

# Documentation for Updates to AOU Full HIPPS Episodes, June-September 2025

By Elma Hsieh

Supervised by Dr. Caitlin Dreisbach

## Changes Made:

*\*Not including slight modifications such as tweaking merging logic to accept new variables, adjusting datetime logic, or CSV file naming.*

- In HIP.py: added trimester information
- In PPS.py: added trimester information
- Created utilities.py: created helper functions to engineer additional variables
- In HIP\_PPS\_Merge.py: integrated helper functions from utilities.py
- In HIPPS.py: added concordance score calculator function from utilities.py

## Expected Outputs from Running Full\_HIPPS\_Episodes.ipynb:

- From HIP.py
  - HIP\_trimester\_events.csv
  - HIP\_trimester\_summary.csv
  - HIP\_trimester\_event\_details.csv
- From PPS.py
  - PPS\_trimester\_events.csv
  - PPS\_trimester\_summary.csv
  - PPS\_trimester\_event\_details.csv
- From Full\_HIPPS\_Episodes.py
  - HIPPS\_Pregnancy\_Episodes.csv

## Additional Variables

1. parity
2. Visit (information in HIP & PPS Trimester CSVs)
3. age\_at\_delivery
4. delivery\_method
5. smoking\_status
6. Trimester (information in HIP & PPS Trimester CSVs)
7. survey\_confirmed\_pregnancy
8. Outcome\_concordance\_score

## Additional Files

1. utilities.py: helper functions
2. Delivery\_Method\_Map\_07232025.csv (version as of July 23rd, 2025): concept IDs mapped to vaginal delivery or cesarean section.

## Explaining utilities.py (where additional variables are engineered)

*All variables are created after the final dataframes are created in each file. For example, trimester information is integrated after getting 'final\_episodes\_w\_length' dataframe in HIP.py; concordance score is calculated after getting 'final\_episodes' dataframe in HIPPS.py; other variables are created after getting 'singles' dataframe in HIP\_PPS\_Merge.py.*

Functions	How they contribute + How they work
label_trimester_events (used in HIP.py and PPS.py)	Links clinical visits ('visit_table') to pregnancy episodes ('episodes_df') by person and episode, checks if visits fall within the estimated start and end of the pregnancy, computes gestational days, and assigns trimester labels. Returns raw trimester-level events and a per-episode summary (counts per trimester).
summarize_trimester_events_by_concept (used in HIP.py and PPS.py)	Provides a more detailed summary of trimester events, grouped by person, episode, trimester, and clinical concept/domain. Useful for analyzing what types of clinical events occur in each trimester.
pull_pregnancy_status (used in HIP_PPS_Merge.py)	Extracts pregnancy-related status information from survey data - patient answers "currently pregnant (concept ID 4299535)" in their Social Determinants of Health Survey.  The concept ID was cross-checked using <a href="#">Athena</a> .
flag_currently_pregnant (used in HIP_PPS_Merge.py)	Of the surveys pulled, flag patients who completed the survey during their pregnancy episode to validate their pregnancy response.
pull_smoking_status (used in HIP_PPS_Merge.py)	Extracts smoking information from Social Determinants of Health survey.
merge_smoking_status (used in HIP_PPS_Merge.py)	Joins smoking status survey responses onto pregnancy episodes, aligning smoking info with each episode.
pull_demographics (used in HIP_PPS_Merge.py)	Pulls DOB. Used later for calculating maternal age at pregnancy end.

add_age_at_merged_end (used in HIP_PPS_Merge.py)	Uses the pulled date of birth to calculate age at 'merged_end' (episode end). Adds this as 'age_at_delivery'.
add_parity_column (used in HIP_PPS_Merge.py)	Computes total number of pregnancies.
pull_delivery_method (used in HIP_PPS_Merge.py)	Pulls EHR records for delivery methods (vaginal or C-section) using concept IDs. See Delivery_Method_Map_07232025.csv for concept IDs used.
merge_delivery_method (used in HIP_PPS_Merge.py)	Merges delivery method data back into the episodes dataframe, aligning mode of delivery with each episode.
calculate_concordance_score (called in HIPPS.py)	Compares HIP vs. PPS detection: assigns 1 point if outcomes match within $\pm 14$ days, and 1 point if gestational age is plausible. Produces a 0–2 concordance score column.

## Interpreting Outcome Concordance Scores:

- **Derived from Jones, et al.'s *Who is Pregnant?*** research paper (bottom right of page 5).
- Score ranges from 0~2
  - **0** = no concordance (neither outcome/timing nor gestational plausibility align)
  - **1** = partial concordance (either outcome/timing *or* gestational plausibility aligns)
  - **2** = strong concordance (both align)

Category	Description and how it's calculated in our algorithm
Timing & Outcome Match (1 point)  PPS detected the same outcome as HIP within $\pm 14$ days.	In the algorithm, we observe the difference between 'HIP_end' and 'PPS_outcome_date' $\pm 14$ days  <b>AND</b>  The 'outcome_match' is TRUE
Gestational Age Plausibility (1 point)  Checked against biologically plausible gestational lengths.	'min_term' $\leq$ 'gestational_age' $\leq$ 'max_term'

## About Delivery\_Method\_Map\_07233035.csv

- This is the first version of the Delivery Method Map, created on July 23rd, 2025
- This dataset is compiled through gathering concept domains via keyword searches
  - How to search concepts in domain: enter your workspace > datasets > click “+” sign next to “select concept sets”
  - Keywords used in the search: vaginal, vaginal delivery, c-section, cesarean, cesarean section
  - Concept domain results are then filtered to only keep concept IDs that explicitly or implicitly indicate relation to vaginal delivery or cesarean section
- The resulting dataset is saved as a CSV and passed into the code for data pulling
- Used in ‘pull\_delivery\_method’ function in utilities.py