

Computational Phenotyping of Operational Context in Florida Skilled Nursing Facilities: A 2023 Analysis of Care Quality

Andre Antonio Alexander Williams

andrewilliams@health.fau.edu

Florida Atlantic University <https://orcid.org/0000-0002-6782-4165>

Research Article

Keywords: Computational Phenotyping, Social Determinants of Health, Skilled Nursing Facility, Machine Learning, Value-Based Reimbursement

Posted Date: October 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11268048/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Background

Current Centers for Medicare & Medicaid Services (CMS) quality evaluations insufficiently account for a skilled nursing facility's (SNF) operational context, potentially penalizing facilities in socially disadvantaged areas.

Objectives

To establish a context-aware framework using computational phenotyping to evaluate how neighborhood-level social determinants of health (SDOH) associate with federal quality metrics.

Research Design:

This retrospective, cross-sectional ecological study applied an unsupervised machine learning pipeline (Partitioning Around Medoids) to merge and analyze 2023 AHRQ and CMS datasets. Generalized Estimating Equations (GEE) were used to evaluate associations between derived phenotypes and quality outcomes, adjusting for facility ownership.

Subjects:

All 696 skilled nursing facilities operating across Florida during the 2023 performance year.

Measures:

Neighborhood-level SDOH variables (64 retained predictors); key facility performance outcomes included 30-day all-cause readmission, Medicare Spending per Beneficiary (MSPB), healthcare-associated infection (HAI) rates, and successful discharge rates.

Results

Computational phenotyping identified two archetypes: Socioeconomically Advantaged ($n = 524$) and Socioeconomically Disadvantaged ($n = 172$) facilities, with the latter heavily clustered in urban coastal corridors. GEE modeling revealed that facilities in the disadvantaged phenotype exhibited a 1.380% higher HAI rate ($p = 0.001$) and a 9.314% lower successful discharge rate ($p < 0.001$) compared to advantaged baselines. Phenotypic differences in readmissions and MSPB were not statistically significant. Non-profit ownership strongly correlated with superior performance across all metrics.

Conclusions

SNFs in disadvantaged neighborhoods face environmental disparities that correlate with higher infection rates and lower discharge success. These findings support a regulatory policy shift toward context-aware risk adjustments and targeted structural subsidies rather than punitive financial measures.

1. Introduction

Under quality evaluation initiatives led by the Centers for Medicare & Medicaid Services (CMS), skilled nursing facilities (SNFs) in the United States are transitioning toward value-based reimbursement models.¹ Financial rewards are now aligned with critical performance benchmarks: the SNF 30-Day All-Cause Readmission Measure, Medicare Spending per Beneficiary (MSPB), healthcare-associated infections (HAI), and the rate of successful discharge to home or community.^{1,2}

While CMS adjusts quality metrics for race, gender, and patient complexity, a critical gap persists: it insufficiently evaluates a facility's operational context.^{3,4} Neighborhood-level factors, including economic, demographic, educational, and physical infrastructure variables, can account for up to 55% of the variance in health outcomes.^{5,6} Failing to integrate ZIP code-level SDOH into CMS evaluation models may inequitably penalize facilities in socially disadvantaged areas.^{7,8}

Existing literature advocates for social risk adjustment in value-based programs,^{9,10} yet traditional linear models alone are insufficient to capture the multidimensional relationships among clinical outcomes and neighborhood deprivation. This study addresses this limitation by introducing computational phenotyping, a machine-learning approach that identifies emergent clusters of facilities based on shared characteristics, to quantify the correlation between a facility's environment and its CMS performance metrics.^{5,11} By integrating 2023 CMS data with the 2023 AHRQ Community-Level Health (CLH) Database for all 696 Florida SNFs, this research pursues three research questions: (RQ1) What computational phenotypes emerge from clustering facilities by neighborhood-level factors? (RQ2) How are these phenotypes distributed geographically across Florida? (RQ3) How do phenotypes correlate with readmission rates, HAI rates, MSPB, and successful discharge rates? Finally, this study highlights critical policy considerations needed to foster equitable quality assessment.

Florida was selected due to its exceptional SNF density (providing statistical robustness for ML clustering), geographic and demographic diversity (encompassing rural communities, retirement hubs, and multicultural cities), and regulatory consistency within a single-state design that isolates ZIP code-level operational context from multi-state governance variation.

