Evaluating Staffing Characteristics, Healthcare-associated Infections, and Provider Characteristics in U.S. Nursing Homes, 2021

Kelly Hatfield

4/7/23

# 1. Summary

The proposed study aims to assess the relationship between overall staffing rates with healthcare-associated infection rates in nursing homes throughout the United States.

# 2. Introduction

## 2.1 General Background Information

Over 15,000 skilled nursing facilities (SNFs) and nursing homes provide residential long term care including both personal and medical services to over a million patients throughout the United States. Due to advanced age and medical complexity of residents in the nursing home setting, this population is at increased risk for significant morbidity and mortality. Previous estimates show that approximately 12% of NH residents had an infection at a given time (Thompson et al., 2020); most commonly urinary tract infection (3.0–5.2%), pneumonia (2.2–4.4%), and cellulitis (1.6–2.0%). In the COVID-19 pandemic, nearly all nursing homes experienced outbreaks of COVID-19, and nursing home residents had disproportionately high morbidity and mortality compared to community dwelling adults. Nursing homes were mentioned as a specific target for reducing the morbidity and mortality related to healthcare associated infections in the 2009 *National Action Plan to Prevent Health Care-Associated Infections: Roadmap to Elimination* (HAI National Action Plan) and subsequent updates published in 2013 and 2018. Most healthcare associated infections (HAIs) are thought to be preventable as they result from the processes and structures related to poor care, including poor adherence to protocols and guidelines.

The Centers for Medicare and Medicaid Services (CMS) reports measures of nursing home quality on their Care Compare website (https://www.medicare.gov/care-compare/). These measures of care include ownership details, health inspection ratings, staffing ratings, and quality measures. Data are searchable and comparable on the website for all participating nursing homes. Aggregate data sets are also available for download.

Federal law requires all nursing homes to provide enough staff to safely care for residents. Nursing home reports daily staffing hours to Medicare in the payroll based journal. Using this information, Medicare calculates a ratio of staffing hours per resident day, the percent of nurse staff that stop working at the facility (turnover), and the number of administrators who have left the facility within a given year and reports the data publicly. CMS also creates a staffing star measure that is adjusted based off the needs of the residents in each home and is compiled using aggregated metrics about the RN staffing levels, total staffing hours, weekend staffing hours, total nurse turnover, RN turnover, and administrator turnover. (https://www.medicare.gov/care-compare/resources/nursing-home/staffing).

CMS also calculates quality measures star rating. This assess different measures of nursing home quality. For this report we are most interested in a new quality measure *Skilled Nursing Facility Healthcare-Associated Infections (HAI) Requiring Hospitalization* that became reported for the first time in October 2022, with data collected throughout FY 2019. The HAI requiring hospitalization metric estimates the risk-standardized rate of HAIs that are acquired during SNF care and result in hospitalization. The measure is calculated on Medicare claims data from fee-for-service beneficiaries residing within a nursing home. The measure requires no additional reporting from nursing home staff, and relies entirely on administrative claim data and is based on principal diagnosis on the Medicare inpatient (IP) claims of SNF residents.

## 2.2 Study Objectives

Previous studies have shown that certain staffing metrics are associated with lower quality metrics in US nursing homes Mukamel, Saliba, Ladd, & Konetzka (2022). However, they were limited because they did not assess healthcare associated infections requiring hospitalization. We aim to assess the association of nursing home staff star ratings with the healthcare associated infection rate in participating nursing homes.

# 3. Methods

## 3.1 Data acquisition

For this project I will utilize data collected and aggregated by the Centers for Medicare Services (CMS) for all CMS certified nursing homes throughout the United States. The data are accessed here: <https://data.cms.gov/provider-data/archived-data/nursing-homes>. There are approximately 15,000 CMS certified nursing homes throughout the US, so we will have a large geographically diverse population to study.

First, the provider information file contains general information on currently active nursing homes, including number of certified beds, quality measure scores, staffing and other information used in the Five-Star Rating System. Data are presented as one row per nursing home. The nurse staffing metrics in this data set include reported and adjusted measures of nurse aide staffing hours per resident per day, licensed professional nurse (LPN) staffing hours per resident per day, registered nurse (RN) staffing hours per resident per day, total nurse staffing hours per resident per day, and total number of nursing staffing hours per resident per day on the weekend. The data also includes turnover data including RN turnover, total staff turn over, and administrator turnover.

