

4/9/2019

Data Quality Report

EndoUse Project Data for Southern Health and Central Intake

2015 – 2019



UNIVERSITY
OF MANITOBA

Rady Faculty of
Health Sciences

Sathi Saha
Manitoba Centre for Health Policy



Contents

Summary	3
Overview of Database	3
Variables Name structure	4
Accuracy: Completeness and Correctness (VIMO Table).....	5
Endouse_southern_201501_201812	5
Endouse_cenintak_201501_201901.....	7
Internal Validity: Trend Analysis	9
Linkability	16
Linkability Over Years.....	16
Phin Type.....	17

This report was based on methods and concepts described in:

Smith M, Lix LM, Azimae M, Enns JE, Orr J, Hong S, Roos LL. Assessing the quality of administrative data for research: a framework from the Manitoba Centre for Health Policy, J Am Med Inform Assoc., September 1, 2017. <https://doi.org/10.1093/jamia/ocx078>

Summary

- Data received for Southern health and Central Intake.
- For Southern health total records: 13985, for Central Intake total records: 90707.
- For Southern Health data available for January 2015 to December 2018.

Overview of Database

Domain	SPDS Table	Dataset Label	Number of Records	Number of Fields
PROJECT	ENDOUSE_CENINTAK_201501_201901	Central Intake - ENDO USE JANUARY 2015 TO JANUARY 2019	90707	14
PROJECT	ENDOUSE_SOUTHERN_201501_201812	SOUTHERN HEALTH - ENDO USE JANUARY 2015 TO DECEMBER 2018	13985	22

Variables Name Structure

This tables represents original variables names (name found in excel or sas tables) and renamed variables.

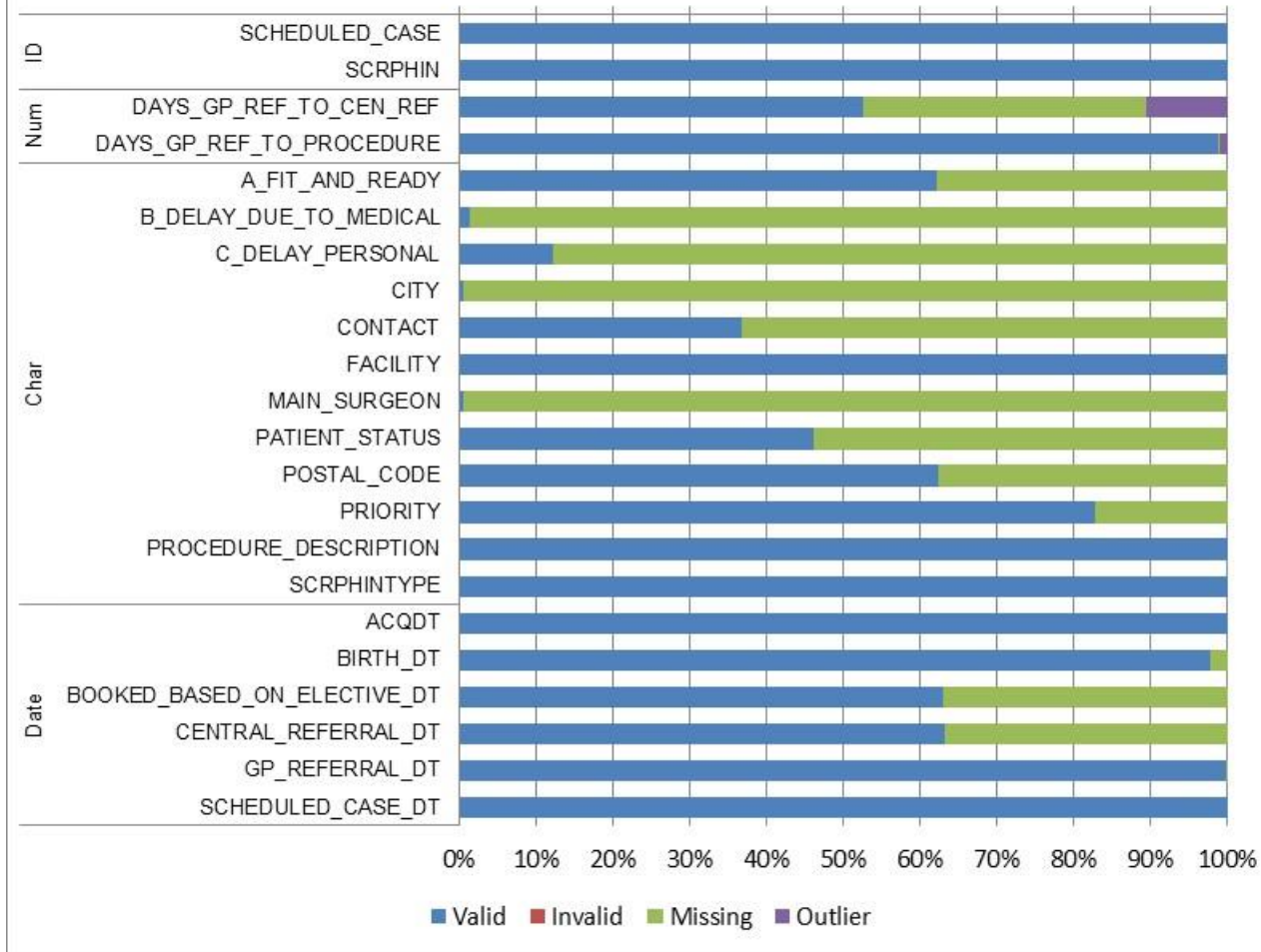
SAS data Installed variables name	Southern Health						Central Intake
	Variables Name						Variables Name
	2015 to Feb 2017	Wait (Mar 2017 to Sep 2017)	STATS (Oct 2017 to April 2018)	Wait (May 2018 to Oct 2018)	Nov 2018	Dec 2018	January 2015 to January 2019
Scheduled_Case	Scheduled Case #	Scheduled Case #	SIMS #	Scheduled Case #	Scheduled Case #	Scheduled Case #	patient_id
Birth_DT	DOB	D.O.B.	D.O.B.	D.O.B.	D.O.B.		
Postal_code		Postal code	Postal code	Postal code	Postal code	Postal code	
Priority	Status	ASA Code		STATUS	STATUS	STATUS	priority_name
Patient_Status		Patient Status	Patient Status	Patient Status	Patient Status	Patient Status	
GP_Referral_DT	Decision to Treat	Referral Date from GP	Referral Date from GP	Referral Date from GP	Referral Date from GP	Referral Date from GP	referral_date
Central_Referral_DT		Date to Central Referral	Date to Central Referral	Date to Central Referral	Date to Central Referral	Date to Central Referral	RequestedDate
Days_gp_ref_to_cen_ref		Wait in Days (B to C)	# DAYS WAITING FROM DATE REFERRAL WRITEN BY GP TO DATE ARRIVED IN CENTRAL REFERRAL OFFICE	Wait in Days - REFERRAL FROM GP TO ARRIVAL AT CENTRAL REFERRAL	Wait in Days (Referral from GP to Arrival in CR Office)	REFERRAL DATE TO ARRIVAL IN CR OFFICE	
Scheduled_Case_DT	Scheduled Case Date	Scheduled Case Date	Scheduled Case Date	Scheduled Case Date	Scheduled Case Date	Scheduled Case Date	appt_date_time
Days_gp_ref_to_procedure	Wait in Days	Wait in Days (B to E)	# DAYS WAITING FROM DATE REFERRAL WRITEN BY GP TO DATE OF PROCEDURE	Wait in Days - REFERRAL FROM GP TO PROCEDURE DATE	Wait in Days (Referral from GP to Procecade Date)	REFERRAL DATE TO PROCEDURE DATE	
Booked_Based_on_Elective_DT		To be Booked by Date	To be Booked by Date	To be Booked by Date	To be Booked by Date	To be Booked by Date	
Procedure_Description	Procedure Description	Procedure Description	Procedure Description	Procedure Description	Procedure Description	Procedure Description	procedure_name
A_Fit_and_Ready		A - Fit and Ready	A - Fit and Ready	A - Fit and Ready	A - Fit and Ready	A - Fit and Ready	
B_Delay_Due_to_Medical		B - Delay Due to Medical	B - Delay Due to Medical	B - Delay Due to Medical	B - Delay Due to Medical	B - Delay Due to Medical	
C_Delay_Personal		C - Delay Personal	C - Delay Personal	C - Delay Personal	C - Delay Personal	C - Delay Personal	
FACILITY	Facility	Facility	Facility	Facility	Facility	Facility	Site
Main Surgeon					Main Surgeon		
City					City		
Contact	Contact						
PhyRequested							PhyRequested
status_name							status_name
type							type
BookedPhyBill							BookedPhyBill
BookedProvider							BookedProvider
ReferringProvider							ReferringProvider
ProcStatus							ProcStatus

