Key Facts for Working with the PHD Data

Update V13

-

Update V12

- Added additional information on Gobeille's impact on the APCD

Update V11

- Moved dataset specific key facts into the brief dataset description

Update V10

- Added more information on OEND data
- Added in a grid of ICD and Procedure code fields in Case Mix

Update V9:

- Added information on MOUD dataset, revised MOUD documentation

Key Facts for Working with the PHD Data

1. The PHD includes data on Massachusetts residents only.
   - While MA residency cannot be easily established in the APCD files, there is a flag on the final demographic dataset (which is a person-level file) to indicate people who are in the APCD who we do not believe are MA residents. We strongly suggest dropping these people from all analyses, as they will not have data available in any other dataset.
2. Due to the court ruling in Gobeille v. Liberty Mutual, self-funded insurance plans are no longer required to report to the APCD starting in 2016. A self-funded (or self-insured) plan is an insurance arrangement in which the employer assumes direct financial responsibility for the costs of enrollees' medical claims. Employers sponsoring self-funded plans typically contract with a third-party administrator or insurer to provide administrative services for the self-funded plan. Motivation to self-insure includes being able to manage claims themselves and saving money by not paying exorbitant rates for businesses deemed risky by insurers.
   - In Massachusetts at the end of 2017, approximately 1.75 million self-insured beneficiaries (that is, 75% of the self-insured) were no longer in the MA APCD. Overall, CHIA has experienced a 42% reduction in all claims submitted to the MA APCD. Looking at the 2020 APCD data this trend has continued, with only about 18% of the original self-insured beneficiaries remaining reported to the APCD. To determine if a service in the APCD was paid for by a self-insured plan, you can utilize DENT_ENROLL_TYPE for dental claims, PHARM_ENROLL_TYPE for pharmacy claims, or MED_ENROLL_TYPE for medical claims. A value of 2 indicates a self-insured plan.
   - Gobeille continues to have an impact on the APCD insurance records in 2023. In 2015 (pre-Gobeille), 20% of all member eligibility records were for individuals with a self-insured insurance type. In 2016 (the start of Gobeille), only 7% of all member eligibility records were for individuals with a self-insured insurance type and by 2019 that number fell to 5%. As of 2023, only 3.9% of member eligibility records are for individuals with a self-insured insurance type.
   - For a detailed analysis of the impact of Gobeille v. Liberty Mutual on MA's APCD data, please visit https://www.chiamass.gov/assets/Uploads/DEMOGRAPHIC-IMPACT-OF-GOBEILLE.pptx and https://www.chiamass.gov/assets/Uploads/User-Workgroup-February-2022.pptx
   - Pre-Gobeille, the CHIA's "Sweet Sixteen" (i.e., the 16 insurance providers that submitted the most data) was led by a private payer; post-Gobeille, CHIA's "Sweet Sixteen" is led by MassHealth due to the loss of private claims
3. Impact of Gobeille on PHD's spine:
   - Within the PHD, everyone is identified by a unique ID which is maintained across years and datasets. This unique ID is based on CHIA's APCD Master Patient Index (MPI), which assigns a single unique surrogate key to each person. To be in the APCD MPI, an individual must be seen in the APCD records at least once at any point in time.
   - Because of the Gobeille ruling, self-funded insurance plans were no longer required to report to state APCDs (previously reporting was compulsory).
   - If an individual has been seen in the APCD at least once, then they will be represented in the APCD MPI. For example, those on self-funded plans before 2016 may no longer have

Key Facts for Working with the PHD Data

their claims in the APCD, but they do have an APCD MPI that can be used to create PHD's ID. However, individuals who are new to the state and have only ever had a self-insured plan that is not reporting to APCD will not be in the APCD MPI and will not have a PHD ID. Without a PHD ID, an individual and their records and not maintained in the PHD.

4. Records where the identifiers did not match to a person in the spine are dropped from the PHD files. Matching rates by dataset are in the document "**PHD_Datasets_Brief Descriptions**"

5. Match Level Breakdown (please note, matching statistics apply at the individual dataset level):
    ο Group 1 – These are members within the spine that received a high score against the DPH input record. Candidates in this group are those that perfectly matched all elements of the DPH input record and those that generally had only one element mismatch. The threshold applied to group 1 candidates has been optimized based upon the input elements and their overall weight in determining a unique individual match.
    ο Group 2 – This group contains members within the spine that received a lower score against the DPH input record yet are considered *possible* matches. Candidates in this group could match on as little as SSN or date of birth and first and last name. These candidates are included for completeness.

6. All dates have been masked. Individuals have different masking numbers. This means you <u>cannot</u> compare the masked dates between individuals (i.e., two people could have an event on the same date, but their masked dates will be different numbers in the PHD).
    ο Within an individual, all dates in the PHD have been treated with the same masking date proxy. This means that all events for an individual can be arranged chronologically. Within an individual's timeline, the larger the masked date, the later in time (i.e., closer to today) the event occurred.
    ο To create time cohorts across individuals, use the month and year variables in datasets.

