## Analytics Assignment

Revati Deshpande

Role: Deployment Strategist at Peregrine

10/6/2024

### Agenda

- 1. Problem Statement & Objectives
- 2. Data Description & Preprocessing
  - a. # Records, Features, Key Variables
  - b. Data transformation (handling missing values, encoding, scaling)
- 3. Exploratory Data Analysis
- 4. Methodology
- 5. Model Implementation
- 6. Results
  - a. Anomaly Distribution, Risk Scoring, Case Studies (Deep-dives)
  - b. Evaluation, Validation
  - c. Limitations
- 7. Conclusion & Recommendations



#### **Executive Summary**



#### Comment#1:

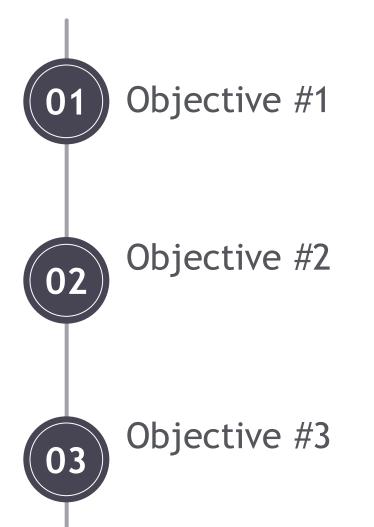
Key Objectives	Metrics



- Insight1
- Insight2
- Insight3
- Insight4

## Problem Statement & Objectives





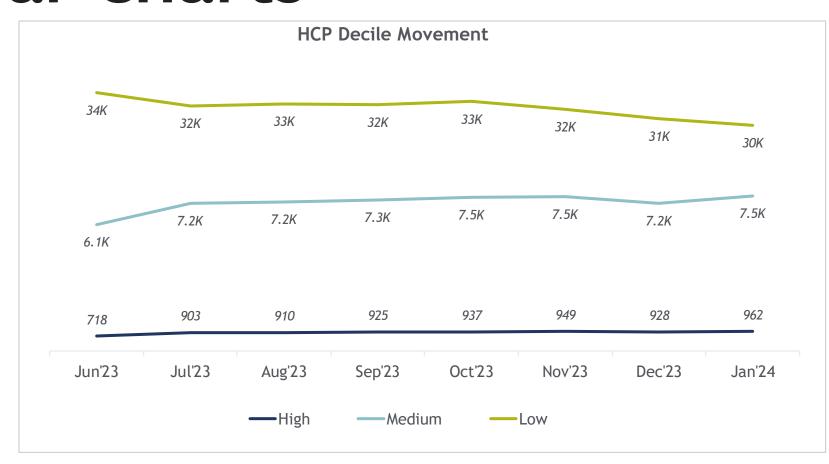
## Data Description & Preprocessing

Drocoss		Super	Insights					
Process		Logistic Regression	Random Forest	IIISIGIILS				
	# Records	78,	600					
	# Features/Columns	1	4					
Description (Raw dataset)	Rinary/Multiclass							
	Handling Missing Values (dropping/imputing)							
	Dropping irrelevant columns							
Preprocessing	Encoding							
	Scaling							
	Feature(s) engineered							
Description	# Records							
(Processed dataset)	# Features/Columns							

## Data Description & Preprocessing

Dwagag		Unsup	Inciabto	
Process		Isolation Forest	Local Outlier Factor	Insights
	# Records	78,	,600	
Description (Days data and)	# Features/Columns	1	14	
(Raw dataset)				
	Handling Missing Values (dropping/imputing)			
	Dropping irrelevant columns			
Preprocessing	Encoding			
	Scaling			
	Feature(s) engineered			
Description	# Records			
(Processed dataset)	# Features/Columns			
	Added Features/Columns			