Accuracy: Completeness and Correctness (VIMO Table)

Endouse_southern_201501_201812

Dataset Label: SOUTHERN HEALTH - ENDO USE JANUARY 2015 TO DECEMBER 2018			Records: 13985			Legend (Potential Data Quality Problems) :						
Dataset Name: Endouse_southern_201501_201812 = No variance or 100% missing value			Period: 2015-2018			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	SCHEDULED_CASE	Scheduled Case #	100.00		.00							
	SCRPHIN	MCHP Scrambled Personal Health Identification Number	100.00		.00							
Num	DAYS_GP_REF_TO_CEN_REF	# Days from date referral from gp to date to central referral office	52.63	.01	36.83	10.54	-1.00	577.00	4.74	.00	24.93	-1 (1 Invalid Obs. in total)
	DAYS_GP_REF_TO_PROCEDURE	# Days from date referral from gp to procedure date	98.93	.01	.10	.97	-226.00	726.00	70.50	51.66	59.12	-226 (1 Invalid Obs. in total)
Top 10 Observed Values												
Char	A_FIT_AND_READY	A - Fit and Ready	62.17		37.83		Y, N					✓
	B_DELAY_DUE_TO_MEDICAL	B - Delay Due to Medical	1.33		98.67		Y, N					
	C_DELAY_PERSONAL	C - Delay Personal	12.14		87.86		Y, N					
	CITY	City	.64		99.36		Winnipeg, Morden, Carman, Altona, Winkler, Darlingford, Miami					
	CONTACT	Contact	36.70		63.30		ABUABD, CLAGER, ROSJAM, JACCHA, NOSMIC, GARSUL,					
	FACILITY	Facility	100.00		.00		BOUNDARY TRAILS HEALTH CENTRE, CARMAN MEMORIAL					
	MAIN_SURGEON	Main Surgeon	.64		99.36		CLAGER, ABUABD, ROBKRI, METJEN, CHECOU, THOELI					
	PATIENT_STATUS	Patient Status	46.06		53.94		OUTPATIENT					
	POSTAL_CODE	Postal code	62.45		37.55		R0G-0J0, R0A-0R0, R0A-1E0, R0A-1T0, R0G-0B0, R0G-1K0, R					
	PRIORITY	Priority	82.75		17.25		ELECTIVE, URGENT					
PROCEDURE_DESCRIPTION	Procedure Description	100.00		.00		COLONOSCOPY LEAN GREEN ELECTIVE, COLONOSCOPY						
SCRPHINTYPE	MCHP process for assigning SCRPHIN	100.00		.00		0, 4						
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2018-12-12	2018-12-12				
	BIRTH_DT	Birth Date	97.74		2.26		26SEP1917	17FEB2009				
	BOOKED_BASED_ON_ELECTIVE_DT	TO BE BOOKED BY: Based on Elective - 180 days; Semi-Urgent - 28 days and Urgent 14 days	62.96	.03	37.01		2010-04-28	6201-06-20				2010-04-28, 2010-06-11, 2201-01-10, 6201-06-20 (4 Invalid Obs. in total)
	CENTRAL_REFERRAL_DT	Date to Central Referral	63.16	.01	36.83		2015-10-15	2019-05-31				2019-01-20, 2019-05-31 (2 Invalid Obs. in total)
	GP_REFERRAL_DT	Referral Date from GP	99.90		.10		2014-04-08	2018-12-13				✓
SCHEDULED_CASE_DT	Scheduled Case Date	100.00		.00		2015-01-02	2018-12-31				✓	

