Hospital Readmission Equity Report

AUTHOR
Markuss Saule

PUBLISHED
September 15, 2025

1 Executive Summary

This analysis explores **30-day readmission** patterns using the public *Diabetes 130-US hospitals (1999–2008)* dataset.

Key messages: - The **overall 30-day readmission rate is about 11%**, consistent with prior analyses of this dataset. - **Race subgroup differences are modest** (roughly 9–12%), with overlapping confidence intervals, suggesting no strong racial disparities. - **Gender shows almost no difference** in readmission. - The strongest predictors of readmission are **clinical complexity**: more diagnoses, longer hospital stays, and greater medication use.

> **Interpretation note:** This dataset is dated and illustrative; findings highlight workflow and methodology, not clinical advice.

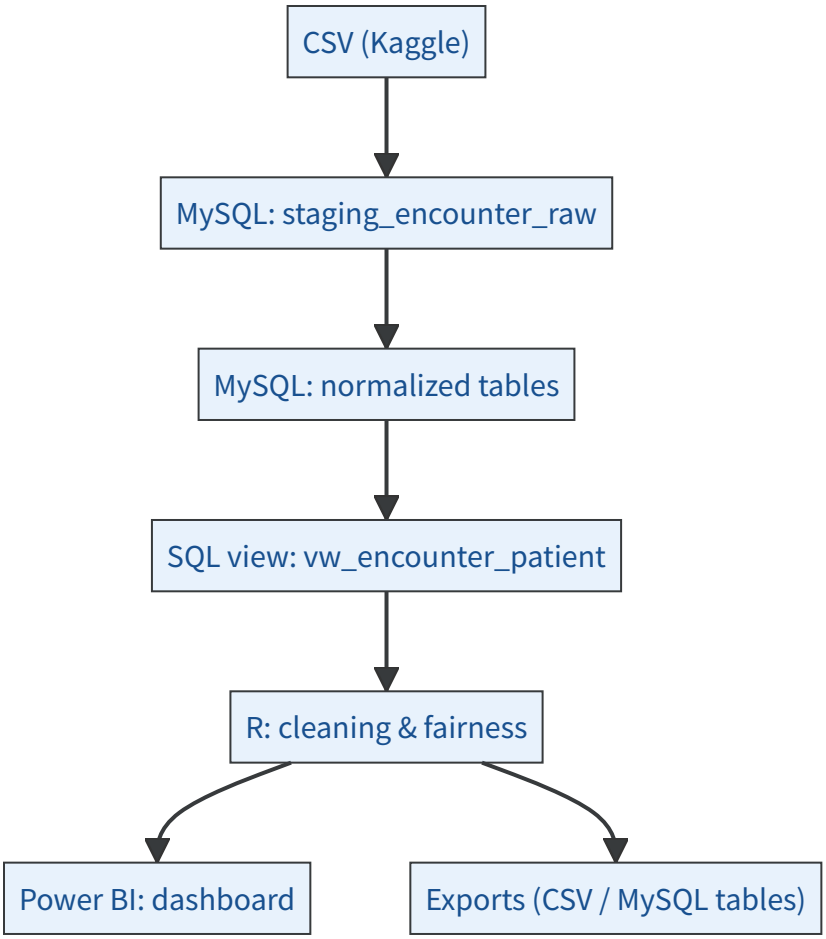
This analysis explores **30-day readmission** patterns using the public *Diabetes 130-US hospitals (1999–2008)* dataset, focusing on **equity across demographic subgroups** and **interpretable drivers** of risk.

- **Overall 30-day readmission rate: 11%**
- **Demographics not significant:** Race and gender coefficients are near 1 with overlapping CIs.
- **Clinical complexity matters:** Longer stays, more diagnoses, and higher medication counts increase odds of readmission.
- **Highest subgroup rate (race): Caucasian — 11%**
- **Gap vs lowest subgroup: 3%**
- **Key signals: time in hospital**, **number of diagnoses**, and **medication intensity** show monotonic associations with readmission.

