POEM TX Cohort Pull

2024-10-07

# 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 | Deduplicating (Taking priority for facility) | 1,035,175 |
| 3 | After removing contradictory records | 1,035,175 |
| 4 | After rolling up overlapping records into episodes | 1,028,143 |
| 5 | After removing repeating episodes | 1,027,238 |
| 6 | After running algorithm | 987,079 |

* A record refers to a unique combination of client\_nbr + clm\_to\_date + outcome\_type (LB/SB) or combination of of client\_nbr + episode\_id once they are rolled up.
* The de-duplicating step is data cleaning: first I take facility records, and then only insert professional records if they do not currently overlap with the facility claims pulled first. This cleans up some messy date math that was occurring.
* Contradictory record is a record such that the same person has a record that same day with a different outcome type (LB vs SB)
* Overlapping records were rolled up according to dates alone. There was no special consideration of discharge codes or special logic for that. So, after cleaning up and sorting the dates, the if the next start\_date > previous end\_date, a new episode is marked. They are then rolled up to collapse the overlaps. This *may catch some transfers* but does not consider them separately.
* Repeating episodes had monthly codes showing a live birth at least once a month for more than six months in a row. These were investigated and removed due to not being able to tell which birth would be considered the correct one.

# Inclusion Exclusion

| **step** | **description** | **record\_count** | **record\_diff** |
| --- | --- | --- | --- |
| 1 | All episodes | 987,079 |  |
| 2 | Exists in enrollment at any point | 986,069 | 1,010 |
| 3 | Age 19-55 | 946,821 | 39,248 |
| 4 | Enrolled month of birth | 946,111 | 710 |
| 5 | Residence | 941,781 | 4,330 |
| 6 | LOS <= 60 | 941,481 | 300 |

## Characteristics after filtering

| **Characteristic** | **N = 941,4811** |
| --- | --- |
| Birth Outcome |  |
| LB | 932,088 (99%) |
| SB | 9,393 (1.0%) |
| Year |  |
| 2019 | 202,527 (22%) |
| 2020 | 198,989 (21%) |
| 2021 | 197,742 (21%) |
| 2022 | 206,817 (22%) |
| 2023 | 135,406 (14%) |
| Race/Ethnicity |  |
| American Indian or Alaskan | 1,753 (0.2%) |
| Asian, Pacific Islander | 18,642 (2.0%) |
| Black, non-Hispanic | 138,987 (15%) |
| Hispanic | 572,366 (61%) |
| Unknown/Other | 48,107 (5.1%) |
| White, non-Hispanic | 161,626 (17%) |
| Age Group |  |
| 15-24 | 317,969 (34%) |
| 25-34 | 496,037 (53%) |
| 35-44 | 125,725 (13%) |
| 45-54 | 1,747 (0.2%) |
| 55-64 | 3 (<0.1%) |
| Hispanic | 572,366 (61%) |
| Enrollment Category (Month of birth episode) |  |
| Emergency | 202,841 (22%) |
| Healthy Texas Women | 1,420 (0.2%) |
| Other | 9 (<0.1%) |
| QMD coverage (Medicare only) | 136 (<0.1%) |
| Regular | 722,486 (77%) |
| Three months prior | 3,473 (0.4%) |
| Unknown (or Not Enrolled) | 11,116 (1.2%) |
| *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
* Race is based on enrollment the month of the episode, so someone might have race recorded elsewhere else if enrollment said “Unknown” that month.

# Race Ethnicity Breakdown

| **race\_desc** | **0** | **1** | **Total** |
| --- | --- | --- | --- |
| American Indian or Alaskan | 1,753 | 0 | 1,753 |
| Asian, Pacific Islander | 18,642 | 0 | 18,642 |
| Black, non-Hispanic | 138,987 | 0 | 138,987 |
| Hispanic | 0 | 572,366 | 572,366 |
| Unknown/Other | 48,107 | 0 | 48,107 |
| White, non-Hispanic | 161,626 | 0 | 161,626 |
| Total | 369,115 | 572,366 | 941,481 |