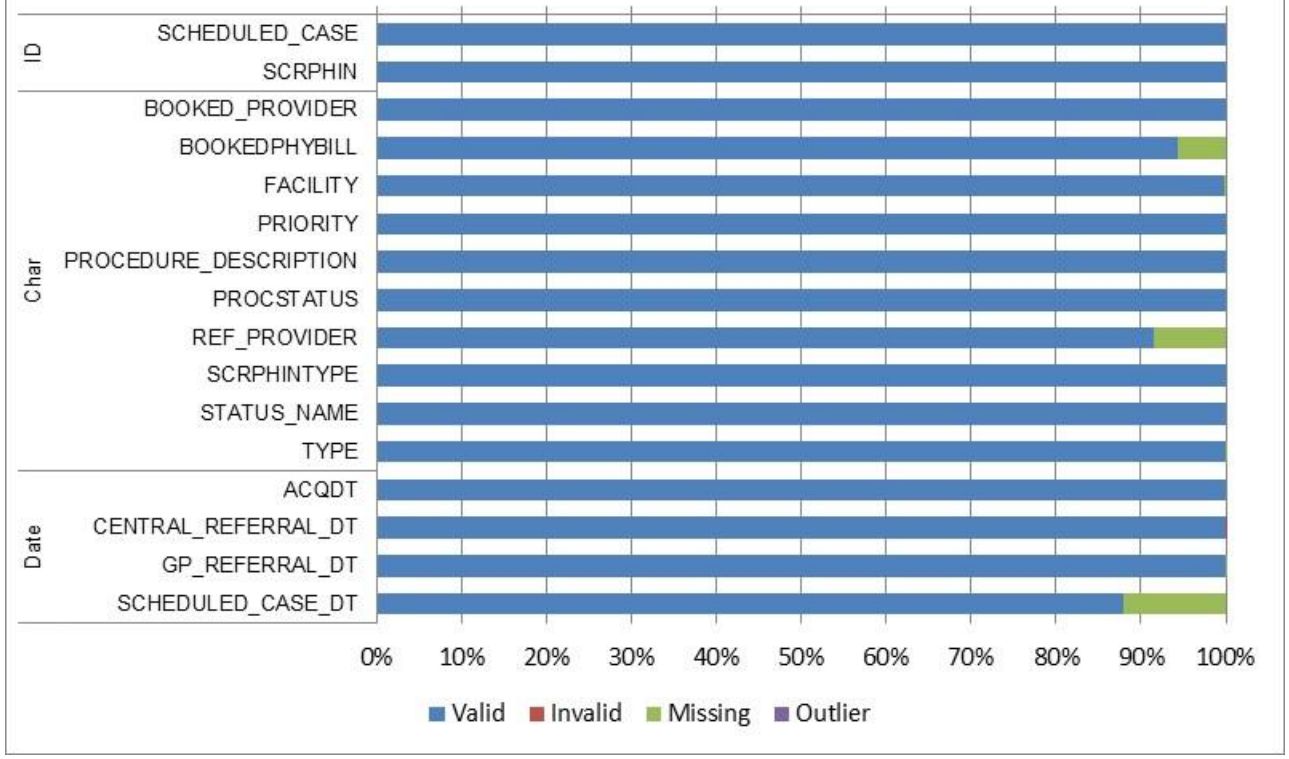
Endouse_southern_201501_201812



Endouse_cenintak_201501_201901

Dataset Label: Central Intake - ENDO USE JANUARY 2015 TO JANUARY 2019		Records: 90707	Legend (Potential Data Quality Problems) :									
Dataset Name: Endouse_cenintak_201501_201901 = No variance or 100% missing value		Period: 2015-2019	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	SCHEDULED_CASE	Scheduled Case #	100.00		.00							
	SCRPHIN	MCHP Scrambled Personal Health Identification Number	100.00		.00							
Top 10 Observed Values												
Char	BOOKED_PROVIDER	Booked Provider Name	100.00		.00		NULNUL, PerDav, SchChr, CanMic, LauChu, MatGeo, BerCha					
	BOOKEDPHYBILL	BookedPhyBill	94.41		5.59		NULL, 1419, 6927, 2096, 1659, 1829, 0493, 2093, 2457, 2200					
	FACILITY	Facility	99.88		.12		HSC, VGH, GGH, SBH, NULL, CGH, SGH, GEN					
	PRIORITY	Priority	100.00		.00		Elective, Semi Urgent, Urgent, Allergy Semi-urgent, Gen Neur					
	PROCEDURE_DESCRIPTION	Procedure Description	100.00		.00		Colonoscopy, Gastroscopy, Gastroscopy/Colonoscopy, Flex-					
	PROCSTATUS	Proc Status	100.00		.00		Booked, Arrived, Cancelled, Noshow					
	REF_PROVIDER	Referring Provider Name	91.56		8.44		CanMic, LauChu, BerCha, PerDav, ManWay, SchChr, DueDo					
	SCRPHINTYPE	MCHP process for assigning SCRPHIN	100.00		.00		0					
	STATUS_NAME	Status name	100.00		.00		Booked, Cancelled, Redirected, Rejected, Waiting, Duplicate,					
	TYPE	Type	99.99		.01		P-Colonoscopy, P-Gastroscopy, NULL, P-Colonoscopy/Gastr					
Date	ACQDT	Date record was acquired at MCHP	100.00		.00		2018-12-12	2018-12-12				
	CENTRAL_REFERRAL_DT	Date to Central Referral	99.92	.08	.00		2014-11-21	9016-08-18				69 invalid obs. out of [19724, 21914] range
	GP_REFERRAL_DT	Referral Date from GP	99.98	.00	.02		2014-11-21	8018-09-28				2077-06-09, 2116-10-18, 7017-09-27, 8018-09-28 (4 Invalid
	SCHEDULED_CASE_DT	Scheduled Case Date	87.85		12.15		2015-10-13	2020-01-07				Obs. in total) ✓

