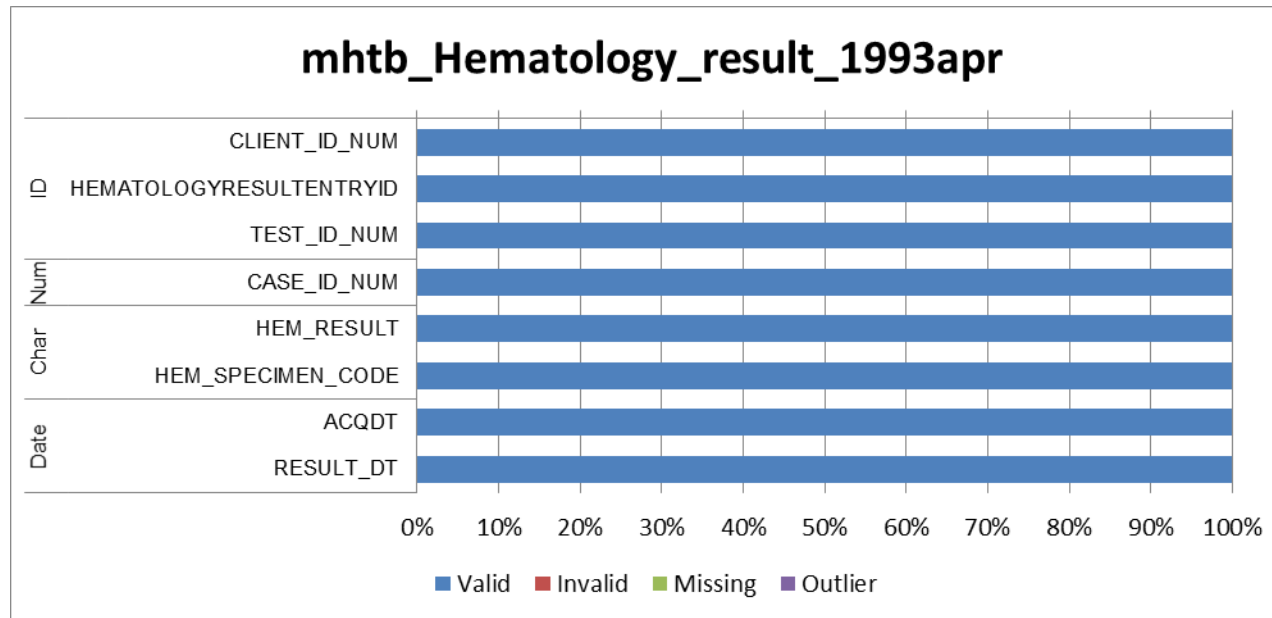
Diagnosis Table

Dataset Label: MH Tuberculosis Registry - diagnosis table: 1993 onward		Records: 19990	Legend (Potential Data Quality Problems) :									
Dataset Name: mhtb_diagnosis_1993apr = No variance or 100% missing value		Period: 1993-2015	None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A						
Legend for comment column												
Blank = variables have not been tested (no formats have been specified for the variables)												
√ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	DIAGNOSISENTRYID	DiagnosisEntryID	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	4.00	.24	.00	.47	
Top 10 Observed Values												
Char	ICD9_CODE	Diagnosis ICD9 code	100.00		.00		mh05, mh06, mh01, 011.0, 011, mh02.0, 012.0, 017.2, 011.2,					√
	PRIME_DIAG	Whether this diagnosis is the primary diagnosis (Y/N)	20.34		79.66		Y, y, 1					
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				



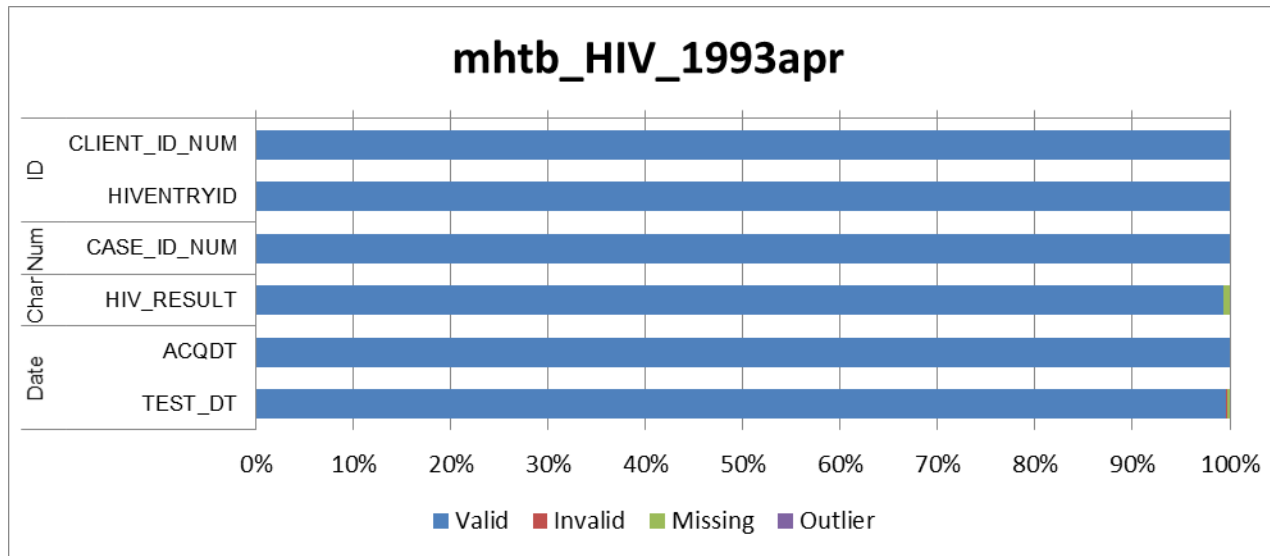
Hematology Result Table

Dataset Label: MH Tuberculosis Registry - Hematology Result: 1993 onward			Records: 36			Legend (Potential Data Quality Problems):						
Dataset Name: mhtb_Hematology_result_1993apr = No variance or 100% missing value			Period: 1993-2015			None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
Legend for comment column												
Blank = variables have not been tested (no formats have been specified for the variables)												
✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	HEMATOLOGYRESULTENTRYID	Auto generated identification number	100.00		.00							
	TEST_ID_NUM	identification number of test. Links to test_number in hematology test table	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	1.00	.53	.81	.51	
Top 10 Observed Values												
Char	HEM_RESULT	Test result	100.00		.00		21, 24, 25, 34, 40, 5, 112, 12, 121, 141					
	HEM_SPECIMEN_CODE	Type of test for specimen	100.00		.00		14, 15, 9, 16, 17, 11, 12					✓
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	RESULT_DT	Test result date	100.00		.00		1996-03-23	1998-01-05				



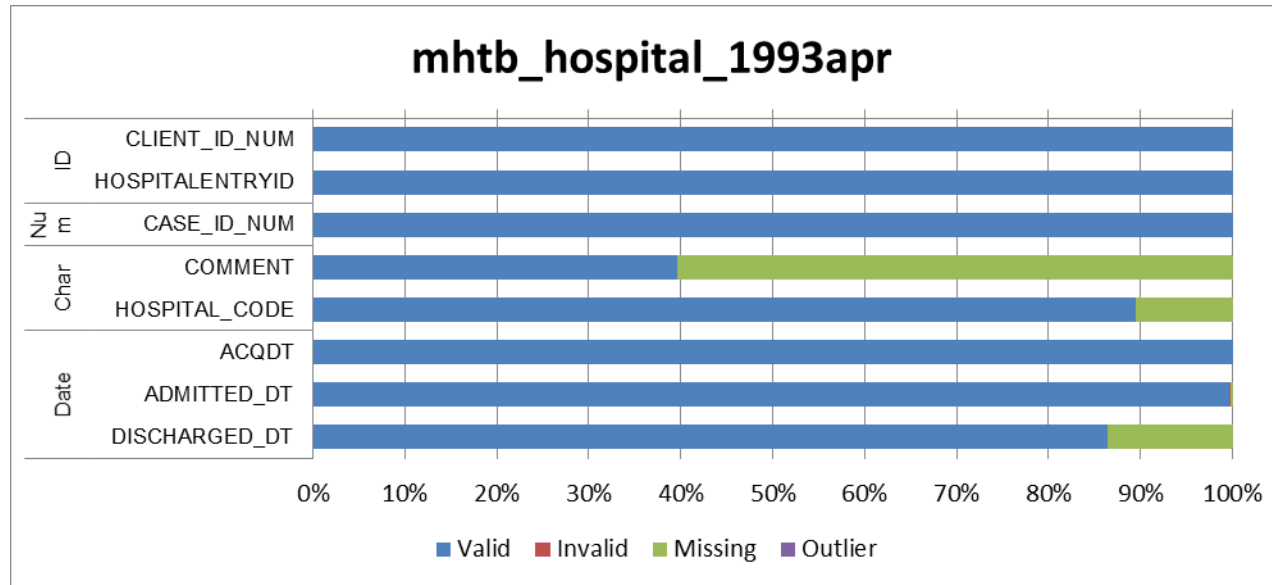
HIV Test Table

Dataset Label: MH Tuberculosis Registry - HIV: 1993 onward		Records: 2894	Legend (Potential Data Quality Problems) :									
Dataset Name: mhtb_HIV_1993apr = No variance or 100% missing value		Period: 1993-2015	None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A						
Legend for comment column												
Blank = variables have not been tested (no formats have been specified for the variables)												
✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	HIVENTRYID	Auto generated identification number	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	10.00	.65	.98	.62	
Top 10 Observed Values												
Date	HIV_RESULT	HIV test result	99.34		.66		N, D, U, P					✓
	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	TEST_DT	Date of HIV test	99.62	.07	.31		1900-01-01	5200-08-24				2041-09-15, 5200-08-24 (2 Invalid Obs. in total)



