POEM TX Cohort Pull

2024-08-13

# Record Counts (pre-Episode Algorithm Logic)

Using the steps laid out in the Episode Algorithm tab of the Data Harmonization Tracker, the count of records (FY19-23 files, filtering starting 2019-01-01):

| **step** | **description** | **record\_count** |
| --- | --- | --- |
| 1 | All DX LB/SB records | 2,043,043 |
| 2 | After removing contradictory records | 2,033,821 |
| 3 | After removing records associated with bundles | 2,032,816 |

* A record refers to a unique combination of [client\_nbr + clm\_to\_date + outcome\_type (LB/SB)]
* Contradictory record is a record such that the same person has a record that same day with a different outcome type (LB vs SB)
* The bundles are series of claims that would typically upset the algorithm and be flagged as false positives. They consist of cases where an individual has a pregnancy code repeating every month for a span of six-months or greater. They would generally need to be inspected on a case by case basis to see what date to assign. UT usually removes these from pregnancy-related analyses.

# All Episodes

After running the algorithm, we find a possible 987489 episodes (N=817916 distinct ids). This is before joining to any sort of enrollment table, so it may contain placeholder ids or those that never had enrollment at any time (no way to verify age, etc.).

If we condition, as we have to eventually, on enrollment (at any point), we get 986477 episodes (N=816927 distinct ids).

### Distribution of cohort criteria (pre-filtering/pre-exclusion):

| **Characteristic** | **N = 986,4771** |
| --- | --- |
| Birth Outcome |  |
| LB | 977,398 (99%) |
| SB | 9,079 (0.9%) |
| Year |  |
| 2019 | 213,038 (22%) |
| 2020 | 208,910 (21%) |
| 2021 | 206,761 (21%) |
| 2022 | 216,405 (22%) |
| 2023 | 141,352 (14%) |
| 2024 | 11 (<0.1%) |
| Age Group |  |
| 0-14 | 1,447 (0.1%) |
| 15-24 | 357,798 (36%) |
| 25-34 | 499,007 (51%) |
| 35-44 | 126,408 (13%) |
| 45-54 | 1,760 (0.2%) |
| 55-64 | 38 (<0.1%) |
| 65-74 | 13 (<0.1%) |
| 75+ | 6 (<0.1%) |
| State (Month of birth episode) |  |
| Non-TX | 4,196 (0.4%) |
| TX | 980,779 (99%) |
| Unknown | 1,502 (0.2%) |
| *1n (%)* | |

* Non-TX is according to UDS zip to ZCTA crosswalk, which includes more zips than the HUD crosswalks (PO Boxes, in particular)
* 0 - 14 age group mostly infants (spot checked enrollment on random records)

## Characteristics after filtering for age (15 to 55 inclusive) and TX residence at time of birth

| **Characteristic** | **N = 979,2891** |
| --- | --- |
| Birth Outcome |  |
| LB | 970,299 (99%) |
| SB | 8,990 (0.9%) |
| Year |  |
| 2019 | 211,278 (22%) |
| 2020 | 207,511 (21%) |
| 2021 | 205,622 (21%) |
| 2022 | 215,027 (22%) |
| 2023 | 139,851 (14%) |
| Race/Ethnicity |  |
| American Indian or Alaskan | 1,814 (0.2%) |
| Asian, Pacific Islander | 18,768 (1.9%) |
| Black, non-Hispanic | 143,756 (15%) |
| Hispanic | 598,316 (61%) |
| Unknown/Other | 50,100 (5.1%) |
| White, non-Hispanic | 166,535 (17%) |
| Age Group |  |
| 15-24 | 355,506 (36%) |
| 25-34 | 496,218 (51%) |
| 35-44 | 125,813 (13%) |
| 45-54 | 1,748 (0.2%) |
| 55-64 | 4 (<0.1%) |
| Enrollment Table (Month of birth episode) |  |
| chip\_enrl | 10,679 (1.1%) |
| enrl | 968,610 (99%) |
| Enrollment Category (Month of birth episode) |  |
| Emergency | 207,969 (21%) |
| Healthy Texas Women | 1,429 (0.1%) |
| Other | 11 (<0.1%) |
| QMD coverage (Medicare only) | 136 (<0.1%) |
| Regular | 755,638 (77%) |
| Three months prior | 3,427 (0.3%) |
| Unknown (or Not Enrolled) | 10,679 (1.1%) |
| State (Month of birth episode) |  |
| TX | 979,289 (100%) |
| *1n (%)* | |

* Age is based on the most common birth date when a person had multiple DOB
* Filtering to TX is based on enrollment the month of the episode, so it implicitly filters out anyone not enrolled the month of their birth. (If that is a problem)
* Race is based on enrollment the month of the episode, so someone might have race recorded elsewhere else if enrollment said “Unknown” that month.

# Dates associated with Pregnancy

Combining all possible dates together (admit, discharge, claim\_from\_date, claim\_to\_date) to see the distribution of dates for a single episode, we see:

| **unique\_dates\_per\_episode** | **episode\_count** |
| --- | --- |
| 1 | 26,385 |
| 2 | 496,653 |
| 3 | 378,723 |
| 4 | 72,788 |
| 5 | 9,219 |
| 6 | 1,698 |
| 7 | 638 |
| 8 | 202 |
| 9 | 60 |
| 10 | 31 |
| 11 | 25 |
| 12 | 7 |
| 13 | 6 |
| 14 | 6 |
| 15 | 11 |
| 16 | 5 |
| 18 | 6 |
| 19 | 3 |
| 22 | 1 |
| 23 | 1 |
| 24 | 1 |
| 25 | 1 |
| 28 | 1 |
| 34 | 1 |
| 36 | 1 |
| 37 | 1 |
| 40 | 1 |
| 50 | 1 |
| 56 | 1 |