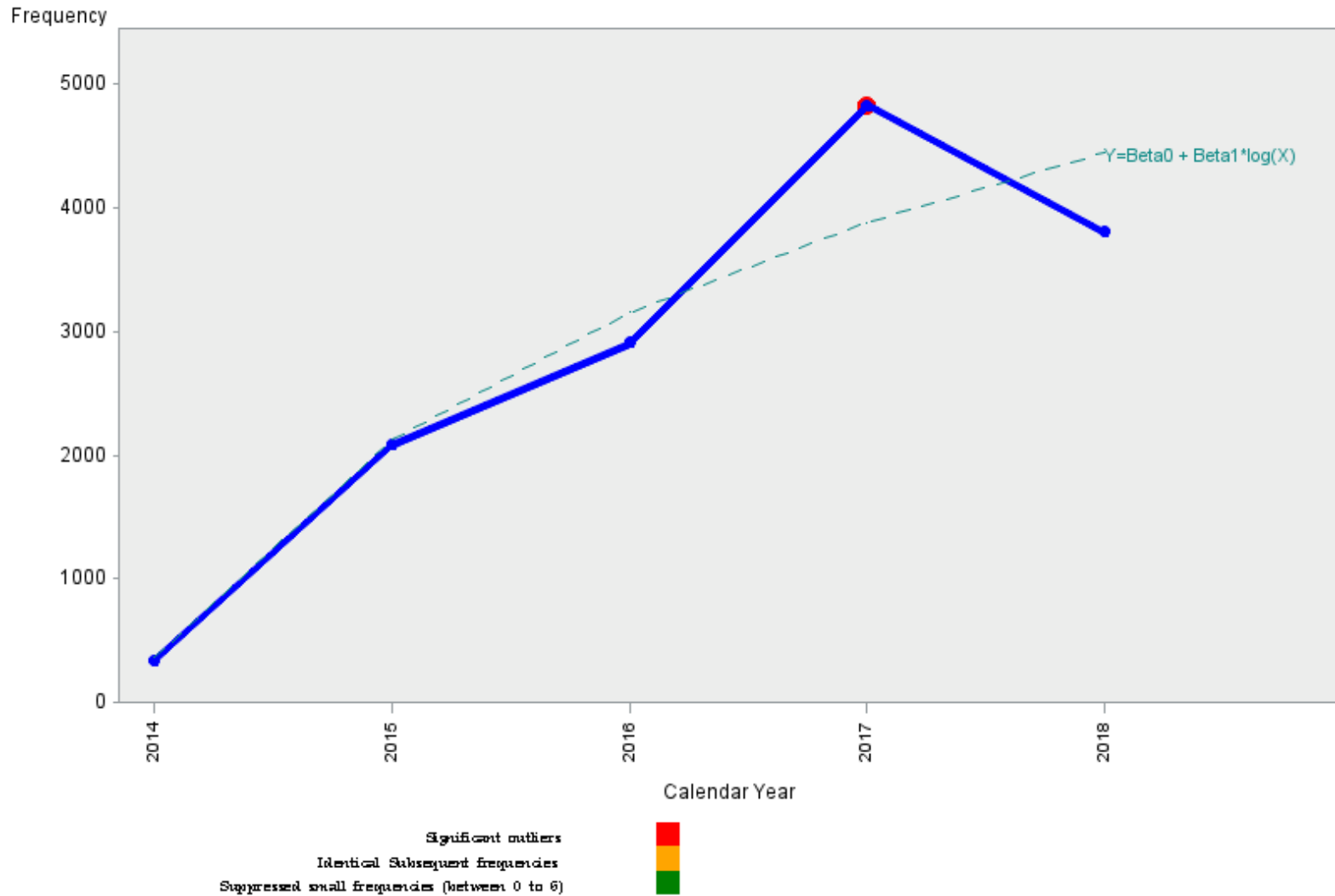
Endouse_cenintak_201501_201901



Internal Validity: Trend Analysis

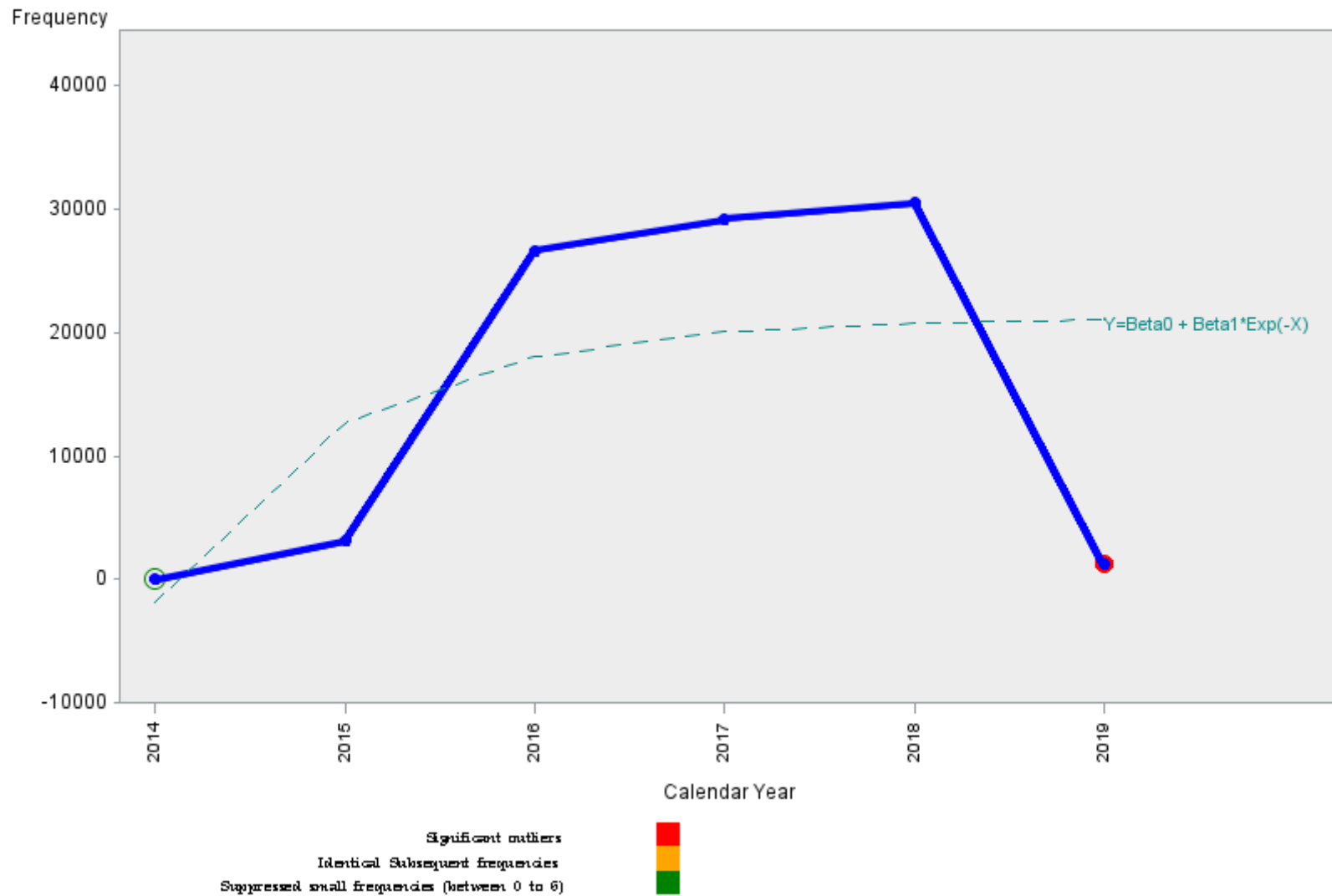
Trend Analysis for Dataset: a_PROJEC.Endouse_southern_201501_201812