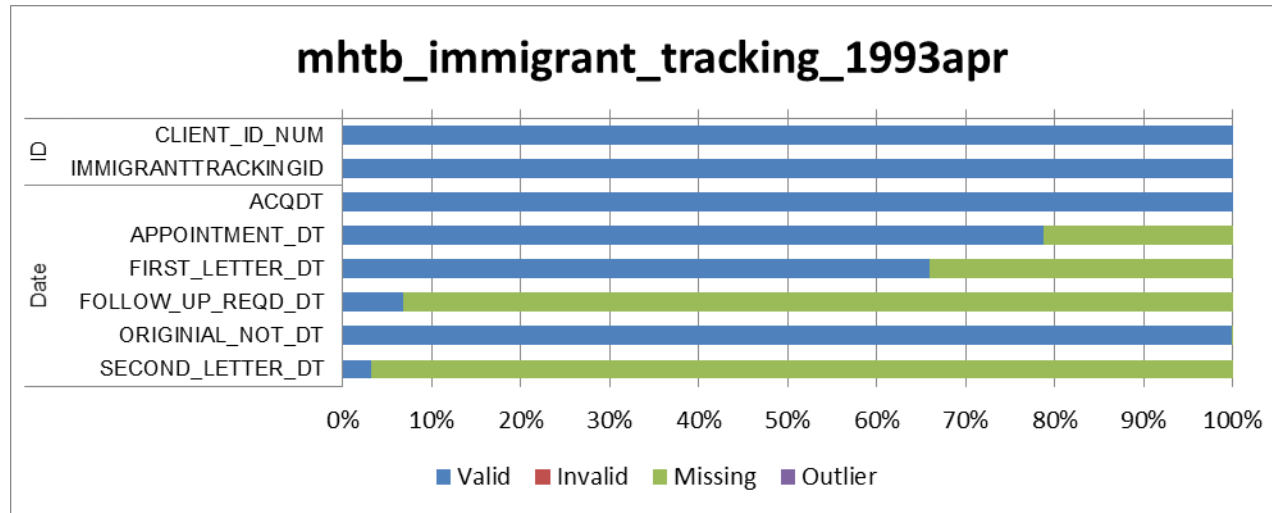
Hospital Table

Dataset Label: MH Tuberculosis Registry - Hospital: 1993 onward		Records: 2428	Legend (Potential Data Quality Problems) :									
Dataset Name: mhtb_hospital_1993apr = No variance or 100% missing value		Period: 1993-2015	None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A						
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	HOSPITALENTRYID	Auto generated identification number	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	10.00	1.02	1.00	.45	
Top 10 Observed Values												
Char	COMMENT	case management comments	39.66		60.34		H-6, Admitted to H6, H-4, CH-4, Deceased, H-7, A-4, D-4, H6					
	HOSPITAL_CODE	Name of hospital	89.50		10.50		1, 12, 3, 2, 13, 5, 4, 7, 8, 10					
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	ADMITTED_DT	Date client admitted to hospital	99.67	.04	.29		1866-10-07	2015-12-02				1866-10-07 (1 Invalid Obs. in total)
	DISCHARGED_DT	Date client discharged from hospital	86.37	.04	13.59		1951-07-03	2200-09-22				2200-09-22 (1 Invalid Obs. in total)



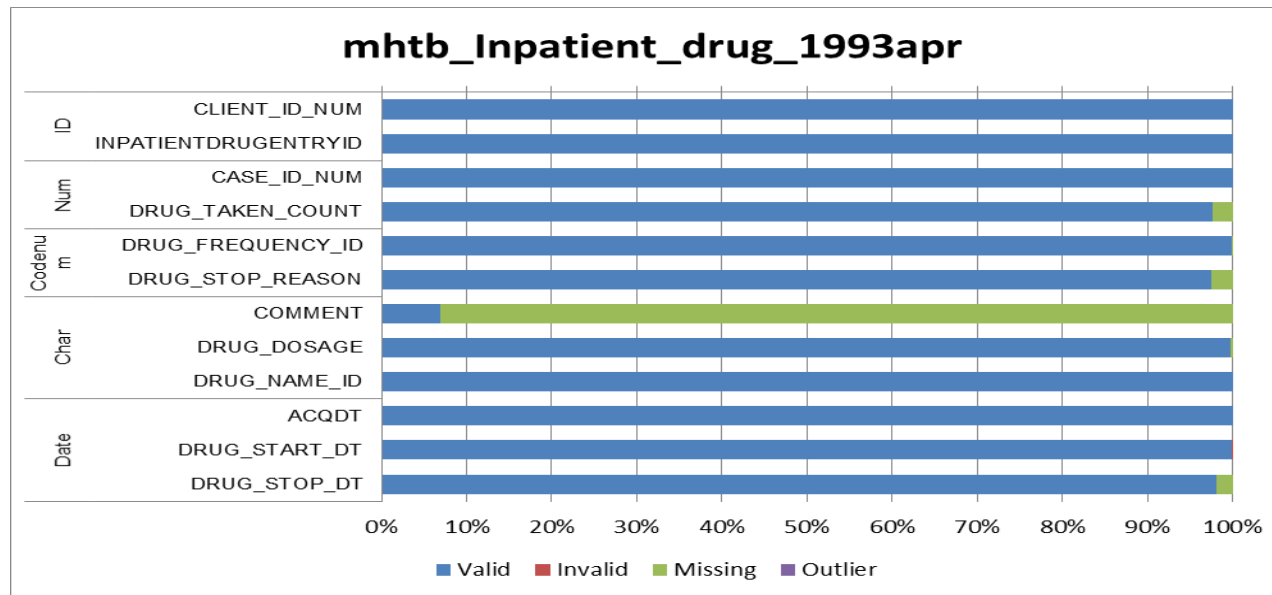
Immigrant Tracking Table

Dataset Label: MH Tuberculosis Registry - Immigrant Tracking: 1993 onward			Records: 4759			Legend (Potential Data Quality Problems) :						
Dataset Name: mhtb_immigrant_tracking_1993apr = No variance or 100% missing value			Period: 1993-2015			None or Minimal < 5%		Moderate 5-30%	Significant > 30%	Unknown or N/A		
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	IMMIGRANTTRACKINGID	ImmigrantTracking ID	100.00		.00							
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	APPOINTMENT_DT	Appointment date	78.69		21.31		1992-09-02	2015-12-02				
	FIRST_LETTER_DT	Date first letter sent	66.00		34.00		1992-07-15	2016-01-07				
	FOLLOW_UP_REQD_DT	Date follow up required	6.83		93.17		2001-08-27	2016-12-01				
	ORIGINAL_NOT_DT	Date unit notified of immigrant	99.89		.11		1992-07-14	2016-01-06				
	SECOND_LETTER_DT	Second letter date (if needed)	3.21	.02	96.76		2004-02-24	2205-12-21				2205-12-21 (1 Invalid Obs. in total)



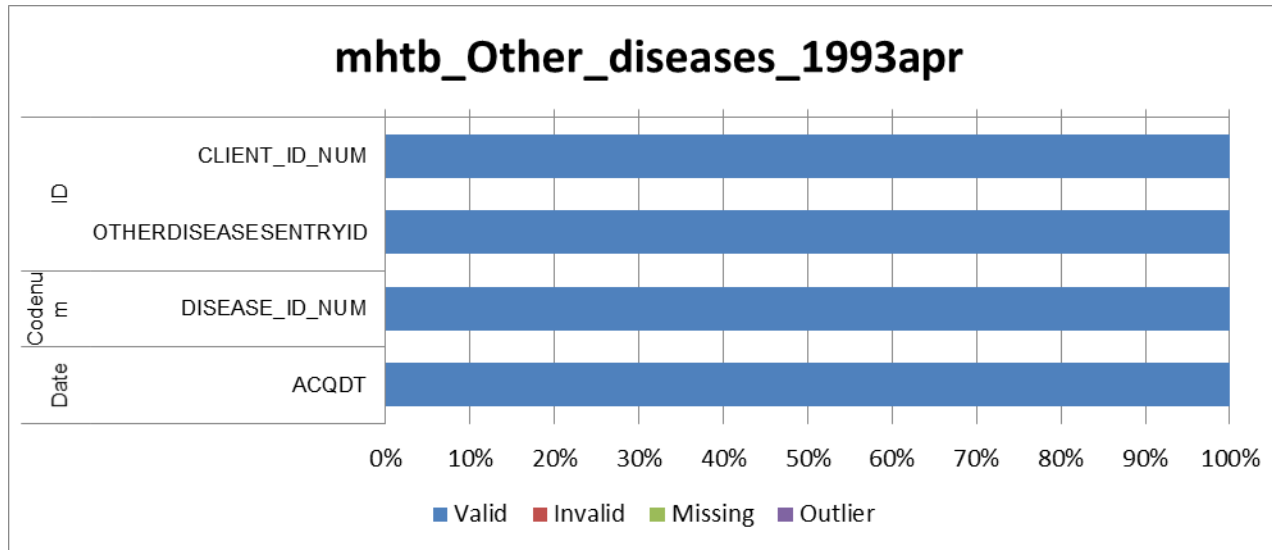
Inpatient Drug Table

Dataset Label: MH Tuberculosis Registry - Inpatient Drug: 1993 onward		Records: 12816	Legend (Potential Data Quality Problems) :									
Dataset Name: mhtb_inpatient_drug_1993apr = No variance or 100% missing value		Period: 1993-2015	None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A						
Legend for comment column												
Blank = variables have not been tested (no formats have been specified for the variables)												
✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	INPATIENTDRUGENTRYID	Auto generated identification number	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	14.00	1.02	1.00	.43	
	DRUG_TAKEN_COUNT	Count of drugs taken	97.71		2.29		0.00	715.00	20.00	14.00	25.12	
Top 10 Observed Values												
Codenum	DRUG_FREQUENCY_ID	Frequency by which drug is to be taken (once per week etc.)	99.94		.06		1, 2, 3, 5, 7, 11, 6, 20, 14, 4					✓
	DRUG_STOP_REASON	Reasons of stopping taking medication	97.45		2.55		11, 8, 64, 2, 41, 76, 51, 7, 71, 1					✓
Char	COMMENT	Case management comments	6.87		93.13		TGH, started in TGH, started at St. Bon Hosp, Victoria Hosp, 2					
	DRUG_DOSAGE	Amount of drug dosage	99.81		.19		600 mg, 300 mg, 1500 mg, 1200 mg, 25 mg, 800 mg, 1000 mg					
	DRUG_NAME_ID	Drug name	100.00		.00		RMP, INH, PZA, EMB, B6, SM, MOX, RIF, AMK, Lev					✓
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	DRUG_START_DT	Start date of drugs	99.98	.02	.00		1952-01-22	2215-06-26				2044-07-16, 2077-06-25, 2215-06-26 (3 Invalid Obs. in total)
	DRUG_STOP_DT	Stop date of drugs	98.10	.02	1.88		1953-01-01	5201-05-10				2215-05-01, 5201-05-10 (2 Invalid Obs. in total)



