POEM TX Cohort Pull

2024-09-05

# Change Log

* removed admit and discharge dates and replaced with clm\_from\_date and clm\_to\_date (too many typos in admit and discharge)
* added priority for facility claims over professional claims. professional claims were inserted ONLY if they did not overlap with a facility claim already present in the dataset.
* added length of stay info

# 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 | 1,033,119 |
| 4 | After rolling up overlapping records into episodes | 1,028,143 |
| 5 | After removing repeating episodes | 1,027,238 |

* 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.

# All Episodes

After running the algorithm, we find a possible 987079 episodes (N=817803 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 986069 episodes (N=816816 distinct ids).

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

| **Characteristic** | **N = 986,0691** |
| --- | --- |
| Birth Outcome |  |
| LB | 976,198 (99%) |
| SB | 9,871 (1.0%) |
| Year |  |
| 2019 | 213,302 (22%) |
| 2020 | 208,859 (21%) |
| 2021 | 206,590 (21%) |
| 2022 | 216,339 (22%) |
| 2023 | 140,979 (14%) |
| Age Group |  |
| 0-14 | 1,449 (0.1%) |
| 15-24 | 357,742 (36%) |
| 25-34 | 498,744 (51%) |
| 35-44 | 126,323 (13%) |
| 45-54 | 1,755 (0.2%) |
| 55-64 | 37 (<0.1%) |
| 65-74 | 13 (<0.1%) |
| 75+ | 6 (<0.1%) |
| State (Month of birth episode) |  |
| Non-TX | 4,182 (0.4%) |
| TX | 980,856 (99%) |
| Unknown | 1,031 (0.1%) |
| *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)

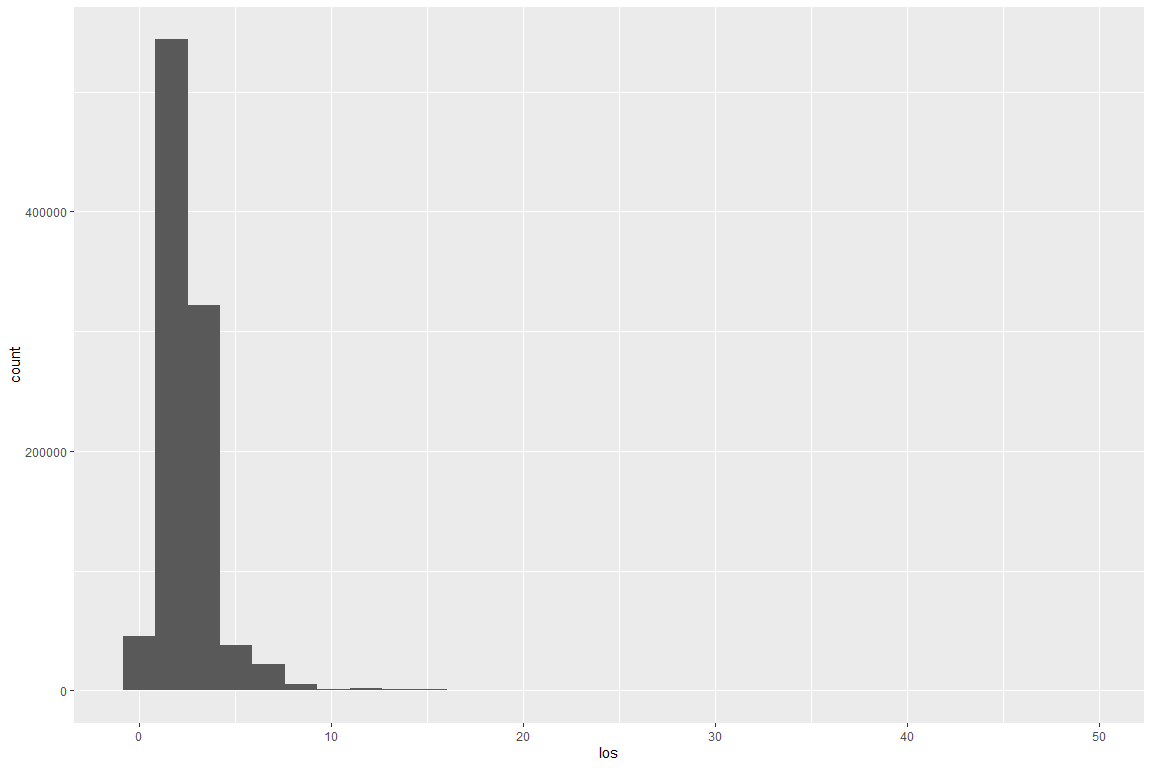
## Characteristics after filtering for age (15 to 55 inclusive)