Date Variable: GP_REFERRAL_DT
By Variable=All Records



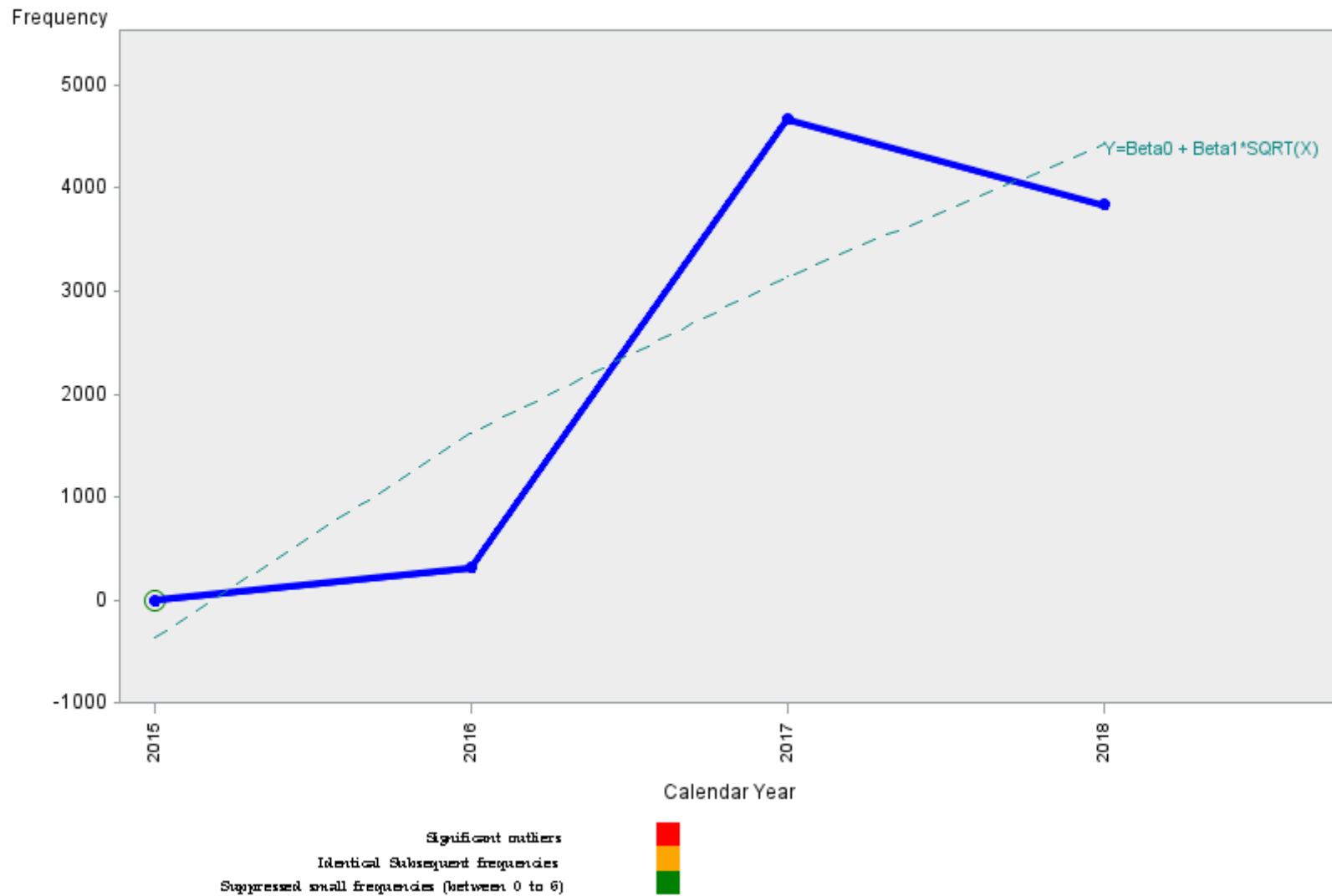
Trend Analysis for Dataset: a_PROJEC.Endouse_oenintak_201501_201901

Date Variable: CP_REFERRAL_DT
By Variable=All Records



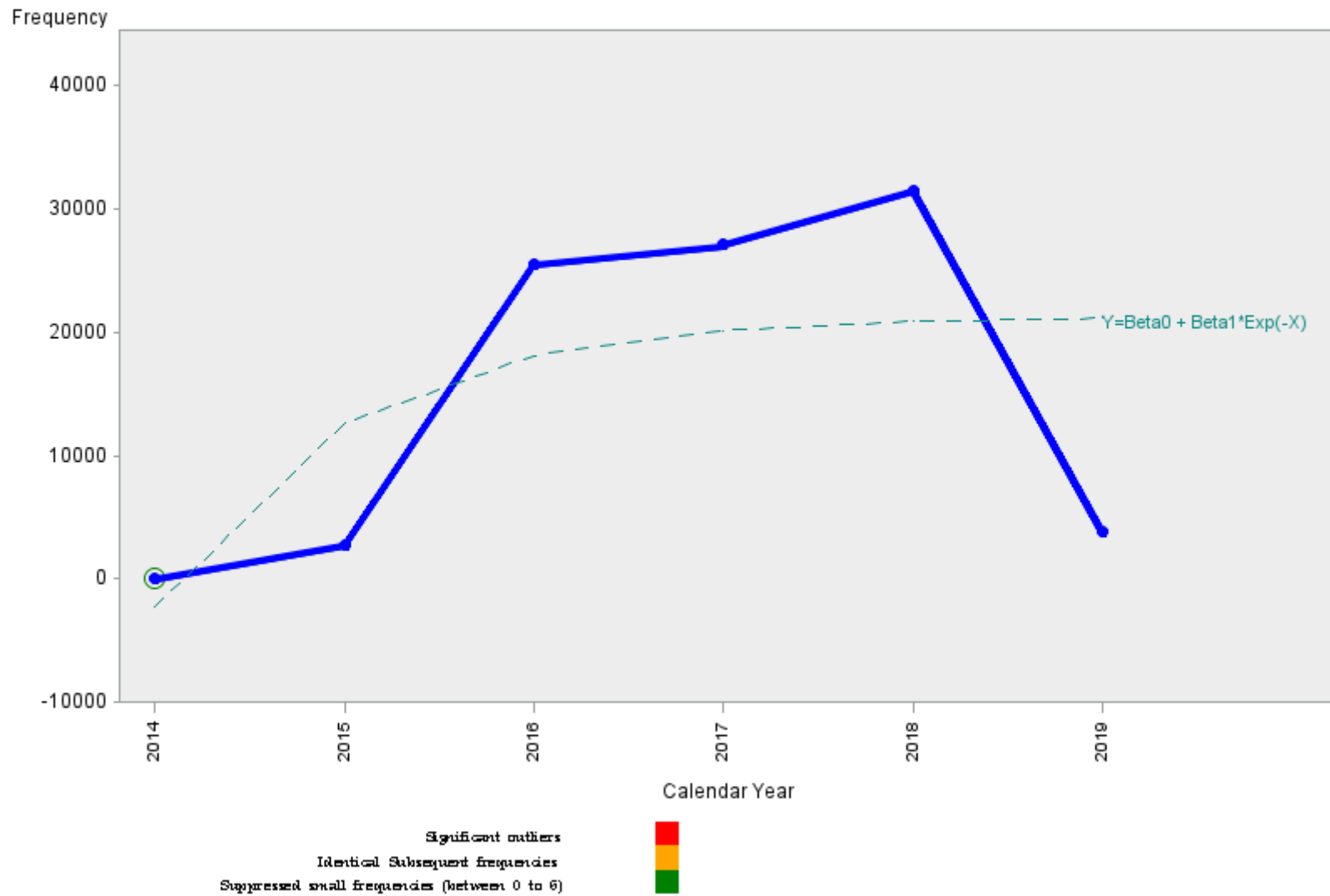
Trend Analysis for Dataset: a_PROJEC.Endouse_southern_201501_201812

