

# Agenda Item V: HPD Data Collection Program Updates

Dionne Evans-Dean, MHA, Chief Data Programs Officer,  
Anthony Tapney, MBA, Assistant Branch Chief,  
Anna Dito, Cost Transparency Section Manager,  
Anna Kriuchkova, MSc, and David Winston, MSc, Research Data Specialists, HCAI

# For Today

- Data Collection Status and Activities
- Data Quality and Completeness
  - [HPD Data Completeness Fact Sheet](#) overview
  - Data Quality Reports and Engagement
  - Expansion of Race Categories in the Data Layout
- Status of Voluntary Self-Funded data
- Dental Data File Submission updates

# Data Collection Status

Type	Status
Commercial	Monthly submissions continue
DHCS	Processed through April 2025
Medicare FFS	Annual data acquired 2018-2023 Quarterly data: 2022 + 2023, 2024 Q1-Q3 with Q4 pending
Dental	Monthly submissions continue
Annual Registration	Completed for plans and submitters. Reopened in April and May for plans and submitters to verify NCP data file submission
Non-Claims Payment (NCP) – <i>Upcoming</i>	Regulations adopted and in effect as of March 25, 2025 <a href="#">NCP Data Layout</a> ™, <a href="#">NCP Fact Sheet</a> , <a href="#">Regulation Text</a>
<a href="#">All Payer Claims Database - Common Data Layout (APCD-CDL</a> ™) version 4.0.1 released on February 24, 2025. HCAI preparing regulations package to adopt new version which incorporates NCP file layout. Anticipate adoption by Q1 2026.	

# Data Collection Activities

- Submitter Group Meetings continue quarterly with last meeting held on July 10.
- Data Quality Reports and Engagement continue with health plans and insurers.
- Implementation continues for NCP data collection including hosting webinar trainings to assist plans/submitters.

# HPD Data Completeness Fact Sheet

*Anna Kriuchkova, MSc, Cost Transparency Section, Research Data Specialist*

# HPD Data Completeness: Fact Sheet

- Data completeness is an important aspect of data quality.
  - A high percentage of data completeness does not measure accuracy of the submitted data
  - Represents the degree to which specific fields are populated with values under expected circumstances (numerator) across all records (denominator) for the expected circumstances
- Data is submitted to HPD monthly and at the service line level.
  - There is an eligibility record for every month a member was enrolled with a plan, and a service line record for every non-denied service line on a claim

The HPD Data Completeness Fact Sheet can be found in the resources section of the HPD Data Access and Release Page: [HPD Data Access and Release Resources](#)

## FILES:

ELIGIBILITY

MEDICAL  
CLAIMS

PROVIDER

PHARMACY  
CLAIMS

*\* DENTAL CLAIMS data was unavailable at the time of the analysis.*

# Fact Sheet Excerpt

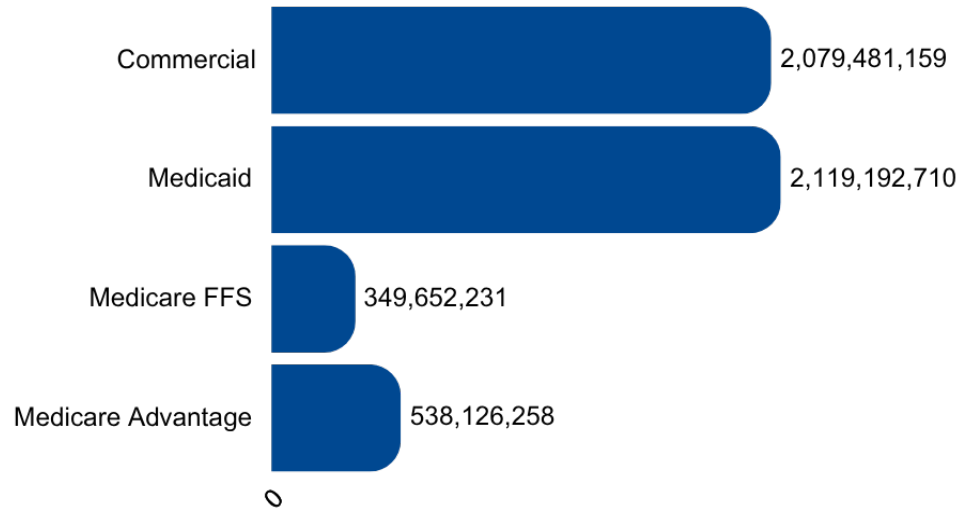
Field	Commercial Range	Commercial	Medi-Cal FFS	Medi-Cal Managed Care	Medicare FFS	Medicare Advantage
Billing Provider NPI	75.9%-100%	97.2%	86%	100%	97.3%	97.2%
Procedure Code (Prof)	99.7%-100%	100%	100%	100%	100%	99.9%
Procedure Code Modifier (1)	16.0%-45.0%	24.8%	48.5%	34.1%	37.2%	25.2%
Date of Service	100%	100%	100%	100%	100%	100%
Rendering Provider NPI	67.6%-100%	97.1%	72.3%	100%	100%	97.3%
Payment Arrangement	100%	100%	100%	100%	100%	100%
Charge Amount (FFS)	86.9%-100%	99.9%	98.4%	N/A	100%	99.8%
Paid Amount (FFS)	93.7%-100%	99.9%	98.5%	N/A	100%	99.8%
Place of Service (Prof)	100%	100%	100%	100%	100%	100%
Type of Bill (Inpatient)	97.8%-100%	99.8%	100%	100%	100%	100%

[Healthcare Payments Database \(HPD\) Data Completeness 2018-2023](#)

# Eligibility File

## Completeness At A Glance

### TOTAL RECORDS BY PAYER TYPE:



Eligibility File: Demographic information for each member eligible for all claim types for one or more days of coverage at any time during the reporting period.