> **Interpretation note:** This is an educational analysis on a public dataset; results are illustrative, not clinical guidance.

2 Data & Methods

We implement a pragmatic analytics pipeline similar to what health systems use:



Outcome: `readmitted` recoded as `readmit_flag = 1` if `"<30"`, else **0**.
Equity: subgroup rates and 95% CIs via normal approximation.
Model: logistic regression (simple, interpretable).

3 Exploratory Overview

Overall snapshot

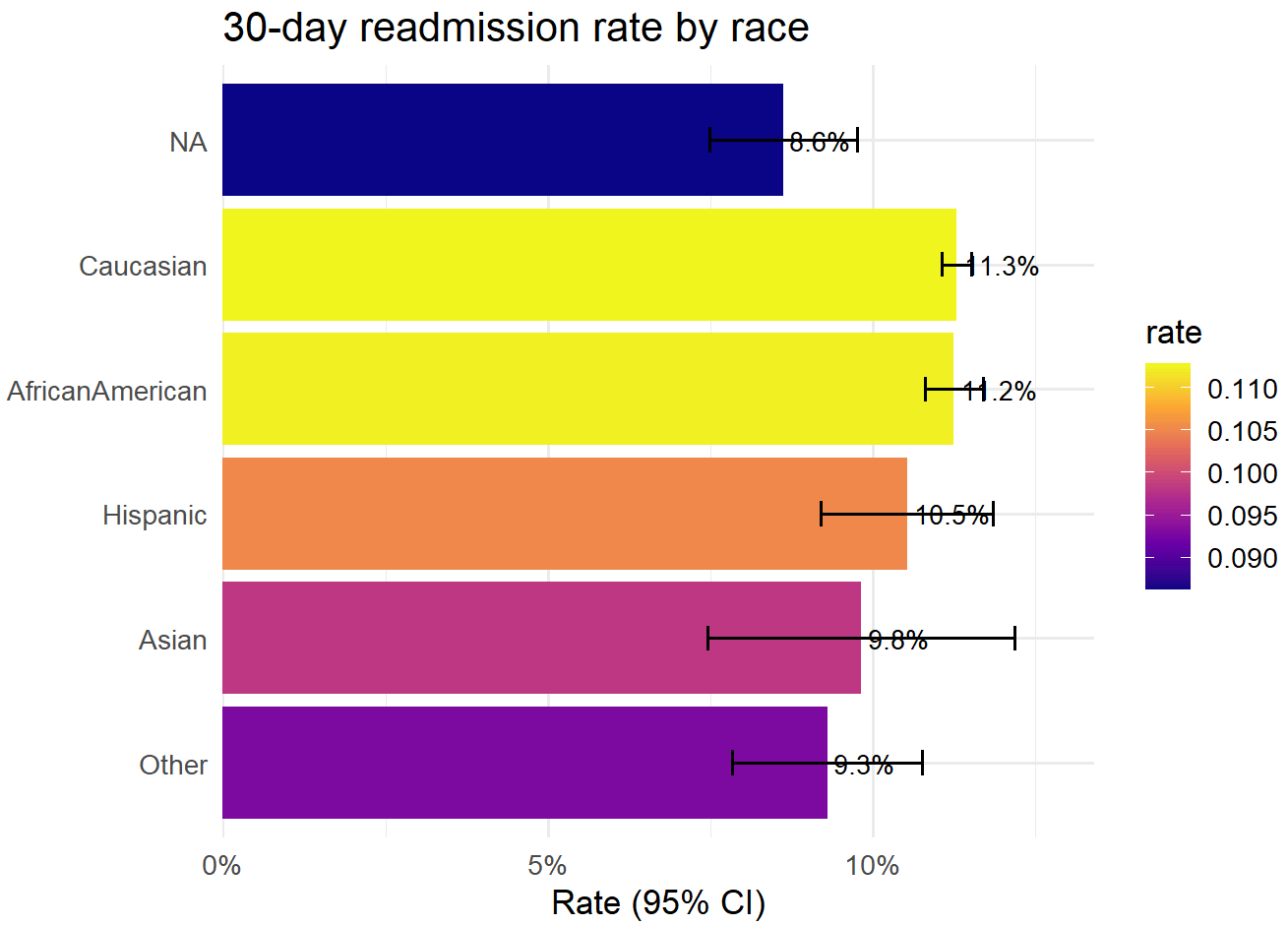
Metric	Value
Encounters	101,766
Overall 30-day readmission rate	11%

