Key Facts for Working with the PHD Data

- 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.
- Records where the identifiers did not match to a person in the APCD are dropped from the PHD files. For matching rates by dataset please see the document "PHD_Datasets_Brief Descriptions"
- Match Level Breakdown (please note, matching statistics apply at the dataset level):
    - Group 1 – These are members within the APCD 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.
    - Group 2 – This group contains members within the APCD 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 name. These candidates are included for completeness.
- 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).
    - 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.
    - To create time cohorts across individuals, use the month and year variables in datasets.
- Datasets that do not cover the full PHD time range of 2011-2019:
    - MATRIS – starts in 2013
    - Toxicology – starts in 2015
- 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.
- Analyzing medications for opioid use disorder
    - Buprenorphine
        - PMP
        - BSAS
        - Toxicology
    - Naltrexone
        - APCD pharmacy records
    - Methadone
        - APCD medical records (procedure codes for methadone administration)
        - APCD pharmacy records
        - BSAS
        - Toxicology

- - DOC flag for pre-release treatment program
  - NOTE: Methadone prescriptions found in PMP are for pain (not addiction treatment)
- Military and Veteran data
  - Data from VA hospitals are not contained in Case Mix hospital data
  - APCD data does not cover TRICARE and the Veterans Health Administration data
  - Prescriptions filled via the VA system were not reported to PMP prior to 1/1/2014
- Tramadol was added to the PMP on 8/18/2014. Gabapentin records are available in PHD data starting 1/1/2018.
- Case Mix hospital data does not cover hospital services rendered to Massachusetts residents at out of state hospitals nor any care rendered at behavioral health hospitals.
- Massachusetts' birth certificate data underwent a major system revision in February 2011; please refer to the document "PHD_Datasets_Brief Descriptions" when analyzing birth certificate data to understand the impacts on specific variables.
- The EI system underwent a major system revision in 2018; please refer to the brief dataset description when analyzing EI data to understand the impacts on specific variables.
- The Vitals Information Partnership (VIP) system is designed to streamline and integrate vital event registration, securely, across the Commonwealth. The deaths application in VIP was launched in November 2014; please refer to the brief dataset description when analyzing death certificate data to understand the impacts on specific variables.
- Donahue Population files: The sum of individual race/ethnicity/age breakdowns will not be equal to the "All races/all ethnicities" or the "total" (if selected) category for a given geography for 2010-2020. "Unknown" race is included in the total, but not in the sub-groups, which is the primary cause of the discrepancy between the total race and the individual races summed. Furthermore, rounding contributes to some of the discrepancy as well, and this is what causes all differences between Hispanic + Non-Hispanic and "All Ethnicities" for the same race group.
- WIC_MOM and WIC_KID files are linked using KID_ID_WIC and MOM_ID_WIC
- In the APCD datasets, not every value submitted has a translated meaning – this is because Insurance Carriers do not always submit data as requested by CHIA. In those cases you will find the following note in the variable 's meta data in the data dictionary - "***For any other value not contained in the list above – those values are as is submitted by the insurance carrier (with unknown translation) ***"

APCD Files Linkage Table

Claims to Insurance Carriers

Dataset Name A	Variable Name A	Dataset Name B	Variable Name B
To link claims data (A) to the APCD PROVIDER data (B) at the Insurance Carrier Level (i.e. linking claims to the correct carrier) use the following:			
APCDPHD.DENTAL	DENT_LINKORGIDPV	APCDPHD.PROVIDER	PROV_ORGID
APCDPHD.MEDICAL	MED_LINKORGIDPV	APCDPHD.PROVIDER	PROV_ORGID
APCDPHD.PHARMACY	PHARM_LINKORGIDPV	APCDPHD.PROVIDER	PROV_ORGID
To link claims data (A) to the APCD PRODUCT data (B) at the Insurance Carrier Level (i.e. linking claims to the correct carrier) use the following:			
APCDPHD.DENTAL	DENT_LINKORGIDPR	APCDPHD.PRODUCT	PROD_ORGID
APCDPHD.MEDICAL	MED_LINKORGIDPR	APCDPHD.PRODUCT	PROD_ORGID
APCDPHD.PHARMACY	PHARM_LINKORGIDPR	APCDPHD.PRODUCT	PROD_ORGID

Claims to Individual Products or Providers

Dataset Name A	Variable Name A	Dataset Name B	Variable Name B
To link claims data (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:			
APCDPHD.DENTAL	DENT_PRODUCT_LINKID	APCDPHD.PRODUCT	PROD_PRODUCT_LINKID
APCDPHD.MEDICAL	MED_PRODUCT_LINKID	APCDPHD.PRODUCT	PROD_PRODUCT_LINKID
APCDPHD.PHARMACY	PHARM_PRODUCT_LINKID	APCDPHD.PRODUCT	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:			
APCDPHD.DENTAL	DENT_SERVICEPROVIDER_LINKID	APCDPHD.PROVIDER	PROV_PROVIDER_LINKID
APCDPHD.MHEE	MHEE_BehavHlthProvider_LINKID	APCDPHD.PROVIDER	PROV_PROVIDER_LINKID
APCDPHD.MHEE	MHEE_LTCProvider_LINKID	APCDPHD.PROVIDER	PROV_PROVIDER_LINKID
APCDPHD.MHEE	MHEE_ManagedCareProvider_LINKID	APCDPHD.PROVIDER	PROV_PROVIDER_LINKID
APCDPHD.MHEE	MHEE_PrimaryCareProvider_LINKID	APCDPHD.PROVIDER	PROV_PROVIDER_LINKID
APCDPHD.MEDICAL	MED_BILLINGPROVIDER_LINKID	APCDPHD.PROVIDER	PROV_PROVIDER_LINKID
APCDPHD.MEDICAL	MED_RENDERINGPROVIDER_LINKID	APCDPHD.PROVIDER	PROV_PROVIDER_LINKID
APCDPHD.MEDICAL	MED_SERVICEPROVIDER_LINKID	APCDPHD.PROVIDER	PROV_PROVIDER_LINKID
APCDPHD.PHARMACY	PHARM_PRESCRIBER_LINKID	APCDPHD.PROVIDER	PROV_PROVIDER_LINKID
APCDPHD.PHARMACY	PHARM_RECIPIENTPCP_LINKID	APCDPHD.PROVIDER	PROV_PROVIDER_LINKID