# **Exploratory Data Analysis - Univariate Analysis Bar charts**



Decile	Jun'23	Jul'23	Aug'23	Sep'23	Oct'23	Nov'23	Dec'23	Jan'24
High	718	185	7	15	12	12	-19	38
Medium	6.1K		0.1K			0.0K	-0.3K	0.5K
Low	34.0K	-1.7K		-0.1K	0.4K		-0.1K	0.3K
No Decile (0,-99)	57.0K	1.3K	0.6K	0.3K	0.6K		-2.5K	3.2K
Total	97.8K	98.6K	99.6K	99.9K	101.0K	100.8K	97.9K	101.9K

#### **温** Insights

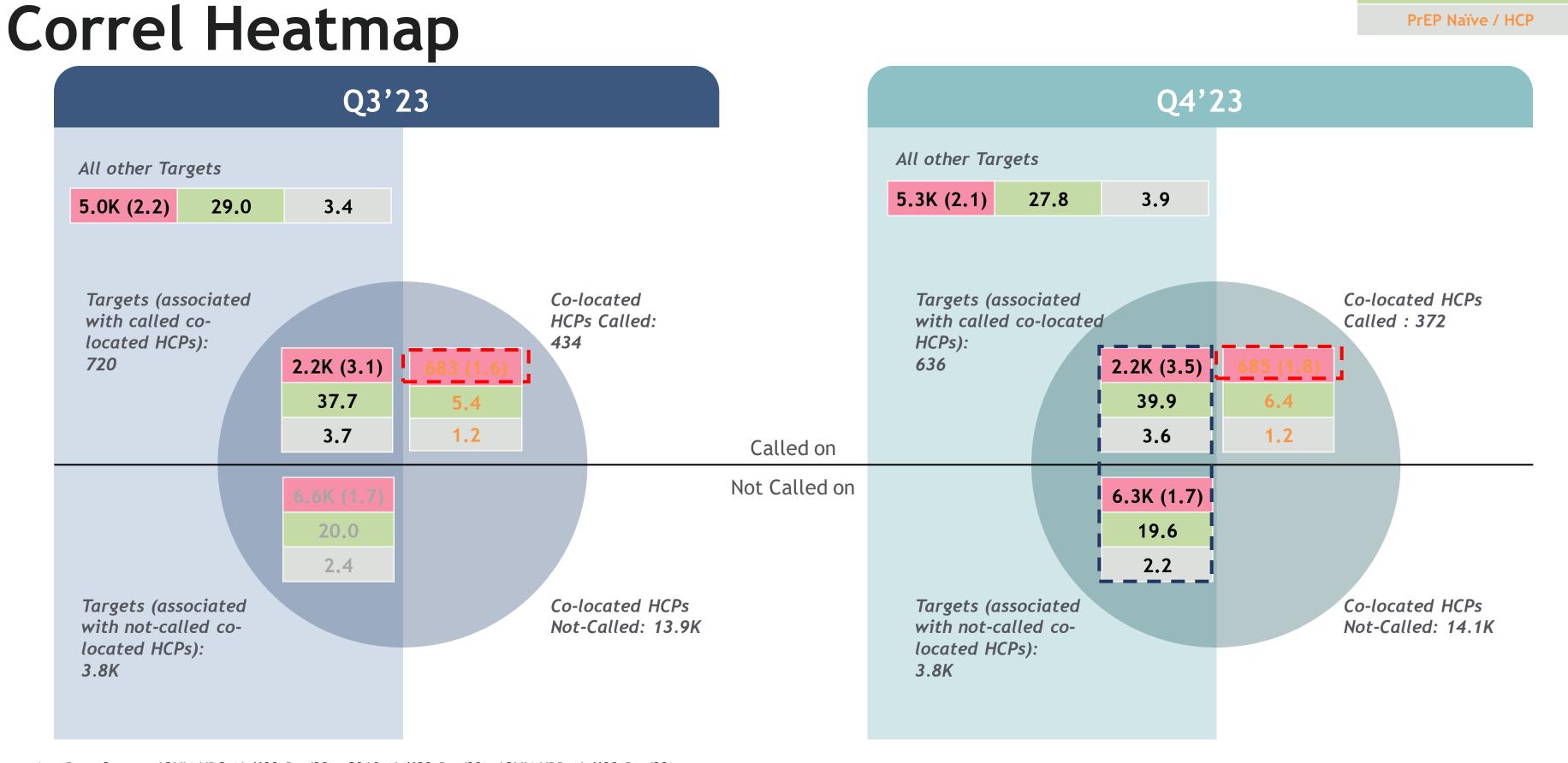
- Insight 1
- Insight 2
- Insight 3
- 1. Footnote 1