Department of Healthcare Access and Information. [Healthcare Payments Database \(HPD\) Data Completeness 2018 – 2023](#). Updated March 19, 2025

- Personally identifiable information (PII) fields such as member's first and last name, address, social security number, and birth date are excluded from standard limited datasets (SLD and SLD+). PII data is only authorized to eligible entities, such as university researchers, for the purposes of data linkage.
- Completion rates for race, ethnicity, and language are calculated as actionable fields, which exclude "unknown", "invalid", and "missing" values.
- Data on commercial plans is captured as a cumulative average of all records.
- The fact sheet includes the most representative fields in each table and should not be regarded as an exhaustive list.
- Mandatory fields would have a higher completion rate due to a higher incentive of regulatory compliance.



# Medical File

## Completeness At A Glance

Medical File: Service-level claims and remittance information for medical services, including member demographics, provider details, payment amounts, clinical diagnosis codes, and procedure codes.

- Fields that are unavailable for some payer types are due to the nature of their service processing.
- Data on commercial plans is captured as a cumulative average of all records.
- The fact sheet includes the most representative fields in each table and should not be regarded as an exhaustive list.
- Mandatory fields would have a higher completion rate due to a higher incentive of regulatory compliance.

### TOTAL RECORDS BY PAYER TYPE:

Field	Commercial Average	Medi-Cal FFS	Medi-Cal Managed Care	Medicare FFS	Medicare Advantage
Total Records Available	1,950,813,845	866,567,577	1,908,712,324	1,813,466,122	842,626,164
Professional Records	1,517,661,940	409,363,992	1,345,168,557	1,013,889,434	575,771,429
Inpatient Records Available	48,667,985	29,490,782	94,368,423	149,137,939	45,065,512
FFS Records Available	859,136,542	866,567,577	N/A	1,813,466,122	190,725,086

Department of Healthcare Access and Information. [Healthcare Payments Database \(HPD\) Data Completeness 2018 – 2023](#). Updated March 19, 2025

# Pharmacy File

## Completeness At A Glance

Pharmacy File: Service-level claims and remittance information for prescription drug claims, including member demographics, provider details, payment amounts, clinical diagnosis codes, and national drug codes (NDC).

- Fields that are unavailable for some payer types are due to the nature of their service processing.
- Data on commercial plans is captured as a cumulative average of all records.
- The fact sheet includes the most representative fields in each table and should not be regarded as an exhaustive list.
- Mandatory fields would have a higher completion rate due to a higher incentive of regulatory compliance.

### TOTAL RECORDS BY PAYER TYPE:

Field	Commercial Average	Medi-Cal FFS	Medi-Cal Managed Care	Medicare FFS	Medicare Advantage
Total Records Available	667,053,656	344,690,864	391,941,661	563,918,922	394,476,462
FFS Records Available	667,041,907	344,690,864	N/A	563,918,922	394,435,734

Department of Healthcare Access and Information. [Healthcare Payments Database \(HPD\) Data Completeness 2018 – 2023](#). Updated March 19, 2025

# Provider File Completeness At A Glance

- Provider File: Demographic and network information about providers associated with eligibility and claims data for a given reporting period.
- Total records: 121,829,559.
- This file has an extended timeline: July 1, 2017 – June 30, 2024 (all preceding files explore data collected between January 1, 2018, and December 31, 2023).
- The completeness measurement does not analyze the completion of fields by payer type for this file since the same provider can be present in data from multiple types of payers.
- HPD does not collect provider files from CMS; HPD cross-references provider data received from other sources to identify providers included on CMS Medicare FFS claims data.
- This document includes the most representative fields in each table and should not be regarded as an exhaustive list.
- Mandatory fields would have a higher completion rate due to a higher incentive of regulatory compliance.