2. Methods

This study employed a retrospective cross-sectional ecological design using secondary analysis of publicly available data. Exemption from federal regulations under 45 CFR 46.104(4ii) was granted by the

XXX Institutional Review Board (IRB). All data processing, machine learning, and statistical analyses were conducted in R (version 4.5.2).

2.1 Data Sources and Linkage

Two publicly available repositories were linked using ZIP codes. The 2023 AHRQ Community-Level Health (CLH) Database provides 346 ZIP code-level SDOH variables organized across five domains: economic context, education, healthcare context, physical infrastructure, and social context (<https://www.ahrq.gov/data/innovations/clh-data.html>). The 2023 CMS SNF Facility-Level Dataset provides facility-specific performance outcome metrics (readmission rates, MSPB, HAI rates, successful discharge rates) and ownership classification (for-profit, non-profit, government) for Florida SNFs (<https://data.cms.gov/provider-data/archived-data/nursing-homes>). These two datasets were merged at the facility level using ZIP codes to link community-level SDOH with facility performance.

2.2 Data Preprocessing and Feature Selection

From the 346 AHRQ CLH variables, only those with $\leq 20\%$ missingness were retained. Missing values were imputed using the missRanger algorithm¹² with 50 decision trees, predictive mean matching ($k = 5$), and all continuous AHRQ variables were standardized (mean-centered and divided by the standard deviation). A multivariate random forest model (randomforestSRC)¹³ was used to concurrently analyze all four outcome variables with 1,000 trees. Predictor influence was quantified via permutation-based Variable Importance (VIMP), standardized and averaged across outcomes; the top 20% of variables by VIMP were retained, yielding 64 candidate predictors for clustering.

2.3 Clustering and AI-Assisted Phenotyping

To address RQ1, we implemented the unsupervised machine learning workflow described above to derive computational phenotypes. The Partitioning Around Medoids (PAM) algorithm was applied to the 64 retained variables using Euclidean distance.¹⁴ The optimal number of clusters was determined by maximizing the average silhouette width. To delineate cluster characteristics, independent-samples t-tests with Benjamini-Hochberg False Discovery Rate (FDR) correction ($\alpha = 0.05$) were conducted across all predictors.¹⁵ The resulting cluster profiles, each variable's mean and semantic label, were exported to a spreadsheet and submitted to Google's Gemini LLM¹⁶ with the following prompt: "Based on the spreadsheet, estimate 1 and estimate 2 contain the means for two unnamed groups. Based on the table, provide a name and definition for each group." The LLM-generated nomenclature was reviewed and verified prior to use.

2.4 Geospatial Visualization

Facility coordinates were converted to spatial features (sf) objects using the st_as_sf function¹⁷ under the World Geodetic System 1984 (WGS 84) EPSG: 4326 coordinate system. Facilities were overlaid on a Florida county-boundary base map¹⁸ within the ggplot2 framework,¹⁹ color-coded by phenotype cluster

(Dark Purple #440154 = Socioeconomically Advantaged; Yellow #FDE725 = Socioeconomically Disadvantaged). Geographic distribution was evaluated descriptively to address RQ2.

2.5 Statistical Modeling

Generalized Estimating Equations (GEE) ²⁰ were used to assess associations between computational phenotype clusters and facility-level performance metrics, with observations clustered at the administrative county level (exchangeable working correlation structure) to account for spatial dependence. A Gaussian family with an identity link was specified. Separate GEE models were fit for each of the four outcomes, with the AI-derived phenotype assignment as the primary predictor and facility ownership type as a structural covariate. This analytical approach addressed research question RQ3. P-values were FDR-adjusted; statistical significance was set at $\alpha = 0.05$.

2.6 Author Reflexivity and Positionality Statement

For transparency, the author acknowledges the personal experiences that informed the study's conceptualization. A key motivation for this research arose from witnessing the profound decline of a close family friend in a post-acute and long-term care facility. Her final days were characterized by severe, largely preventable complications, including advanced pressure ulcers, while clinical management remained primarily focused on palliative pain control rather than proactive, system-level intervention.

The emotional impact of this experience highlights the urgent need to examine structural vulnerabilities within nursing homes and skilled nursing facilities. Although this study uses rigorous, objective machine learning and statistical methods to evaluate institutional performance, it is driven by a commitment to understanding operational contexts and advocating for equitable, context-sensitive systemic support. Ultimately, this work seeks to improve the dignity and quality of care for vulnerable populations.

