Social Vulnerability is Associated with Prehospital Transport Outcome Among Older Adults in New York State

Kaleb J. Frierson (kjf2152)[1], MPH Candidate
Certificate in Applied Biostatistics & Health Data Science

First Reader: Zachary Mannes (zlm2106)[1,2] PhD, MPH

[1]Department of Epidemiology, Columbia University Mailman School of Public Health

[2]Department of Emergency Medicine, Columbia University Medical Center

Submitted in partial fulfillment of the requirements for the degree of Master of Public Health in the Department of Epidemiology at the Columbia University Mailman School of Public Health.

April 2025

Table of Contents

AGKNOWLEDGEMENTS

Dear reader,

My thesis project wouldn't have been obtainable without the contribution and input of a significant list of individuals. First and foremost, I would like to express my gratitude to Zachary Mannes who took on the role of my first reader and mentor throughout my first formal epidemiologic study. Without the quick intervention of Peter Brodie and Jacob DeMay following recent federal obstructions in data delivery, this project would have ceased. They helped me file with New York State and obtain the dataset that fed my analysis. Susan (Susie) Burnett, a beloved friend and colleague with an impressive history of prehospital and academic service, has offered her listening ear and thought provoking input since the birth of this project. Andrea Dávalos, Laura Eierman, and Steven Broyles – excellent mentors, scientists, friends, and people in general – were all professors of mine at SUNY Cortland and largely shaped my identity as a scientist and integrity for doing good science. My closest friends – some of whom have been involved in their own dissertations this year – have offered a much needed social support system throughout this journey. Thank you Brendan, George, Jack, Julia, Nehal, Nicholas, Sara, Sydney, and Taylor for your friendship and inspiration. Last, but certainly not least, my partner Jenna. She has been my measuring stick for success and my motivation to do right by this world since knowing her.

This work is dedicated to my family. I am the first of them to earn a Bachelors Degree, and now a Masters Degree. It is their shoulders that I stand on and to them that I owe each of my achievements.

Thank you, the reader, for taking the time to learn about what I am passionate about. Zach, Susie, Jack, and I plan to address some of the limitations in this work and submit a manuscript for publication using a more comprehensive dataset soon! I urge you to critique my methods, hold me to robust standards, and use those findings that you feel are meaningful to affect change.

ABSTRACT

Objectives: To assess how social vulnerability influences emergency medical service (EMS) utilization and transport outcomes among older adults, including fall-related complaints and racial/ethnic disparities.

Methods: We conducted a cross-sectional study of 1,268,070 EMS activations among adults aged 65 years or older using 2023 statewide data from the New York State Department of Health. Social Vulnerability Index (SVI) scores were linked to scene ZIP codes and categorized as low, moderate, or high. We examined dispatch for a fall-related incident and whether the patient was transported to the hospital. Logistic regression models, stratified by race/ethnicity and adjusted for age and sex, assessed associations between SVI and each outcome.

Results: Fall-related EMS calls were less common in high-SVI areas compared to low-SVI areas, among White patients (OR=0.63, 95% CI: 0.62–0.64) and non-White patients (OR=0.61, 95% CI: 0.57–0.64). Overall, older adults in moderate and high-SVI areas had higher odds of non-transport, especially among non-White patients (OR=2.10, 95% CI: 1.99–2.23), compared to low-SVI areas. However, among fall-related calls, patients in moderate and high-SVI areas were more likely to be transported than those in low-SVI areas.

Conclusions: Social vulnerability and race shape EMS utilization and transport outcomes among older adults. Non-transport is more likely in high-SVI communities. While most EMS encounters for older adults occur in high-SVI areas, we found the opposite of fall-related complaints. These findings may reflect structural barriers to accessing prehospital health resources. They illustrate the need for EMS-specific strategies that address disparities, including enhanced education on social determinants of health, improved data collection, and targeted interventions to ensure equitable prehospital care delivery.

INTRODUCTION

In the United States (US), Emergency Medical Services (EMS) clinicians respond to over 17 million healthcare calls annually[1]. Emergencies among older adults result in a disproportionate number of those activations[2]. One common reason for EMS activation among older adults (people aged 65 years or older) is fall-related incidents which is a leading cause of injury-related deaths in this age group[3]. In 2014 and 2015, falls resulted in nearly 17% of EMS activations among people 65 and older[4]. Approximately 28.7% of U.S. older adults fell in 2014, resulting in approximately $23 billion[5] in healthcare costs. Beyond immediate physical harm, falls can lead to long-term complications including disability, loss of independence, and increased likelihood of nursing home admissions[3–6].

Approximately 5% of all EMS activations result in patient refusal or non-transport after evaluation[7]. Among older adults, this is particularly relevant in fall-related incidents, where EMS frequently responds to "lift-assist" calls – activations in which the patient is assisted after a fall but ultimately not transported to a hospital. While perceived as benign, lift-assist calls carry significant clinical risk. Older adults who refuse transport following a “lift-assist” experience elevated rates of subsequent morbidity and mortality, including increased risk of recurrent falls, hospitalization, and death[8]. The decision to leave a patient at home after a fall-related EMS activation may therefore represent a critical factor in determining health outcomes, particularly for patients who are socially or medically disadvantaged.

Understanding the factors influencing EMS utilization and transport decisions is critical, particularly as the U.S. Census Bureau estimates that by 2030, one in five residents will be of retirement age or older[9]. Disparities in healthcare access, utilization, and outcomes are well-documented across various clinical settings. Social Determinants of Health (SDoH) are recognized as key drivers of health outcomes and intervention points for reducing disparities[10], but their role in shaping patterns of EMS utilization and decision-making remains underexplored. While EMS clinicians are positioned to document information about an individual’s SDoH, they often lack the training and resources to do so, reflected by a lack of relevant data points in EMS records[11].

The Social Vulnerability Index (SVI), developed by the U.S. Centers for Disease Control and Prevention (CDC), is one method to quantify neighborhood-level SDoH. The SVI combines 16 variables of four domains: socioeconomic status, household composition and disability, racial/ethnic minority status and language proficiency, and housing and transportation characteristics[12]. Higher SVI scores indicate greater vulnerability and have been linked to adverse health outcomes and limited access to healthcare services[12–15]. Despite its use to identify the impact of disparities on various clinical outcomes[13,16–18], we’re unaware of prior studies assessing the association of SVI in prehospital falls or transport outcome among older adults.

Using EMS records from the New York State Department of Health (NYS DOH) from 2023, we aimed to characterize EMS activations among older adults as a function of social vulnerability. Specifically, we tested whether older adults residing in zip codes with high SVI scores had a disproportionate burden of fall-related EMS activations and subsequent non-transport to hospitals. We hypothesized that older adults receiving EMS care in high-SVI areas would experience more EMS activations for falls, though have lower rates of hospital transport than those receiving care in low-SVI areas.

METHODS

Study Design and Data Source

This cross-sectional study of patients aged 65 years or older used 2023 EMS activation data obtained from the NYS DOH Division of State Emergency Medical Services. This data captures EMS activations across NYS and includes non-identifying patient demographics, dispatch complaints, and transport outcomes (transport to hospital/non-transport). This study was exempt from Columbia University Institutional Review Board review as all data was de-identified. The initial dataset contained 1,490,959 EMS activations. We removed duplicate activations and excluded those with missing SVI data, transport information, or sex. This yielded a final analytic sample of 1,268,070 EMS encounters (Figure S2).

Exposure

The primary exposure variable was the zip code SVI score. SVI was operationalized into three ordinal categories: Low SVI (0.0–0.5), Moderate SVI (0.5–0.8), and High SVI (>0.8). This categorization facilitates interpretation and aligns with prior research[13,14].

Outcomes

Two outcomes were examined: fall (yes/no), assessed by whether EMS dispatch was for a fall-related complaint, and transport outcome (not transported vs. transported to hospital). Fall-related activations were identified using the "Complaint Reported by Dispatch" field, restricting to activations coded as either fall or medical alarm. Transport outcome was determined using the "Disposition" field.

Analysis

Descriptive statistics were calculated to characterize the study population by SVI category. Categorical variables were summarized as counts and percentages, and age was reported using the median and interquartile range. Differences across SVI categories were assessed using chi-square tests for categorical variables and Kruskal-Wallis tests for age.

Three logistic regression models were used to examine the relationship between SVI and our primary outcomes. The first set of models examined the relationship between SVI and the likelihood of falls. The second set of models examined the relationship between SVI and transport outcome. The third set of models examined the relationship between SVI and transport outcome among fall cases. All models controlled for patient age (continuous) and sex (male vs. female). Although race is a component of the SVI, we stratified all models by race (White vs. non-White). Prior research either suggests that the relationship between area-level vulnerability and health outcomes may differ across racial groups or controls for race in statistical modelling[14,16,17,19,20].