Date Variable: CENTRAL_REFERRAL_DT
By Variable=All Records



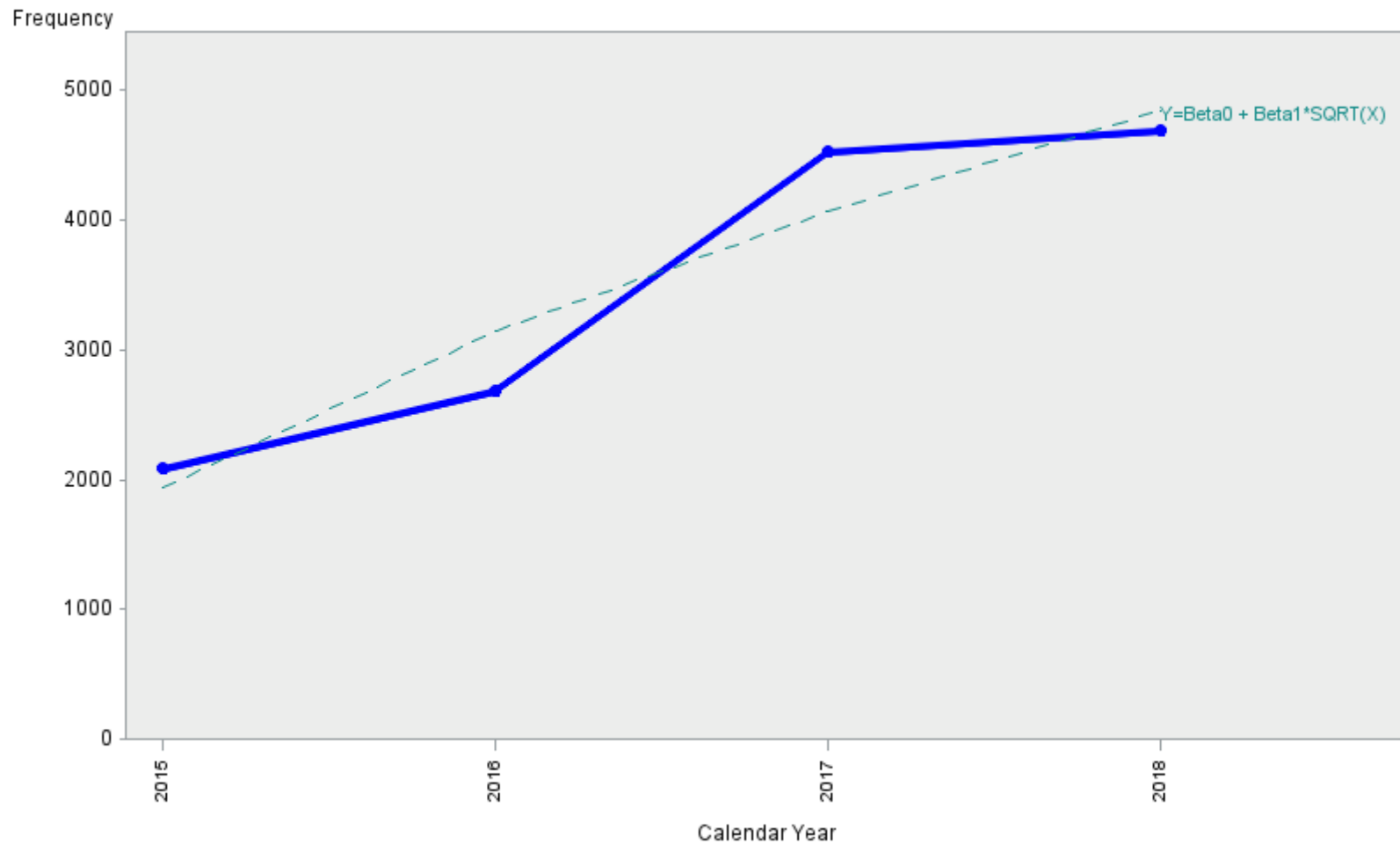
Trend Analysis for Dataset: a_PROJEC.Endouse_oenintak_201501_201901