Other Diseases Table

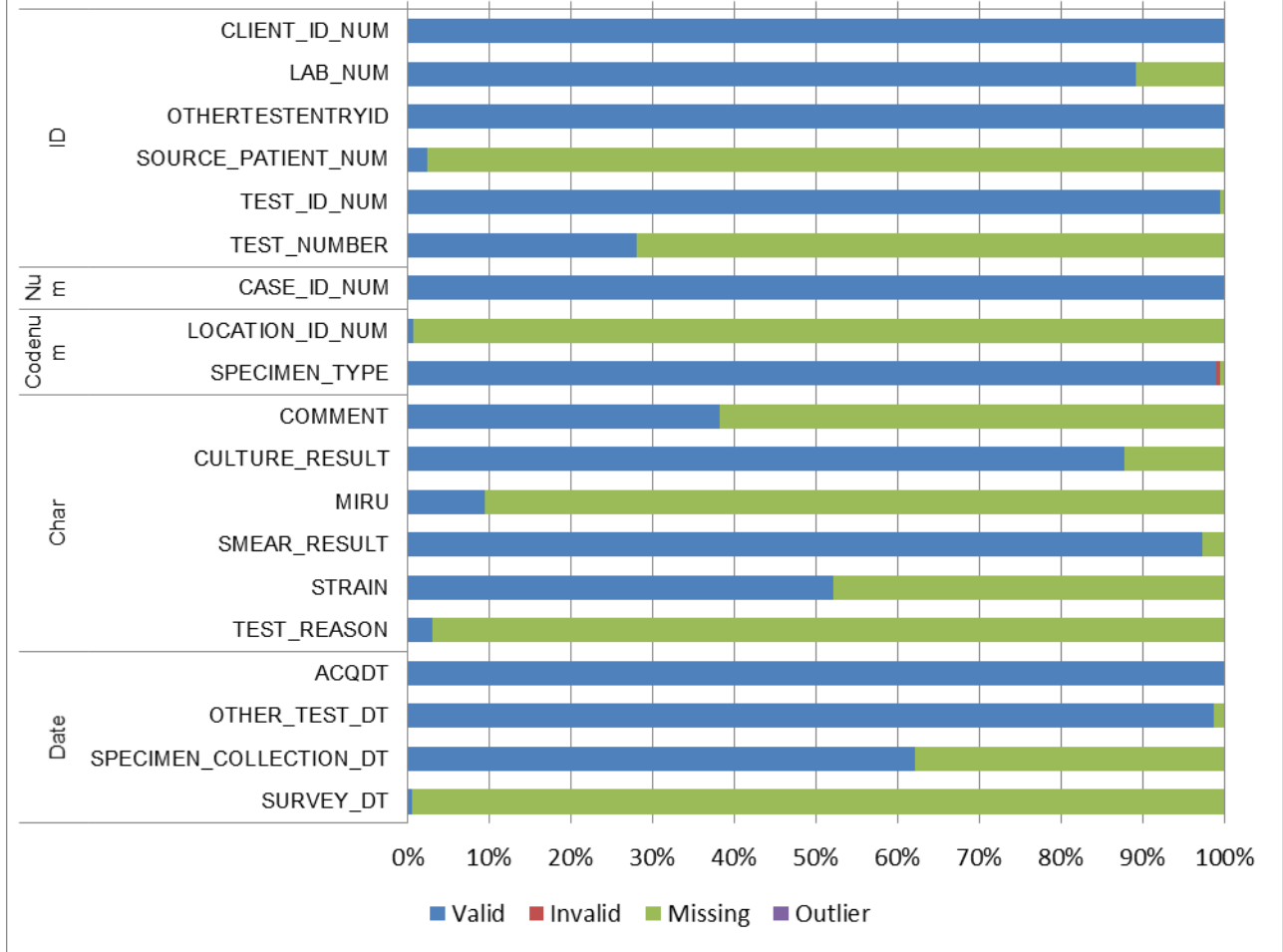
Dataset Label: MH Tuberculosis Registry - Other Diseases: 1993 onward		Records: 941	Legend (Potential Data Quality Problems) :									
Dataset Name: mhtb_Other_diseases_1993apr = No variance or 100% missing value		Period: 1993-2015	None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A						
Legend for comment column												
Blank = variables have not been tested (no formats have been specified for the variables)												
✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	OTHERDISEASESENTRYID	Auto generated identification number	100.00		.00							
Top 10 Observed Values												
Codenum	DISEASE_ID_NUM	Any other diseases client has been diagnosed with by specific disease	100.00		.00		2, 9, 19, 5, 7, 18, 4, 31, 1, 3					✓
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				



Other Test Table

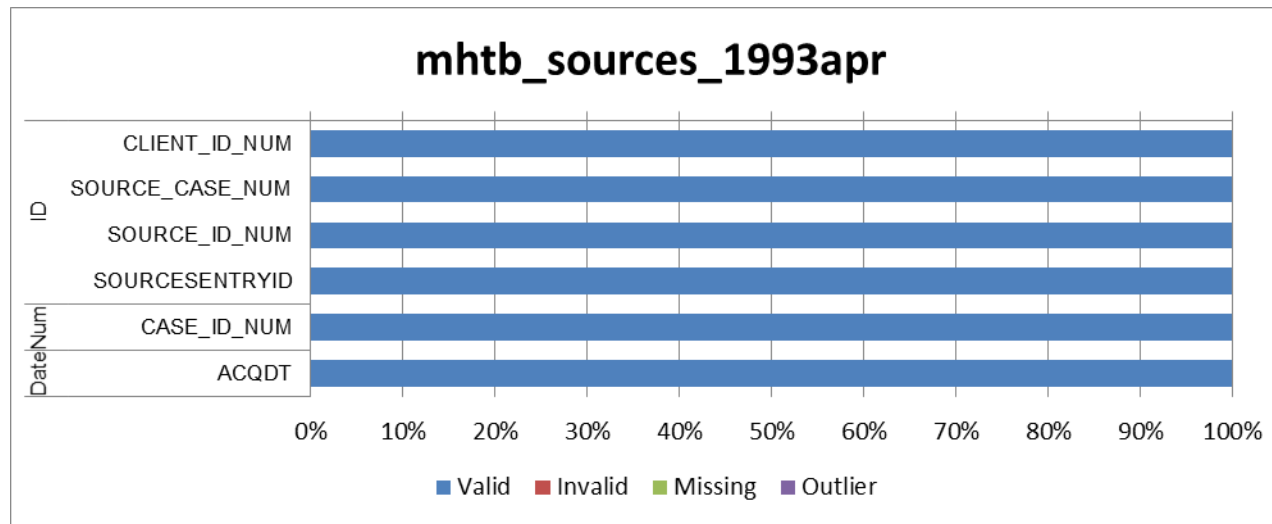
Dataset Label: MH Tuberculosis Registry - Other Test: 1993 onward			Records: 14879				Legend (Potential Data Quality Problems) :					
Dataset Name: mhtb_Other_test_1993apr = No variance or 100% missing value			Period: 1993-2015				None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A		
Legend for comment column												
Blank = variables have not been tested (no formats have been specified for the variables)												
✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	LAB_NUM	Unique number of lab test sent to	89.14		10.86							
	OTHERTESTENTRYID	Auto generated identification number	100.00		.00							
	SOURCE_PATIENT_NUM	Surveillance number	2.38		97.62							
	TEST_ID_NUM	Unique ID number for test	99.51		.49							
	TEST_NUMBER	Unique number of test	28.12		71.88							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	10.00	.88	1.00	.52	
Top 10 Observed Values												
Codenum	LOCATION_ID_NUM	Location of client	.65	.03	99.33		87, 86, 161, 399, 61, 81, 88, 89, 60, 92				60, 387 (4 Invalid Obs. in total)	
	SPECIMEN_TYPE	Type of specimen	99.13	.42	.45		1, 2, 4, 3, 5, 9, 6, 29, 23, 43				0 (62 Invalid Obs. in total)	
Char	COMMENT	Case management comments	38.29		61.71		WRHA, SBGH, necrotizing granuloma, caseating granuloma,					
	CULTURE_RESULT	Result of culture: N - negative, P - positive	87.77		12.23		P, N, U					
	MIRU	MIRU	9.41		90.59		223125153322, 227625173433, 254326223432, 22431514332					
	SMEAR_RESULT	Result of smear: N - negative, P - positive	97.29		2.71		N, P, U					
	STRAIN	Strain	52.18		47.82		M tuberculosis, M avium, M. Tuberculosis, M gordonae, M tub					
	TEST_REASON	Reason for test - contact or TB survey	3.11		96.89		C, S				✓	
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				2044-09-11, 2215-06-15, 2215-10-15 (3 Invalid Obs. in total)
	OTHER_TEST_DT	Date of other test	98.71	.02	1.27		1967-09-14	2215-10-15				
	SPECIMEN_COLLECTION_DT	Date of specimen collection	62.14		37.86		1996-09-01	2016-12-20				
	SURVEY_DT	Date of survey	.47		99.53		1995-01-23	2007-12-03				