All data cleaning, analyses, and table building were performed in R version 4.4.3 [21] using the following packages: tidyverse[22], gtsummary[23], dagitty[24], ggdag[25], DiagrammeR[26], DiagrammeRsvg[27], rsvg[28], broom[29], labelled[30]. An annotated code for each of these steps is available in the supplement (S3).

RESULTS

Sample Characteristics

A total of 1,268,070 EMS activations were included in the analytic sample (Table 1). Approximately 16% of activations were classified as Low SVI, 28% as Moderate SVI, and 56% as High SVI. The median age of patients decreased with increasing SVI, from 80 years (IQR: 73–87) among those residing in Low SVI areas to 77 years (IQR: 70–84) among those in High SVI areas ($p < 0.001$). Across all SVI groups, a slight majority of activations involved female patients, ranging from 55% to 56% ($p < 0.001$). The racial composition varied notably by SVI category. Among activations from Low SVI areas, 88% involved White patients compared to 52% in High SVI areas, where non-White patients accounted for nearly half (48%) of activations ($p < 0.001$). The proportion of activations identified as fall-related was highest among the Low SVI group (14%) and decreased progressively across Moderate (11%) and High SVI groups (7%) ($p < 0.001$). Similarly, the proportion of activations resulting in non-transport was higher among those from Moderate (11%) and High SVI (11%) areas compared to Low SVI areas (9%) ($p < 0.001$).

Association between SVI and fall-related incidents

Among White patients, residing in areas with Moderate SVI was not associated with the odds of a fall-related EMS dispatch compared to those in Low SVI areas (OR = 1.00; 95% CI: 0.98–1.01; Table 2). However, White patients in High SVI areas had 37% lower odds of a fall-related EMS dispatch compared to those in Low SVI areas (OR = 0.63; 95% CI: 0.62–0.64; $p < 0.001$; Table 2). Among Non-White patients, Moderate SVI was associated with 15% higher odds of a fall-related EMS dispatch (OR = 1.15; 95% CI: 1.09–1.22; $p < 0.001$; Table 2), whereas High SVI was associated with 39% lower odds (OR = 0.61; 95% CI: 0.57–0.64; $p < 0.001$; Table 2) compared to Low SVI areas.

Association between SVI and EMS transportation

White patients in Moderate and High SVI areas had 25% higher odds of non-transport compared to those in Low SVI areas (Moderate SVI: OR = 1.25; 95% CI: 1.23–1.27; High SVI: OR = 1.25; 95% CI: 1.22–1.27; both $p < 0.001$; Table 3). Among non-White patients, the association was stronger. Compared to those in Low SVI areas, patients in Moderate SVI areas had 53% higher odds of non-transport (OR = 1.53; 95% CI: 1.43–1.63; $p < 0.001$; Table 3), and those in High

SVI areas had more than twice the odds of non-transport (OR = 2.10; 95% CI: 1.99–2.23; p < 0.001; Table 3).

Association between SVI and EMS transportation among patients with a fall-related incident

Among patients experiencing a fall-related EMS dispatch, White patients residing in Moderate SVI areas had 15% higher odds of non-transport compared to those in Low SVI areas (OR = 1.15; 95% CI: 1.11–1.19; p < 0.001; Table 4). However, White patients in High SVI areas had 22% lower odds of non-transport (OR = 0.78; 95% CI: 0.75–0.81; p < 0.001; Table 4). In contrast, among non-White patients with fall-related activations, both Moderate and High SVI were associated with lower odds of non-transport (Moderate SVI: OR = 0.77; 95% CI: 0.66–0.91; p = 0.001; High SVI: OR = 0.47; 95% CI: 0.40–0.54; p < 0.001; Table 4).

Table 1. Descriptive statistics summarized by level of social vulnerability.

Characteristic	Low SV N = 205,961[1]	Moderate SV N = 352,133[1]	High SV N = 709,976[1]	p-value[2]
Age	80 (73, 87)	79 (72, 86)	77 (70, 84)	<0.001
Sex				<0.001
Female	115,143 (56%)	197,323 (56%)	389,891 (55%)	
Male	90,818 (44%)	154,810 (44%)	320,085 (45%)	
Race/Ethnicity				<0.001
Non-White	24,218 (12%)	74,111 (21%)	343,394 (48%)	
White	181,743 (88%)	278,022 (79%)	366,582 (52%)	
Dispatch Complaint is Fall				<0.001
No	176,689 (86%)	305,132 (87%)	660,052 (93%)	
Yes	29,272 (14%)	47,001 (13%)	49,924 (7.0%)	
Transport Outcome				<0.001
Transport	187,316 (91%)	313,976 (89%)	632,089 (89%)	
No Transport	18,645 (9.1%)	38,157 (11%)	77,887 (11%)	

[1]Median (Q1, Q3); n (%) [2]Kruskal-Wallis rank sum test; Pearson's Chi-squared test

Table 2. Outputs of the race/ethnicity stratified logistic regression models assessing the relationship between SVI and fall events, holding age and gender constant.

	White (N = 826,347)			Non-White (N = 441,723)		
Characteristic	**OR**[1]	**95% CI**[1]	**p-value**	**OR**[1]	**95% CI**[1]	**p-value**
Social Vulnerability Index (SVI)						
Low SV	—	—		—	—	
Moderate SV	1.00	0.98, 1.01	0.6	1.15	1.09, 1.22	<0.001
High SV	0.63	0.62, 0.64	<0.001	0.61	0.57, 0.64	<0.001
Age (years)	1.02	1.02, 1.02	<0.001	1.01	1.01, 1.02	<0.001
Sex						
Female	—	—		—	—	
Male	0.86	0.85, 0.87	<0.001	0.98	0.95, 1.00	0.078

[1]OR = Odds Ratio, CI = Confidence Interval

Table 3. Outputs of the race/ethnicity stratified logistic regression models assessing the relationship between SVI and hospital non-transport, holding age and gender constant.

	White (N = 826,347)			Non-White (N = 441,723)		
Characteristic	**OR**[1]	**95% CI**[1]	**p-value**	**OR**[1]	**95% CI**[1]	**p-value**
Social Vulnerability Index (SVI)						
Low SV	—	—		—	—	
Moderate SV	1.25	1.23, 1.27	<0.001	1.53	1.43, 1.63	<0.001
High SV	1.25	1.22, 1.27	<0.001	2.10	1.99, 2.23	<0.001
Age (years)	1.00	1.00, 1.00	0.4	1.00	1.00, 1.00	<0.001
Sex						
Female	—	—		—	—	
Male	0.95	0.93, 0.96	<0.001	0.83	0.82, 0.85	<0.001

[1]OR = Odds Ratio, CI = Confidence Interval

Table 4. Outputs of the race/ethnicity stratified logistic regression models assessing the relationship between SVI and hospital non-transport among fall events, holding age and gender constant.

	White (N = 104,570)			Non-White (N = 21,627)		
Characteristic	**OR**[1]	**95% CI**[1]	**p-value**	**OR**[1]	**95% CI**[1]	**p-value**
Social Vulnerability Index (SVI)						
Low SV	—	—		—	—	
Moderate SV	1.15	1.11, 1.19	<0.001	0.77	0.66, 0.91	0.001
High SV	0.78	0.75, 0.81	<0.001	0.47	0.40, 0.54	<0.001
Age (years)	0.99	0.99, 0.99	<0.001	0.99	0.99, 1.00	0.023
Sex						
Female	—	—		—	—	
Male	1.17	1.14, 1.21	<0.001	0.95	0.86, 1.04	0.3

[1]OR = Odds Ratio, CI = Confidence Interval

DISCUSSION & LIMITATIONS

Our study examined how social vulnerability, as measured by the SVI, was associated with EMS dispatch for fall-related incidents and transport outcomes among older adults in New York State. We found that EMS activations among older adults were more frequently located in high-SVI areas, which accounted for over half of all calls. The racial composition of encounters varied by SVI: while 88% of patients in low-SVI areas were White, this proportion dropped to 52% in high-SVI areas. Hospital transport was less likely in moderate and high-SVI areas compared to low-SVI areas. Fall-related encounters among White patients were less common in moderate and high-SVI areas, whereas among non-White patients, such encounters were less likely in high-SVI areas but more common in moderate-SVI areas. Among patients who experienced a fall, those in moderate and high-SVI areas were more likely to be transported to the hospital than those in low-SVI areas. These findings underscore the need to address structural and social barriers that shape prehospital care delivery. EMS systems require enhanced resources – including training, education, and community-specific strategies – to ensure equitable and effective responses across all communities.