Date Variable: CENTRAL_REFERRAL_DT
By Variable=All Records



Trend Analysis for Dataset: a_PROJEC.Endouse_southern_201501_201812

Date Variable: SCHEDULED_CASE_DT
By Variable=All Records

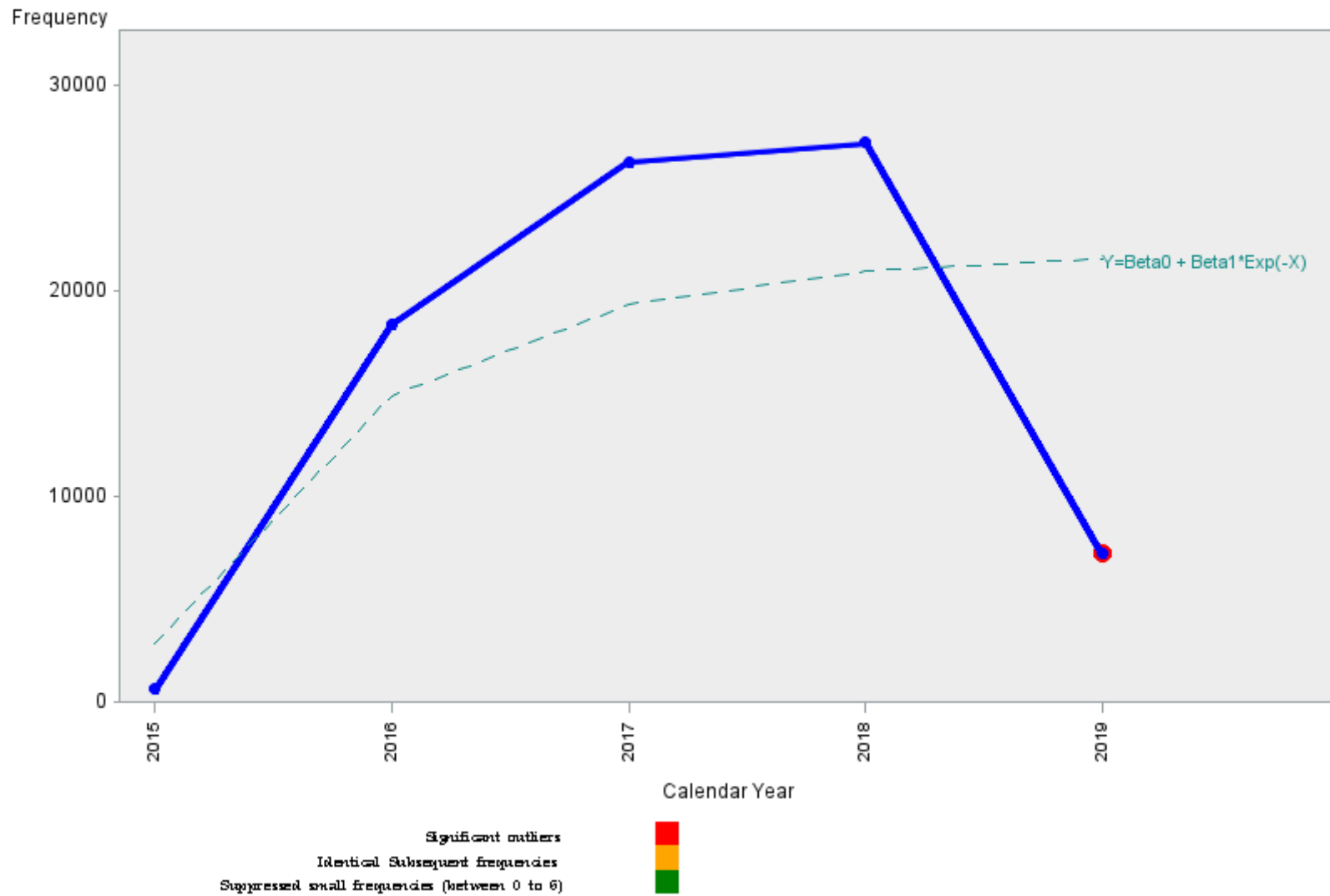


Significant outliers
Identical Subsequent frequencies
Suppressed small frequencies (between 0 to 6)