3. Results

Table 1 presents descriptive statistics for all 696 Florida SNFs included in the 2023 performance year. The mean successful discharge rate was 42.0% (SD = 14.6), the mean 30-day readmission rate was 11.4% (SD = 3.82), the mean HAI rate was 8.46% (SD = 4.03), and the mean MSPB was 1.07 (SD = 0.154). The majority of facilities were for-profit (534; 76.7%), followed by non-profit (143; 20.5%) and government (19; 2.7%).

Table 1
Descriptive Statistics for All Florida SNFs (n = 696), 2023 Performance Year

Category	Indicator	Overall (n = 696)
Readmission Rate	Mean (SD)	11.4 (3.82)
	Median [Min, Max]	11.1 [0, 26.3]
MSPB Score	Mean (SD)	1.07 (0.154)
	Median [Min, Max]	1.05 [0.520, 1.68]
Successful Discharge Rate	Mean (SD)	42.0 (14.6)
	Median [Min, Max]	42.1 [3.51, 83.3]
HAI Rate	Mean (SD)	8.46 (4.03)
	Median [Min, Max]	8.00 [0, 36.5]
Ownership Type:	For profit	534 (76.7%)
	Government	19 (2.7%)
	Non-profit	143 (20.5%)
<i>The following statistical summary details the operational and performance characteristics of 696 Florida skilled nursing facilities during the 2023 performance year. The data encompasses distributions of facility ownership alongside core performance metrics, including Medicare Spending per Beneficiary (MSPB), readmission rates, healthcare-associated infection (HAI) rates, and successful discharge rates, presented through central tendency measures (mean, SD, median, and range) for continuous variables and frequency distributions for categorical markers.</i>		

3.1 Computational Phenotype Characterization

Sixty-four variables were retained after feature selection. PAM clustering yielded a two-cluster solution with an overall average silhouette width of 0.26 (Cluster 1: 0.34; Cluster 2: 0.06). Although t-tests indicated significant mean differences across 59 of 64 variables ($p < .05$), the low silhouette widths indicate substantial boundary overlap; clusters should therefore be interpreted as thematic archetypes rather than discrete, non-overlapping entities. The LLM analysis identified two phenotypes: the Socioeconomically Advantaged group (n = 524; 75.29%) and the Socioeconomically Disadvantaged group (n = 172; 24.71%). Supplemental Table 1 presents the LLM results. Table 2 presents selected SDOH characteristics for each cluster.

Table 2
Selected SDOH Characteristics of Machine Learning–Derived Computational Phenotypes

Domain	Indicator	Socioeconomically Advantaged (n = 524)	Socioeconomically Disadvantaged (n = 172)
Demographics	Non-Hispanic White	67.36%	25.59%
	Hispanic	14.85%	37.81%
	Black	10.88%	31.80%
	Foreign-born	13.15%	34.64%
	English-only speakers	82.84%	54.26%
Economic Status	Homeownership	68.86%	48.71%
	Renters	31.14%	51.23%
	Median Income (Black residents)	\$62,664	\$48,076
	SNAP use below poverty threshold	3.71%	9.69%
Health & Social	Medicaid / means-tested coverage	12.20%	23.21%
	Below poverty threshold	17.58%	29.12%
Physical Infrastructure	High-density housing (10 + units)	16.45%	30.20%
	Housing units built before 1979	35.58%	53.76%
<i>This table summarizes key demographic, economic, health, and infrastructure indicators used to distinguish between the two computational phenotypes, Socioeconomically Advantaged (n = 524) and Socioeconomically Disadvantaged (n = 172), of Florida skilled nursing facilities, with all differences marked as statistically significant ($p < .05$, FDR-adjusted).</i>			

3.2 Geospatial Distribution of Phenotypes

Figure 1 illustrates the non-random spatial distribution of SNF phenotypes across Florida. Socioeconomically Disadvantaged facilities are heavily concentrated in Southeast Florida, Miami-Dade (94.44%), Broward (72.73%), and Palm Beach (44.44%) counties, and in central urban corridors (Hillsborough and Osceola counties; 40.0% each). In contrast, Pinellas County, despite having the state's highest SNF volume (n = 65), shows only a 4.62% prevalence of disadvantaged facilities, and Brevard, Lake, Pasco, and Sarasota counties contain no disadvantaged-phenotype facilities.