White older adults encountering EMS in areas with moderate or high social vulnerability had lower odds of fall-related dispatch complaints compared to those in low-SVI areas. Among non-White individuals, fall-related dispatches were less likely to occur in high-SVI areas but more likely to occur in moderate-SVI areas compared to low-SVI areas. Notably, our data only capture falls for which EMS was activated and therefore don't capture the true incidence. The lower rate of EMS encounters in high-SVI areas may reflect underutilization rather than a lower burden of falls. Individuals in the highest vulnerability areas may face more frequent experiences of healthcare related discrimination and stigma, which can foster mistrust and discourage use of health services[31]. Cost-related concerns may also contribute – individuals with limited income, insurance coverage, or existing medical debt may be dissuaded from seeking medical care from fear of ambulance or subsequent hospital costs. Stroke care research by the American Heart Association has shown that some patients and family members are sometimes reluctant to activate the EMS system due to a lack of autonomy in hospital choice and high costs, among other factors[32]. Further, individuals exposed to high social vulnerability are likely the most resource strained of our cohort and could lack access to mobile phones[33], or have lower levels of health literacy which can impair someones ability to recognize when EMS care is needed[34]. Relative to people in high-SVI areas, those in moderate-SVI areas may experience less barriers to utilizing prehospital resources.

Patients in moderate- and high-SVI areas were less likely to be transported following EMS activation, and among non-White individuals, the odds of non-transprot were more than twixe those of their counterparts encountering EMS in low-SVI areas. These findings suggest that transport to the hospital is administered disproportionately by EMS to older adults based on social vulnerability and race. Racial disparities have been documented among other EMS interventions, including pain management[35] and chemical restraint[36]. EMS clinicians may be unconsciously biased by a patient's demographics or previous notions about the neighborhood to which they are dispatched, resulting in more clinician-initiated refusal of transport. Patient initiatied refusals may also be more common in high-SVI communities due to aforementioned structural barriers like cost or trust in healthcare institutions. Additional research should explore the differences in patient and clinician initiated refusal/non-transport decisions in order to better understand disparities and identify intervention points.

Among older adults who experienced a fall-related dispatch complaint, patterns of EMS transport differed from overall findings. In contrast to the general trend of higher non-transport rates in high-SVI areas, patients with fall-related complaints in moderate and high-SVI areas were more likely to be transported to the hospital. Stratified analyses showed that, among White patients, those in moderate-SVI areas were more likely to be transported, while those in high-SVI areas were less likely to refuse care compared to those in low-SVI areas. For non-White patients, residing in moderate- or high-SVI areas was associated with significantly lower odds of non-transport. These findings may reflect the higher clinical acuity or seriousness of fall-related incidents among socially vulnerable populations. Older adults in high-SVI areas are more likely to have comorbid conditions, such as frailty[37] or cardiovascular disease[14,19], that could increase the risk of injury and need for transport. Cognitive impairment – such as that caused by Alzheimer's disease, which has been associated with high social vulnerability[38] – may limit a patient's ability to refuse transport to the hospital. It is also possible that some fall-related EMS activations occurred in nursing homes or assisted living facilities, where the standard of care often necessitates hospital transport. Because our analysis was limited to scene zip code and did not account for scene type or residential status, further research is needed to better characterize how clinical, cognitive, and contextual factors influence transport decisions in this population.

Our study has several limitations. Our models account for some potential confounding by age, race and ethnicity, and sex; but unmeasured variables such as scene location type, urbanicity, comorbidities, and patient acuity were unavailable in our data and could prove meaningful in interpreting these data. Race and ethnicity were coded by EMS clinicians at the scene and may be subject to misclassification or bias. SVI was operationalized at the scene ZIP code level, which may not fully reflect an individual's exposure to social vulnerability. Further, our outcome variables – particularly transport outcome – are broad and do not distinguish between patient-initiated and provider-recommended non-transport decisions, which could reflect different dynamics.

We provide evidence that social vulnerability is associated with patterns of EMS utilization and transport disposition among older adults, with important variation by race and dispatch complaint. We found that older individuals encountering EMS in areas with higher social vulnerability were less likely to be taken to a hospital. These findings have important implications for EMS systems, which serve clinical and public health roles for their communities. Public health agencies and EMS administrators should invest in strategies that improve trust, equity, and follow-through in prehospital care, particularly in socially vulnerable communities. Policymakers and regulators should reform education standards to include public health topics such as the SDoH and epidemiologic disparities in care. Future research should examine patient-level motivations for refusing transport, provider-level motivations for urging patients to go (or not to go) to the hospital, and test targeted interventions to reduce disparities in EMS care.

REFERENCES

1. Wang HE, Mann NC, Jacobson KE, et al. National Characteristics of Emergency Medical Services Responses in the United States. *Prehospital Emergency Care*. 2013;17(1):8-14. doi:10.3109/10903127.2012.722178

2. Duong HV, Herrera LN, Moore JX, et al. National Characteristics of Emergency Medical Services Responses for Older Adults in the United States. *Prehospital Emergency Care*. 2018;22(1):7-14. doi:10.1080/10903127.2017.1347223

3. Bergen G, Stevens MR, Burns ER. Falls and Fall Injuries Among Adults Aged ≥65 Years — United States, 2014. *MMWR Morb Mortal Wkly Rep*. 2016;65(37):993-998. doi:10.15585/mmwr.mm6537a2

4. Faul M, Stevens JA, Sasser SM, et al. Older Adult Falls Seen by Emergency Medical Service Providers. *American Journal of Preventive Medicine*. 2016;50(6):719-726. doi:10.1016/j.amepre.2015.12.011

5. Ambrose AF, Paul G, Hausdorff JM. Risk factors for falls among older adults: A review of the literature. *Maturitas*. 2013;75(1):51-61. doi:10.1016/j.maturitas.2013.02.009

6. Hamilton MP, Bellolio F, Jeffery MM, et al. Risk of falls is associated with 30-day mortality among older adults in the emergency department. *The American Journal of Emergency Medicine*. 2024;79:122-126. doi:10.1016/j.ajem.2024.02.020

7. Knight S, Olson LM, Cook LJ, Mann NC, Corneli HM, Dean JM. Against all advice: an analysis of out-of-hospital refusals of care. *Ann Emerg Med*. 2003;42(5):689-696. doi:10.1016/S0196064403005249

8. Leggatt L, Van Aarsen K, Columbus M, et al. Morbidity and Mortality Associated with Prehospital “Lift-assist” Calls. *Prehospital Emergency Care*. 2017;21(5):556-562. doi:10.1080/10903127.2017.1308607

9. Bureau, U. C. Bureau, U. C. Older Population and Aging. United States Census Bureau. Published online 2023. https://www.census.gov/topics/population/older-aging.html

10. United States Centers for Disease Control & Prevention. Why Addresing SDoH is Important. Published online 2024. https://www.cdc.gov/about/priorities/why-is-addressing-sdoh-important.html

11. Burnett S, Stemerman R, Innes J, Kaisler M, Crowe R, Clemency B. Social Determinants of Health in EMS Records: A Mixed-methods Analysis Using Natural Language Processing and Qualitative Content Analysis. *Western Journal of Emergency Medicine*. 2013;24. doi:10.5811/WESTJEM.59070

12. Centers for Disease Control and Prevention. Social Vulnerability Index. Published online 2024. https://www.atsdr.cdc.gov/place-health/php/svi/index.html

13. Higginbotham JK, Segovia LM, Rohm KL, Anderson CM, Breitenstein SM. Social Vulnerability Index and Health Outcomes in the United States: A Systematic Review. *Family & Community Health*. 2025;48(2):81-96. doi:10.1097/FCH.0000000000000421

14. Ibrahim R, Sainbayar E, Pham HN, et al. Social Vulnerability Index and Cardiovascular Disease Care Continuum. *JACC: Advances*. 2024;3(7):100858. doi:10.1016/j.jacadv.2024.100858

15. NOAA. In Harm's Way: Hurricane Ida's Impact on Socially Vulnerable Communities. 2022. https://www.nesdis.noaa.gov/news/harms-way-hurricane-idas-impact-socially-vulnerable-communities

16. Duncan AJ, Bloomsburg SJ, Ahmeti M. "Utility of social vulnerability index in trauma: A systematic review." *Injury*. 2024;55(12):112016. doi:10.1016/j.injury.2024.112016

