NEXT STEPs: Nursing Home Data Sources Compendium Peripheral Data Sources

Data Source	Description	Nursing Home Relevance	Link for Additional Info	Example Manuscript
Area Deprivation Index (ADI)	The Area Deprivation Index (ADI) is a neighborhood-level measure of socioeconomic disadvantage developed using U.S. Census data. It incorporates 17 indicators of education, employment, housing quality, and income to assign rankings of deprivation at the census block group level. Higher ADI scores reflect greater disadvantage, and the index has been widely used to study health disparities, access to care, and policy targeting.	The ADI is valuable in nursing home research for capturing community-level context that may influence residents' health outcomes, facility resources, and geographic patterns of care. It has been previously used to identify the neighborhood disadvantage of the communities in which nursing homes or skilled nursing facilities are located. It is publicly available and can be linked to other datasets using geographic identifiers.	https://www.neighborhoodatlas.medicine.wisc.edu/	Travers JL, Hade EM, Friedman S, Raval A, Hadson K, Falvey JR. Staffing and Antipsychotic Medication Use in Nursing Homes and Neighborhood Deprivation. JAMA Netw Open. 2024 Apr 1;7(4):e248322. doi: 10.1001/jamanetworkopen.2024.8322. PMID: 38656575; PMCID: PMC11043897. URL: https://pubmed.ncbi.nlm.nih.gov/38656575/
Area Health Resource File (AHRF)	The Area Health Resource File (AHRF), maintained by the Health Resources and Services Administration (HRSA), is a comprehensive county-level database on health care resources, providers, facilities, and population characteristics across the United States. It compiles information from over 50 sources, including the American Medical Association, Census Bureau, CMS, and others, into a standardized, annually updated dataset. Key variables include counts of health professionals, hospital and facility characteristics, economic indicators, and population demographics. The AHRF is widely used to examine geographic variation in health services availability, provider supply, and health system capacity.	It is especially useful for contextualizing facility environments and linking resident or facility outcomes to broader county-level health care resources. Investigators have previously used the AHRF to measure factors related to county-level post-acute care supply.	https://data.hrsa.gov/topics/health-workforce/ahrf	Yang MT, Temkin-Greener H, Veazie P, Cai S. Post-acute care transitions during COVID-19: Racial, ethnic, and socioeconomic differences in older adults with Alzheimer's disease and related dementia. J Am Geriatr Soc. 2024 Jul;72(7):2006-2016. doi: 10.1111/jgs.18884. Epub 2024 Mar 27. PMID: 38539279; PMCID: PMC11226367. URL: https://pubmed.ncbi.nlm.nih.gov/38539279/
Health Resources and Services Administration (HRSA) Provider Relief Fund Data	The HRSA Provider Relief Fund Data document federal financial support distributed to health care providers, including nursing homes, during the COVID-19 pandemic under the Coronavirus Aid, Relief, and Economic Security (CARES) Act. The dataset includes details on payment amounts, recipient organizations, provider types, and geographic distribution of funds. Funds were allocated largely based on total revenue and Medicare revenue. The purpose of the data is to increase transparency and accountability in how relief funds were allocated to stabilize providers and maintain access to care during the public health emergency. The dataset has been updated regularly and is publicly available through HRSA's open data platform.	These data are particularly useful for evaluating how financial relief was distributed across facilities, assessing equity in funding allocation, and examining the relationship between financial support and resident or facility outcomes.	https://data.cdc.gov/Administrative/HHS-Provider-Relief-Fund/kh8y-3es6/about_data	Orewa GN, Weech-Maldonado R, Lord J, Davlyatov G, Becker D, Feldman SS. COVID-19 Pandemic Impact on Nursing Homes' Financial Performance. Inquiry. 2024 Jan-Dec;61:469580241240698. doi: 10.1177/00469580241240698. PMID: 38515246; PMCID: PMC10958812. URL: https://pubmed.ncbi.nlm.nih.gov/38515246/