7. Suppression of values between 1-10. The PHD cell suppression policy is not to publish or presenting tables with cell sizes less than 11 to anyone who is not an authorized user of the data. Values of 0 CAN be published.

8. Datasets that do not start in 2011:
    ο HIV – starts in 2013
    ο MATRIS – starts in 2013
    ο APCD – starts in 2014
    ο Toxicology – starts in 2014
    ο DMH RAP – starts in 2016
    ο BSAS HOC -- starts in 2019
    ο COVID MIIS – starts in 2020
    ο COVID MAVEN – starts in 2020

9. Search the PHD for any variable or name – use the **PHD2.0 Analytic Data Dictionaries.pdf - Edit-Find** use the magnifying glass to enter the search term.

10. **Demographic Variables.** The PHD has race, sex and age variables for almost every dataset. Researchers can determine how they chose the primary value based on the objectives of their research. There are variables in the Spine: RaceEth_Many and Sex_Many to flag individuals who have different values across the PHD. In addition, we have added in final variables created with a tiering methodology to reduce processing time. You do not need to use these variables, and

Key Facts for Working with the PHD Data

we encourage you to resolve these demographic variables as best makes sense to your group. These variables are FINAL_RE and FINAL_SEX.

11. When working with ICD codes in any of the PHD datasets, please remember that most medical institutions in the US transitioned from using ICD 9 to ICD 10 on October 1, 2015. V codes in ICD 9 are also used in ICD 10 but do not mean the same thing – make sure to consider the date (by searching for ICD9 before 10.1.2025 and ICD10 After 10.1.2015) if you are working with ICD and V codes.
    - Death records in the US have been implementing ICD 10 codes since January 1, 1999.
    - All ICD 9 and 10 diagnosis code variables in the PHD (found in the following datasets: APCD, CaseMix, Deaths, DMH, EI_DiagnosesTable, Fetal Deaths, and MATRIS) do not contain decimals.

12. Identify dual eligible individuals in the PHD:
    - Dual eligibles are individuals who receive both Medicare and Medicaid benefits
    - The two programs cover many of the same services, but Medicare pays first for the Medicare-covered services that are also covered by Medicaid
    - To find these individuals you must use the PHDAPCD.MHEE file: they are those where MHEE_PopCategoryCat=1 and (MHEE_MedicarePtAFlag=1 OR MHEE_MedicareAdvantageFlag =1 OR MHEE_MedicarePtBFlag=1)
    - Regarding claims data, when Medicare pays for Medicare services for dually eligible members, many of those claims are then forwarded by Medicare to MassHealth for potential additional payment to the providers. This is because Medicare Fee For Service (FFS) does not include member cost sharing, but duals are eligible for having that cost sharing covered by MassHealth (e.g., $100 claim where Medicare pays $80 and the member cost sharing is $20 – MassHealth will pay up to $20 on behalf of the member). Most of these claims will end up paying $0, particularly when Medicare 80% coverage rate is already more than the MassHealth payment rate for the service. But the utilization will be in the MassHealth FFS claims data for most Medicare services (notable exceptions are Hospice and some home health). For SCO and One Care, there should be no separate Medicare claims. PACE members have no claims at all.