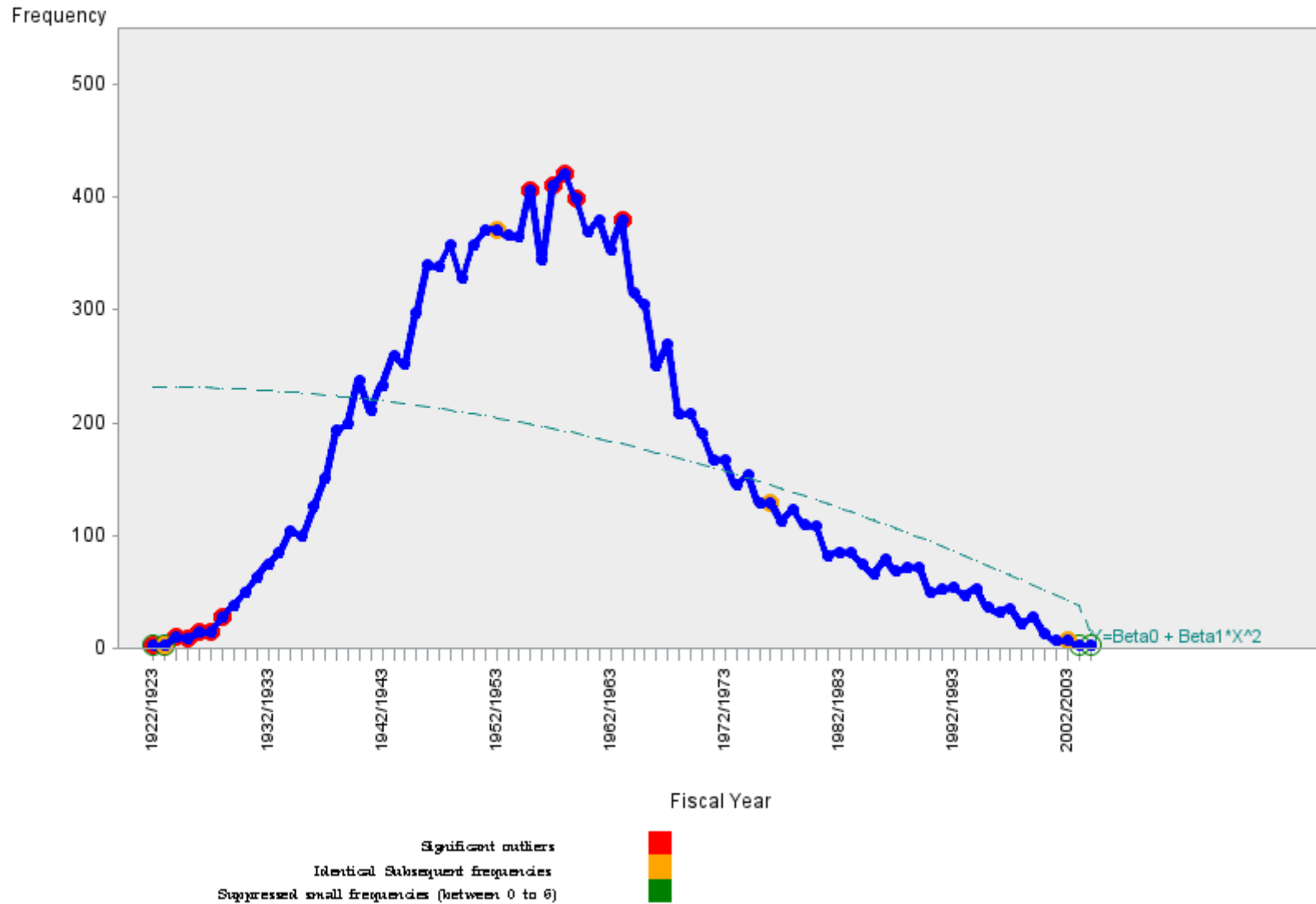


Trend Analysis for Dataset: a_PROJEC.Endouse_oenintak_201501_201901

Date Variable: SCHEDULED_CASE_DT
By Variable=All Records



Trend Analysis for Dataset: a_PROJEC.Endouse_southern_201501_201812
 Date Variable: BIRTH_DT



Linkability

For Southern Health

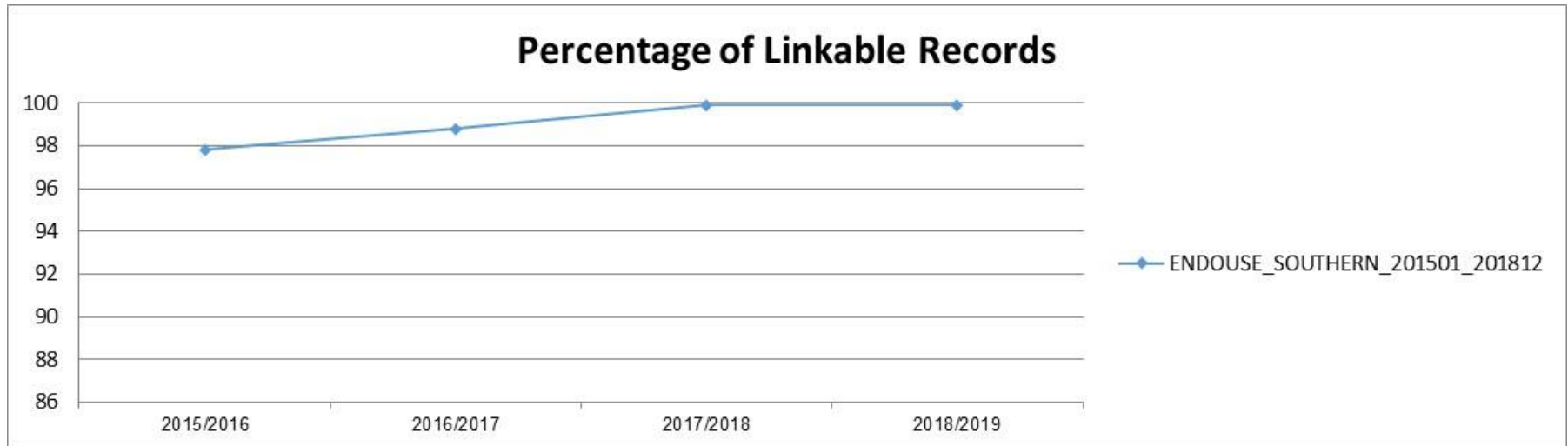
Dataset	Total Number of Records	Number of Linkable Records	% Linkable Records	Number of Linkable Individuals
ENDOUSE_SOUTHERN_201501_201812	13985	13873	99.2	12650

For Central Intake

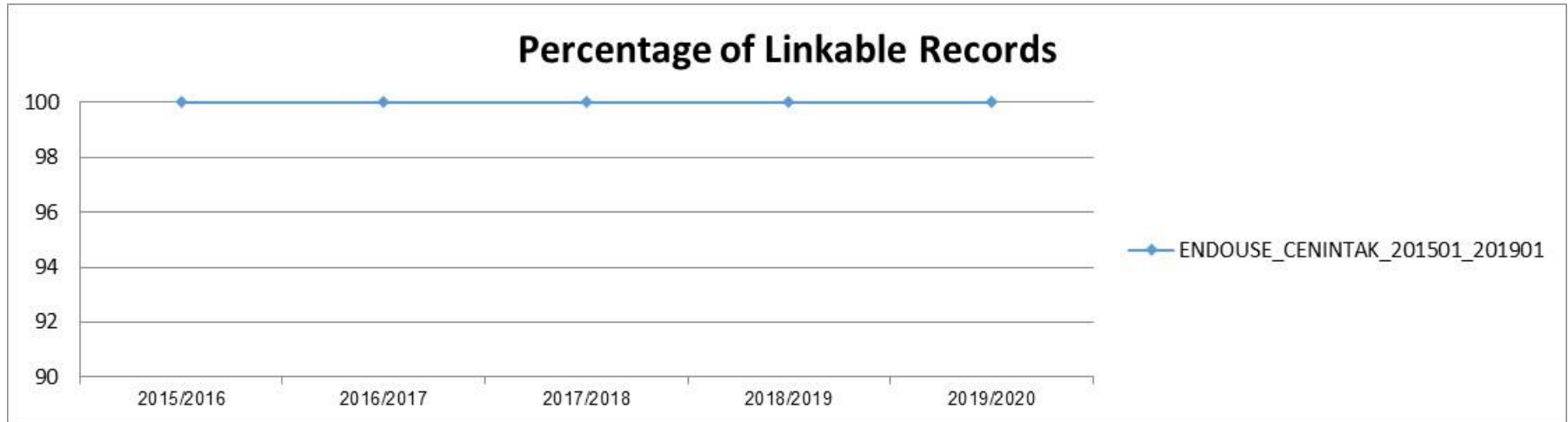
Dataset	Total Number of Records	Number of Linkable Records	% Linkable Records	Number of Linkable Individuals
ENDOUSE_CENINTAK_201501_201901	90707	90707	100	67621

Linkability Over Years

For Southern Health



For Central Intake



Phin Type

For Southern Health

SCRPHINTYPE	ENDOUSE_SOUTHERN_201501_201812
0 MH verified against concurrent registries	99.2
4 MCHP db specific ScrPHIN - No MH found	0.8

For Central Intake

SCRPHINTYPE	ENDOUSE_CENINTAK_201501_201901
0 MH verified against concurrent registries	100