Healthcare Cost Report Information System (HCRIS)	The Healthcare Cost Report Information System (HCRIS), maintained by the Centers for Medicare & Medicaid Services (CMS), compiles annual Medicare cost report data submitted by hospitals, skilled nursing facilities, home health agencies, and other provider types. For nursing homes, these reports include detailed financial information such as revenues, expenses, payer mix, staffing costs, and utilization measures. The dataset is widely used to assess financial performance, operating margins, and the economic pressures faced by providers. Because reporting is mandatory for Medicare- and Medicaid-certified facilities, HCRIS offers a near-universal source of financial data, although accuracy can vary depending on how facilities prepare their reports.	HCRIS is valuable for examining trends in facility solvency, exploring cost structures, and linking financial performance to quality and resident outcomes.	https://www.cms.gov/data-research/statistics-trends-and-reports/cost-reports	Travers JL, McGarry BE, Friedman S, Holaday LW, Ross JS, Lopez L, Chen K. Association of Receipt of Paycheck Protection Program Loans With Staffing Patterns Among US Nursing Homes. JAMA Netw Open. 2023 Jul 3;6(7):e2326122. doi: 10.1001/jamanetworkopen.2023.26122. PMID: 37498597; PMCID: PMC10375300. URL: https://pubmed.ncbi.nlm.nih.gov/37498597/
Kaiser Family Foundation State Health Facts	The Kaiser Family Foundation (KFF) State Health Facts is an online data resource that compiles and disseminates a wide range of state-level health and health policy indicators. It draws from multiple sources, including federal surveys, administrative datasets, and KFF's own analyses, to provide accessible, regularly updated statistics. Topics covered include health care costs, insurance coverage, Medicaid and Medicare enrollment, demographics, public health, and provider capacity. Its interactive platform allows for straightforward cross-state comparisons, making it a valuable tool for policy evaluation and comparative research.	KFF State Health Facts is particularly useful for placing facility- or resident-level findings in a broader policy context, such as state differences in Medicaid reimbursement rates, long-term services and supports, or demographic characteristics of older adults.	https://www.kff.org/statedata/	Ritter AZ, Freed S, Coe NB. Younger Individuals Increase Their Use of Nursing Homes Following ACA Medicaid Expansion. J Am Med Dir Assoc. 2022 May;23(5):852-857.e5. doi: 10.1016/j.jamda.2021.08.020. Epub 2021 Sep 21. PMID: 34555342; PMCID: PMC11027188. URL: https://pubmed.ncbi.nlm.nih.gov/34555342/
Medicare Data on Provider Practice and Specialty (MD-PPAS)	The Medicare Data on Provider Practice and Specialty (MD-PPAS) is a CMS dataset that provides information on the characteristics, specialties, and practice affiliations of physicians and other health care professionals who bill Medicare. It combines individual-level data from the Provider Enrollment, Chain, and Ownership System (PECOS) with Medicare claims, allowing researchers to examine providers' demographics, specialty training, practice locations, and group affiliations. MD-PPAS is particularly useful for studying provider supply, distribution across geographic areas, and patterns of practice organization, such as the growth of multispecialty groups.	These data can be leveraged to identify which providers deliver care to nursing home residents, to analyze access to specialists, and to study the structure of practices serving older adults. It can also be used to identify clinicians who primarily practice in nursing homes (i.e., "specialists" in nursing home care). A key strength of MD-PPAS is its national scope and longitudinal updates, though its coverage is limited to Medicare-participating clinicians.	https://resdac.org/cms-data/files/md-ppas	Riester MR, Ryskina KL, White EM, Hayes KN, Harris DA, Zullo AR. Approaches to Identify Nursing Home Specialists Using Medicare Claims Data. Med Care. 2025 Jul 1;63(7):520-528. doi: 10.1097/MLR.0000000000002161. Epub 2025 Apr 28. PMID: 40307676; PMCID: PMC12185168. URL: https://pubmed.ncbi.nlm.nih.gov/40307676/