Footnote 2

## Exploratory Data Analysis - Multivariate Analysi

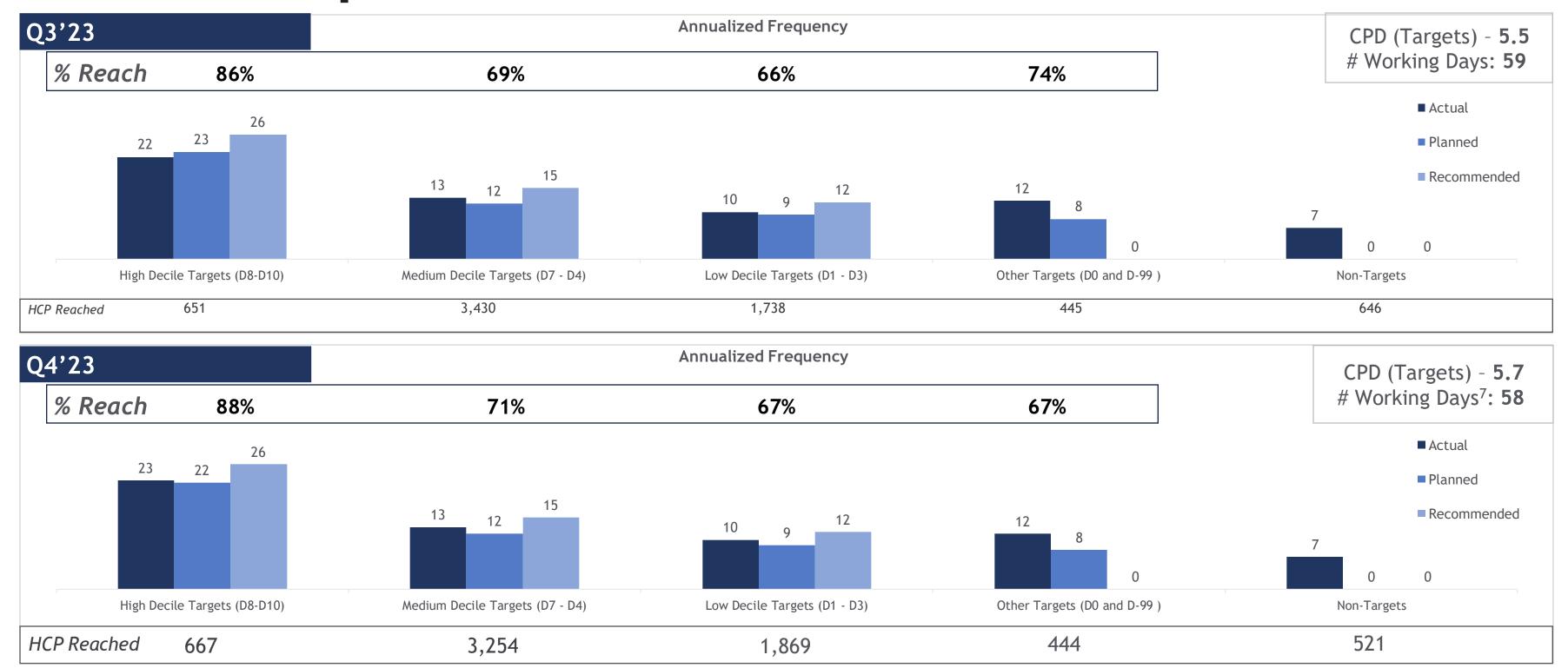
D4P TRx / HCP

PrEP Naïve / HCP



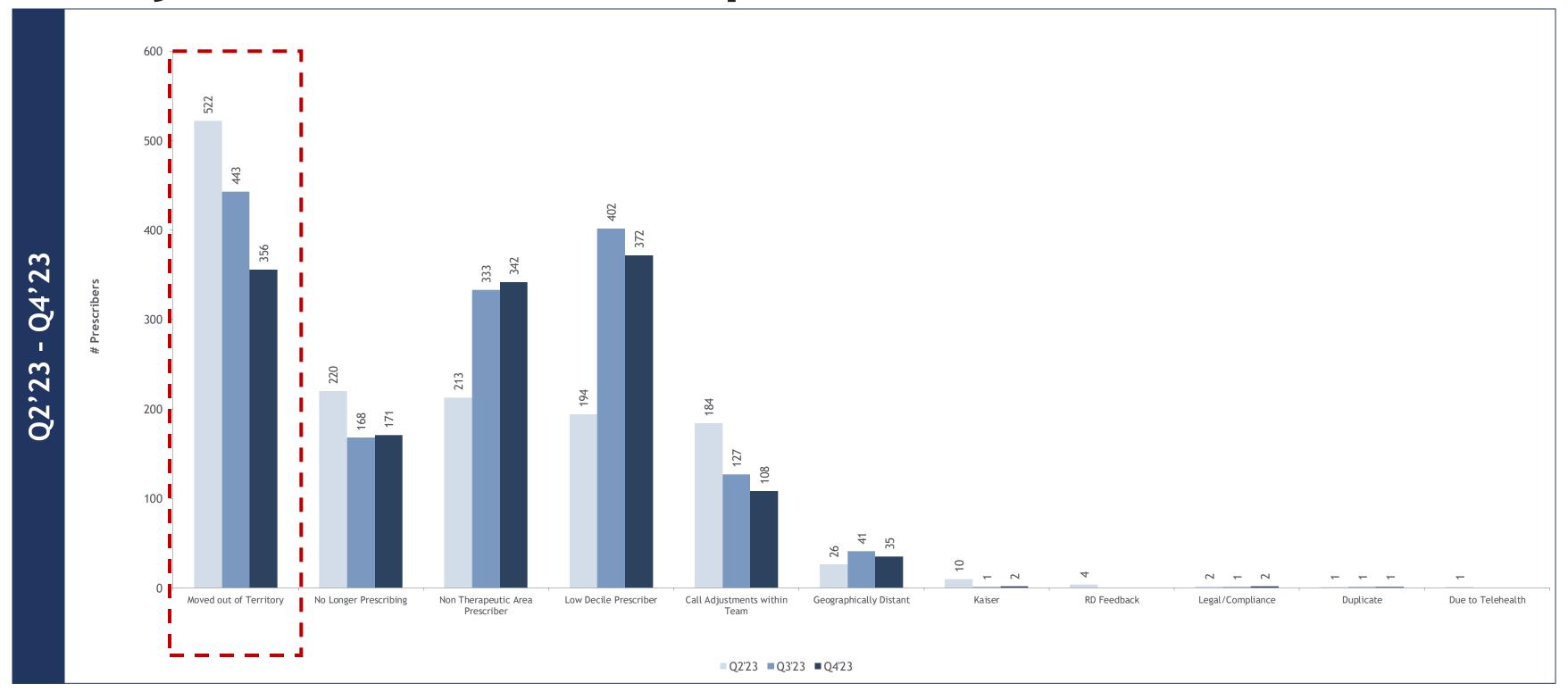
- Data Source: IQVIA XPO (Jul'23-Dec'23); G360 (Jul'23-Dec'23); IQVIA XPD (Jul'23-Dec'23)
- Q4'23 and Q3'23 target Lists are considered
  - 3. Non-Target HCPs having PrEP TRx or have been Called in R12M (till Dec'23) have been considered for Co located HCP universe