Second, we addend the provider information with aggregated data reported as part ofthe Skilled Nursing Facility Quality Reporting Program. CMS aggregates data from the over 15,000 certified nursing homes for a number of quality metrics. They provide these metrics on their Care Compare site referenced above.

The primary QRP metric of interest for this study is titled S\_039\_01: Percentage of patients who acquired a healthcare-associated infection during their SNF stay that resulted in hospitalization. This percentage is shown both as the observed rate and as a risk standardized rate. The HAI requiring hospitalization metric is also given as a categorical variable, where the nursing home is compared to all other nursing homes in that time period and categorized as average, significantly worse than average, or significantly better than average.

## 3.2 Data cleaning

Data from both data sets will be merged together using a unique facility identifier, the CMS Certification Number (CCN). We will subset our data set to include only nursing homes with complete provider information and HAI requiring hospitalization metrics. Nursing homes with missing data will be excluded from analysis.

## 3.3 Statistical analysis

First, we describe the distribution of adjusted average total staffing and risk standardized rate of HAIs requiring hospitalization by nursing home. We also describe the frequency and percent of the categorical staffing rating and HAI categorization (better, worse, or same as average). Second, we consider bivariate associations of the risk standardized infection rate by the staffing rating variable.

Third, we create multivariable models that assess the relationship between the risk standardized HAI rate (outcome) and the total average staffing per resident per day and staffing rating, while adjusting for important nursing home characteristics as identified through the literature including: ownership, number of health violation, health inspection score, and average number of residents per day. For our multivariable models, we aim to determine if the total average staffing per resident per day and staffing rating are independently associated with the outcome.

To create our multivariable models and assess their ability for prediction, we first divided our full data cohort into a training (70% of nursing homes) and test data set (30% of nursing homes). We used our training data to determine covariates included in our final model. Then, we applied our final model to our test data subset. We compared parameter estimates and conclusions, to ensure that the model fits similarly in both subsets of data.

# 4. Results

## 4.1 Exploratory/Descriptive analysis

From all data, we identified 10,502 nursing homes with complete data to include in our analysis. Complete description of the cohort included in the study in the Supplementary Files.

Figure 1

[Table 1](#tbl-summarytable) shows a summary of the data.

Note the loading of the data providing a **relative** path using the ../../ notation. (Two dots means a folder up). You never want to specify an **absolute** path like C:\ahandel\myproject\results\ because if you share this with someone, it won’t work for them since they don’t have that path. You can also use the here R package to create paths. See examples of that below.

Table 1: Data summary table.