3.3 GEE Regression Results

Table 3 presents consolidated GEE results for all four outcomes. Facilities in the Socioeconomically Disadvantaged cluster demonstrated significantly higher HAI rates (1.380%; $p = 0.001$) and significantly lower successful discharge rates (-9.314% ; $p < 0.001$) compared to the Socioeconomically Advantaged reference group. No significant associations were found for readmission rates (0.515%; $p = 0.182$) or MSPB (0.032; $p = 0.063$). Non-profit facilities showed statistically significant improvements over for-profit facilities across all four outcomes; government facilities showed significantly lower HAI rates and higher discharge rates than for-profit SNFs.

Table 3
Consolidated GEE Regression Results for All Four Outcomes, Adjusted for Ownership Type

Term	Readmission Rate Estimate (p-value)	MSPB Estimate (p-value)	HAI Rate Estimate (p-value)	Successful Discharge Estimate (p-value)
Intercept	11.433 (< 0.001)	1.069 (< 0.001)	8.288 (< 0.001)	43.041 (< 0.001)
Cluster: Socioeconomically Advantaged	Reference	Reference	Reference	Reference
Cluster: Socioeconomically Disadvantaged	0.515 (0.182)	0.032 (0.063)	1.380 (0.001)*	-9.314 (< 0.001)*
Ownership: For-Profit	Reference	Reference	Reference	Reference
Ownership: Government	-1.070 (0.128)	-0.003 (0.958)	-2.690 (< 0.001)*	7.846 (0.010)*
Ownership: Non-profit	-1.311 (0.003)*	-0.052 (0.040)*	-1.270 (0.001)*	8.346 (< 0.001)*
<i>This table presents the Generalized Estimating Equation (GEE) regression estimates and associated FDR-adjusted p-values for four performance metrics: 30-day all-cause readmission rate, Medicare Spending per Beneficiary (MSPB), healthcare-associated infection (HAI) rate, and successful discharge rate, evaluating the impact of computational phenotype (Socioeconomically Advantaged vs. Disadvantaged) while adjusting for facility ownership type (For-profit, Government, Non-profit). * implies the p-value is less than 0.05.</i>				

4. Discussion

This computational phenotyping study establishes an empirical foundation for quantifying the relationship between a facility's operational context and its CMS quality performance. By transcending broad socioeconomic indices, the analysis delineates data-derived groupings, the Socioeconomically Advantaged and Socioeconomically Disadvantaged phenotypes, that correspond to structurally distinct ZIP code environments (RQ1). The PAM algorithm identified these clusters based on meaningful statistical differences across 59 SDOH variables, though the modest silhouette widths (0.26 overall) indicate boundary fuzziness characteristic of SDOH-based data, where variables exist along a continuum

rather than within natural, siloed categories. These clusters should therefore be interpreted as thematic archetypes, not discrete entities.

Geospatial analysis (RQ2) confirms a non-random geographic pattern: Socioeconomically Disadvantaged facilities are disproportionately concentrated in Southeast Florida (Miami-Dade, Broward, and Palm Beach counties) and in central urban corridors, while being nearly absent in suburban and rural counties such as Brevard, Pasco, and Sarasota.

In response to RQ3, the Socioeconomically Disadvantaged phenotype is associated with significantly higher HAI rates and lower successful discharge rates, but not with readmission rates or MSPB. The absence of significant associations for the latter two metrics is theoretically plausible: readmission rates and MSPB are assessed longitudinally and may be more sensitive to individual-level patient factors, such as social support in the patient's home ZIP code, rather than to the facility's neighborhood environment.^{2,21} Further investigation is warranted to determine whether policy, patient-level context, or facility-specific frameworks predominate for these outcomes.

The analysis also underscores the critical role of ownership structure. Non-profit facilities demonstrated significantly better performance across all four metrics, and government facilities showed significantly lower HAI rates and higher discharge rates than for-profit entities, which constitute over 76% of Florida's SNF landscape.

By integrating an extensive array of SDOH variables, this research provides a more detailed and statistically robust evaluation of a facility's operational environment. This methodology surpasses conventional techniques that often rely on a limited set of factors or on ambiguous composite indicators with poor interpretability.