17. Mah JC, Penwarden JL, Pott H, Theou O, Andrew MK. Social vulnerability indices: a scoping review. *BMC Public Health*. 2023;23(1):1253. doi:10.1186/s12889-023-16097-6

18. Tran T, Rousseau MA, Farris DP, Bauer C, Nelson KC, Doan HQ. The social vulnerability index as a risk stratification tool for health disparity research in cancer patients: a scoping review. *Cancer Causes Control*. 2023;34(5):407-420. doi:10.1007/s10552-023-01683-1

19. Jain V, Al Rifai M, Khan SU, et al. Association Between Social Vulnerability Index and Cardiovascular Disease: A Behavioral Risk Factor Surveillance System Study. *Journal of the American Heart Association*. 2022;11(15):e024414. doi:10.1161/JAHA.121.024414

20. Rahman SN, Long JB, Westvold SJ, et al. Area Vulnerability and Disparities in Therapy for Patients With Metastatic Renal Cell Carcinoma. *JAMA Netw Open*. 2024;7(4):e248747. doi:10.1001/jamanetworkopen.2024.8747

21. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2025. https://www.R-project.org/

22. Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. *Journal of Open Source Software*. 2019;4(43):1686. doi:10.21105/joss.01686

23. Sjoberg DD, Whiting K, Curry M, Lavery JA, Larmarange J. Reproducible Summary Tables with the gtsummary Package. *The R Journal*. 2021;13(1):570-580. doi:10.32614/RJ-2021-053

24. Textor J, Zander B van der, Gilthorpe MS, Liśkiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package "dagitty." *International Journal of Epidemiology*. 2016;45(6):1887-1894. doi:10.1093/ije/dyw341

25. Barrett M. *Ggdag: Analyze and Create Elegant Directed Acyclic Graphs*.; 2024. https://CRAN.R-project.org/package=ggdag

26. Iannone R, Roy O. *DiagrammeR: Graph/Network Visualization*.; 2024. https://CRAN.R-project.org/package=DiagrammeR

27. Iannone R. *DiagrammeRsvg: Export DiagrammeR Graphviz Graphs as SVG*.; 2016. https://CRAN.R-project.org/package=DiagrammeRsvg

28. Ooms J. *Rsvg: Render SVG Images into PDF, PNG, (Encapsulated) PostScript, or Bitmap Arrays*.; 2025. https://CRAN.R-project.org/package=rsvg

29. Robinson D, Hayes A, Couch S. *Broom: Convert Statistical Objects into Tidy Tibbles*.; 2024. https://CRAN.R-project.org/package=broom

30. Larmarange J. *Labelled: Manipulating Labelled Data*.; 2025. https://CRAN.R-project.org/package=labelled

31. Shukla M, Schilt-Solberg M, Gibson-Scipio W. Medical Mistrust: A Concept Analysis. *Nursing Reports*. 2025;15(3). doi:10.3390/nursrep15030103

32. Ramirez M, Bedgio R, Ramos V, et al. Abstract WMP93: Addressing Stroke Patient and Family Reasons for Not Calling 911. *Stroke*. 50(Suppl_1):AWMP93-AWMP93. doi:10.1161/str.50.suppl_1.WMP93

33. Emily A. Vogels. *Digital Divide Persists Even as Americans with Lower Incomes Make Gains in Tech Adoption*. Pew Research Center; 2021. Accessed April 18, 2025. https://www.pewresearch.org/short-reads/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/#:~:text=Roughly%20a%20quarter%20of%20adults,incomes%20are%20not%20tablet%20owners.

34. Lastrucci V, Lorini C, Caini S, Florence Health Literacy Research Group, Bonaccorsi G. Health literacy as a mediator of the relationship between socioeconomic status and health: A cross-sectional study in a population-based sample in Florence. *PLOS ONE*. 2019;14(12):e0227007. doi:10.1371/journal.pone.0227007

35. Crowe RP, Kennel J, Fernandez AR, et al. Racial, Ethnic, and Socioeconomic Disparities in Out-of-Hospital Pain Management for Patients With Long Bone Fractures. *Annals of Emergency Medicine*. 2023;82(5):535-545. doi:10.1016/j.annemergmed.2023.03.035

36. Bongiorno DM, Peters GA, Samuels-Kalow ME, et al. Racial and Ethnic Disparities in EMS Use of Restraints and Sedation for Patients With Behavioral Health Emergencies. *JAMA Network Open*. 2025;8(3):e251281-e251281. doi:10.1001/jamanetworkopen.2025.1281

37. Andrew MK, Mitnitski AB, Rockwood K. Social Vulnerability, Frailty and Mortality in Elderly People. *PLOS ONE*. 2008;3(5):e2232. doi:10.1371/journal.pone.0002232

38. Desai P, Bond J, Dhana A, et al. The Social Vulnerability Index and Incidence of Alzheimer Disease in a Population-Based Sample of Older Adults. *Neurology*. 2025;104(8):e213464. doi:10.1212/WNL.0000000000213464

SUPPLEMENTAL MATERIALS

S1. Directed Acyclic Graphs

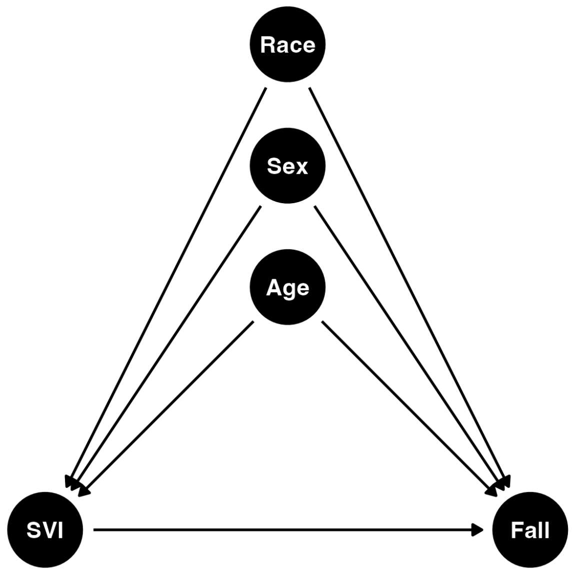

Figure S1. Directed Acyclic Graph demonstrating the hypothesized relationship of SVI on falling confounded by race, sex, and age.

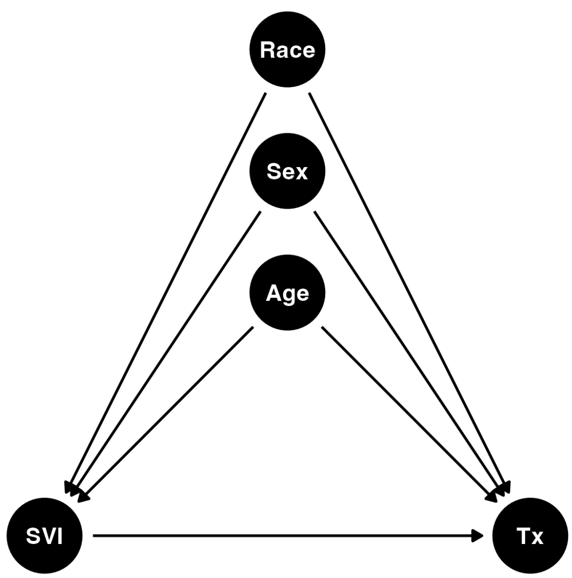

Figure S2. Directed Acyclic Graph demonstrating the hypothesized relationship of SVI on Tx (transport outcome/disposition) confounded by race, sex, and age.

S2. Study Population Flow Diagram

Figure S3. Study Population Flow Diagram

S3. R Markdown Code

The below document includes all of the data cleaning, variable defining, exploratory analysis, and formal analytic process for this study.