mhtb_Other_test_1993apr



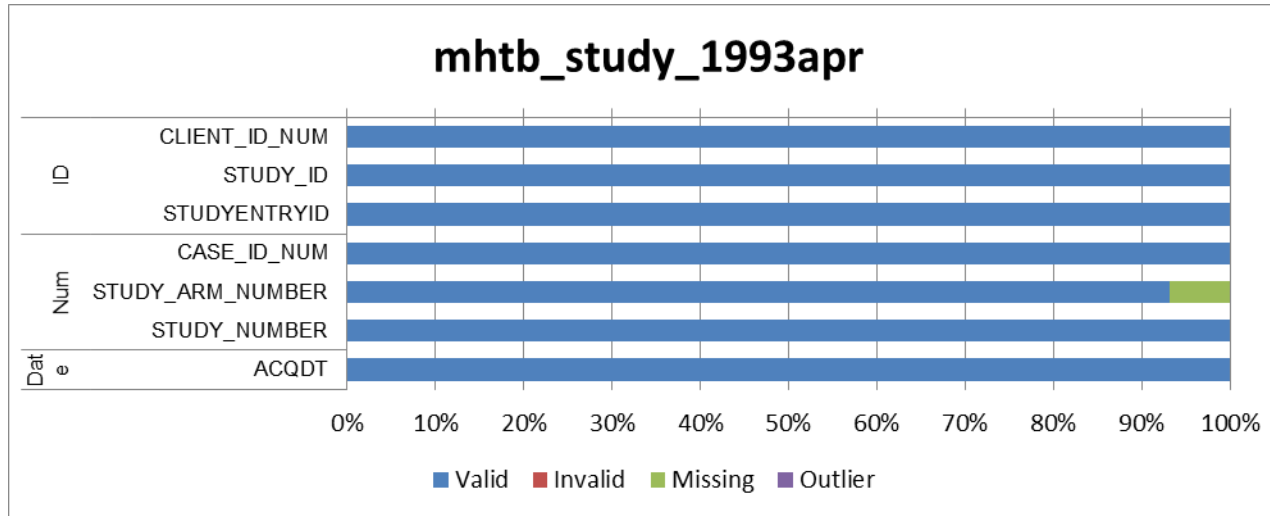
Sources table

Dataset Label: MH Tuberculosis Registry - Sources: 1993 onward				Records: 53460			Legend (Potential Data Quality Problems) :					
Dataset Name: mhtb_sources_1993apr = No variance or 100% missing value				Period: 1993-2015			None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A		
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in the client table	100.00		.00							
	SOURCE_CASE_NUM	Source case number	100.00		.00							
	SOURCE_ID_NUM	Source ID number	100.00		.00							
	SOURCEENTRYID	SourcesEntryID	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	5.00	.02	.00	.15	
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				



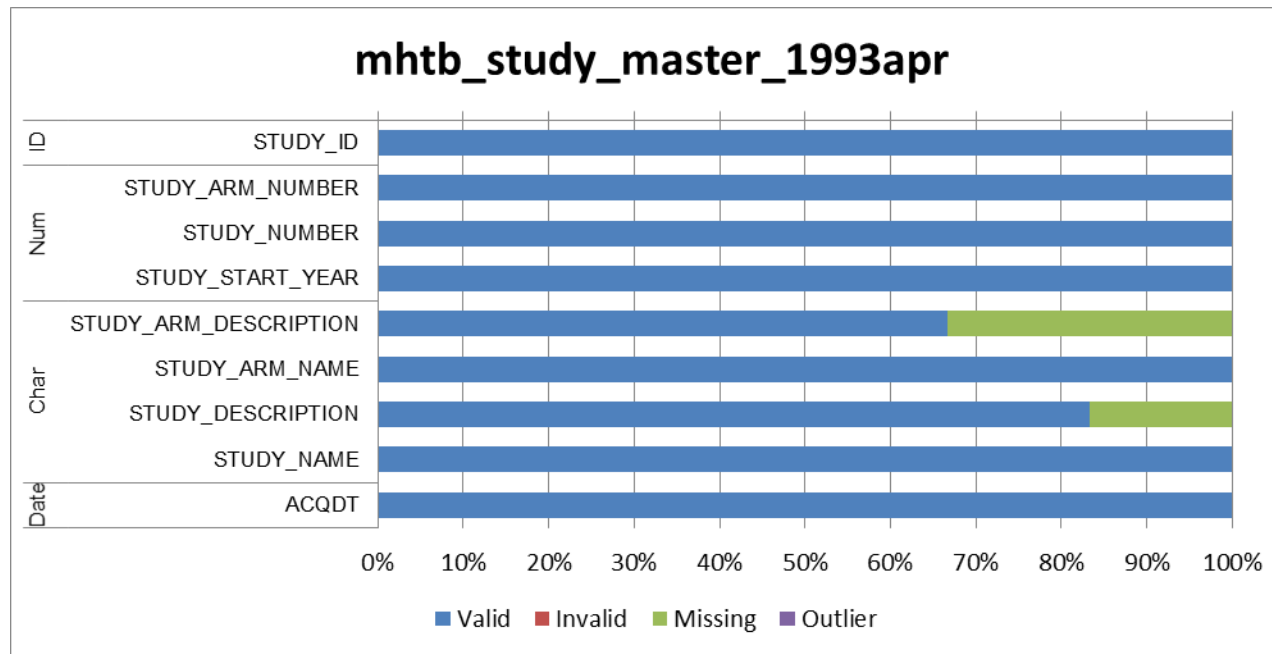
Study Table

Dataset Label: MH Tuberculosis Registry - Study: 1993 onward				Records: 73		Legend (Potential Data Quality Problems) :						
Dataset Name: mhtb_study_1993apr = No variance or 100% missing value				Period: 1993-2015		None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	STUDY_ID	Study identification number. Links to study name, description and start date	100.00		.00							
	STUDYENTRYID	Auto generated identification number	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	3.00	.96	1.00	.51	
	STUDY_ARM_NUMBER	Study arm of the study	93.15		6.85		1.00	5.00	1.65	1.47	.84	
	STUDY_NUMBER	Study number to which study arm links	100.00		.00		1.00	4.00	1.23	1.00	.68	
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				



Study Master Table

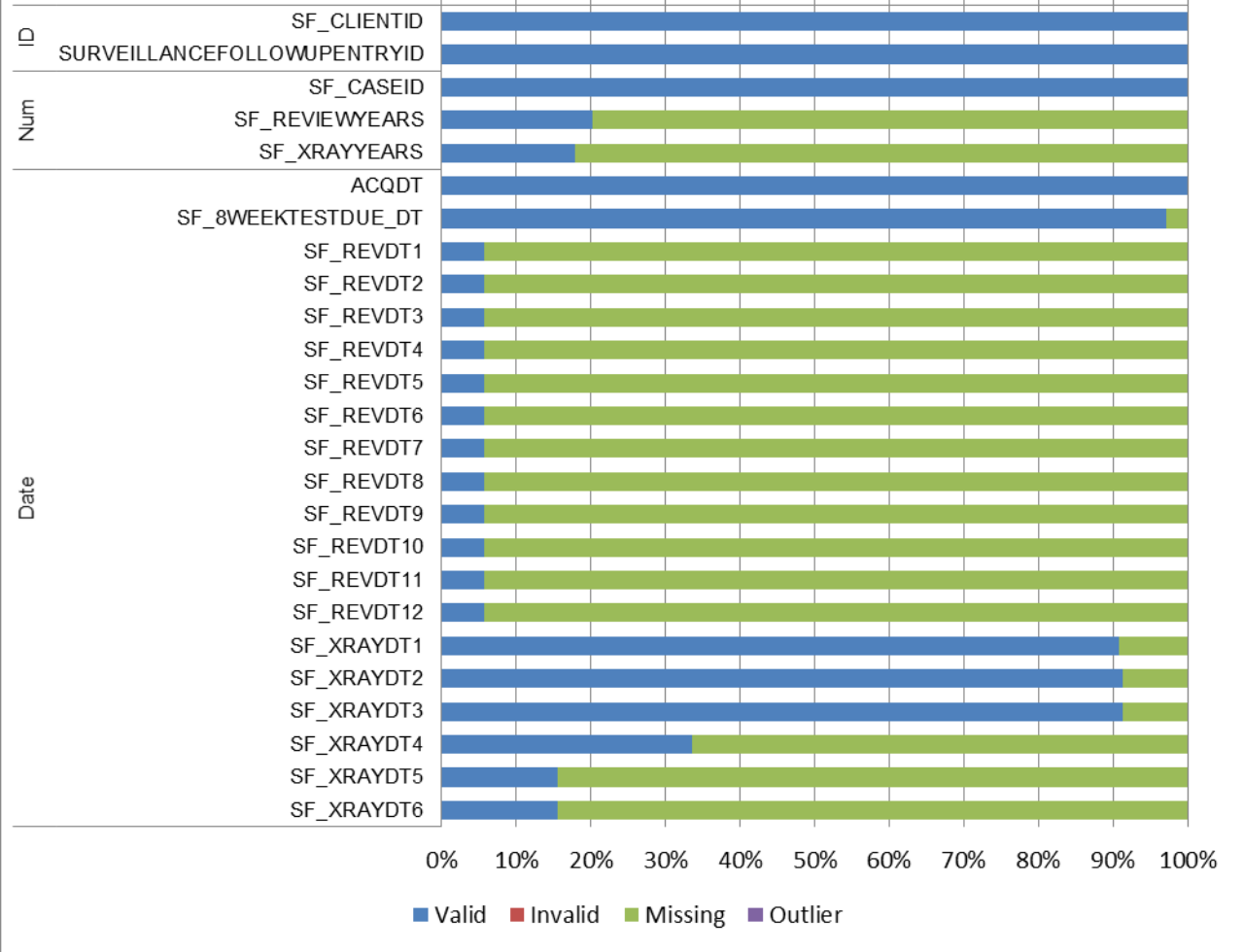
Dataset Label: MH Tuberculosis Registry - Study Master: 1993 onward		Records: 6	Legend (Potential Data Quality Problems) :									
Dataset Name: mhtb_study_master_1993apr = No variance or 100% missing value		Period: 1993-2015	None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A						
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	STUDY_ID	Study identification number. Links to study name, description and start date	100.00		.00							
Num	STUDY_ARM_NUMBER	Study arm of the study	100.00		.00		0.00	5.00	2.50	2.50	1.87	
	STUDY_NUMBER	Study number to which study arm links	100.00		.00		0.00	4.00	1.83	1.50	1.47	
	STUDY_START_YEAR	Year in which study started	100.00		.00		0.00	1999.00	1664.50	1997.00	815.44	
Top 10 Observed Values												
Char	STUDY_ARM_DESCRIPTION	Description of study arm	66.67		33.33		A Randomized, double-blind, phase I safety and toleration study					
	STUDY_ARM_NAME	Name of study arm	100.00		.00		INH & Rifampin, NH/Rifabutin/PZA/STREP, RIF/EMB/PZA, RPT					
	STUDY_DESCRIPTION	Brief description of study	83.33		16.67		Efficacy and Safety of Once-Weekly Rifapentine and Isoniazid C					
	STUDY_NAME	Name of study	100.00		.00		CDC/Atlanta #200-94-0807, 3RPT/INH or 9 MTHS INH, TBTC St					
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				



