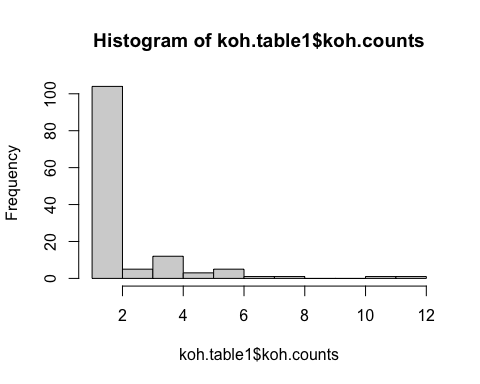
Capstone EDA

Gabby Lopez

2025-02-02

# Descriptive Statistics

Changing any cases where PreDM=1 and T2DM=1 to PreDM=0 for the descriptive statistics

distribution of KOH attendance 

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 1.000 1.000 2.045 2.000 12.000

## 1 2 3 4 5 6 7 8 11 12   
## 82 22 5 12 3 5 1 1 1 1

## Primary Objective - Hypertension pre and post analysis datasets

subsetting koh patients with hypertension

subsetting marshallese non-koh pts with HTN

## Primary objective hypertension longitudinal analysis datasets

koh patients

marshallese non-koh patients

## Primary Objective - Diabetes pre and post analysis datasets

subsetting koh patients with prediabetes or T2DM

subsetting marshallese non-koh pts with diabetes

## Primary objective Diabetes longitudinal analysis datasets

koh pts

marshallese non-koh pts

make table 1 dataset for primary objective 249 unique patients 127 from bp and 221 from a1c

table 1 for primary objective population

## Adding missing grouping variables: `UniqueIdentifier`

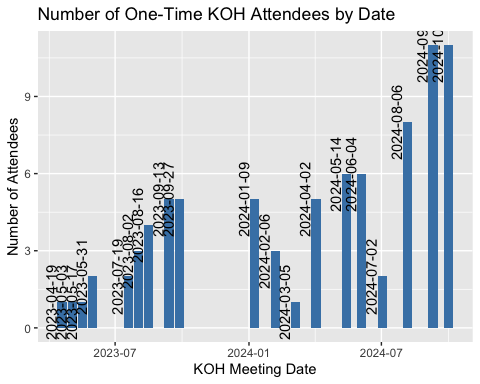
| **Baseline Characteristics** | **KOH Participant**  N = 71 | **Non-Participant**  N = 178 |
| --- | --- | --- |
| **Age** | 62 (32, 80) | 56 (19, 89) |
| **Sex** |  |  |
| Female | 45 (63%) | 107 (60%) |
| Male | 26 (37%) | 71 (40%) |
| **KOH Meetings Attended** | 2.00 (1.00, 11.00) | 0.00 (0.00, 0.00) |
| **Hypertension** | 42 (59%) | 98 (55%) |
| **Pre-Diabetes** | 3 (4.2%) | 8 (4.5%) |
| **Type II Diabetes** | 68 (96%) | 164 (92%) |
| **Hypertension and T2DM** | 42 (59%) | 92 (52%) |
| **Clinic Location** |  |  |
| Maple | 14 (23%) | 67 (38%) |
| Market | 46 (77%) | 111 (62%) |
| Missing | 11 | 0 |
| **Housing Status** |  |  |
| Housed | 28 (39%) | 108 (61%) |
| Doubling Up | 43 (61%) | 62 (35%) |
| Permanent Supportive Housing | 0 (0%) | 5 (2.8%) |
| Transitional | 0 (0%) | 1 (0.6%) |
| Other | 0 (0%) | 2 (1.1%) |
| Unknown | 0 (0%) | 0 (0%) |
| **Income Level** | 0 (0, 294) | 43 (0, 295) |
| Missing | 1 | 5 |
| **Risk Score** | 2 (0, 9) | 2 (0, 12) |
| Missing | 0 | 9 |

table 1 for all Marshallese and non-Marshallese

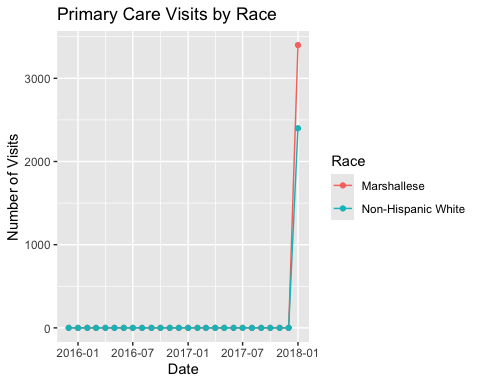
| **Baseline Characteristics** | **Overall**  N = 27,6871 | **Marshallese**  N = 869 | **Non-Marshallese**  N = 26818 |
| --- | --- | --- | --- |
| **Age** |  |  |  |
| Median (Min, Max) | 41 (18, 102) | 46 (18, 89) | 41 (18, 102) |
| **Sex** |  |  |  |
| Female | 16,223 (59%) | 576 (66%) | 15,647 (58%) |
| Male | 11,464 (41%) | 293 (34%) | 11,171 (42%) |
| **Hypertension** | 6,639 (24%) | 211 (24%) | 6,428 (24%) |
| **Pre-Diabetes** | 1,306 (4.7%) | 35 (4.0%) | 1,271 (4.7%) |
| **Type II Diabetes** | 3,045 (11%) | 386 (44%) | 2,659 (9.9%) |
| **Hypertension and T2DM** | 2,708 (9.8%) | 190 (22%) | 2,518 (9.4%) |
| **Clinic Location** |  |  |  |
| Maple | 19,016 (69%) | 429 (52%) | 18,587 (69%) |
| Market | 8,622 (31%) | 391 (48%) | 8,231 (31%) |
| Missing | 49 | 49 | 0 |
| **Housing Status** |  |  |  |
| Housed | 20,628 (75%) | 547 (63%) | 20,081 (75%) |
| Doubling Up | 5,585 (20%) | 304 (35%) | 5,281 (20%) |
| Permanent Supportive Housing | 350 (1.3%) | 7 (0.8%) | 343 (1.3%) |
| Transitional | 626 (2.3%) | 4 (0.5%) | 622 (2.3%) |
| Other | 444 (1.6%) | 6 (0.7%) | 438 (1.6%) |
| Unknown | 54 (0.2%) | 1 (0.1%) | 53 (0.2%) |
| **Income Level** |  |  |  |
| Median (Min, Max) | 117 (0, 25,568) | 43 (0, 461) | 120 (0, 25,568) |
| Missing | 4,607 | 68 | 4,539 |
| **Risk Score** |  |  |  |
| Median (Min, Max) | 1 (0, 22) | 1 (0, 12) | 1 (0, 22) |
| Missing | 3,148 | 73 | 3,075 |
| 1n (%) | | | |