DATA CLEANING

Calling Libraries

```
library(tidyverse)
library(lubridate)
library(hms)
```

Import

Bringing in CSV from NYS, naming it `raw_data`:

```
raw_data =
  read_csv("Data Files/2025-0037 (Kaleb)_Export.csv")
```

Encountered some parsing issues, investigate:

```
problems(raw_data)

# Just a format issue with date/time, will fix it after column names.
```

Looking at column names and making note of NEMSIS fields:

```
colnames(raw_data)
nrow(raw_data)
```

I used previous information to make myself a data dictionary. Now cleaning up column names to simply be the field code:

```
cleaned_colnames = names(raw_data) |>
  str_extract("\\(([^)]+)\\)") |>    # extract text inside parentheses
  str_remove_all("[()]")             # remove parentheses

names(raw_data) = cleaned_colnames # assign new column names

colnames(raw_data) # confirms done correctly
```

Preliminary Variable Changes

Now that I have changed column names I will:

- Fix the names of shared 3.4/3.5 variables to take on one label
- Convert time variables such that R doesn't treat them as characters
- Create new time variables

```r
raw_data =
  raw_data |>
  rename(
    eTimes.17 = `3.4=itTimes.016/3.5=eTimes.17`,
    eResponse.15 = `3.4=eResponse.15/3.5=itResponse.115`,
    eDisposition.12 = `3.4=eDisposition.12/3.5=itDisposition.112`,
    eDisposition.23 = `3.4=eDisposition.23/3.5=itDisposition.123`
  )
```

Selecting which time variables to drop:

```r
fix_time =
  raw_data |>
  select(-c(
    eTimes.02, eTimes.04, eTimes.08, eTimes.10, eTimes.15, eTimes.17))

head(fix_time, 15)

# Convert all eTimes columns to datetime format
fix_time = fix_time |>
  mutate(
    across(eTimes.01:eTimes.16, ~ parse_date_time(.x, orders = "mdy HMS"))
  )
```

Now checking that I did that correctly:

```r
head(fix_time)
```

Creating a date column and dropping date from each time column:

```r
fix_time =
  fix_time |>
  mutate(
    eTimes_date = as_date(eTimes.03)  # Extracts date from eTimes.03
  )

fix_time = fix_time |>
  mutate(
    across(eTimes.01:eTimes.16, ~ as_hms(.x))  # Keep only time
  )
```

Creating new time variables including:

- eTimes_resp time from dispatch to scene
- eTimes_scenetime_tx scence to transport time
- eTimes_cl unit dispatched to in service time

```r
fix_time = fix_time |>
  mutate(
    eTimes_resp = as.numeric(eTimes.06 - eTimes.03) / 60,  # Response time (m
inutes)
```

```
    eTimes_scenetime_tx = as.numeric(eTimes.09 - eTimes.06) / 60,  # Scene-to
-transport time (minutes)
    eTimes_cl = as.numeric(eTimes.13 - eTimes.03) / 60  # Call-to-clear time
(minutes)
  )
```

Checking Restriction Criteria

Checking that data was restricted by age as requested and looking at distribution of ages:

```
summary(fix_time$ePatient.15)
hist(fix_time$ePatient.15)

# looks good
```

Missing and Duplicates

Looking at missing data by variable:

```
missing_summary =
  fix_time |>
  summarize(across(everything(), ~sum(is.na(.)))) |>
  pivot_longer(cols = everything(), names_to = "Column", values_to = "Missing
_Count") |>
  mutate(
    Missing_Percentage = round(Missing_Count / nrow(fix_time)*100, 2)  # Calc
ulate percentage
  ) |>
  arrange(desc(Missing_Percentage))  # Sort by highest missingness

print(missing_summary)
```

Scene Incident Zip Code (eScene.19) has the least amount of missing data, so we will link SVI with that.

Taking a look zip code values in `raw_data`:

```
response_counts_zip =
  fix_time |>
  count(eScene.19, sort = TRUE) |>
  arrange(eScene.19)

print(response_counts_zip)
```

Excitingly, ePatient.14 (race) is only missing in about 0.12 percent of rows.

Now lets take a look at duplicates

Looking to see how many of these duplicates have identical data for each row entered:

```r
exact_duplicates =
  fix_time |>
  group_by(across(everything())) |>
  filter(n() > 1) |>
  ungroup()

print(nrow(exact_duplicates))

# about 5000 true duplicates that provide no additional data, will delete the
se duplicates.
```

Keeping only unique rows, naming new dataset *dedup_data*:

```r
dedup_data =
  fix_time |>
  distinct()

nrow(dedup_data)

# this worked correctly

write_csv(dedup_data, "Data Files/deduplicated_nys_data.csv")
```

Isolating rows with duplicate eRecord.01 to inspect more closely:

```r
duplicates =
  dedup_data |>
  group_by(eRecord.01) |>
  filter(n() > 1) |>
  ungroup()

write.csv(duplicates, "duplicate_eRecord.csv", row.names = FALSE)

# viewing outside of the RMD as to not cause computational delays
#eDisposition.18 is one of the problems
```

Looks like field providers can put in more than one entry for eDisposition.18 which descibes whether they use lights and sirens to get to the hospital. eDisposition.17 says emergent or non-emergent, which I think is enough information for my purposes. So, I am going to drop the column and then de-duplicate again and see how much that helps:

```r
dedup_data =
  dedup_data |>
  select(-eDisposition.18)

dedup_data =
  dedup_data |>
  distinct()

nrow(dedup_data)
```

```
write_csv(dedup_data, "Data Files/deduplicated_nys_data.csv")
```

I fixed about 5k duplicates but there is still more, below I see how much more duplicate data there is:

```
duplicates_2 =
  dedup_data |>
  group_by(eRecord.01) |>
  filter(n() > 1) |>
  ungroup()

write_csv(duplicates_2, "Data Files/duplicate(2)_eRecord.csv")

head(duplicates_2, 20)
```

Count by incident number:

```
duplicate_counts =
  dedup_data |>
  count(eRecord.01) |>
  arrange(desc(n)) |>
  filter(n > 1)

total_duplicates =
  dedup_data |>
  group_by(eRecord.01) |>
  filter(n() > 1) |>
  nrow()

print(duplicate_counts)
print(total_duplicates)
```

Trying to identify if there are eRecord.01 repeats that actually have different columns, despite using the distinct() function earlier, I still have rows that are identical on closer inspection. I think this has to do with some of the variables being character variables.

```
duplicates_with_differences = dedup_data |>
  group_by(eRecord.01) |>
  filter(n() > 1) |>  # only duplicates
  summarise(across(everything(), n_distinct)) |>
  filter_if(is.numeric, any_vars(. > 1)) |>  #column has >1 unique value
  pull(eRecord.01)  # pull unique eRecord.01 values

#view this outside of rmd for computational efficiency
```

Look at specific eRecord.01's shown in the output:

```
filtered_data = dedup_data |>
  filter(eRecord.01 == "009317a9c9324271ae14af9301078835")
```

```
filtered_data
```

Some of the times are entered with same minutes but one hour apart (maybe these charts were locked and then edited and then re-sent to State). I think it is safe to keep only the first row of the ~ 40k repeats in this dataset.

```
dedup_data =
  dedup_data |>
  distinct(eRecord.01, .keep_all = TRUE)  # keep only one row per eRec.01

nrow(dedup_data)
```

Re-assess: did this work?

```
total_duplicates_again =
  dedup_data |>
  group_by(eRecord.01) |>
  filter(n() > 1) |>
  nrow()

print(total_duplicates_again)
```

This worked - no more duplicates in the data set *dedup_data*

```
write_csv(dedup_data, "Data Files/deduplicated_nys_data.csv")
```

Merging with SVI Data

```
svi =
  read_csv("Data Files/NewYork_ZipCode.csv")

colnames(svi)

RPLTHEMES =
  svi |>
  group_by(FIPS) |>
  summarize(RPL_THEMES)

print(RPLTHEMES)
```

Rename eScene.19 as FIPS

```
dedup_data =
  dedup_data |>
  rename(FIPS = eScene.19)
```

Treat FIPS as character variable

```
dedup_data =
  dedup_data |>
  mutate(FIPS = as.character(FIPS))
```

Left merge:

```
merged_data =
  dedup_data |>
  left_join(svi |>
              select(FIPS,
                     RPL_THEME1,
                     RPL_THEME2,
                     RPL_THEME3,
                     RPL_THEME4,
                     RPL_THEMES),
            by = "FIPS")

head(merged_data)
nrow(merged_data)
```

Save the merged_data as a csv so that I don't have to do all that again:

```
write_csv(merged_data, "Data Files/analytic.data.csv")
```

CASE DEFINITIONS

Library Calling

```
library(tidyverse)
library(broom)
library(ggplot2)
library(stringr)
library(nnet)
library(gtsummary)
```

Loading Data

```
data =
  read_csv("Data Files/analytic.data.csv")
```

Confirming Dataset Characteristics

```
nrow(data)
ncol(data)
colnames(data)
```

I will be using eDispatch.01, eDisposition.12 (outcomes), and RPL_THEMES (exposures) so I will check those for missingness (I did check for missingness in analytic dataset preparation and I believe these fields are good but want to be sure):

```
missing_summary =
  data |>
  select(eDispatch.01, eDisposition.12, RPL_THEMES) |>
  summarize(across(everything(), ~ sum(is.na(.)) / n() * 100))

print(missing_summary)
```

Percent missingess is less than 10% for all three fields.