4 Readmission by Race (Equity Perspective)

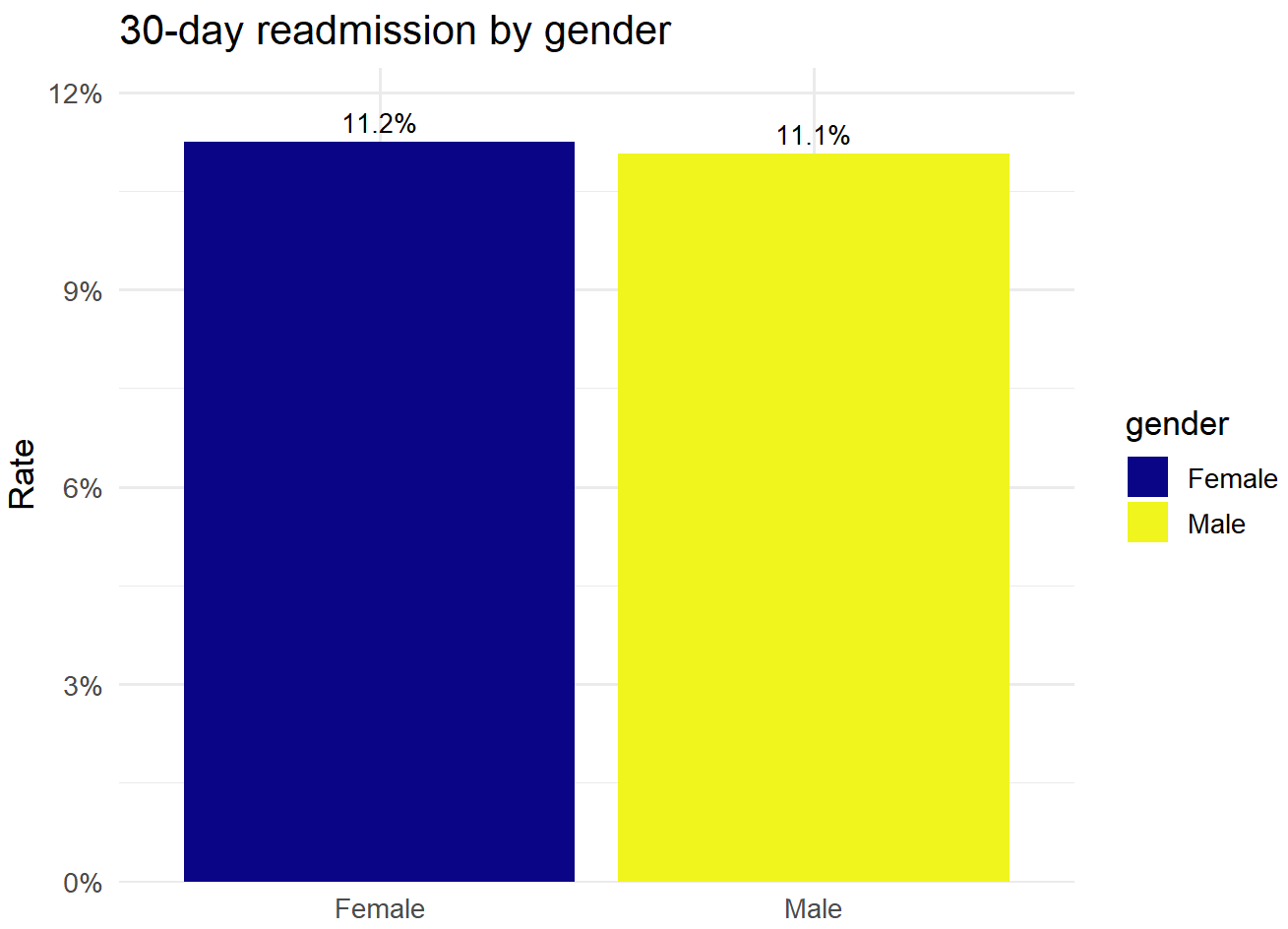
Readmission by race with 95% CI

race	n	rate	se	lcl	ucl
Caucasian	76043	11.284%	0.0011474	11.060%	11.51%
AfricanAmerican	19165	11.239%	0.0022815	10.792%	11.69%

race	n	rate	se	lcl	ucl
Hispanic	2063	10.519%	0.0067546	9.195%	11.84%
Asian	611	9.820%	0.0120390	7.460%	12.18%
Other	1517	9.295%	0.0074549	7.834%	10.76%
NA	2367	8.619%	0.0057683	7.488%	9.75%