The current investigation reinforces and broadens the empirical understanding of the nexus between community-level factors and SNF quality benchmarks. Prior research indicated that incremental increases in neighborhood-level poverty and higher proportions of minority residents correlate with substantial reductions in CMS quality scores;⁷ these observations are mirrored in the significantly elevated HAI rates and diminished successful discharge rates identified within this study's Socioeconomically Disadvantaged archetype. Furthermore, the two-cluster framework generated via PAM aligns with unsupervised machine learning methodologies and supports the validation that SDOH-based computational phenotyping reveals distinct outcome profiles.⁵ The lack of statistical significance regarding readmission rates supports previous findings, suggesting that 30-day readmission risk is primarily influenced by individual-level patient complexities and longitudinal social support rather than a facility's immediate neighborhood infrastructure.^{22,23} Additionally, the superior performance of non-profit entities over their for-profit counterparts observed in this analysis is consistent with research showing improved functional outcomes and lower hospitalization risk in general healthcare non-profit settings.²⁴ Collectively, this literature supports the premise that while neighborhood-level operational context is a

robust predictor of HAI and discharge success, risk adjustment of readmission metrics requires more granular, patient-level integration of SDOH.

The neighborhood-level variances identified between the two AI-generated clusters align with established literature regarding residential segregation and social determinants of health. Predominantly non-Hispanic/Latino white ZIP codes typically demonstrate preferable economic indicators, such as elevated median incomes and homeownership, relative to more racially diverse communities, reflecting deep-seated socio-spatial stratification.²⁵ Conversely, areas with significant minority representation often face structural economic burdens, including higher housing density.²⁶ Empirical evidence suggests that these racially patterned profiles are associated with decreased access to resources, accounting for the clustering of socioeconomic advantage and disadvantage observed in this study's Florida SNF data.

A major limitation of this study is that cluster phenotypes and statistical analyses are restricted to the ZIP code level. Consequently, these findings describe a facility's operational environment rather than individual resident SDOH. Furthermore, LLM-derived nomenclature is nondeterministic and potentially ambiguous, which may limit the replicability and generalization of cluster names. While PAM clustering identified two groups with statistically significant differences, silhouette widths suggest a weak structural partition. These groups likely represent emergent trends rather than discrete entities, a common finding in SDOH-based analysis where variables exist along a continuum. The gap between high statistical significance and modest silhouette scores indicates distinct group centers but significant boundary overlap. Thus, these clusters should be viewed as thematic archetypes. Future research should incorporate additional features to better refine these subgroups.

Finally, this study maintains methodological neutrality, affirming the worth of every community without stigmatization. The primary objective is mapping systemic disparities to help policymakers implement targeted support, ensuring equitable and measurable health outcomes across all ZIP codes in Florida. All ZIP codes matter.

4.1 Strategic Policy Implications

To address inequities within CMS evaluation frameworks, these findings advocate for a data-driven, context-aware policy approach. Because punitive financial mechanisms are inadequate for facilities functioning under significant structural strain, targeted interventions are proposed to foster equitable quality assessment:

1. Establishment of an Intermediate Policy Tier: Using the proximity of facilities to phenotype medoids, an intermediate tier can be delineated. This group includes facilities in the Socioeconomically Advantaged cluster whose Euclidean distance to the Socioeconomically Disadvantaged medoid falls within that cluster's maximum radius. To ensure fairness, these facilities should be included in the same intervention frameworks as those designated for the disadvantaged cluster.
2. Workforce Stability and Retention Grants: Given the established link between staff turnover and lower discharge success rates,²⁷ which this analysis found to be 9.314% lower in Socioeconomically Disadvantaged clusters, performance-based grants are recommended to stabilize the workforce and enhance discharge outcomes, with the facilities in the intermediate tier also receiving these grants.
3. Infection Preventionist (IP) Staffing Adequacy and Training Grants: CMS currently mandates that all certified SNFs appoint an IP, but only on at least a part-time basis, and does not specify a minimum staffing ratio or adjust requirements based on facility size, patient acuity, or neighborhood disadvantage.^{28,29} Because HAI incidence tracks with IP staffing adequacy and experience,³⁰ targeted grants should fund risk-adjusted IP staffing (e.g., 1.0 full-time equivalent [FTE] for facilities exceeding 100 beds, 0.5 FTE for smaller facilities) and specialized infection prevention and control certification for facilities in the Socioeconomically Disadvantaged or intermediate tier. The approach strengthens enforcement and resourcing of an existing federal requirement rather than establishing a new type of regulation.
4. Region-Specific Federal Support: Direct federal funding should be allocated to counties where at least 50% of SNFs are Socioeconomically Disadvantaged or included in the intermediate tier, such as Miami-Dade and Broward counties.