Should I drop the rows with missing RPL_THEMES (?)

```
data |>
  filter(is.na(RPL_THEMES)) |>
  count(FIPS)
```

Most of the zip codes have very little missingness for `RPL_THEMES` and those with high missingness are not in NYS (Vermont, PA, etc). Dropping rows with missing value for RPL_THEMES.

There are also missing values listed as -999, chaning those to NA before dropping:

```
data =
  data |>
  mutate(RPL_THEMES = if_else(RPL_THEMES == -999, NA_real_, RPL_THEMES))

data =
  data |>
  drop_na(RPL_THEMES)
nrow(data)
```

Exploring SVI: Exposure Definitions

Using `quantile` to break SVI into quartiles:

```
# Define quartile breaks
svi_breaks =
  quantile(
    data$RPL_THEMES, probs = c(0, 0.25, 0.50, 0.75, 1), na.rm = TRUE)

# Assign quartiles
data =
  data |>
  mutate(svi_quartile = cut(RPL_THEMES, breaks = svi_breaks, include.lowest =
TRUE, labels = c(1, 2, 3, 4)))
```

Look at all activations again by RPL_THEMES:

```
svi_summary =
  data |>
  group_by(svi_quartile) |>
  summarize(total_cases = n()) |>
  mutate(prop = total_cases / sum(total_cases) * 100)

svi_summary
```

Checking distribution of RPL_THEMES:

```
data |>
  summarize(unique_svi = n_distinct(RPL_THEMES))
```

Graphically:

```
histplot =
  data |>
  ggplot(aes(x = RPL_THEMES)) +
  geom_histogram(binwidth = 0.05, fill = "blue", alpha = 0.6) +
  labs(title = "Distribution of SVI (rpl_themes)", x = "SVI Score", y = "Coun
t") +
  theme_minimal()

histplot
```

Look at bar plot of quartiles:

```
svi_quartile_plot =
  data |>
  ggplot(aes(x = as.factor(svi_quartile))) +
  geom_bar(fill = "blue", alpha = 0.6) +
  labs(title = "Distribution of Cases by SVI Quartile",
       x = "SVI Quartile (1 = Lowest Vulnerability, 4 = Highest)",
       y = "Number of EMS Calls") +
  theme_minimal()

svi_quartile_plot

quantile(data$RPL_THEMES, probs = c(0, 0.25, 0.50, 0.75, 1), na.rm = TRUE)
```

EXPOSURE DEFINITION

I am going to define categories myself:

```
data = data |>
  mutate(svi_category = case_when(
    RPL_THEMES <= 0.5 ~ "1: Low SV",
    RPL_THEMES > 0.5 & RPL_THEMES <= 0.8 ~ "2: Moderate SV",
    RPL_THEMES > 0.8 ~ "3: High SV"
  ))

svi_category_plot =
  data |>
  ggplot(aes(x = as.factor(svi_category))) +
  geom_bar(fill = "blue", alpha = 0.6) +
  labs(title = "Distribution of Cases by SVI Quartile",
       x = "SVI Category (1 = Lowest Vulnerability, 3 = Highest)",
       y = "Number of EMS Calls") +
  theme_minimal()
```

```
svi_category_plot
```

This looks much better.

Exploring Falls: Outcome Definitions

```
unique(data$eDispatch.01)
```

CASE DEFINITION

```
# create case as a binary variable including falls and medical alarms
data =
  data |>
  mutate(fall_case = if_else(eDispatch.01 %in% c("Falls", "Medical Alarm"), 1
, 0))

fall_summary =
  data |>
  group_by(svi_category) |>
  summarise(total_cases = sum(fall_case, na.rm = TRUE)) |>  # Sum counts of f
all cases
  arrange(svi_category)

print(fall_summary)
```

Percent cases of calls in each SVI category:

```
fall_summary =
  data |>
  group_by(svi_category) |>
  summarise(
    total_cases = sum(fall_case, na.rm = TRUE),
    total_calls = n(),
    fall_rate = (total_cases / total_calls) * 100  # Percent of calls that ar
e falls
  ) |>
  arrange(svi_category)

print(fall_summary)
```

Visualizing fall counts per SVI category:

```
fall_summary |>
  ggplot(aes(x = svi_category, y = total_cases)) +
  geom_bar(stat = "identity", fill = "blue", alpha = 0.6) +
  labs(title = "Total Fall Cases by SVI Category",
       x = "SVI Category",
       y = "Number of Fall Cases") +
  theme_minimal()
```

Exploring Disposition: Outcome Defintions

```r
# Summarize disposition frequencies by SVI quartile
disposition_summary =
  data |>
  group_by(svi_category, eDisposition.12) |>
  summarise(n = n(), .groups = "drop") |>
  group_by(svi_category) |>
  mutate(prop = n / sum(n) * 100) # I think this is of total not of category
of SVI?


disp_plot =
  disposition_summary |>
  ggplot(aes(x = as.factor(svi_category), y = prop, fill = eDisposition.12))
+
  geom_col(position = "dodge") +
  labs(
    title = "EMS Disposition by SVI Quartile",
    x = "SVI Category (1 = Low, 3 = High)",
    y = "Percentage of Dispositions",
    fill = "Disposition"
  ) +
  theme_minimal()


print(disposition_summary)
disp_plot
```

Among cases:

```r
# Count disposition types by SVI category among fall cases
disposition_summary =
  data |>
  filter(fall_case == 1) |>  # Only include fall cases
  group_by(svi_category, eDisposition.12) |>
  summarise(total_cases = n(), .groups = "drop") |>
  arrange(svi_category, desc(total_cases))

print(disposition_summary)
```

I am going to re-categorize disposition:

```r
# obtaining list of response options
unique(data$eDisposition.12)
```

CASE DEFINITION

Re-cateogorizing disposition:

```
data =
  data |>
  mutate(disposition_category = case_when(
    eDisposition.12 %in% c(
      "Treated, Transported by this EMS Unit",
      "Treated, Transfer Care to another EMS Unit",
      "Treated, Transported by Law",
      "Transported to Landing Zone, Care Transferred",
      "Treated, Transported by Private Vehicle"
    ) ~ "1: Transported",

    eDisposition.12 %in% c(
      "Refused Evaluation/Care (Without Transport)",
      "Refused Evaluation/Care (With Transport)",
      "Patient Treated, Released (AMA)",
      "Treated, Released (per protocol)",
      "Assist, Public"  # Now included in Refusal/AMA
    ) ~ "2: Refusal/AMA",

    TRUE ~ "3: Other/Unknown"
  ))|>
  filter(disposition_category != "3: Other/Unknown")

nrow(data)

variables_defined_data = data

write_csv(variables_defined_data, "Data Files/variables_defined_data.csv")
```

Look at distribution of new categories:

```
data |> count(disposition_category)

edisp_plot =
  data |>
ggplot(aes(x = disposition_category)) +
  geom_bar(fill = "blue", alpha = 0.6) +
  labs(title = "Distribution of EMS Dispositions",
       x = "Disposition Category",
       y = "Number of Cases") +
  theme_minimal()

edisp_plot
```

Disposition category among cases:

```
# Count disposition types by SVI category among fall cases
disposition_summary =
  data |>
  filter(fall_case == 1) |>  # Only include fall cases
```

```r
  group_by(svi_category, disposition_category) |>
  summarise(total_cases = n(), .groups = "drop") |>
  arrange(svi_category, desc(total_cases))

print(disposition_summary)
```

Considering Confounders

What might confound the relationship between SVI and Fall/Disposition?

- **Patient Acuity** *lots of missingness in only possible proxy in dataset*
- **Serious Medical Event** *we lack this data*
- **Past Medical History** *we lack this data*
- **Age**
- **Sex**
- **Race/Ethnicity**

Model could look like:

$$\log(\text{Fall Case}) = \beta_0 + \beta_1 \text{SVI Category} + \beta_2 \text{Age} + \beta_3 \text{Sex} + \beta_4 \text{Race} + \beta_5 \text{Initial Acuity}$$

Where: - $P(\text{Fall Case} = 1)$ is the probability of a fall case. - β_0 is the intercept. - $\beta_1, \ldots, \beta_5$ are the regression coefficients.