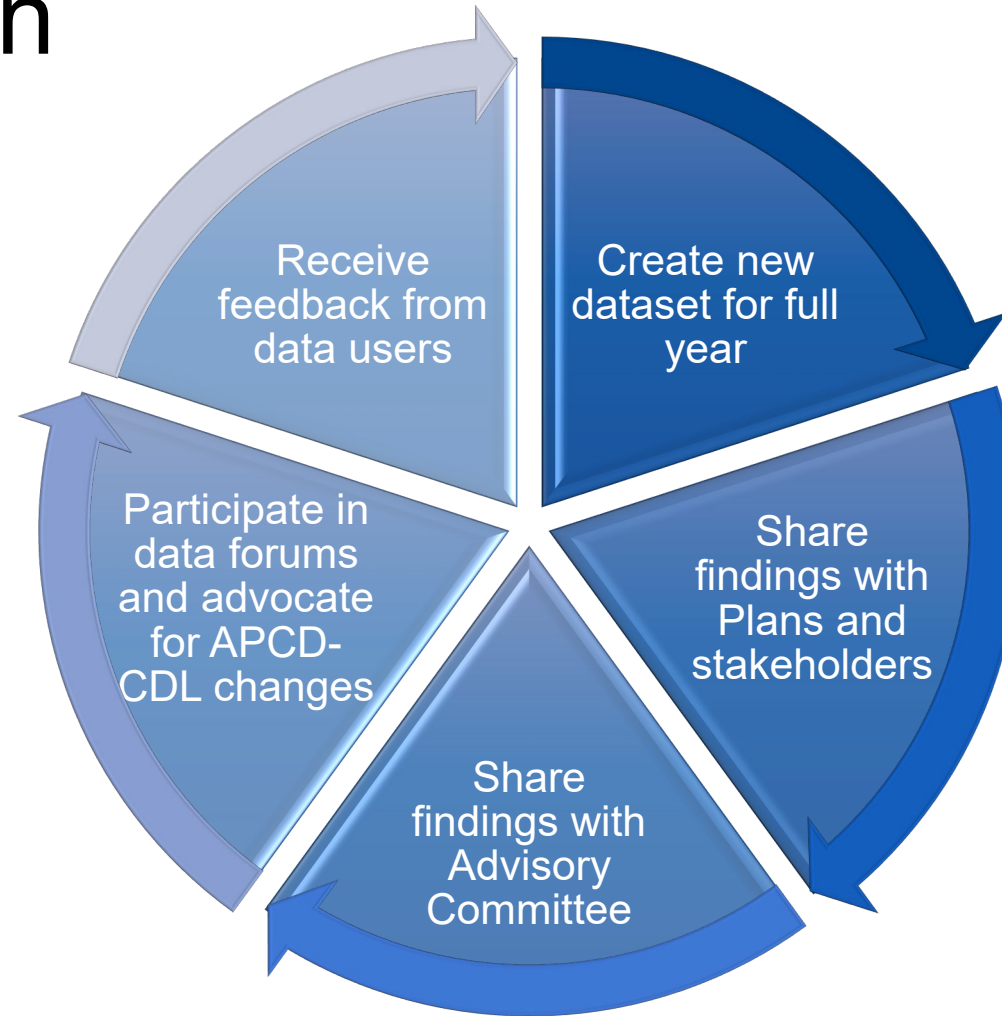
Department of Healthcare Access and Information. [Healthcare Payments Database \(HPD\) Data Completeness 2018 – 2023](#). Updated March 19, 2025

# Data Quality Reports and Engagement

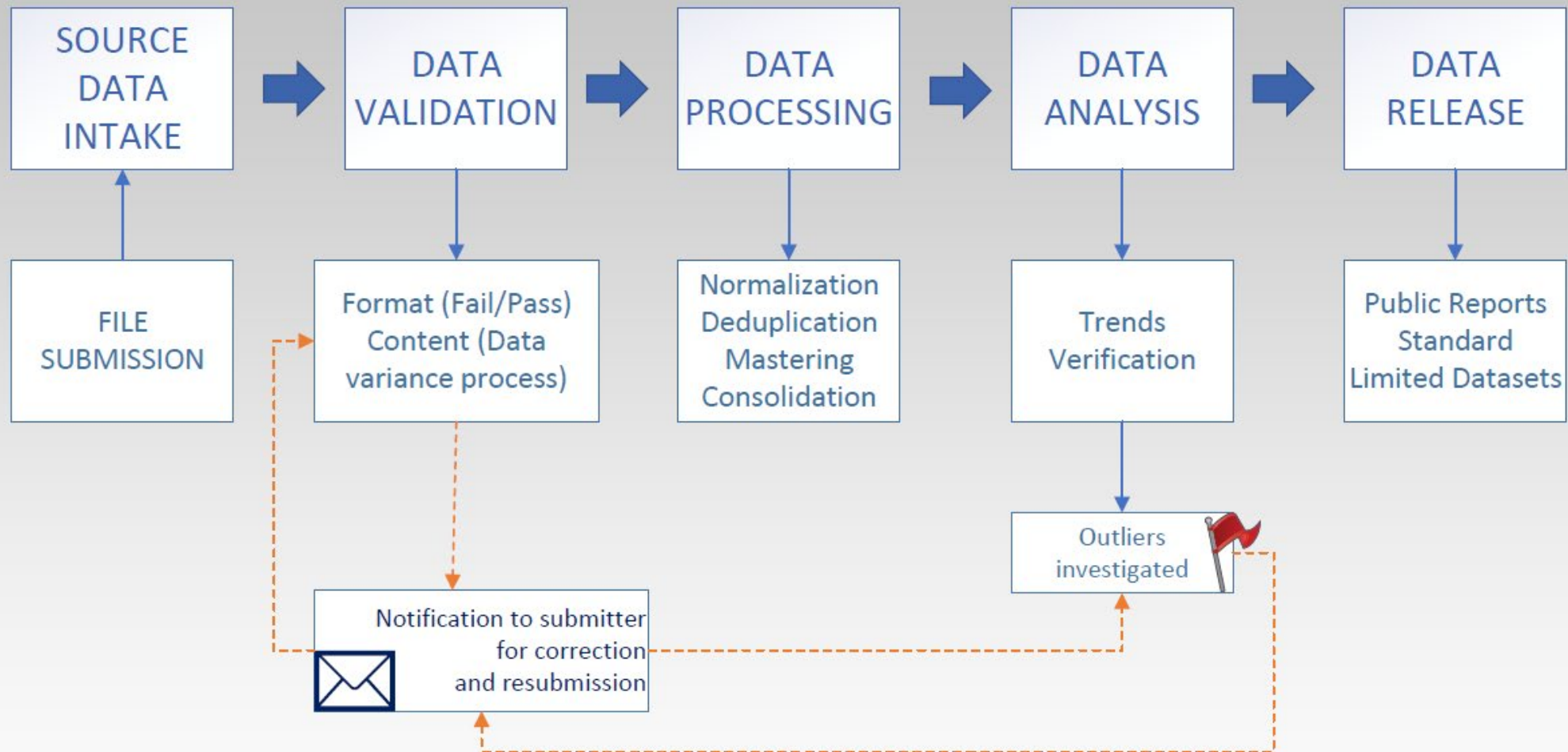
*Anna Dito, Cost Transparency Section Manager, HCAI*