Surveillance Follow up Table

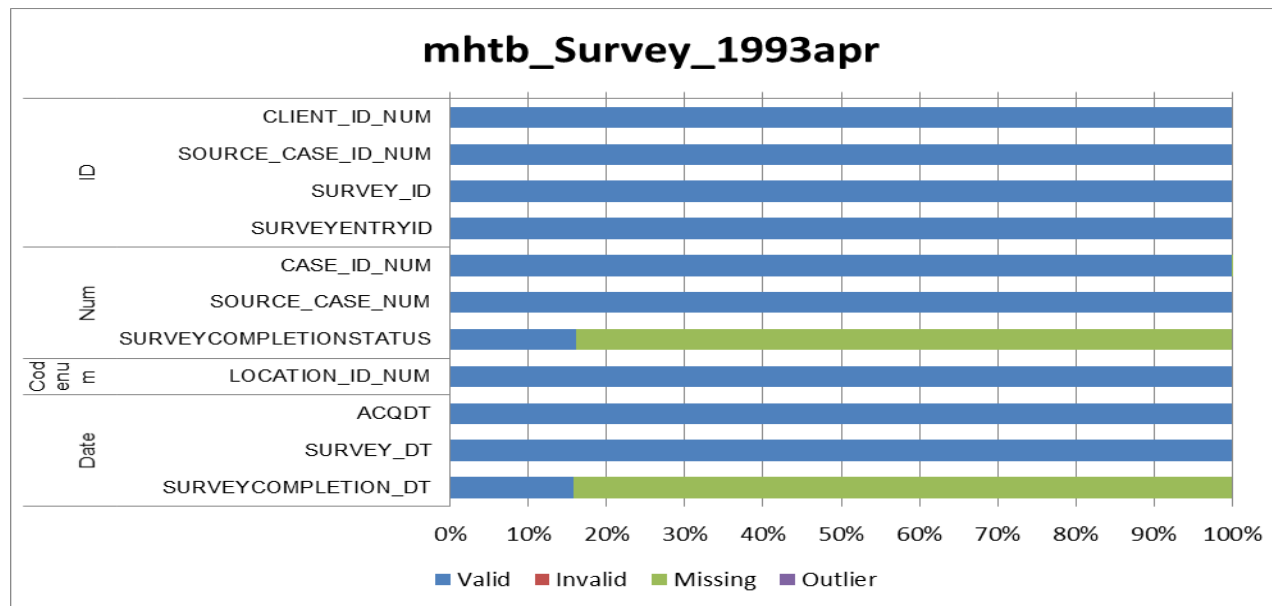
Dataset Label: MH Tuberculosis Registry - Surveillance Follow up: 1993 onward			Records: 173				Legend (Potential Data Quality Problems) :					
Dataset Name: mhbt_surveil_followup_1993apr = No variance or 100% missing value			Period: 1993-2015				None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A		
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	SF_CLIENTID	Links to main client id table. Each client should have one unique identifier	100.00		.00							
	SURVEILLANCEFOLLOWUPENTRYID	Auto generated identification number	100.00		.00							
Num	SF_CASEID	Links to main case identification via client ID	100.00		.00		1.00	2.00	1.08	1.00	.26	
	SF_REVIEWYEARS	Number of years in which surveillance is to take place	20.23		79.77		3.00	3.00	3.00	3.00	.00	
	SF_XRAYYEARS	Number of years for x-rays	17.92		82.08		2.00	3.00	2.87	3.00	.34	
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	SF_8WEEKTESTDUE_DT	8 week test due date	97.11		2.89		2005-02-23	2010-09-05				
	SF_REVDT1	Surveillance dates	5.78		94.22		2006-03-01	2007-12-01				
	SF_REVDT2	Surveillance dates	5.78		94.22		2006-06-01	2008-03-01				
	SF_REVDT3	Surveillance dates	5.78		94.22		2006-09-01	2008-06-01				
	SF_REVDT4	Surveillance dates	5.78		94.22		2006-12-01	2008-09-01				
	SF_REVDT5	Surveillance dates	5.78		94.22		2007-03-01	2008-12-01				
	SF_REVDT6	Surveillance dates	5.78		94.22		2007-06-01	2009-03-01				
	SF_REVDT7	Surveillance dates	5.78		94.22		2007-09-01	2009-06-01				
	SF_REVDT8	Surveillance dates	5.78		94.22		2007-12-01	2009-09-01				
	SF_REVDT9	Surveillance dates	5.78		94.22		2008-03-01	2009-12-01				
	SF_REVDT10	Surveillance dates	5.78		94.22		2008-06-01	2010-03-01				
	SF_REVDT11	Surveillance dates	5.78		94.22		2008-09-01	2010-06-01				
	SF_REVDT12	Surveillance dates	5.78		94.22		2008-12-01	2010-09-01				
	SF_XRAYDT1	X Ray due dates	90.75		9.25		2005-03-01	2011-01-01				
	SF_XRAYDT2	X Ray due dates	91.33		8.67		2005-09-01	2011-07-01				
	SF_XRAYDT3	X Ray due dates	91.33		8.67		2006-09-01	2012-07-01				
	SF_XRAYDT4	X Ray due dates	33.53		66.47		2007-09-01	2013-07-01				
SF_XRAYDT5	X Ray due dates	15.61		84.39		2008-07-01	2010-04-01					
SF_XRAYDT6	X Ray due dates	15.61		84.39		2009-01-01	2010-10-01					