| skim\_type | skim\_variable | n\_missing | complete\_rate | character.min | character.max | character.empty | character.n\_unique | character.whitespace | numeric.mean | numeric.sd | numeric.p0 | numeric.p25 | numeric.p50 | numeric.p75 | numeric.p100 | numeric.hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| character | cms\_certification\_number\_ccn | 0 | 1.0000000 | 6 | 6 | 0 | 10502 | 0 | NA | NA | NA | NA | NA | NA | NA | NA |
| character | Provider.State | 0 | 1.0000000 | 2 | 2 | 0 | 44 | 0 | NA | NA | NA | NA | NA | NA | NA | NA |
| character | Ownership.Type | 0 | 1.0000000 | 17 | 38 | 0 | 13 | 0 | NA | NA | NA | NA | NA | NA | NA | NA |
| character | state | 0 | 1.0000000 | 2 | 2 | 0 | 44 | 0 | NA | NA | NA | NA | NA | NA | NA | NA |
| character | COMP\_PERF | 0 | 1.0000000 | 28 | 35 | 0 | 3 | 0 | NA | NA | NA | NA | NA | NA | NA | NA |
| numeric | Average.Number.of.Residents.per.Day | 18 | 0.9982860 | NA | NA | NA | NA | NA | 86.151450 | 49.5040330 | 1.00000 | 54.700000 | 78.400000 | 104.500000 | 708.00000 | ▇▁▁▁▁ |
| numeric | Adjusted.Total.Nurse.Staffing.Hours.per.Resident.per.Day | 262 | 0.9750524 | NA | NA | NA | NA | NA | 3.666884 | 0.8969378 | 0.03568 | 3.106247 | 3.541045 | 4.070308 | 11.63592 | ▁▇▁▁▁ |
| numeric | Staffing.Rating | 154 | 0.9853361 | NA | NA | NA | NA | NA | 2.503576 | 1.3003879 | 1.00000 | 1.000000 | 2.000000 | 4.000000 | 5.00000 | ▇▆▆▅▂ |
| numeric | Total.Weighted.Health.Survey.Score | 28 | 0.9973338 | NA | NA | NA | NA | NA | 68.595685 | 78.9925164 | 0.00000 | 22.667000 | 44.000000 | 85.625250 | 1099.67000 | ▇▁▁▁▁ |
| numeric | Number.of.Citations.from.Infection.Control.Inspections | 12 | 0.9988574 | NA | NA | NA | NA | NA | 1.550524 | 2.6305139 | 0.00000 | 0.000000 | 1.000000 | 2.000000 | 43.00000 | ▇▁▁▁▁ |
| numeric | HAI\_039\_Observed | 0 | 1.0000000 | NA | NA | NA | NA | NA | 7.449629 | 7.3505220 | 0.00000 | 3.000000 | 5.000000 | 10.000000 | 97.00000 | ▇▁▁▁▁ |
| numeric | HAI\_039\_Obs\_Rate | 0 | 1.0000000 | NA | NA | NA | NA | NA | 7.854995 | 4.9329914 | 0.00000 | 4.350000 | 7.140000 | 10.470000 | 50.00000 | ▇▃▁▁▁ |
| numeric | HAI\_039\_Denominator | 0 | 1.0000000 | NA | NA | NA | NA | NA | 99.189392 | 86.3619371 | 25.00000 | 44.000000 | 72.000000 | 121.000000 | 1129.00000 | ▇▁▁▁▁ |
| numeric | HAI\_039\_RS\_Rate | 0 | 1.0000000 | NA | NA | NA | NA | NA | 7.677486 | 1.8052256 | 3.12000 | 6.390000 | 7.410000 | 8.680000 | 21.66000 | ▅▇▁▁▁ |

## 4.2 Basic statistical analysis

*To get some further insight into your data, if reasonable you could compute simple statistics (e.g. simple models with 1 predictor) to look for associations between your outcome(s) and each individual predictor variable. Though note that unless you pre-specified the outcome and main exposure, any “p<0.05 means statistical significance” interpretation is not valid.*

[Figure 1](#fig-result) shows a scatterplot figure produced by one of the R scripts.

|  |
| --- |
| Figure 1: Height and weight stratified by gender. |

## 4.3 Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You want the code to produce figures and data ready for display as tables, and save those. Then you load them here.*

Example [Table 2](#tbl-resulttable2) shows a summary of a linear model fit.

Table 2: Linear model fit table.

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | 149.2726967 | 23.3823360 | 6.3839942 | 0.0013962 |
| Weight | 0.2623972 | 0.3512436 | 0.7470519 | 0.4886517 |
| GenderM | -2.1244913 | 15.5488953 | -0.1366329 | 0.8966520 |
| GenderO | -4.7644739 | 19.0114155 | -0.2506112 | 0.8120871 |

# 5. Discussion

## 5.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 5.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 5.3 Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

This paper (Leek & Peng, 2015) discusses types of analyses.

These papers (**mckay2020?**; **mckay2020a?**) are good examples of papers published using a fully reproducible setup similar to the one shown in this template.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like, I just used the generic word references.bib but giving it a more descriptive name is probably better.

# 6. References

Leek, J. T., & Peng, R. D. (2015). Statistics. What is the question? *Science (New York, N.Y.)*, *347*(6228), 1314–1315. <https://doi.org/10.1126/science.aaa6146>

Mukamel, D. B., Saliba, D., Ladd, H., & Konetzka, R. T. (2022). Daily variation in nursing home staffing and its association with quality measures. *JAMA Network Open*, *5*(3), e222051–e222051.

Mukamel, D. B., Saliba, D., Ladd, H., & Konetzka, R. T. (2023). Association of staffing instability with quality of nursing home care. *JAMA Network Open*, *6*(1), e2250389–e2250389.

Thompson, N. D., Penna, A., Eure, T. R., Bamberg, W. M., Barney, G., Barter, D., et al.others. (2020). Epidemiology of antibiotic use for urinary tract infection in nursing home residents. *Journal of the American Medical Directors Association*, *21*(1), 91–96.