*David Winston, MSc, Research Data Specialist, HCAI*

# Intended Approach for Ongoing Data Quality Collaboration

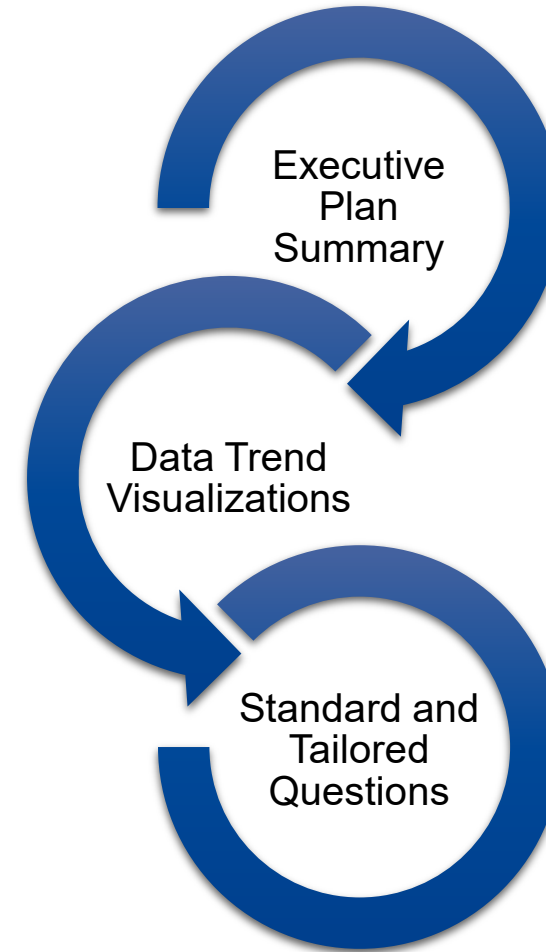


# HPD DATA COLLECTION PROCESS

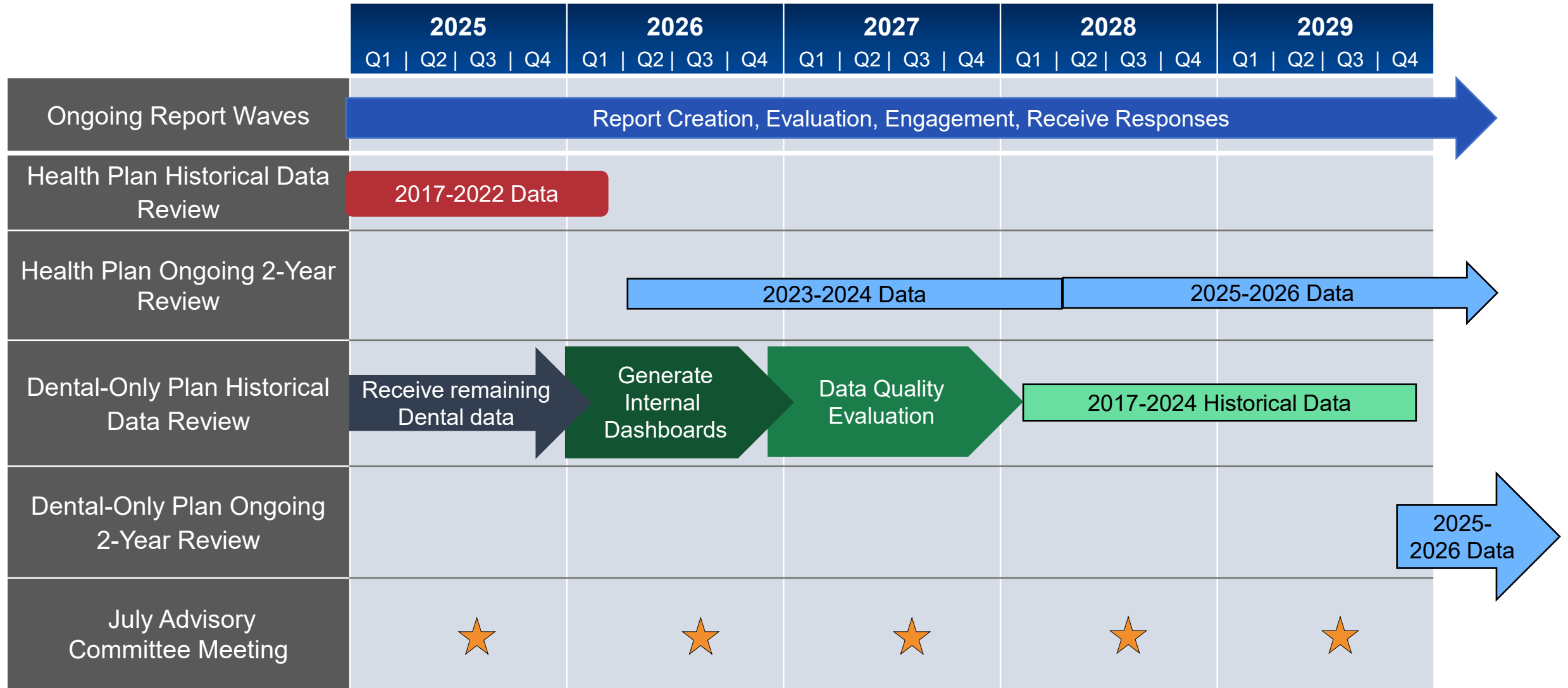


# Data Quality Reports (DQR) and Engagement

- Initiated reports and engagement with plans in May 2024
- Opportunity for plans to validate data in HPD, feedback loop with plans to share experiences and best practices
- Developed standard and tailored questions for each plan to gain further insight on data quality



# Data Quality Report Roadmap





# What Are We Evaluating?

- Validating covered lives counts for 2017-2022. Assessing historical data completeness for 2020-2022.
- Data has a time lag to assess full years. Upon completion of historical analyses, the next engagement planned for 2023-2024 years.
- Questions are designed to share the plan's data in HPD and gather more information on:
  - Data Completeness
  - Methods of data collection
  - Implementation of demographic data tracking
  - Preference of aggregation for public reporting on data completion.

# Evaluating Commercial Data Completeness of Key Metrics (2018-2023)

MEMBER ELIGIBILITY FILE	MEDICAL CLAIMS FILE (Services)	MEDICAL CLAIMS FILE (Claims/Encounter)	PHARMACY CLAIMS FILE
Race	Principal Diagnosis	Rendering Provider NPI	NDC Drug Code
Ethnicity	Procedure Code	Payment Arrangement	Pharmacy NPI
Language	Procedure Modifier 1	Capitated Claims with non-zero Allowed Amount	Prescribing NPI
	Revenue Code	Service Date	
		Place of Service	
		Type of Bill	

	Very High Completion Rate
	Somewhat High Completion Rate
	Lower Completion Rate