mhtb_surveil_followup_1993apr



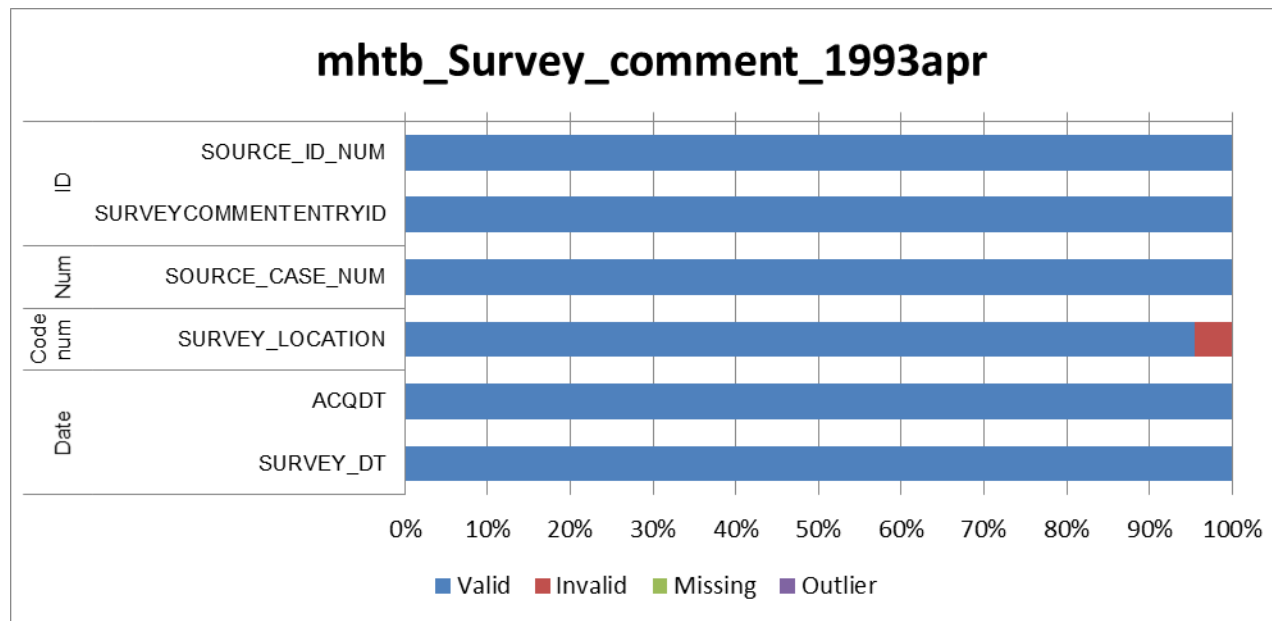
Survey Table

Dataset Label: MH Tuberculosis Registry - Survey: 1993 onward				Records: 37743		Legend (Potential Data Quality Problems) :						
Dataset Name: mhtb_Survey_1993apr = No variance or 100% missing value				Period: 1993-2015		None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables) ✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Links to individual client ID number in client table	100.00		.00							
	SOURCE_CASE_ID_NUM	ID number of source case if that is reason for survey	100.00		.00							
	SURVEY_ID	Identification number of this survey	100.00		.00							
	SURVEYENTRYID	SurveyEntryID	100.00		.00							
Num	CASE_ID_NUM	Case ID number of this client if this client became a case	99.97		.03		0.00	3.00	.00	.00	.03	
	SOURCE_CASE_NUM	Case number of source case	100.00		.00		0.00	4.00	.34	.00	.53	
	SURVEYCOMPLETIONSTATUS	SurveyCompletionStatus	16.01		83.99		0.00	3.00	2.94	3.00	.41	
Top 10 Observed Values												
Codenum	LOCATION_ID_NUM	Location where survey took place	100.00		.00		18, 386, 117, 41, 1, 42, 145, 36, 408, 43					✓
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	SURVEY_DT	Date of survey	100.00		.00		1902-12-23	2007-12-03				
	SURVEYCOMPLETION_DT	Date survey completed	15.69		84.31		2005-09-30	2010-05-31				



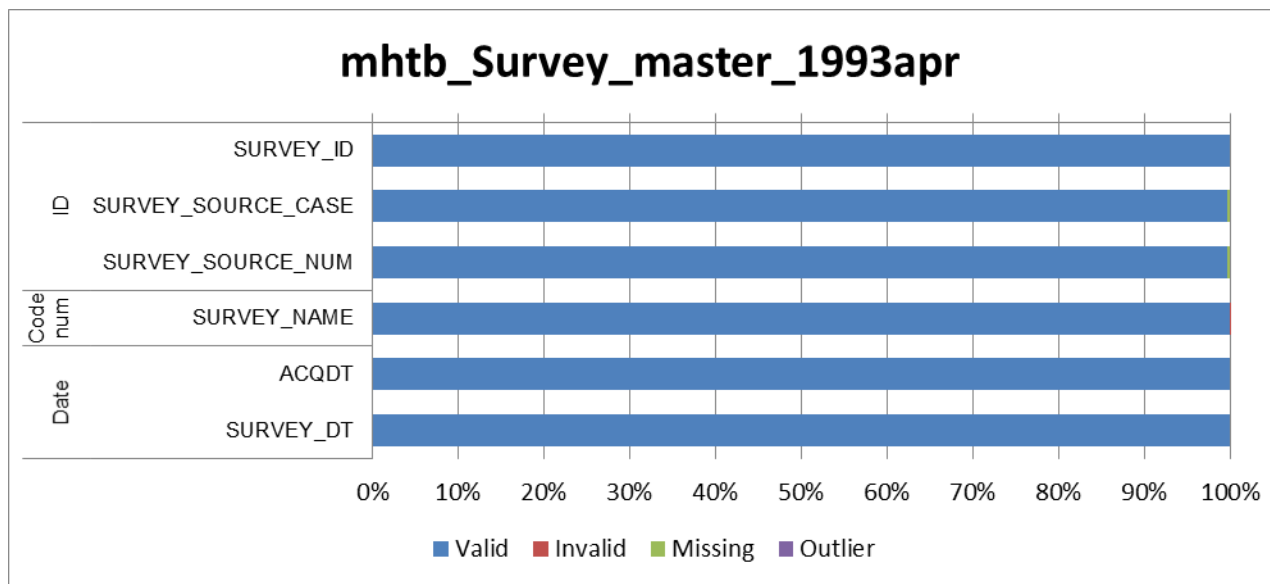
Survey Comment Table

Dataset Label: MH Tuberculosis Registry - Survey Comment: 1993 onward			Records: 44				Legend (Potential Data Quality Problems) :					
Dataset Name: mhtb_Survey_comment_1993apr			Period: 1993-2015				None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A		
= No variance or 100% missing value											Legend for comment column	
											Blank = variables have not been tested (no formats have been specified for the variables)	
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	SOURCE_ID_NUM	ID number of source case if that is reason for survey	100.00		.00							
	SURVEYCOMMENTENTRYID	SurveyCommentEntryID	100.00		.00							
Num	SOURCE_CASE_NUM	Case number of the source case	100.00		.00		0.00	3.00	1.05	1.00	.53	
Top 10 Observed Values												
Date Codenum	SURVEY_LOCATION	Location where survey took place	95.45	4.55	.00		59, 60, 65, 69, 73, 77, 81, 82, 85, 86					59, 60 (2 Invalid Obs. in total)
	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
Date	SURVEY_DT	Date of survey	100.00		.00		1995-11-01	1999-11-01				



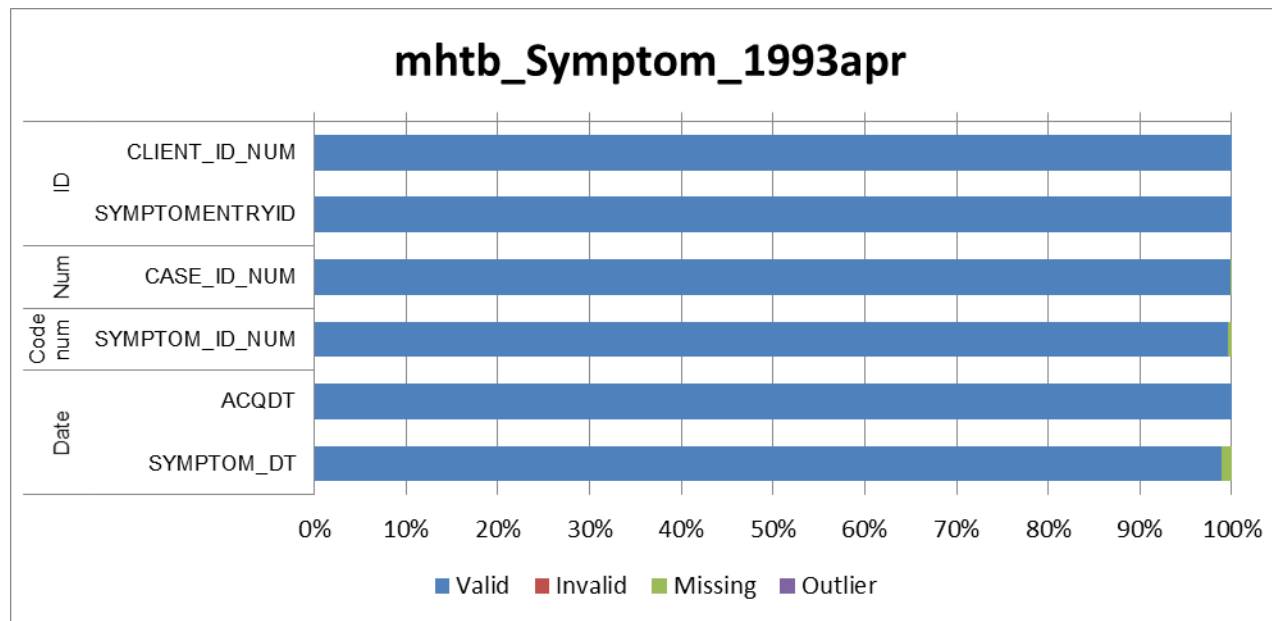
Survey Master Table

Dataset Label: MH Tuberculosis Registry - Survey Master: 1993 onward			Records: 1026			Legend (Potential Data Quality Problems) :						
Dataset Name: mhtb_Survey_master_1993apr = No variance or 100% missing value			Period: 1993-2015			None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	SURVEY_ID	Id number of this survey	100.00		.00							
	SURVEY_SOURCE_CASE	Links to "source case number" in survey table	99.71		.29							
	SURVEY_SOURCE_NUM	Links to "survey source case id number" in survey table	99.71		.29							
Top 10 Observed Values												
Date Codenum	SURVEY_NAME	Links to survey table where this field is called survey location ID	99.90	.10	.00		41, 42, 1, 19, 36, 43, 17, 44, 24, 33				804 (1 Invalid Obs. in total)	
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	SURVEY_DT	Date of survey	100.00		.00		1902-12-23	2007-12-03				