APCD
APCD File Linkage Tables

Key Facts for Working with the PHD Data

	APCD Files Linkage Table		
	Claims to Individual Products or Providers		
Dataset Name A	**Variable Name A**	**Dataset Name B**	**Variable Name B**
To link the ME Monthly file (A) to the ME Full file (B), use the following:			
PHDAPCD.ME_MTH	ME_MEMELGID & ME_SUBCONTROLID	PHDAPCD.ME	ME_MEMELGID & ME_SUBCONTROLID
To link claims data or Member Eligibility (A) to the APCD PRODUCT data (B) at the Individual Insurance Product Level (i.e., linking claims to the correct product that covered that claim) use the following:			
PHDAPCD.DENTAL	DENT_LINKORGIDPR & DENT_PRODUCT_LINKID	PHDAPCD.PRODUCT	PROD_ORGID & PROD_PRODUCT_LINKID
PHDAPCD.MEDICAL	MED_LINKORGIDPR & MED_PRODUCT_LINKID	PHDAPCD.PRODUCT	PROD_ORGID & PROD_PRODUCT_LINKID
PHDAPCD.PHARMACY	PHARM_LINKORGIDPR & PHARM_PRODUCT_LINKID	PHDAPCD.PRODUCT	PROD_ORGID & PROD_PRODUCT_LINKID
PHDAPCD.ME	ME_LINKORGIDPR & ME_PRODUCT_LINKID	PHDAPCD.PRODUCT	PROD_ORGID & PROD_PRODUCT_LINKID
To link claims data (A) to the APCD PROVIDER data (B) at the Individual Provider Level (i.e., linking claims to the correct provider that covered that claim as categorized) use the following:			
PHDAPCD.DENTAL	DENT_LINKORGIDPV & DENT_SERVICEPROVIDER_LINKID	PHDAPCD.PROVIDER	PROV_ORGID & PROV_PROVIDER_LINKID
PHDAPCD.MEDICAL	MED_LINKORGIDPV & MED_BILLINGPROVIDER_LINKID	PHDAPCD.PROVIDER	PROV_ORGID & PROV_PROVIDER_LINKID
PHDAPCD.MEDICAL	MED_LINKORGIDPV & MED_RENDERINGPROVIDER_LINKID	PHDAPCD.PROVIDER	PROV_ORGID & PROV_PROVIDER_LINKID
PHDAPCD.MEDICAL	MED_LINKORGIDPV & MED_SERVICEPROVIDER_LINKID	PHDAPCD.PROVIDER	PROV_ORGID & PROV_PROVIDER_LINKID
PHDAPCD.PHARMACY	PHARM_LINKORGIDPV & PHARM_PRESCRIBER_LINKID	PHDAPCD.PROVIDER	PROV_ORGID & PROV_PROVIDER_LINKID
PHDAPCD.PHARMACY	PHARM_LINKORGIDPV & PHARM_RECIPIENTPCP_LINKID	PHDAPCD.PROVIDER	PROV_ORGID & PROV_PROVIDER_LINKID
PHDAPCD.MHEE	MHEE_BehavHlthProvider_LINKID	PHDAPCD.PROVIDER (where PROV_ORGID = 3156)	PROV_PROVIDER_LINKID
PHDAPCD.MHEE	MHEE_LTCProvider_LINKID	PHDAPCD.PROVIDER (where PROV_ORGID = 3156)	PROV_PROVIDER_LINKID
PHDAPCD.MHEE	MHEE_ManagedCareProvider_LINKID	PHDAPCD.PROVIDER (where PROV_ORGID = 3156)	PROV_PROVIDER_LINKID
PHDAPCD.MHEE	MHEE_PrimaryCareProvider_LINKID	PHDAPCD.PROVIDER (where PROV_ORGID = 3156)	PROV_PROVIDER_LINKID

Key Facts for Working with the PHD Data

o

Case Mix

- o For acute care hospitalizations, Case Mix is the best source of data – after Gobeille impacted APCD, Case Mix reports about a 55% higher monthly discharge volume than APCD does for inpatient hospital stays
- o In Case Mix, newborns always have a separate discharge record from their birthing parent. In APCD, there are instances where the newborn's claims are rolled into their birthing parent's claims (i.e., the newborn does not get separate claims).

Dataset	Diagnosis code fields	Procedure code fields - ICD	Procedure code fields - CPT/HCPCS
ED	ED_DIAG1, ED_PRINCIPLE_ECODE		
ED_DIAG	ED_DIAG		
ED_PROC		ED_PROC	ED_PROC
HD	HD_DIAG1, HD_ECODE		
HD_DIAG	HD_DIAG	HD_PROC1	
HD_PROC		HD_PROC	HD_PROC

Key Facts for Working with the PHD Data

OO	OO_DIAG1-OO_DIAG16, OO_PrincipalExternal _CauseCode	OO_PROC1-OO_PROC4	OO_CPT1-OO_CPT10

Finding MOUD

- o Analyzing medications for opioid use disorder (MOUD). Full documentation and variables can be found posted on the PHD Teams External Channel (location: Files -> OUD & MOUD -> MOUD variables.xlsx). Alternatively, you may want to use the PHDSPINE.MOUD dataset
  - Buprenorphine
    - i. PMP
    - ii. APCD medical records (procedure codes for buprenorphine/MAT administration)
    - iii. APCD Pharmacy records
    - iv. HOC – BSAS HOC treatment data
  - Methadone
    - v. APCD medical records (procedure codes for methadone administration for OUD). Please note any methadone in APCD pharmacy records is for pain.
    - vi. BSAS treatment records (use METHADONE_BSAS, not the program names to find methadone)
    - vii. HOC – BSAS HOC treatment data
    - viii. NOTE: Methadone prescriptions found in PMP are for pain (not addiction treatment) and those in tox may be for pain
  - Naltrexone
    - i. APCD pharmacy records
    - ii. APCD medical records (procedure codes for naltrexone administration for OUD).
    - iii. DOC – flag for MATRI (MATRI is a treatment program where naltrexone is given, Please note MATRI offers wraparound services, linking clients with services, and MAT. Not all MATRI clients receive MAT.)
    - iv. HOC – BSAS HOC treatment data