Source: [Healthcare Payments Database \(HPD\)](#)  
[Data Completeness Fact Sheet 2018-2023](#)

# Data Collection Method – Race, Ethnicity, and Language, 2020-2022

<u>Demographic Element</u>	<u>Collection Method</u>	<u>Plan Size</u>		
		Small	Medium	Large
Ethnicity	Self-Reported			1
	Combination	4	2	1
Language	Self-Reported	1		
	Combination	3	2	1
Race	Imputed			1
	Self-Reported			1
	Combination	4	2	1

- Most of the 8 plans responding, of all sizes, use a **combination of approaches** to obtain race, ethnicity, and language data
  - Self-report: information entered directly by patient
  - Imputation: information populated based on other sources, such as provider electronic medical records
  - Reported on someone's behalf: authorized representative of the patient provided the information
- Plurality of data directly from patient
- Next most common was imputed from medical records

# Race and Ethnicity Data Separation, 2020-2022

- The APCD-CDL requires that race and ethnicity fields are reported separately.
- Most plans will have separate race and ethnicity reporting by the end of this year. Of the 10 plans responding:
  - Five plans are collecting race/ethnicity separately as of 2024 and one more is on track to separate the data this year
  - Three other plans have made progress but will not have separate data by year end
  - Two plans are not collecting race/ethnicity data

Responses from 10 Plans (multiple responses allowed)	Total	Small	Medium	Large
Collected separately	3	0	1	2
Combined for 2020-22, separate for 2024	2	1	1	0
Combined field with separate tracking	2	2	0	0
Data partially separated	1	1	0	0
Not collecting	2	0	2	0