So need to look at Acuity (eDisposition.19) *or* transport mode (emergency/non emergency eDisposition.16)

Looking at counts of each disposition field answer:

```r
# Count occurrences of eDisposition.19 and eDisposition.16
eDisposition_19_summary =
  data |>
  count(eDisposition.19, sort = TRUE)

eDisposition_16_summary =
  data |>
  count(eDisposition.16, sort = TRUE)

eDisposition_17_summary =
  data |>
  count(eDisposition.17, sort = TRUE)

eDisposition_19_summary
eDisposition_16_summary
eDisposition_17_summary

eDisposition_17_summary =
  data |>
  filter(disposition_category == "1: Transported") |>  # Filter for transport
```

```
ed cases
  count(eDisposition.17, sort = TRUE)  # Count unique values

# Print results
print(eDisposition_17_summary)
```

Might not even be necessary to control for this variable given its not the best proxy for acuity.

Might only end up controlling for age, sex, race/eth

Lets see what happens:

```
# mess with ePatient.14 first - race/eth variable might not be working right.

race_variable_data =
  data |>
  mutate(race = if_else(
    str_count(ePatient.14, "\\S+") >= 4,  # Count number of words
    word(ePatient.14, 1, 4),  # Extract first 4 words if there are at least 4
    ePatient.14  # Otherwise, keep original value
  ))

race_summary = race_variable_data |> count(race, sort = TRUE)

first_column_list = race_summary |> pull(1)  # Extracts the first column as a
vector

print(first_column_list)

#recoding race

race_variable_data =
  race_variable_data |>
  mutate(
    race_clean = case_when(
      str_detect(race, "White") & !str_detect(race, "Black|Asian|Native|Hispa
nic") ~ "White",
      str_detect(race, "Black|African American") ~ "Black",
      str_detect(race, "Hispanic|Latino") ~ "Hispanic",
      str_detect(race, "Asian|Hawaiian|Pacific") ~ "Asian",
      is.na(race) | str_detect(race, "Not Recorded|Not Applicable") ~ "Unknow
n",
      TRUE ~ "Other/Multiple"
    )
  )

race_variable_data |> count(race_clean, sort = TRUE)
```

renaming data

```
new_data = race_variable_data
```

writing a model

```
# Fit logistic regression model
fall_model = glm(
  fall_case ~ svi_category + ePatient.15 + ePatient.13 + race_clean,
  family = binomial,
  data = new_data
)
```

taking a look

```
summary(fall_model)
```

I am going to - rename ePatient.15 to age and then scale it so that we look at each change in SD of age instead of one unit increases in age - put unknown/others for gender and for race into one category - make sure svi_category is a factor with reference being 1

Scaling Age:

```
new_data =
  new_data |>
  rename(age_years = ePatient.15) |>
  mutate(age_scaled = scale(age_years))
```

Re-categorizing race:

```
new_data =
  new_data |>
  mutate(race_clean = case_when(
    race_clean == "Unknown" ~ "Other",
    race_clean == "Other/Multiple" ~ "Other",
    TRUE ~ race_clean
  ))

new_data |> count(race_clean, sort = TRUE)

new_data =
  new_data |>
  rename(raceeth = race_clean)
```

Re-categorizing gender:

```
new_data =
  new_data |>
  mutate(gender = case_when(
    ePatient.13 %in% c("Male", "Female") ~ ePatient.13,  # Keep valid values
    ePatient.13 %in% c("Not Applicable", "Not Recorded", "Unknown (Unable to
Determine)") ~ "Unknown",
    TRUE ~ "Unknown"  # Catch any other odd values
```

```
  ))

new_data |> count(gender, sort= TRUE)
```

Dropping unknowns:

```
new_data =
  new_data |>
  filter(gender != "Unknown")
nrow(new_data)
```

Making sure SVI is factor with low as baseline:

```
new_data =
  new_data |>
  mutate(svi_category = fct_relevel(svi_category, "1: Low SV"))
```

Making raceeth a factor with white as reference:

```
new_data =
  new_data |>
  mutate(raceeth = fct_relevel(raceeth, "White"))
```

running glm see how this looks now:

```
fall_model_1 =
  glm(fall_case ~ svi_category + age_scaled + gender + raceeth,
  family = binomial,
  data = new_data
)

summary(fall_model_1)

fall_model_results = tidy(fall_model_1, exponentiate = TRUE, conf.int = TRUE)

print(fall_model_results)
```

Now I am going to look at transport outcomes:

Making sure disposition_outcome is a factor

```
new_data =
  new_data |>
  mutate(disposition_category = as.factor(disposition_category))

new_data =
  new_data |>
  mutate(age_scaled = as.numeric(age_scaled)) # causing problems saving datas
et

write_csv(new_data, "Data Files/analytic_dataset_20250320.csv")
```

Now writing binary logistic regression model for all activation disposition:

```
disposition_model = glm(disposition_category ~ svi_category + age_scaled + ge
nder + raceeth,
                        family = binomial,
                        data = new_data
)
```

look at it:

```
summary(disposition_model)
```

Extract OR:

```
disposition_model_results = tidy(disposition_model, exponentiate = TRUE, conf
.int = TRUE)

print(disposition_model_results)
```

Visualizations

Forest plots for model fall events

glm model:

```
fall_model_results |>
  filter(term != "(Intercept)") |>
  ggplot(aes(x = reorder(term, estimate), y = estimate)) +
  geom_point(color = "blue") +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width = 0.2) +
  coord_flip() +
  labs(title = "Odds Ratios for Fall Events",
       x = "Predictor",
       y = "Odds Ratio") +
  theme_minimal()
```

Maybe I don't need to keep other/unknown in the model, my concern is that disposition is documented so variably, that inevitably some of the patients in the other/unknown metric were those who were left at home by EMS and documented as a 'lift assist' instead of a true fall victim.

FORMAL ANALYSIS

Library Calling

```
library(tidyverse)
library(gtsummary)
library(dagitty)
library(ggdag)
```

```r
library(DiagrammeR)
library(DiagrammeRsvg)
library(rsvg)
library(ggplot2)
library(broom)
library(purrr)
library(labelled)
```

Introduction

To this point, I have cleaned the dataset now saved as `analytic_dataset_20250320`, defined exposure/outcome variables, and made an analytic plan.

flow diagram

```r
# Save as SVG first
DiagrammeRsvg::export_svg(grViz("
digraph study_flow {
  graph [layout = dot, rankdir = TB]

  node [shape = box, style = filled, fillcolor = lightgray]

  A [label = 'Raw Dataset\\n(n = 1,490,959)']
  B [label = 'Without Duplicate Rows\\n(n = 1,425,732)']
  C [label = 'Not Missing SVI Score\\n(n = 1,324,811)']
  D [label = 'Not Missing Disposition\\n(n = 1,276,971)']
  E [label = 'Not Missing Gender - Analytic Dataset\\n(n = 1,268,070)']
  F1 [label = 'Dispatch Complaint Fall/Medical Alarm\\n(n = 126,197)']
  F2 [label = 'Disposition Category Refusal/AMA\\n(n = 134,689)']

  A -> B
  B -> C
  C -> D
  D -> E
  E -> F1
  E -> F2
}
")) |>
charToRaw() |>
rsvg::rsvg_png("figure1.png")
```

Figure Study flow diagram shows number of rows included in final analytic data set (Not Missing Gender) as well as number of activations with outcome of interest fall and outcome of interest disposition refusal. ## DAG 1

```r
dag1 = dagitty("dag {
  SVI -> Fall
  Age -> SVI
```

```
  Age -> Fall
  Sex -> SVI
  Sex -> Fall
  Race -> SVI
  Race -> Fall
}")

# Set custom coordinates
coordinates(dag1) = list(
  x = c(SVI = 0, Fall = 2, Age = 1, Sex = 1, Race = 1),
  y = c(SVI = 0, Fall = 0, Age = 1, Sex = 1.5, Race = 2)
)

# Plot with custom layout
dag_1 = ggdag(dag1, text = TRUE) +
  coord_fixed() +
  theme_dag()
```

DAG 2

```
dag2 = dagitty("dag {
  SVI -> Tx
  Age -> SVI
  Age -> Tx
  Sex -> SVI
  Sex -> Tx
  Race -> SVI
  Race -> Tx
}")

# Set custom coordinates
coordinates(dag1) = list(
  x = c(SVI = 0, Tx = 2, Age = 1, Sex = 1, Race = 1),
  y = c(SVI = 0, Tx = 0, Age = 1, Sex = 1.5, Race = 2)
)

# Plot with custom layout
dag_2 = ggdag(dag1, text = TRUE) +
  coord_fixed() +
  theme_dag()
```

save DAGs

```
# DAG 1
ggsave("dag1_svi_fall.png",
       plot = dag_1,  # or replace with the actual object if saved as `p1` et
c.
       width = 6, height = 4, dpi = 300)
```

```
ggsave("dag2_svi_transport.png",
       plot = dag_2,
       width = 6, height = 4, dpi = 300)
```