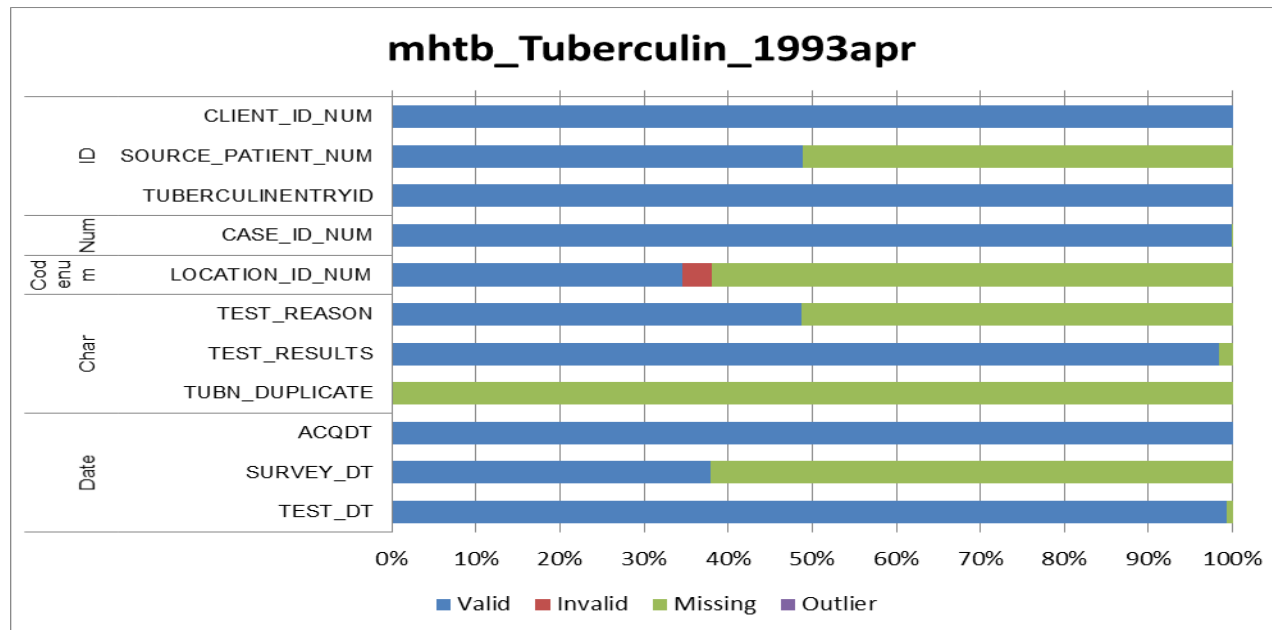
Symptom Table

Dataset Label: MH Tuberculosis Registry - Symptom: 1993 onward			Records: 11814				Legend (Potential Data Quality Problems) :					
Dataset Name: mhtb_Symptom_1993apr = No variance or 100% missing value			Period: 1993-2015				None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A		
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	SYMPTOMETRYID	SymptomEntryID	100.00		.00							
Num	CASE_ID_NUM	Specific case number	99.85		.15		0.00	3.00	.03	.00	.17	
	Top 10 Observed Values											
Date Codenum	SYMPTOM_ID_NUM	Type of symptoms: ex cough, loss of appetite etc.	99.63	.02	.36		39, 1, 4, 12, 2, 3, 15, 9, 10, 14				0 (2 Invalid Obs. in total)	
	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
Date	SYMPTOM_DT	Symptom onset date	98.90	.02	1.08		1984-12-01	2214-12-01				2212-08-15, 2214-12-01 (2 Invalid Obs. in total)



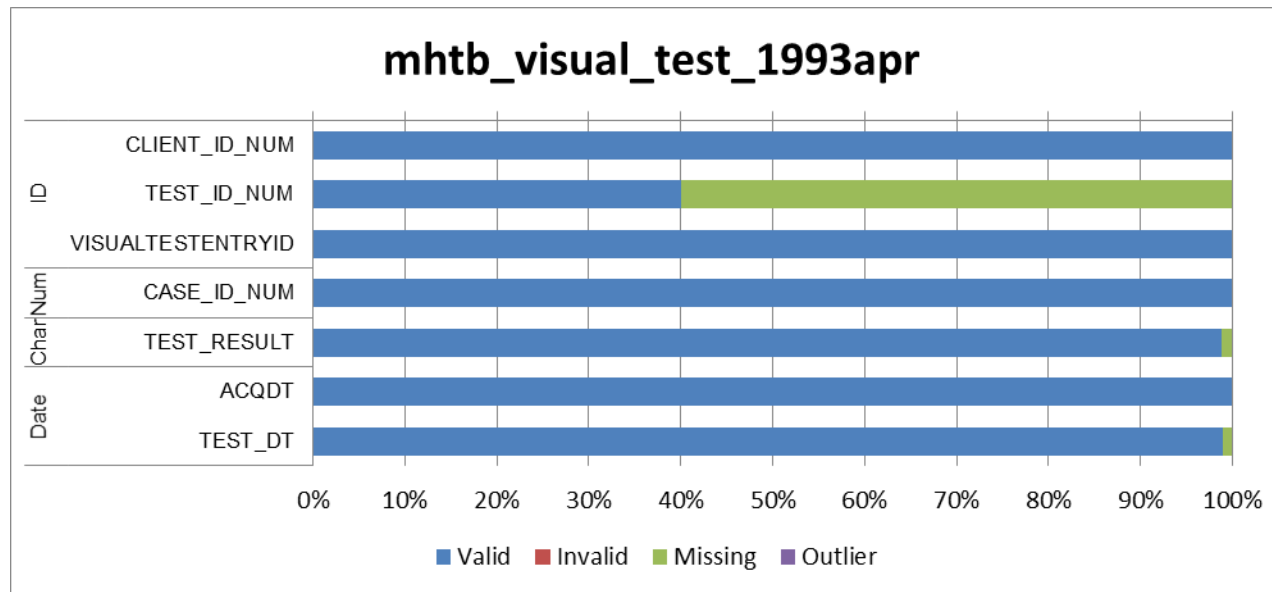
Tuberculin Table

Dataset Label: MH Tuberculosis Registry - Tuberculin: 1993 onward			Records: 71905			Legend (Potential Data Quality Problems) :						
Dataset Name: mhtb_Tuberculin_1993apr = No variance or 100% missing value			Period: 1993-2015			None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables) ✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	SOURCE_PATIENT_NUM	ID number of source case if that is reason for survey	48.95		51.05							
	TUBERCULINENTRYID	TuberculinEntryID	100.00		.00							
Num	CASE_ID_NUM	Specific case number	99.98		.02		0.00	370.00	.02	.00	1.67	
Top 10 Observed Values												
Codenum	LOCATION_ID_NUM	Location where survey took place	34.60	3.46	61.93		41, 386, 42, 44, 211, 1, 408, 81, 404, 43				0, 59, 60, 190, ... (2489 Invalid Obs. in total)	
Char	TEST_REASON	Reason for test: contact or TB survey	48.73		51.27		S, C				✓	
	TEST_RESULTS	Test result	98.40		1.60		0 mm, 10 mm, 20 mm, 15 mm, DNA, 12 mm, 18 mm, 14 mm					
	TUBN_DUPLICATE	Tuberculin duplicate	.07		99.93		Y, N					
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	SURVEY_DT	Date of survey	37.85		62.15		1902-12-23	2014-12-22				
	TEST_DT	Date of test	99.23		.77		1902-09-13	2016-01-08				



Visual test table

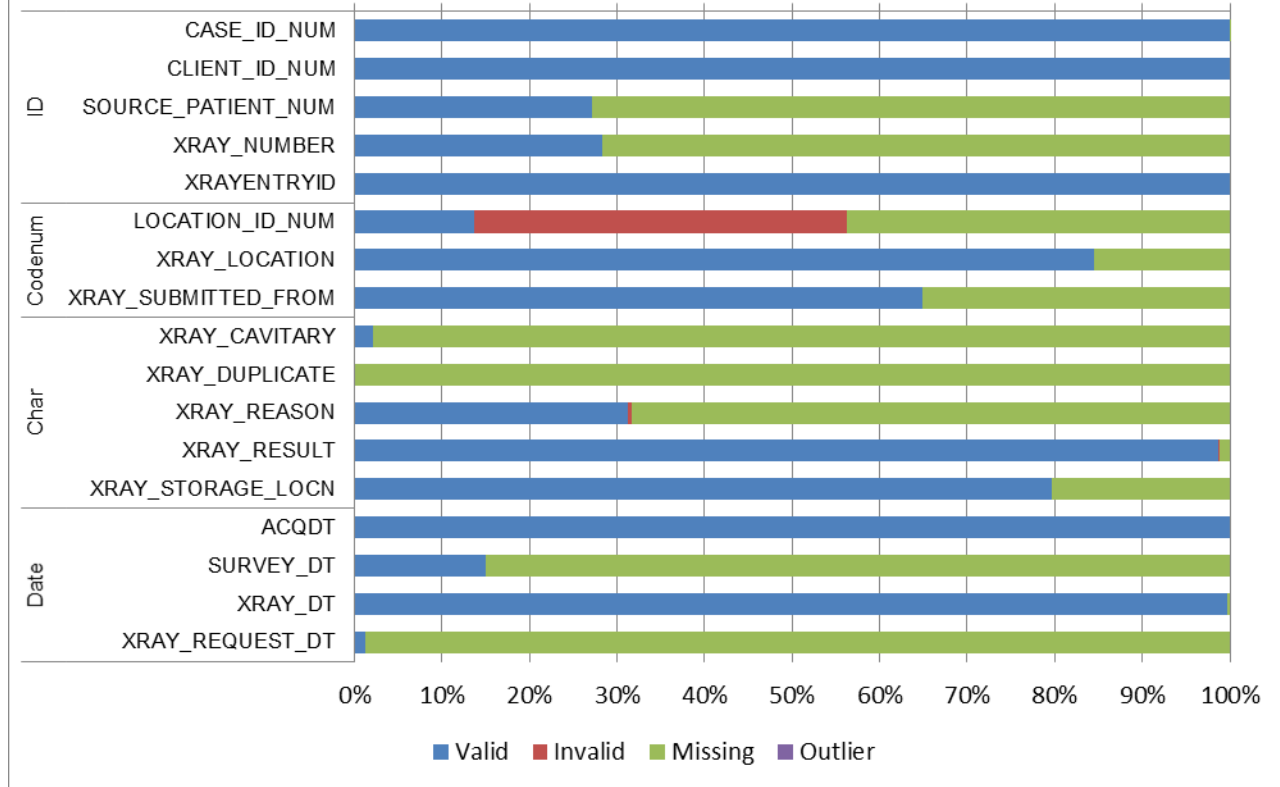
Dataset Label: MH Tuberculosis Registry - Visual Test: 1993 onward			Records: 485			Legend (Potential Data Quality Problems) :						
Dataset Name: mhtb_visual_test_1993apr = No variance or 100% missing value			Period: 1993-2015			None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
						Legend for comment column Blank = variables have not been tested (no formats have been specified for the variables)						
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	TEST_ID_NUM	Unique number of test	40.00		60.00							
	VISUALTESTENTRYID	Auto generated identification number	100.00		.00							
Num	CASE_ID_NUM	Specific case number	100.00		.00		0.00	10.00	.98	1.00	.55	
Top 10 Observed Values												
Char	TEST_RESULT	Test result	98.76		1.24		15/15, 14/15, 13/15, 11/15, 12/15, 10/15, 9/15, 7/15, 8/15, 3/1					
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	TEST_DT	Date of visual exam	98.97		1.03		1969-03-03	2015-02-05				