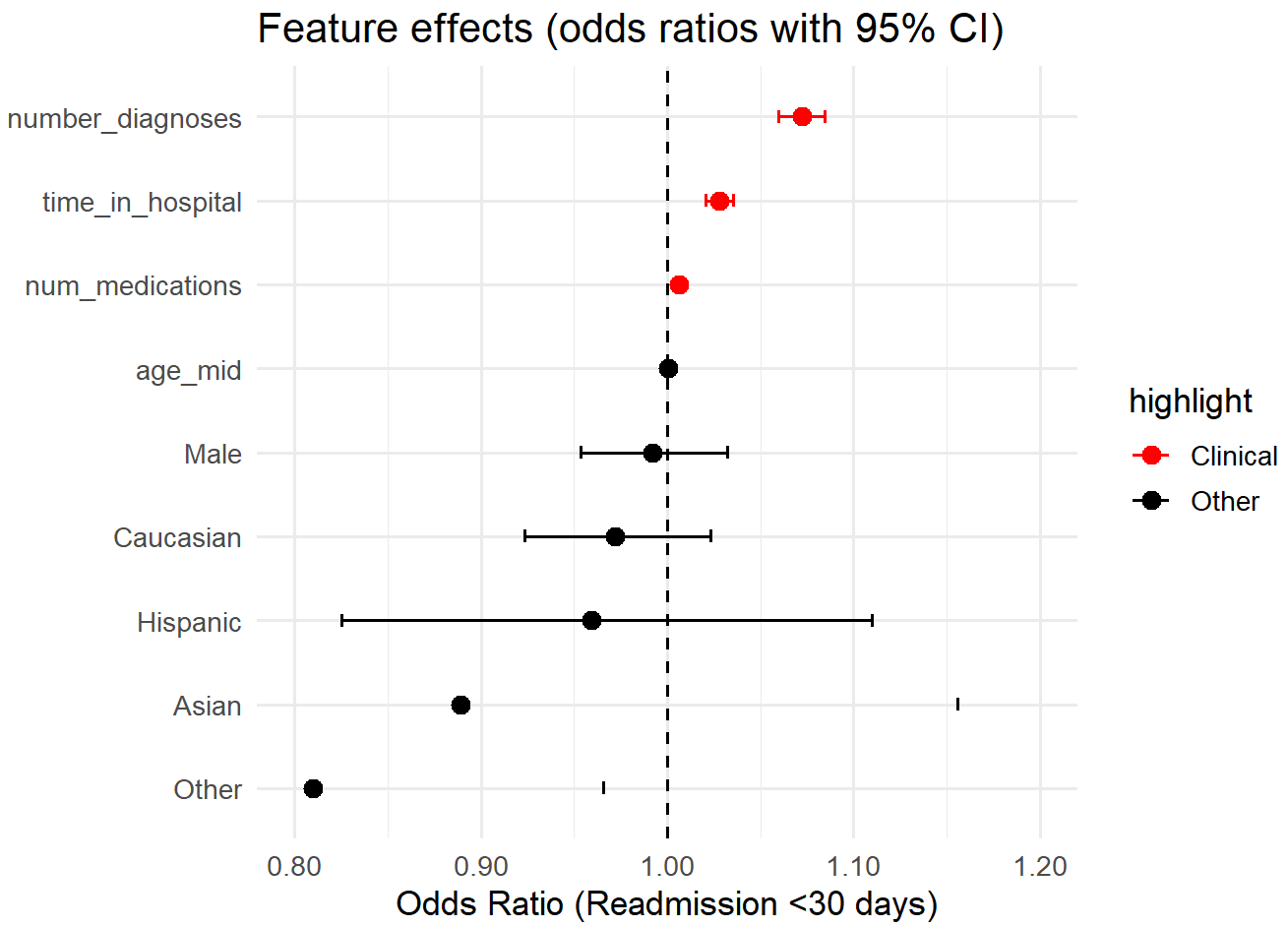
5 Readmission by Gender



6 Logistic Regression (Interpretable Model)

Logistic regression (odds ratios)

Feature	Odds Ratio	95% CI (low)	95% CI (high)
(Intercept)	0.060	0.053	0.067
Asian	0.889	0.672	1.156
Caucasian	0.972	0.923	1.023
Hispanic	0.959	0.825	1.110
Other	0.810	0.674	0.966
Male	0.992	0.953	1.032
age_mid	1.000	0.999	1.002
time_in_hospital	1.028	1.020	1.035
num_medications	1.006	1.003	1.009
number_diagnoses	1.072	1.060	1.085



7 Exports for Power BI

Preview: Subgroup Equity Export

race	n	rate	se	lcl	ucl
Caucasian	76043	0.1128	0.0011474	0.1106	0.1151
AfricanAmerican	19165	0.1124	0.0022815	0.1079	0.1169
Hispanic	2063	0.1052	0.0067546	0.0919	0.1184
Asian	611	0.0982	0.0120390	0.0746	0.1218
Other	1517	0.0929	0.0074549	0.0783	0.1076
NA	2367	0.0862	0.0057683	0.0749	0.0975

Preview: Encounter Slim Export

encounter_id	patient_id	race	gender	age_group	time_in_hospital	num_medications	number_diagnoses	re
24437208	135	Caucasian	Female	[50-60)	8	33	8	<
26264286	135	Caucasian	Female	[50-60)	3	14	5	>
29758806	378	Caucasian	Female	[50-60)	2	11	3	N