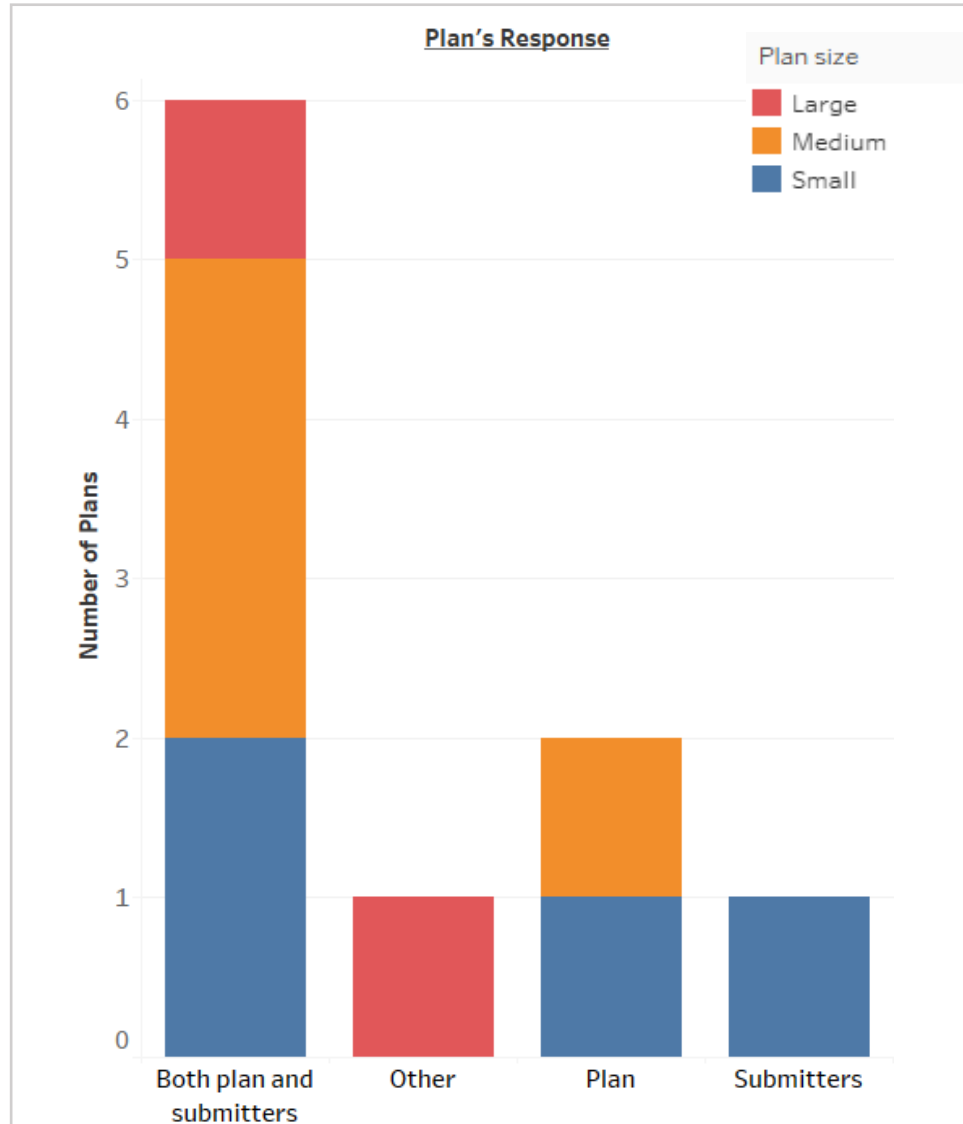
# Reasons for Missing or Incomplete Data, 2020-2022

N=10 Plan Responses (multiple responses allowed)	Race	Ethnicity	Language
Not a required field	4	4	4
Data integration issue	3	3	2
Combined R/E fields	2	1	N/A
No race code for Hispanic / Latino / South American	1	N/A	N/A
Use Race data when Ethnicity data is missing	N/A	2	N/A
Ethnicity data does not include option for “Not Hispanic or Latino”	N/A	1	N/A
Did not collect	2	2	0
Average completion	46%	68%	89%

Among the 10 plans that responded with reasons that race, ethnicity, and/or language data is missing or incomplete, the main explanations were:

- Data is not available: the plan does not collect the data or the fields are optional for members
- Challenges with data integration: data comes from multiple systems that are difficult to synchronize
- Challenges with data collection approach: race and ethnicity fields are combined, fields do not align with common data layout

# Plans' Preference for Public Completeness Data Aggregation



Plans were asked about their preference of data aggregation for public reporting on data completeness.

- Aggregation method options were:
  - Plan
  - Submitters
  - Plan and Submitters
  - Other
- Most plans preferred showing completeness data at both plan and submitter level, regardless of plan size.
  - Second most popular choice was at plan level
  - One small plan recommended to share data at the submitter level.
  - One plan selected "other" to suggest rolling data up to higher plan level.