This research introduces a computational phenotyping framework to quantify how neighborhood-level operational context associates with CMS quality benchmarks for Florida SNFs. Facilities located in Socioeconomically Disadvantaged ZIP codes, which are characterized by higher minority representation, increased Medicaid enrollment, and lower homeownership, demonstrate higher HAI rates and lower successful discharge rates. These facilities are heavily concentrated in Southeast and Central Florida. By providing a measurable foundation for context-aware risk adjustment, this framework enables a policy shift from punitive measures toward data-driven, equity-focused structural support.

Abbreviations

AHRQ

Agency for Healthcare Research and Quality

CLH

Community-Level Health

CMS

Centers for Medicare & Medicaid Services

FDR

False Discovery Rate

GEE
Generalized Estimating Equations
HAI
Healthcare-Associated Infections
IP
Infection Preventionist
IRB
Institutional Review Board
LLM
Large Language Model
MSPB
Medicare Spending per Beneficiary
PAM
Partitioning Around Medoids
SDOH
Social Determinants of Health
SNF
Skilled Nursing Facility
VIMP
Variable Importance
WGS 84
World Geodetic System 1984.

Declarations

Funding:

The author received no financial support for this article's research, writing, or publication.

Competing Interests: The author is an employee of Florida Atlantic University and reports no financial interests or potential conflicts of interest.

Contributions: The author contributed to the study's conception or design and to the acquisition, analysis, or interpretation of the data. The author drafted or critically revised the manuscript and gave final approval of the version submitted for publication. The author agrees to be accountable for all aspects of the work, ensuring integrity and accuracy.

Declaration of Generative AI Use: During manuscript preparation, the author used Google Gemini (gemini-3.1-pro-preview) for cluster nomenclature and language editing, Grammarly (v1.2.234.1829), and Writefull (v2025.50.0). The author reviewed and edited all AI-assisted content and takes full responsibility for the final publication.

Acknowledgements:

Financial Disclosure: This work, including its research and subsequent publication, was conducted without any external financial support.

Competing Interests: The author is an employee of Florida Atlantic University and reports no financial interests or potential conflicts of interest.

Data Availability:

Data are openly available at <https://www.ahrq.gov/data/innovations/clh-data.html> and <https://data.cms.gov/provider-data/archived-data/nursing-homes>.

References

1. Wheeler S, Lin SC, Hammond G et al (2025) Penalties in 2024 in Medicare's Skilled Nursing Facility Value-Based Purchasing Program. *J Am Geriatr Soc* 73(10):3021–3030. 10.1111/jgs.70025
2. Burke RE, Xu Y, Rose L (2022) Skilled Nursing Facility Performance and Readmission Rates Under Value-Based Purchasing. *JAMA Netw Open* 5(2). 10.1001/jamanetworkopen.2022.0721
3. Neuman MD, Wirtalla C, Werner RM (2014) Association Between Skilled Nursing Facility Quality Indicators and Hospital Readmissions. *JAMA* 312(15):1542–1542. 10.1001/jama.2014.13513
4. Nong P, Hamasha R, Platt J (2024) Equity and AI Governance at Academic Medical Centers. *Am J Manag Care* 30. 10.37765/ajmc.2024.89555
5. Bhavnani SK, Zhang W, Bao D et al (2023) Subtyping Social Determinants of Health in *All of Us*: Network Analysis and Visualization Approach. *Health Informatics*. Preprint posted online January 28. 10.1101/2023.01.27.23285125
6. Khanna N (2020) Complexities in Integrating Social Risk Assessment into Health Care Delivery. *J Am Board Fam Med* 33(2):179–181. 10.3122/jabfm.2020.02.200016
7. Park YJ, Martin EG (2017) Geographic Disparities in Access to Nursing Home Services: Assessing Fiscal Stress and Quality of Care. *Health Serv Res* 53:2932–2951. 10.1111/1475-6773.12801
8. Zhang Y, Li J, Yu J, Braun R, Casalino LP (2021) Social Determinants of Health and Geographic Variation in Medicare per Beneficiary Spending. *JAMA Netw Open* 4(6). 10.1001/jamanetworkopen.2021.13212
9. Johnston KJ, Wen H, Schootman M, Maddox KEJ (2019) Association of Patient Social, Cognitive, and Functional Risk Factors with Preventable Hospitalizations: Implications for Physician Value-Based Payment. *J Gen Intern Med* 34(8):1645–1652. 10.1007/s11606-019-05009-3
10. Rogstad TL, Gupta S, Connolly J, Shrank WH, Roberts ET (2022) Social Risk Adjustment In The Hospital Readmissions Reduction Program: A Systematic Review And Implications For Policy. *Health Aff (Millwood)* 41(9):1307–1315. 10.1377/hlthaff.2022.00614

