

## Codebook

### Texas APS: Client Data

<b>Dataset name:</b>	codebook_df
<b>Dataset size:</b>	18.8 Mb
<b>Column count:</b>	9
<b>Row count:</b>	378,604
<b>Updated date:</b>	2025-04-04

#### Description:

From pre-processed 'Clients.xlsx' file from Texas APS

#### Column Attributes:

1	Column name:	id		
	Column description:	Cross-set unique subject ID		
	Source information:	DETECT		
	Column type:	Categorical		
	Data type:	Integer		
	Unique non-missing value count:	370,825		
	Missing value count:	0		
	Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
	1	1	279475	7
	2	1	65735	8
	3	1	48360	9
	4	1	284591	12
	5	1	51282	18
2	Column name:	id_aps		

Column description: APS within-set unique subject ID

Source information: DETECT

Column type: Categorical

Data type: Integer

Unique non-missing value count: 370,825

Missing value count: 0

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
51	1	101795	7
80	1	109430	8
81	1	1799	9
82	1	104303	12
83	1	5	18

<b>3</b>	<b>Column name:</b>	<b>aps_client_num</b>
----------	---------------------	-----------------------

Column description: Anonymized Proxy for APS Client ID

Source information: DETECT

Column type: Categorical

Data type: Integer

Unique non-missing value count: 378,418

Missing value count: 0

Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency
1	1	259546	3
2	1	285045	3
3	1	285736	3
4	1	286483	3
5	1	345950	3

<b>4</b>	<b>Column name:</b>	<b>age_90cat</b>
----------	---------------------	------------------

Column description: If subject age anonymized into "90+"

Source information: DETECT

Column type: Categorical

Data type: Logical

Unique non-missing value count: 2  
Missing value count: 28,599

Categories	Frequency	Cumulative Frequency	Percent
FALSE	321,879	321,879	85.02
TRUE	28,126	350,005	7.43
Missing	28,599	378,604	7.55

<b>5</b>	<b>Column name:</b>	<b>dob_year</b>
----------	---------------------	-----------------

Column description: Participant's birth year  
Source information: MedStar ePCR Records  
Column type: Time  
Data type: Numeric  
Unique non-missing value count: 86  
Missing value count: 56,725

Statistic	Value	Frequency	Percentage
Minimum	1933	5,061	1.34
Mode	1947	10,646	2.81
Maximum	2022	2	0.00

<b>6</b>	<b>Column name:</b>	<b>country</b>
----------	---------------------	----------------

Column description: Participant's country of residence  
Source information: Texas APS  
Column type: Categorical  
Data type: Character  
Unique non-missing value count: 8  
Missing value count: 378,466

Categories	Frequency	Cumulative Frequency	Percent
canada	2	2	0.00
great britain	1	3	0.00
honduras	1	4	0.00

lagos	1	5	0.00
mexico	130	135	0.03
peru	1	136	0.00
russia	1	137	0.00
taiwan	1	138	0.00
Missing	378,466	378,604	99.96

7	Column name:	state			
	Column description:	Participant's state of residence, 2 character abbreviation			
	Source information:	Texas APS			
	Column type:	Categorical			
	Data type:	Character			
	Unique non-missing value count:	52			
	Missing value count:	29,025			
Categories with Smallest Values		Frequency	Categories with Largest Values		Frequency
	vt	2		ca	240
	dc	4		la	254
	de	4		ok	273
	sd	4		Missing	29025
	hi	5		tx	346828

8	Column name:	zip3_hipaa		
	Column description:	3 digit HIPAA compliant ZIP Code		
	Source information:	DETECT		
	Column type:	Character		
	Data type:	Character		
	Unique non-missing value count:	608		
	Missing value count:	40,579		
	Categories with Smallest Values	Frequency	Categories with Largest Values	Frequency

009	1	750	15105
012	1	785	15772
017	1	782	26872
019	1	770	35412
031	1	Missing	40579

<b>9</b>	<b>Column name:</b>	<b>pobox</b>
----------	---------------------	--------------

Column description: Was a PO Box address in the original "street address" field?

Source information: DETECT

Column type: Categorical

Data type: Logical

Unique non-missing value count: 2

Missing value count: 0

Categories	Frequency	Cumulative Frequency	Percent
FALSE	377,576	377,576	99.73
TRUE	1,028	378,604	0.27