# Exploratory Data Analysis - Multivariate Analysis Scatter/Pair plots

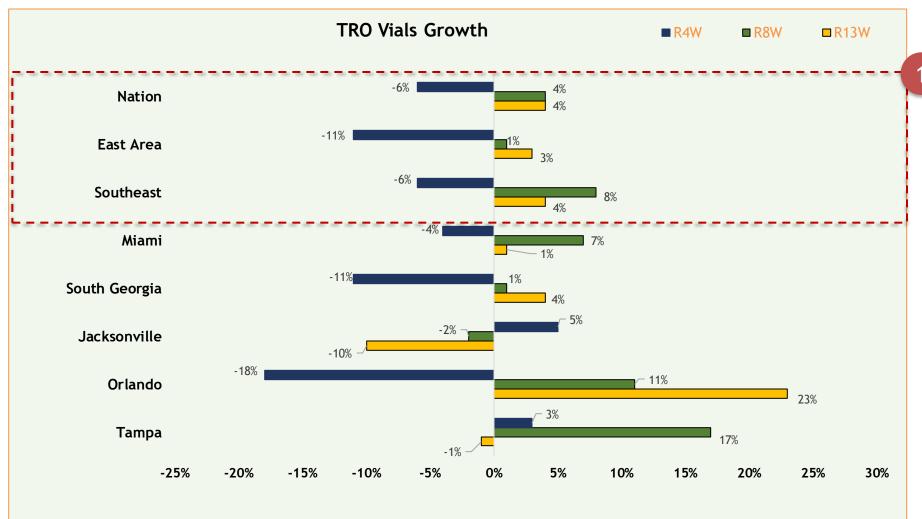


- 1. Time Period Sales on Q4'23 WE 12.29, Calls based on respective quarters.
- 2. Annualized Frequency based on Called HCPs in the bucket
- 3. Vacant Territories have been excluded from Actual Calls/HCP calculation

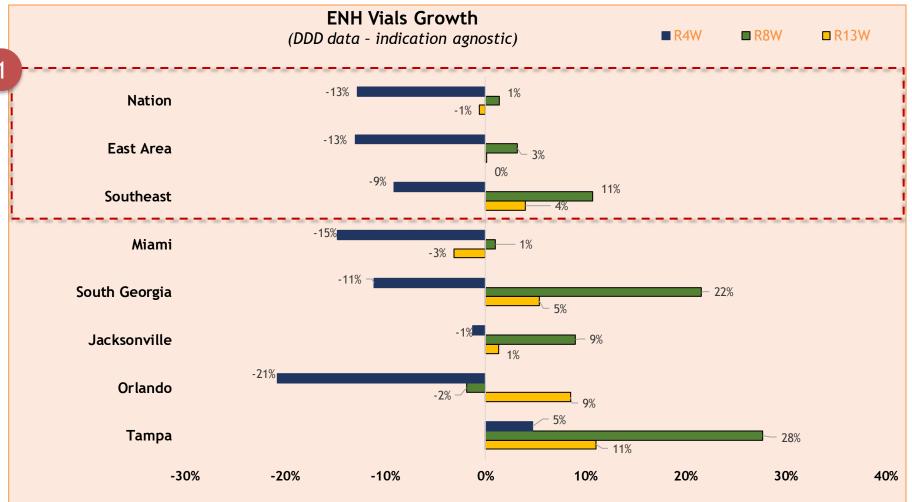
# Exploratory Data Analysis - Multivariate Analysis Side-by-side Scatter/Pair plots