11. Irvin JA, Kondrich AA, Ko M et al (2020) Incorporating machine learning and social determinants of health indicators into prospective risk adjustment for health plan payments. *BMC Public Health* 20(1):608. 10.1186/s12889-020-08735-0
12. Mayer M (2024) *missRanger: Fast Imputation of Missing Values*. 10.32614/CRAN.package.missRanger
13. Ishwaran H, Kogalur UB (2019) Fast unified random forests for survival, regression, and classification (RF-SRC). *R Package Version*. ;2(1)
14. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2026) *Cluster: Cluster Analysis Basics and Extensions*. <https://CRAN.R-project.org/package=cluster>
15. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Stat Methodol* 57(1):289–300. 10.1111/j.2517-6161.1995.tb02031.x
16. Google, Gemini (2026) Published online <https://gemini.google.com/>
17. Pebesma E (2018) Simple Features for R: Standardized Support for Spatial Vector Data. *R J* 10(1):439–446. 10.32614/RJ-2018-009
18. Walker K (2025) *Tigris: Load Census TIGER/Line Shapefiles*. 10.32614/CRAN.package.tigris
19. Wickham H (2016) *Ggplot2: Elegant Graphics for Data Analysis*. Springer-, New York. <https://ggplot2.tidyverse.org>
20. McDaniel LS, Henderson NC, Rathouz PJ (2013) Fast pure R implementation of GEE: application of the Matrix package. *R J* 5(1):181–187
21. Kaknes DA, Barbour TB, Ji W, Stein SM, Naylor MD, Hanlon AL (2024) Interventional Analytics in Skilled Nursing Facilities Associated With Reduced Readmissions. *Am J Manag Care*. ;30(6 Spec No):SP478.
22. Bensken WP, Alberti PM, Koroukian SM (2021) Health-Related Social Needs and Increased Readmission Rates: Findings from the Nationwide Readmissions Database. *J Gen Intern Med* 36(5):1173–1180. 10.1007/s11606-021-06646-3
23. Kind A, Jencks S, Brock J et al (2014) Neighborhood Socioeconomic Disadvantage and 30-Day Rehospitalization. *Ann Intern Med* 161(11):765–774. 10.7326/m13-2946
24. Ronald LA, McGregor MJ, Harrington C, Pollock AM, Lexchin J (2016) Observational Evidence of For-Profit Delivery and Inferior Nursing Home Care: When Is There Enough Evidence for Policy Change? *PLoS Med* 13(4). 10.1371/journal.pmed.1001995
25. Swope CB, Hernández D (2019) Housing as a determinant of health equity: A conceptual model. *Soc Sci Med* 243:112571–112571. 10.1016/j.socscimed.2019.112571
26. Yang Y, Cho A, Nguyen QC, Nsoesie EO (2023) Association of Neighborhood Racial and Ethnic Composition and Historical Redlining With Built Environment Indicators Derived From Street View Images in the US. *JAMA Netw Open* 6(1). 10.1001/jamanetworkopen.2022.51201

27. Shen K, McGarry BE, Gandhi AD (2023) Health Care Staff Turnover and Quality of Care at Nursing Homes. *JAMA Intern Med* 183(11):1247. 10.1001/jamainternmed.2023.5225
28. Herzig C, Stone PW, Castle NG, Pogorzelska-Maziarz M, Larson E, Dick AW (2015) Infection Prevention and Control Programs in US Nursing Homes: Results of a National Survey. *J Am Med Dir Assoc* 17(1):85–88. 10.1016/j.jamda.2015.10.017
29. Stone PW, Agarwal M, Pogorzelska-Maziarz M (2020) Infection preventionist staffing in nursing homes. *Am J Infect Control* 48(3):330–332. 10.1016/j.ajic.2019.12.010
30. Clifford RJ, Newhart D, Laguio-Vila MR, Gutowski JL, Bronstein MZ, Lesho EP (2022) Infection preventionist staffing levels and rates of 10 types of healthcare-associated infections: A 9-year ambidirectional observation. *Infect Control Hosp Epidemiol* 43(11):1641–1646. 10.1017/ice.2021.507