| **Characteristic** | **N = 984,5671** |
| --- | --- |
| Birth Outcome |  |
| LB | 974,713 (99%) |
| SB | 9,854 (1.0%) |
| Year |  |
| 2019 | 212,899 (22%) |
| 2020 | 208,527 (21%) |
| 2021 | 206,286 (21%) |
| 2022 | 216,031 (22%) |
| 2023 | 140,824 (14%) |
| Race/Ethnicity |  |
| American Indian or Alaskan | 1,847 (0.2%) |
| Asian, Pacific Islander | 18,891 (1.9%) |
| Black, non-Hispanic | 144,702 (15%) |
| Hispanic | 600,056 (61%) |
| Unknown/Other | 51,267 (5.2%) |
| White, non-Hispanic | 167,804 (17%) |
| Age Group |  |
| 15-24 | 357,742 (36%) |
| 25-34 | 498,744 (51%) |
| 35-44 | 126,323 (13%) |
| 45-54 | 1,755 (0.2%) |
| 55-64 | 3 (<0.1%) |
| Enrollment Category (Month of birth episode) |  |
| Emergency | 207,430 (21%) |
| Healthy Texas Women | 1,432 (0.1%) |
| Other | 10 (<0.1%) |
| QMD coverage (Medicare only) | 141 (<0.1%) |
| Regular | 759,694 (77%) |
| Three months prior | 3,681 (0.4%) |
| Unknown (or Not Enrolled) | 12,179 (1.2%) |
| State (Month of birth episode) |  |
| Non-TX | 4,180 (0.4%) |
| TX | 979,365 (99%) |
| Unknown | 1,022 (0.1%) |
| Length of Stay | 2 (2, 3) |
| Facility/Professional Claims |  |
| F | 943,244 (96%) |
| P | 41,323 (4.2%) |
| *1n (%); Median (IQR)* | |

* 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.

# Length of Stay

LOS is calculated as end\_date - start\_date, so if an episode’s end\_date is the same as the start\_date, los is marked as 0 for now. We can collapse that into 1 or more, but need to make a group decision on that. The graph below only shows the distribution for those under 50 days (there were 513 individuals with a LOS > 50 days)