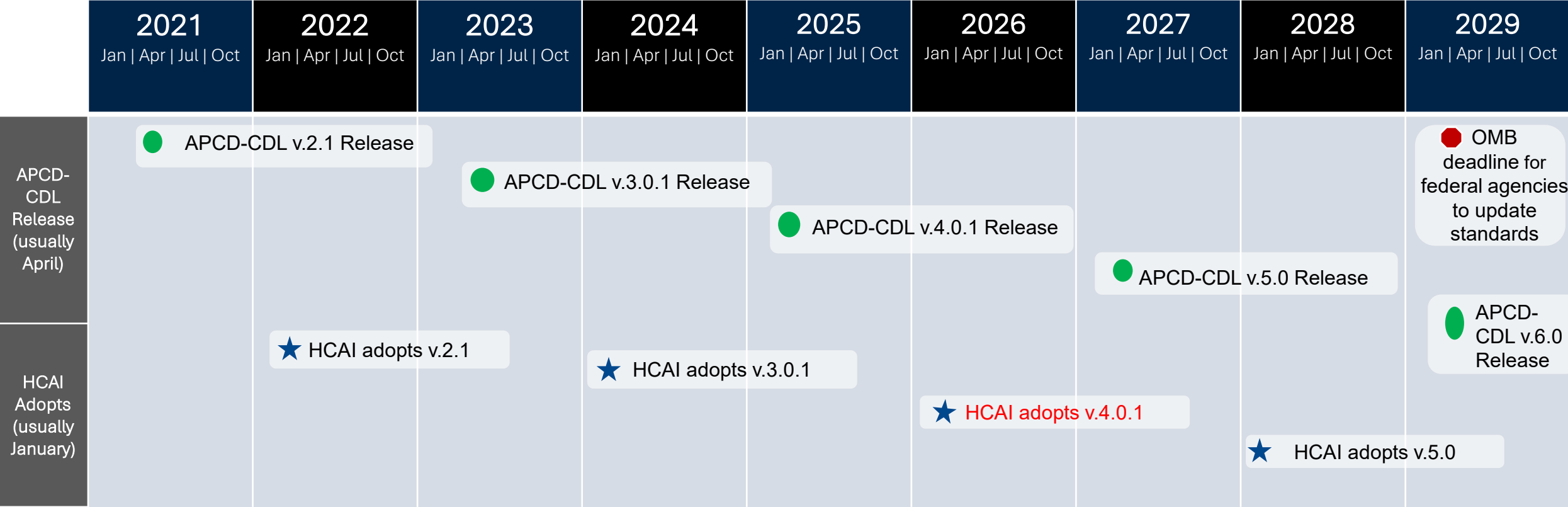
# APCD-CDL v3.0.1

## Preliminary Results of Expansion of Race Categories

*Anna Kriuchkova, MSc, Cost Transparency Section, Research  
Data Specialist*

# APCD-CDL Data Collection Timeline

- APCD-CDL release and adoption is on a 2-year cycle (exception: annual releases in 2023 and 2024)
- HCAI updates regulations prior to adoption, so adoption lags release.
- Two cycles remain before the OMB deadline.





# Race Category Reporting Among Commercial Plans: 2023 (6 categories) vs. 2024 (146 categories)

## METHODOLOGY

Comparative analysis of race category reporting for commercial submitters in 2023 (old race categorization with 6 groups) vs 2024 (post expanding to 146 categories). Primary focus was on actionable race field completeness rates, as well as percentage of missing and unknown values reported for the years in question. Note that actionable rates exclude missing, invalid, and unknown values.

Plans have been stratified by number of reported covered lives:

- Large: 1,000,000 +
- Medium: 100,000 – 999, 999
- Small: under 100,000

## OBSERVATIONS

- Two large plans (from a cohort of 9 plans) and two medium plans (from the cohort of 11 plans) have shown the most tangible improvement in actionable race field completeness rates, increasing their numbers by 45-50+ percents.
- Increase in completeness is aligned with their use of more detailed race categories.
- No significant changes have been observed for small plans.
- 3 large plans have reported over 70 race categories.

# Comparative Analysis Of Race Category Reporting Among Commercial Plans

## (2023-2024)

2023-2024 Actionable Completeness/Missing Values/Unknown Values  
Distribution for Race



As illustrated by the graphics to the left, there is a direct correlation between introduction of detailed race categories and improvement in reporting quality for the race field. Large and medium plans have shown rapid increases in actionable completeness rates and drops in missing and unknown values. We are anticipating this trend to become more pronounced as we receive more data from submitters.