Figures

Geographic Distribution of Nursing Facilities by Cluster Group in Florida

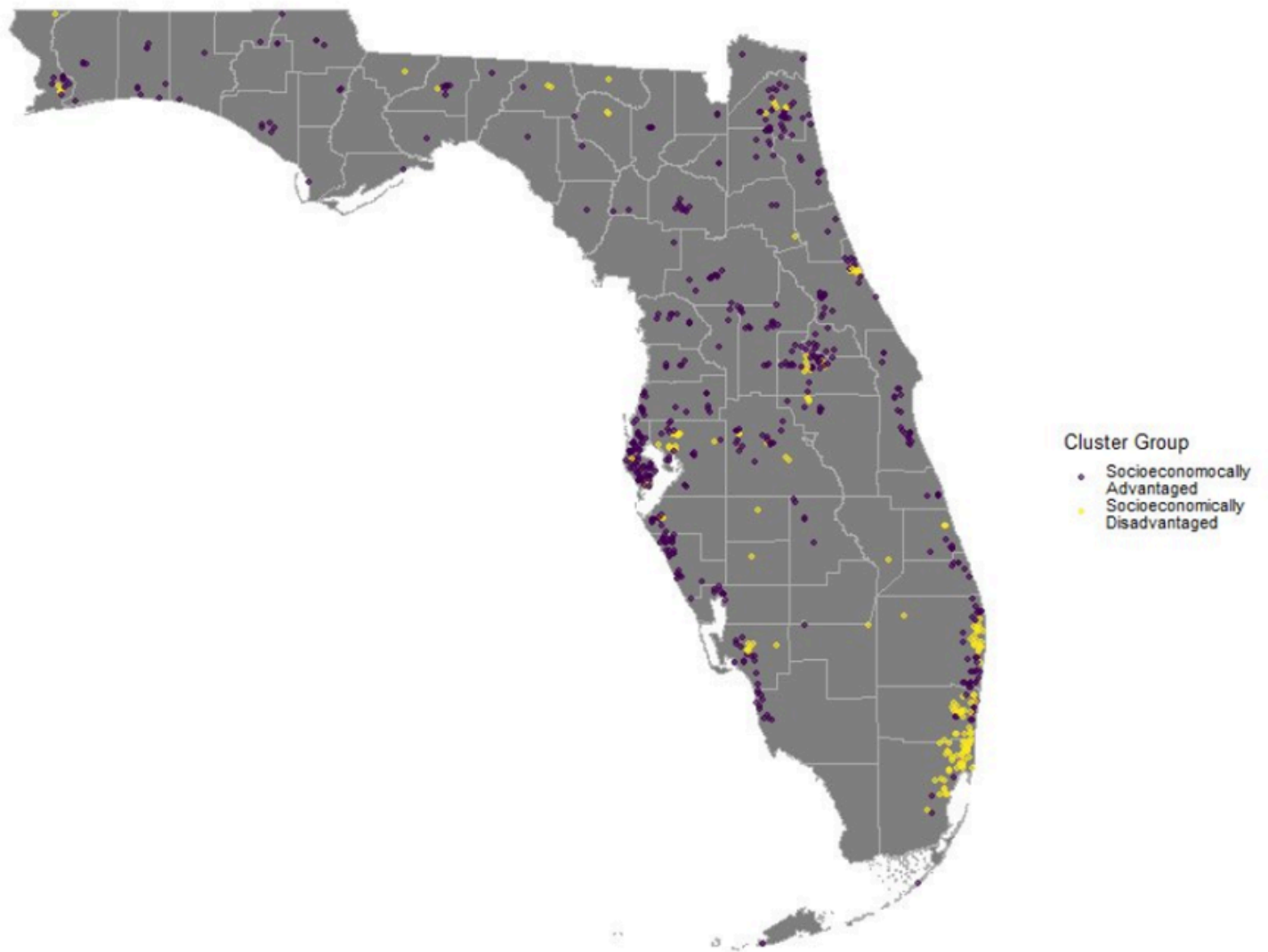


Figure 1

Geospatial distribution of Florida Skilled Nursing Facilities (SNFs) categorized by computational phenotype: Socioeconomically Advantaged (Dark Purple, #440154) and Socioeconomically Disadvantaged (Yellow, #FDE725). The map illustrates the concentration of Socioeconomically Disadvantaged facilities in major urban coastal corridors (Southeast Florida) and their relative sparsity in the rural interior and Panhandle.