Xray Table

Dataset Label: MH Tuberculosis Registry - Xray: 1993 onward			Records: 77634			Legend (Potential Data Quality Problems) :						
Dataset Name: mhtb_Xray_1993apr = No variance or 100% missing value			Period: 1993-2015			None or Minimal < 5%	Moderate 5-30%	Significant > 30%	Unknown or N/A			
Legend for comment column												
Blank = variables have not been tested (no formats have been specified for the variables)												
✓ = Variables have been tested against the associated formats and no invalid values found												
Type	Variable Name	Variable Label	Valid	Invalid	Missing	Outlier	Min	Max	Mean	Median	STD	Comment
ID	CASE_ID_NUM	Specific case number	99.99		.01							
	CLIENT_ID_NUM	Link to individual client ID number in client table	100.00		.00							
	SOURCE_PATIENT_NUM	ID number of source case if that is reason for survey	27.08		72.92							
	XRAY_NUMBER	Unique identifier for Xray	28.26		71.74							
	XRAYENTRYID	Auto generated identification number	100.00		.00							
Top 10 Observed Values												
Codenum	LOCATION_ID_NUM	Location where survey took place	13.73	42.58	43.69		0, 36, 117, 18, 1, 387, 170, 377, 41, 43				0, 59, 60, 190, ... (33059 Invalid Obs. in total)	
	XRAY_LOCATION	Location on body where Xray taken	84.41		15.59		1, 7, 4, 10, 11, 9, 14, 2, 5, 6				✓	
	XRAY_SUBMITTED_FROM	Location from which Xray submitted	64.88		35.12		14, 34, 57, 24, 8, 36, 127, 33, 11, 77				✓	
Char	XRAY_CAVITARY	Cavitary, Non-cavitary or not specified	2.11	.00	97.89		N, C, S, 1				1 (2 Invalid Obs. in total)	
	XRAY_DUPLICATE	Whether or not Xray is a duplicate	.06		99.94		D, N					
	XRAY_REASON	Reason for Xray: Contact or TB survey	31.29	.37	68.33		C, S, R, U				R, U (291 Invalid Obs. in total)	
	XRAY_RESULT	X Ray result: Normal, Abnormal, Spoiled etc.	98.71	.12	1.17		N, A, P, Z, U, S, R, D, L				S (95 Invalid Obs. in total)	
	XRAY_STORAGE_LOCN	Location where Xray stored	79.65		20.35		S, H, M				✓	
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2016-03-11	2016-03-11				
	SURVEY_DT	Date of survey	14.98	.02	85.00		1899-12-30	2199-11-28				1899-12-30, 2199-11-28 (16 Invalid Obs. in total)
	XRAY_DT	Date of XRay	99.73	.01	.27		1901-01-01	5201-02-14				6 invalid obs. out of [1900-01-01, 2016-03-11] range
	XRAY_REQUEST_DT	Date of Xray request	1.30		98.70		1991-10-30	2015-12-22				✓

mhtb_Xray_1993apr



Referential Integrity Check

Primary Key: client_id_num

Primary Table	Duplicate	Missing	Total Records
MHTB_CLIENT_1993APR	0	0	72422

Primary Key: client_id_num

Primary Table	Duplicate	Missing	Total Records
MHTB_CLIENT_1993APR	0	0	72422

Foreign Key: sf_clientid

Foreign Table	Orphan Values	Total Records
MHTB_SURVEIL_FOLLOWUP_1993APR	2	173

Foreign Key: client_id_num

Foreign Table	Orphan Values	Total Records
MHTB_CASE_1993APR	0	4128
MHTB_ADDRESS_1993APR	0	76074
MHTB_CONTACT_1993APR	4	47880
MHTB_CONTACTLIST_COMMNT_1993APR	2	427
MHTB_CLINIC_VISIT_1993APR	2	846
MHTB_DAILY_DRUG_1993APR	3	8468
MHTB_DIAGNOSIS_1993APR	15	19990
MHTB_DOT_DRUG_DOSES_1993APR	0	49047
MHTB_DOT_DRUG_LIST_1993APR	3	30498
MHTB_DOT_DRUG_MASTER_1993APR	1	9003
MHTB_DRUG_REACTION_1993APR	1	13009
MHTB_DRUG_SUMMARY_1993APR	5	8364
MHTB_HEMATOLOGY_RESULT_1993APR	0	36
MHTB_HIV_1993APR	1	2894
MHTB_HOSPITAL_1993APR	6	2428
MHTB_IMMIGRANT_TRACKING_1993APR	0	4759
MHTB_INPATIENT_DRUG_1993APR	5	12816
MHTB_OTHER_DISEASES_1993APR	0	941
MHTB_OTHER_TEST_1993APR	7	14879
MHTB_SOURCES_1993APR	49	53460
MHTB_STUDY_1993APR	0	73
MHTB_SURVEY_1993APR	10	37743
MHTB_SYMPTOM_1993APR	15	11814
MHTB_TUBERCULIN_1993APR	43	71905
MHTB_VISUAL_TEST_1993APR	0	485
MHTB_XRAY_1993APR	49	77634

Primary Key: client_id_num case_id_num

Primary Table	Filter Condition	Duplicate	Missing	Total Records
MHTB_CASE_1993APR	WHERE=(CASE_ID_NUM > 0)	0	0	4073

Foreign Key: client_id_num case_id_num

Foreign Table	Filter Condition	Orphan Values	Total Records
MHTB_CONTACT_1993APR	WHERE=(CASE_ID_NUM > 0)	41	47230
MHTB_CONTACTLIST_COMMNT_1993APR	WHERE=(CASE_ID_NUM > 0)	11	409
MHTB_CLINIC_VISIT_1993APR	WHERE=(CASE_ID_NUM > 0)	6	311
MHTB_DAILY_DRUG_1993APR	WHERE=(CASE_ID_NUM > 0)	22	1275
MHTB_DIAGNOSIS_1993APR	WHERE=(CASE_ID_NUM > 0)	16	4579
MHTB_DOT_DRUG_DOSES_1993APR	WHERE=(CASE_ID_NUM > 0)	17	19419
MHTB_DOT_DRUG_LIST_1993APR	WHERE=(CASE_ID_NUM > 0)	21	21408
MHTB_DOT_DRUG_MASTER_1993APR	WHERE=(CASE_ID_NUM > 0)	12	3488
MHTB_DRUG_REACTION_1993APR	WHERE=(CASE_ID_NUM > 0)	15	12862
MHTB_DRUG_SUMMARY_1993APR	WHERE=(CASE_ID_NUM > 0)	61	2377
MHTB_HEMATOLOGY_RESULT_1993APR	WHERE=(CASE_ID_NUM > 0)	0	19
MHTB_HIV_1993APR	WHERE=(CASE_ID_NUM > 0)	4	1744
MHTB_HOSPITAL_1993APR	WHERE=(CASE_ID_NUM > 0)	87	2297
MHTB_INPATIENT_DRUG_1993APR	WHERE=(CASE_ID_NUM > 0)	10	12010
MHTB_OTHER_TEST_1993APR	WHERE=(CASE_ID_NUM > 0)	22	12115
MHTB_SOURCES_1993APR	WHERE=(CASE_ID_NUM > 0)	109	729
MHTB_STUDY_1993APR	WHERE=(CASE_ID_NUM > 0)	0	64
MHTB_SURVEY_1993APR	WHERE=(CASE_ID_NUM > 0)	7	29
MHTB_SYMPTOM_1993APR	WHERE=(CASE_ID_NUM > 0)	2	281
MHTB_TUBERCULIN_1993APR	WHERE=(CASE_ID_NUM > 0)	183	750
MHTB_VISUAL_TEST_1993APR	WHERE=(CASE_ID_NUM > 0)	1	448
MHTB_XRAY_1993APR	WHERE=(CASE_ID_NUM > 0)	158	8574

Linkability

Dataset	Total Number of Records	Number of Linkable Records	% Linkable Records	Number of Linkable Individuals
MHTB_CLIENT_1993APR	72422	35856	49.51	35737

Phin Type

FILEPHINTYPE	MHTB_CLIENT_1993APR
0 MH verified against concurrent registries	49.51
4 MCHP db specific ScrPHIN - No MH found	50.49

Linkability for active case and contact individuals

	Total Number of Records	Number of Linkable Records	% Linkable Records	Number of Linkable Individuals
Active Case	4128	2956	71.61	2755
Contact	47861	29808	62.28	18512

Agreement

Dataset Name	Degree of Agreement with Registry - Sex (Kappa Statistic)	Degree of Agreement with Registry - DOB (kappa Statistic)
MHTB_CLIENT_1993APR	0.9668	0.9713

Internal Validity: Trend Analysis

Trend Analysis for Dataset: health.mhfb_client_1993apr
Date Variable: added_dt