encounter_id	patient_id	race	gender	age_group	time_in_hospital	num_medications	number_diagnoses	re
189899286	729	Caucasian	Female	[80-90)	4	23	9	N
64331490	774	Caucasian	Female	[80-90)	3	20	9	N
14824206	927	AfricanAmerican	Female	[30-40)	5	5	3	N
8380170	1152	AfricanAmerican	Female	[50-60)	6	13	2	>3
30180318	1152	AfricanAmerican	Female	[50-60)	6	15	6	>3
55533660	1152	AfricanAmerican	Female	[50-60)	10	19	9	>3
80742510	1152	AfricanAmerican	Female	[50-60)	8	16	2	>3

8 Findings & Interpretation

- **Demographic predictors (race, gender) were not statistically significant** after adjustment; odds ratios were ~1 with overlapping 95% CIs.
- **Clinical complexity** (time in hospital, number of diagnoses, medication count) showed positive, statistically significant associations with readmission.

Summary of results: - The **overall readmission rate is ~11%**. - **Race subgroup rates range from ~9% to 12%**, with overlapping CIs → differences are modest and not clearly significant. - **Gender rates are nearly identical.** - Logistic regression confirms that **clinical complexity (time in hospital, number of diagnoses, medications)**, not demographics, drives readmission risk.