Descriptive Statistics

read in data

```
data =
  read_csv("Data Files/analytic_dataset_20250320.csv")
```

table 1

```
variable_changes =
  data |>
   mutate(
    svi_category =
      recode(svi_category,
                          "1: Low SV" = "Low SV",
                          "2: Moderate SV" = "Moderate SV",
                          "3: High SV" = "High SV"),
    disposition_category =
      recode(disposition_category,
             "1: Transported" = "Transport",
             "2: Refusal/AMA" = "No Transport"),
    fall_case =
      factor(fall_case, levels = c(0,1),
             labels = c("No", "Yes"), exclude = NULL)
  )

variable_changes =
  variable_changes |>
  mutate(svi_category = factor(svi_category,
                          levels = c("Low SV", "Moderate SV", "High SV"))
  )

# create race_binary
variable_changes =
  variable_changes |>
  mutate(race_binary = ifelse(raceeth == "White", "White", "Non-White"))

variable_changes =
  variable_changes |>
  mutate(
    disposition_category =
      factor(disposition_category, levels = c("Transport", "No Transport"))
  )
```

```
vars = c("age_years", "gender", "race_binary", "fall_case", "disposition_cate
gory")


# Create the table
table1 =
  variable_changes |>
  select(all_of(vars), svi_category) |>
  tbl_summary(
    by = svi_category,
    statistic = list(
      age_years ~ "{median} ({p25}, {p75})",
      all_categorical() ~ "{n} ({p}%)"
    ),
    label = list(
      age_years ~ "Age",
      gender ~ "Gender",
      race_binary ~ "Race/Ethnicity",
      fall_case ~ "Dispatch Complaint is Fall",
      disposition_category ~ "Disposition"
    ),
    type = list(fall_case ~ "categorical"),
    missing = "no"
  ) |>
  add_p() |>
  modify_header(label = "**Characteristic**") |>
  bold_labels()


table1 |>
  as_gt() |>
  gt::tab_options(table.font.size = "small",
                  table.width = "100%")

table1 |>
  as_flex_table() |>
  flextable::save_as_docx(path = "Table1.docx")
```

Regression Analysis

Three models, for each hold age and gender constant, stratify by race/ethnicity:

- Assess the relationship between social vulnerability and fall.
- Assess the relationship between social vulnerability and disposition.
- Assess the relationship between social vulnerability and disposition among fall cases.

adding variable labels

```
variable_changes =
  variable_changes |>
  set_variable_labels(
    svi_category = "Social Vulnerability Index (SVI)",
    age_years = "Age (years)",
    gender = "Gender",
    race_binary = "Race/Ethnicity",
    disposition_category = "Disposition",
    fall_case = "Dispatch Complaint is Fall"
  )
```

writing models

```
# Model 1

model1_white = glm(fall_case ~ svi_category + age_years + gender, data = vari
able_changes |>
                     filter(race_binary == "White"), family = binomial)

model1_nonwhite = glm(fall_case ~ svi_category + age_years + gender, data = v
ariable_changes |>
                        filter(race_binary == "Non-White"), family = binomial
)

# Model 2

model2_white = glm(disposition_category ~ svi_category + age_years + gender,
data = variable_changes |>
                     filter(race_binary == "White"), family = binomial)

model2_nonwhite = glm(disposition_category ~ svi_category + age_years + gende
r, data = variable_changes |>
                        filter(race_binary == "Non-White"), family = binomial
)

# Model 3

model3_white = glm(disposition_category ~ svi_category + age_years + gender,
data = variable_changes |>
                     filter(race_binary == "White" & fall_case == "Yes"), fam
ily = binomial)

model3_nonwhite = glm(disposition_category ~ svi_category + age_years + gende
r, data = variable_changes|>
                        filter(race_binary == "Non-White" & fall_case == "Yes
"), family = binomial)
```

building tables

```r
# Model 1 activation fall and svi

tbl_model1_white = tbl_regression(model1_white, exponentiate = TRUE)

tbl_model1_nonwhite = tbl_regression(model1_nonwhite, exponentiate = TRUE)

# Model 2 all activation disposition and svi

tbl_model2_white = tbl_regression(model2_white, exponentiate = TRUE)

tbl_model2_nonwhite = tbl_regression(model2_nonwhite, exponentiate = TRUE)

# Model 3 among cases only disopsition and svi

tbl_model3_white = tbl_regression(model3_white, exponentiate = TRUE)

tbl_model3_nonwhite = tbl_regression(model3_nonwhite, exponentiate = TRUE)
```

viewing tables

```r
tbl_model1_white
tbl_model1_nonwhite
tbl_model2_white
tbl_model2_nonwhite
tbl_model3_white
tbl_model3_nonwhite
```

sample sizes

```r
# Table 2: Fall Outcome
table2_white = variable_changes %>%
  filter(race_binary == "White") %>%
  count(svi_category, gender, name = "n") %>%
  mutate(model = "Table 2", race = "White")

table2_nonwhite = variable_changes %>%
  filter(race_binary == "Non-White") %>%
  count(svi_category, gender, name = "n") %>%
  mutate(model = "Table 2", race = "Non-White")

# Table 3: Disposition Outcome (All Cases)
table3_white = variable_changes %>%
  filter(race_binary == "White") %>%
  count(svi_category, gender, name = "n") %>%
  mutate(model = "Table 3", race = "White")

table3_nonwhite = variable_changes %>%
  filter(race_binary == "Non-White") %>%
```

```r
  count(svi_category, gender, name = "n") %>%
  mutate(model = "Table 3", race = "Non-White")

# Table 4: Disposition Outcome (Fall Cases Only)
table4_white = variable_changes %>%
  filter(race_binary == "White", fall_case == "Yes") %>%
  count(svi_category, gender, name = "n") %>%
  mutate(model = "Table 4", race = "White")

table4_nonwhite = variable_changes %>%
  filter(race_binary == "Non-White", fall_case == "Yes") %>%
  count(svi_category, gender, name = "n") %>%
  mutate(model = "Table 4", race = "Non-White")

# Combine all into one table
sample_size_details = bind_rows(
  table2_white, table2_nonwhite,
  table3_white, table3_nonwhite,
  table4_white, table4_nonwhite
)

# View
print(sample_size_details)

# Total sample sizes per model
n_table2_white = nrow(variable_changes %>% filter(race_binary == "White"))
n_table2_nonwhite = nrow(variable_changes %>% filter(race_binary == "Non-Whit
e"))

n_table3_white = n_table2_white
n_table3_nonwhite = n_table2_nonwhite

n_table4_white = nrow(variable_changes %>% filter(race_binary == "White", fal
l_case == "Yes"))
n_table4_nonwhite = nrow(variable_changes %>% filter(race_binary == "Non-Whit
e", fall_case == "Yes"))

# Display total Ns
total_ns = tibble(
  Model = c("Table 2", "Table 2", "Table 3", "Table 3", "Table 4", "Table 4")
,
  Race = c("White", "Non-White", "White", "Non-White", "White", "Non-White"),
  Total_N = c(n_table2_white, n_table2_nonwhite,
              n_table3_white, n_table3_nonwhite,
              n_table4_white, n_table4_nonwhite)
)

print(total_ns)
```

stacking tables

```
stacked_model1 = tbl_merge(
  tbls = list(tbl_model1_white, tbl_model1_nonwhite),
  tab_spanner = c("**White**", "**Non-White**")
)

stacked_model2 = tbl_merge(
  tbls = list(tbl_model2_white, tbl_model2_nonwhite),
  tab_spanner = c("**White**", "**Non-White**")
)

stacked_model3 = tbl_merge(
  tbls = list(tbl_model3_white, tbl_model3_nonwhite),
  tab_spanner = c("**White**", "**Non-White**")
)

stacked_model1
stacked_model2
stacked_model3
```

saving tables

```
# Save all stacked models into Word document
stacked_model1 |>
  as_flex_table() |>
  flextable::save_as_docx(path = "Model1_table.docx")

stacked_model2 |>
  as_flex_table() |>
  flextable::save_as_docx(path = "Model2_table.docx")

stacked_model3 |>
  as_flex_table() |>
  flextable::save_as_docx(path = "Model3_table.docx")
```