# Status of Voluntary Self-Funded Data Collection

*David Winston, MSc, Cost Transparency Section, Research Data  
Specialist*

# ERISA Self-Funded Participation is Low

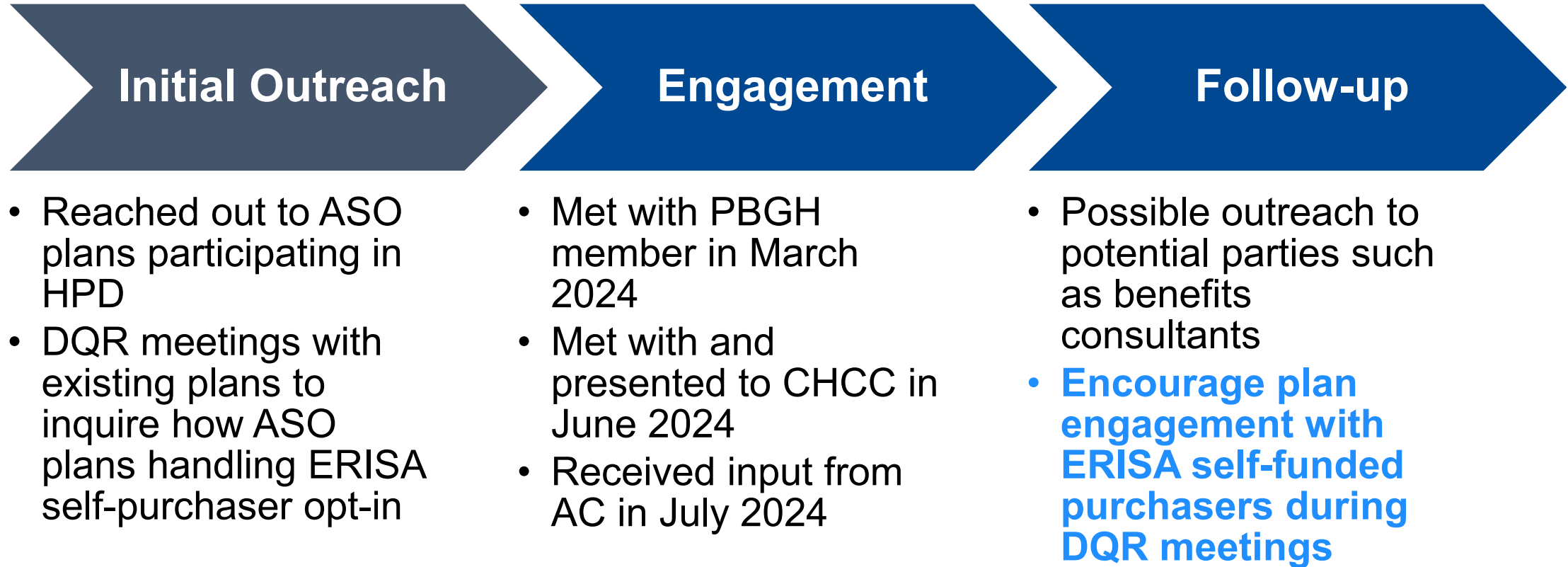
Estimate of Self-Funded Lives, in Millions

	2021	2022	2023
<b>State-Wide</b>			
ERISA	4.3	4.5	4.6
Non-ERISA	1.2	1.3	1.4
<b>Total, State-Wide</b>	<b>5.5</b>	<b>5.8</b>	<b>6.0</b>
<b>In HPD System</b>			
ERISA	0.2	0.3	0.3
Percent of State-Wide ERISA	5%	7%	6.5%
Non-ERISA	0.8	0.8	0.9
<b>Total, HPD System</b>	<b>1.0</b>	<b>1.1</b>	<b>1.2</b>

- No definitive figures on the number of ERISA self-funded lives in CA – but estimated to be 4.6 million
- Challenging to estimate ERISA vs. non-ERISA in HPD
- Estimated: ~300k (6.5%) of 4.6 million ERISA self-funded lives are in HPD 2023 data

Source for 2021-2022 data: Department of Health Care Access and Information, [Health Care Payments Data Program Report](#), March 2024.

# Outreach Plans and Progress



# Results of Outreach (Ongoing)



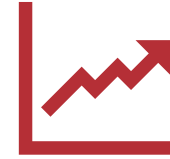
## **Difficulties in collecting ERISA data:**

- As of 2023, HPD has received just 6.5% covered lives of ERISA data
- Relies on private self-funded plans opting-in to sharing HPD data
  - Often requires employers to pay to opt-in to sharing data to HPD



## **Of the 6 plans with ERISA data, two have responded**

- Waiting on further responses to analyze data trends



## **Potential improvements in ERISA data collection:**

- HCAI's DQR outreach meetings with submitters highlight the value in collecting this data

# Dental Data File Submission Updates

*Anna Dito, Cost Transparency Section, Manager*

# Dental Data File Submission Status

- Two types of dental data file submitters;
  - 21 health plans with 8 delegated dental submitters
  - 21 dental-only plans, associated with 25 delegated dental submitters
- Majority of dental data submitters are in production, and have sent historical and ongoing monthly data
- Next steps – evaluate data completeness and quality of dental files



# Dental-Only Plan Status

Dental Plan Summary Status	Total Number of Plans	Total Covered Lives *
Required to submit Dental Only Data	21	8,173,560
2025 Registration Complete	17	7,879,800
Submitted Test Data	14	7,642,142
Submitted Production Data	13	7,287,065
Outreached related to member threshold	7	617,172
Responded to Outreach	6	540,696
Zero response to Outreach	1	59,828

\*Total Covered Lives count is based on the Department of Managed Health (DMHC) and the California Department of Insurance (CDI) data for the month of December 2023.

# Public Comment