# One-Time KOH Attendance Distribution

## Adding missing grouping variables: `UniqueIdentifier`



# Preliminary Parallel Trends EDA



## Encounters Table

i think i still need to adjust this table a bit but it’s fine for now. I removed visit types (Health Equity and Other) that had 0 visits for multiple years as well. 2016, 2017, and 2025 didn’t have much data either so I removed them. Then I removed years 2018 and 2019 just to make the table a bit smaller since it was so wide.

|  | **2020** | | **2021** | | **2022** | | **2023** | | **2024** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mean Visits per Patient** | **Marshallese**  N = 3541 | **Non-Marshallese**  N = 148491 | **Marshallese**  N = 4071 | **Non-Marshallese**  N = 163001 | **Marshallese**  N = 5041 | **Non-Marshallese**  N = 190101 | **Marshallese**  N = 6721 | **Non-Marshallese**  N = 220291 | **Marshallese**  N = 7321 | **Non-Marshallese**  N = 217531 |
| **Primary Care** | 2.1 (0.0, 23.0) | 2.6 (0.0, 81.0) | 3.8 (0.0, 25.0) | 5.9 (0.0, 217.0) | 1.8 (0.0, 18.0) | 2.9 (0.0, 108.0) | 2.0 (0.0, 35.0) | 3.0 (0.0, 85.0) | 2.1 (0.0, 24.0) | 3.1 (0.0, 99.0) |
| **Emergency** | 0.6 (0.0, 11.0) | 0.9 (0.0, 44.0) | 1.1 (0.0, 16.0) | 1.5 (0.0, 101.0) | 0.7 (0.0, 16.0) | 0.9 (0.0, 66.0) | 0.8 (0.0, 13.0) | 0.9 (0.0, 76.0) | 0.8 (0.0, 12.0) | 1.0 (0.0, 153.0) |
| **Urgent Care** | 0.5 (0.0, 5.0) | 0.6 (0.0, 14.0) | 0.9 (0.0, 12.0) | 1.1 (0.0, 32.0) | 0.4 (0.0, 7.0) | 0.4 (0.0, 16.0) | 0.4 (0.0, 10.0) | 0.4 (0.0, 16.0) | 0.4 (0.0, 9.0) | 0.4 (0.0, 21.0) |
| **Behavioral Health** | 0.0 (0.0, 3.0) | 0.7 (0.0, 49.0) | 0.1 (0.0, 8.0) | 1.3 (0.0, 96.0) | 0.0 (0.0, 4.0) | 0.8 (0.0, 83.0) | 0.0 (0.0, 3.0) | 0.7 (0.0, 101.0) | 0.0 (0.0, 8.0) | 0.9 (0.0, 106.0) |
| **Pharmacy** | 0.6 (0.0, 14.0) | 0.2 (0.0, 71.0) | 1.3 (0.0, 38.0) | 0.4 (0.0, 117.0) | 1.0 (0.0, 18.0) | 0.2 (0.0, 58.0) | 1.1 (0.0, 20.0) | 0.3 (0.0, 57.0) | 1.1 (0.0, 17.0) | 0.3 (0.0, 57.0) |
| **Dental** | 0.5 (0.0, 12.0) | 0.8 (0.0, 26.0) | 1.3 (0.0, 16.0) | 2.1 (0.0, 52.0) | 0.9 (0.0, 12.0) | 1.1 (0.0, 27.0) | 1.2 (0.0, 14.0) | 1.3 (0.0, 33.0) | 1.3 (0.0, 15.0) | 1.3 (0.0, 25.0) |
| **Multispecialty** | 0.6 (0.0, 29.0) | 0.2 (0.0, 25.0) | 1.3 (0.0, 49.0) | 0.5 (0.0, 56.0) | 1.0 (0.0, 24.0) | 0.3 (0.0, 41.0) | 0.7 (0.0, 23.0) | 0.2 (0.0, 28.0) | 0.9 (0.0, 23.0) | 0.2 (0.0, 31.0) |
| 1Mean (Minimum, Maximum) | | | | | | | | | | |