- Overall 30-day readmission rate: **11%**
- Highest subgroup (race): **Caucasian — 11%**
- Gap vs lowest subgroup: **3%**
- Strongest predictors: **time in hospital**, **number of diagnoses**, **medication burden**.

> **Note:** These are illustrative only, not clinical recommendations.

9 Limitations

- Dataset covers 1999–2008, and may not reflect today's clinical practices.
- Subgroup labels are limited; race/ethnicity categories are coarse.
- Readmission coding may not align with modern standards.

10 Next Steps

1. Add A1C & glucose metrics to models.
2. Build risk-adjusted fairness metrics.
3. Export dashboard-ready tables with subgroup CIs.
4. Explore oncology datasets to connect personal mission to Mayo's work.

11 Appendix

11.1 SQL: Analysis View

```
CREATE OR REPLACE VIEW vw_encounter_patient AS
SELECT
  e.encounter_id,
  p.patient_id,
  r.code AS race,
  g.code AS gender,
  a.label AS age_group,
  e.time_in_hospital,
  e.num_medications,
  e.number_diagnoses,
  cr.code AS readmitted
FROM encounter e
JOIN patient p              ON p.patient_id = e.patient_id
LEFT JOIN code_race r       ON r.race_id = p.race_id
LEFT JOIN code_gender g     ON g.gender_id = p.gender_id
LEFT JOIN code_age_group a  ON a.age_group_id = p.age_group_id
LEFT JOIN code_readmitted cr ON cr.readmitted_id = e.readmitted_id;
```

11.2 Session Info

```
R version 4.4.1 (2024-06-14 ucrt)
Platform: x86_64-w64-mingw32/x64
Running under: Windows 11 x64 (build 26100)

Matrix products: default


locale:
[1] LC_COLLATE=English_United States.utf8
[2] LC_CTYPE=English_United States.utf8
[3] LC_MONETARY=English_United States.utf8
[4] LC_NUMERIC=C
[5] LC_TIME=English_United States.utf8

time zone: America/Denver
tzcode source: internal

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] kableExtra_1.4.0 knitr_1.48       readr_2.1.5      stringr_1.5.1
 [5] scales_1.4.0     broom_1.0.6      ggplot2_4.0.0    forcats_1.0.0
```

```
 [9] tidyr_1.3.1      dplyr_1.1.4      RMariaDB_1.3.4   DBI_1.2.3

loaded via a namespace (and not attached):
 [1] utf8_1.2.4         generics_0.1.3     xml2_1.3.6         stringi_1.8.4
 [5] hms_1.1.3          digest_0.6.37      magrittr_2.0.3     timechange_0.3.0
 [9] evaluate_0.24.0    grid_4.4.1         RColorBrewer_1.1-3 fastmap_1.2.0
[13] jsonlite_1.8.8     backports_1.5.0    purrr_1.0.2        fansi_1.0.6
[17] viridisLite_0.4.2  textshaping_0.4.0  cli_3.6.3          crayon_1.5.3
[21] rlang_1.1.4        bit64_4.5.2        withr_3.0.1        yaml_2.3.10
[25] parallel_4.4.1     tools_4.4.1        tzdb_0.4.0         vctrs_0.6.5
[29] R6_2.5.1           lubridate_1.9.3    lifecycle_1.0.4    bit_4.5.0
[33] vroom_1.6.5        pkgconfig_2.0.3    pillar_1.9.0       gtable_0.3.6
[37] glue_1.7.0         systemfonts_1.2.3  highr_0.11         xfun_0.47
[41] tibble_3.2.1       tidyselect_1.2.1   rstudioapi_0.16.0  farver_2.1.2
[45] htmltools_0.5.8.1  labeling_0.4.3     rmarkdown_2.28     svglite_2.2.1
[49] compiler_4.4.1     S7_0.2.0
```