| **los** | **count\_** |
| --- | --- |
| 0 | 45,861 |
| 1 | 127,267 |
| 2 | 416,813 |
| 3 | 237,310 |
| 4 | 84,710 |
| 5 | 37,729 |
| 6 | 15,727 |
| 7 | 6,628 |
| 8 | 3,215 |
| 9 | 1,970 |
| 10 | 1,166 |
| 11 | 956 |
| 12 | 718 |
| 13 | 601 |
| 14 | 519 |
| 15 | 417 |
| 16 | 365 |
| 17 | 347 |
| 18 | 330 |
| 19 | 286 |
| 20 | 240 |
| 21 | 223 |
| 22 | 183 |
| 23 | 181 |
| 24 | 137 |
| 25 | 142 |
| 26 | 129 |
| 27 | 124 |
| 28 | 98 |
| 29 | 104 |
| 30 | 105 |
| 31 | 102 |
| 32 | 86 |
| 33 | 91 |
| 34 | 60 |
| 35 | 61 |
| 36 | 62 |
| 37 | 51 |
| 38 | 49 |
| 39 | 43 |
| 40 | 46 |
| 41 | 44 |
| 42 | 44 |
| 43 | 38 |
| 44 | 35 |
| 45 | 35 |
| 46 | 26 |
| 47 | 22 |
| 48 | 31 |
| 49 | 28 |
| 50 | 24 |
| 51 | 26 |
| 52 | 24 |
| 53 | 14 |
| 54 | 11 |
| 55 | 23 |
| 56 | 16 |
| 57 | 11 |
| 58 | 15 |
| 59 | 18 |
| 60 | 20 |
| 61 | 15 |
| 62 | 12 |
| 63 | 14 |
| 64 | 14 |
| 65 | 13 |
| 66 | 8 |
| 67 | 8 |
| 68 | 3 |
| 69 | 9 |
| 70 | 7 |
| 71 | 3 |
| 72 | 5 |
| 73 | 6 |
| 74 | 7 |
| 75 | 5 |
| 76 | 2 |
| 77 | 8 |
| 78 | 6 |
| 79 | 5 |
| 80 | 2 |
| 81 | 5 |
| 82 | 4 |
| 83 | 4 |
| 84 | 3 |
| 85 | 3 |
| 86 | 6 |
| 87 | 4 |
| 88 | 3 |
| 89 | 3 |
| 90 | 3 |
| 91 | 3 |
| 92 | 3 |
| 93 | 2 |
| 94 | 3 |
| 96 | 1 |
| 97 | 1 |
| 98 | 3 |
| 99 | 2 |
| 101 | 1 |
| 102 | 1 |
| 103 | 1 |
| 104 | 3 |
| 105 | 1 |
| 109 | 1 |
| 110 | 2 |
| 111 | 1 |
| 112 | 1 |
| 116 | 1 |
| 117 | 1 |
| 118 | 1 |
| 119 | 2 |
| 121 | 1 |
| 123 | 1 |
| 124 | 2 |
| 129 | 1 |
| 132 | 2 |
| 136 | 1 |
| 137 | 1 |
| 139 | 1 |
| 140 | 2 |
| 142 | 1 |
| 143 | 1 |
| 152 | 2 |
| 154 | 1 |
| 155 | 3 |
| 157 | 1 |
| 159 | 1 |
| 162 | 1 |
| 163 | 1 |
| 167 | 1 |
| 168 | 1 |
| 169 | 1 |
| 171 | 1 |
| 175 | 1 |
| 182 | 1 |
| 183 | 2 |
| 185 | 1 |
| 188 | 4 |
| 191 | 2 |
| 192 | 1 |
| 198 | 1 |
| 204 | 1 |
| 205 | 1 |
| 206 | 1 |
| 210 | 1 |
| 217 | 1 |
| 218 | 1 |
| 221 | 1 |
| 222 | 1 |
| 224 | 1 |
| 225 | 1 |
| 230 | 1 |
| 231 | 1 |
| 234 | 1 |
| 236 | 2 |
| 244 | 1 |
| 259 | 1 |
| 277 | 1 |
| 306 | 1 |
| 307 | 1 |
| 365 | 1 |
| 366 | 2 |
| 367 | 8 |
| 368 | 9 |
| 369 | 4 |
| 370 | 1 |
| 373 | 1 |

# Discharge

| **discharge\_code** | **c** |
| --- | --- |
| 01 | 939,589 |
| Missing | 42,186 |
| 07 | 1,771 |
| 02 | 761 |
| 06 | 582 |
| 81 | 184 |
| 30 | 174 |
| 21 | 140 |
| 65 | 97 |
| 70 | 79 |
| 20 | 77 |
| 62 | 69 |
| 09 | 62 |
| 50 | 59 |
| Other | 43 |
| 05 | 42 |
| 03 | 33 |
| 04 | 28 |
| 43 | 25 |
| 63 | 23 |
| 66 | 12 |
| 88 | 11 |
| 95 | 11 |
| 51 | 11 |