#### **EDA 4 - Comparison Boxplots**

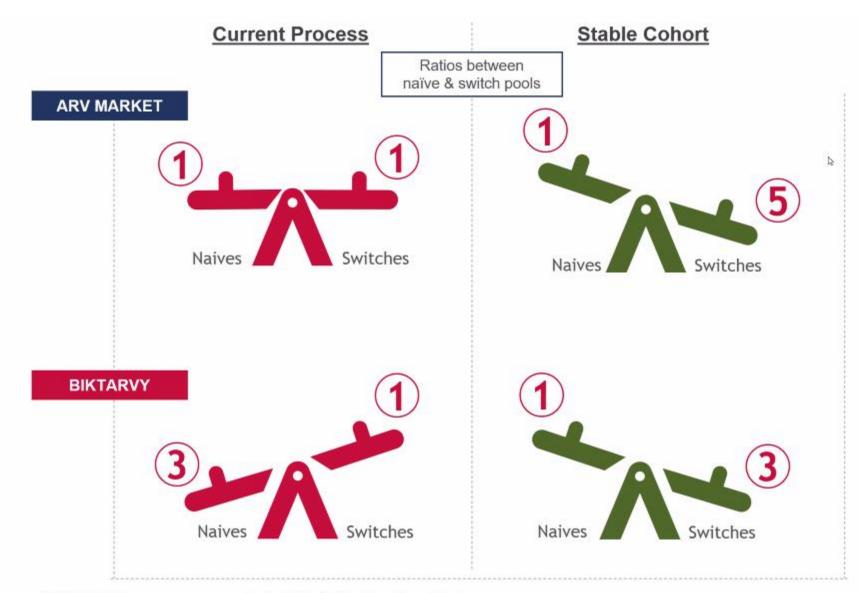


T		% Contribution (R13W)					
Territory's contribution to Region	Q1'24 mBC Potential Contribution	TRO Vials (867 data)	Enhertu Vials (DDD data - indication agnostic)				
East to Nation	51%	52%	50%				
Southeast to East	19%	20%	19%				
Miami	24%	34%	29%				
South Georgia	15%	19%	18%				
Jacksonville	15%	9%	12%				
Orlando	19%	18%	17%				
Tampa	27%	20%	23%				

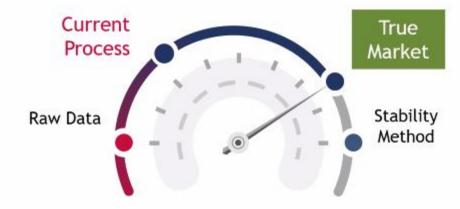


- R13W TRO growth in Southeast is at par with East and Nation, but R13W ENH growth is higher
- 2 Growth opportunity in Jacksonville and Tampa as TRO R13W sales contribution lags ENH and mBC potential contribution

- Source: Trodelvy Sales 867 data till Dec'23 ; Enhertu Sales DDD data (indication agnostic) till Dec'23
- R13W: 10/21 01/19, R8W: 11/25-01/19, R4W: 12/23-01/19
- mBC Potential is calculated as per Q1'24 account segmentation



Bo Bao



#### Stable Cohort Controls for limited longitudinal trackability of HIV patients

- The inherently stable patients tend to be slightly older, and marginally higher in commercial / Medicare
- Given the strong tie-back to epi modeling + CDC numbers for new incidences, we believe the true market to be much closer to one determined with stability method.



### Methodology

#### We recommend xxx

**Completed** high-level testing on the data, taking yy as a **directional metric** to measure impact. Below are our recommendations to the business users:

- ✓ Conduct thorough testing and impact analysis to ensure data and metrics meet requirements.
- ✓ Retrofit processes as needed to align with updated data.
- ✓ Assess impact of updated data.

Completed			
	In Progress		
		Pending	
			Pending

## Model Implementation - Performance Assessment

Itaration	Classes	Model 1							lu ai mh ta			
Iteration		Accuracy	Precision	Recall	F1-score	Support	Accuracy	Precision	Recall	F1-score	Support	Insights
Iteration 1	Low											
	Moderate											
	High											
Iteration 2	Low											
	Moderate											
	High											

<u>a</u> In

Insights

Insight 1

## Model Implementation - Performance Assessment

	n Classes	Model 1					ı			
Iteration										Insights
Iteration 1	Low									
	Moderate									
	High									
Iteration 2	Low									
	Moderate									
	High									

**國** Insights

• Insight 1

### Results

#### <u>国</u> Insights

- Most influential features:
  - Feature importance
- Business rules for classification:
  - BR1
  - BR2

#### Conclusion & Recommendations

#### Conclusion:



## Thank You