National Conference of State Legislature or American Health Planning Association Certificate of Need (CON) State Laws Data	The Certificate of Need (CON) State Laws Data, compiled by organizations such as the National Conference of State Legislatures (NCSL) and the American Health Planning Association, track state-level regulations governing the establishment and expansion of health care facilities and services. These laws typically require providers to obtain state approval before making major capital investments, expanding bed capacity, or introducing new services. The dataset provides information on which states have CON laws, the specific health services and facility types covered, and historical changes in regulatory environments.	CON data are especially relevant because many states regulate the supply of nursing home beds, influencing market entry, competition, and geographic availability of long-term care. Researchers use these data to examine how regulatory policies shape access, costs, quality, and investment in nursing homes and related post-acute services.	https://www.ncsl.org/health/certificate-of-need-state-laws	Rahman M, Tyler D, Thomas KS, Grabowski DC, Mor V. Higher Medicare SNF care utilization by dual-eligible beneficiaries: can Medicaid long-term care policies be the answer? Health Serv Res. 2015 Feb;50(1):161-79. doi: 10.1111/1475-6773.12204. Epub 2014 Jul 22. PMID: 25047831; PMCID: PMC4319876. URL: https://pubmed.ncbi.nlm.nih.gov/25047831/

Data Source	Description	Nursing Home Relevance	Link for Additional Info	Example Manuscript
National Hospital Discharge Survey (NHDS) / National Hospital Care Survey (NHCS)	The National Hospital Discharge Survey (NHDS), conducted by the National Center for Health Statistics until 2010, provided nationally representative data on inpatient hospital discharges, including patient demographics, diagnoses, procedures, lengths of stay, and discharge dispositions. It has since been replaced by the National Hospital Care Survey (NHCS), which expands on NHDS by integrating electronic health records, hospital billing data, and administrative sources from both inpatient and emergency department settings. NHCS offers a more contemporary and comprehensive view of hospital utilization, care processes, and patient outcomes, including linkages to the National Death Index for mortality follow-up. While NHDS provides historical benchmarks, NHCS now serves as the ongoing source of hospital discharge and care data.	These surveys are valuable for understanding the characteristics of hospitalizations that precede nursing home admission, the conditions leading to post-acute care needs, and patterns of discharge to nursing homes versus other settings.	https://www.cdc.gov/nchs/nhcs/index.html	Ahmed AA, Hays CI, Liu B, Aban IB, Sims RV, Aronow WS, Ritchie CS, Ahmed A. Predictors of in-hospital mortality among hospitalized nursing home residents: an analysis of the National Hospital Discharge Surveys 2005-2006. J Am Med Dir Assoc. 2010 Jan;11(1):52-8. doi: 10.1016/j.jamda.2009.08.003. Epub 2009 Nov 25. PMID: 20129215; PMCID: PMC2818085. URL: https://pubmed.ncbi.nlm.nih.gov/20129215/
Outcome and Assessment Information Set (OASIS)	The Outcome and Assessment Information Set (OASIS) is a standardized assessment tool used by Medicare-certified home health agencies to collect comprehensive data on all adult patients receiving skilled home health care. It captures patient-level information on sociodemographic characteristics, clinical status, functional abilities, health service needs, and outcomes over time. OASIS is primarily used by CMS for quality monitoring, public reporting, and reimbursement under the Home Health Prospective Payment System, but it is also widely used in research. Limitations include its restriction to home health recipients and potential variability in coding practices across agencies.	OASIS provides critical insight into post-acute care trajectories, enabling comparisons between patients discharged from hospitals to home health versus to nursing facilities. It is particularly valuable for examining care transitions, functional recovery, and outcomes across different post-acute care settings.	https://resdac.org/cms-data/files/oasis	Shi S, Olivieri-Mui B, Oh G, McCarthy E, Kim DH. Analysis of Functional Recovery in Older Adults Discharged to Skilled Nursing Facilities and Then Home. JAMA Netw Open. 2022 Aug 1;5(8):e2225452. doi: 10.1001/jamanetworkopen.2022.25452. PMID: 36006647; PMCID: PMC9412223. URL: https://pubmed.ncbi.nlm.nih.gov/36006647/
Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI)	The Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI) is a standardized dataset required for all Medicare-certified inpatient rehabilitation facilities (IRFs). It collects detailed patient-level information on demographics, clinical conditions, functional and cognitive status, and therapy utilization from admission through discharge. These data are central to Medicare's Prospective Payment System for IRFs, determining case-mix groups and reimbursement rates. For researchers, the IRF-PAI provides valuable insights into rehabilitation intensity, functional outcomes, and care patterns among patients who may alternatively be discharged to skilled nursing facilities. The dataset's strengths include standardized national coverage and detailed functional measures, though access is restricted and primarily available through CMS data use agreements.	IRF-PAI allows for comparison of post-acute care trajectories across settings, helping to understand which patients receive inpatient rehabilitation versus nursing home–based rehabilitation and the outcomes associated with each.	https://resdac.org/cms-data/files/irf-pai#:~:text=The%20Inpatient%20Rehabilitation%20Facility%20Patient,to%20measure%20quality%20of%20care.	Li CY, Haas A, Pritchard KT, Karmarkar A, Kuo YF, Hreha K, Ottenbacher KJ. Functional Status Across Post-Acute Settings is Associated With 30-Day and 90-Day Hospital Readmissions. J Am Med Dir Assoc. 2021 Dec;22(12):2447-2453.e5. doi: 10.1016/j.jamda.2021.07.039. Epub 2021 Aug 30. PMID: 34473961; PMCID: PMC8627458. URL: https://pubmed.ncbi.nlm.nih.gov/34473961/
Small Business Administration (SBA)	The Small Business Administration (SBA) provides data on federal loan and grant programs intended to support small businesses, including those in the health care sector. During the COVID-19 pandemic, SBA programs such as the Paycheck Protection Program (PPP) and Economic Injury Disaster Loans (EIDL) became especially important sources of financial relief for nursing homes and other long-term care providers. These datasets include information on loan size, recipient organizations, industry classification, and geographic distribution. Although the SBA data provide transparency into program distribution, limitations include variability in reporting and lack of detail on how funds were ultimately used by facilities.	SBA data are useful for understanding the extent to which facilities accessed emergency financing, how such support varied across markets, and the potential relationship between financial relief and facility staffing, solvency, or quality of care.	https://data.sba.gov/	Travers JL, McGarry BE, Friedman S, Holaday LW, Ross JS, Lopez L, Chen K. Association of Receipt of Paycheck Protection Program Loans With Staffing Patterns Among US Nursing Homes. JAMA Netw Open. 2023 Jul 3;6(7):e2326122. doi: 10.1001/jamanetworkopen.2023.26122. PMID: 37498597; PMCID: PMC10375300. URL: https://pubmed.ncbi.nlm.nih.gov/37498597/
State Medicaid Data	State Medicaid Data encompass a variety of administrative datasets collected and reported by individual state Medicaid programs, often in partnership with the Centers for Medicare & Medicaid Services (CMS). These data include beneficiary enrollment files, claims for services, pharmacy dispensing records, provider participation, and state-specific policy information such as waiver programs and reimbursement structures. They are particularly valuable for studying populations that rely heavily on Medicaid, including long-stay nursing home residents and individuals receiving long-term services and supports (LTSS). However, data availability, formatting, and completeness vary considerably across states, and researchers often need to navigate state-specific approval processes and documentation to obtain access.	State Medicaid data enable analyses of resident demographics, patterns of care use, prescription drug utilization, and state policy impacts on access and quality.	https://data.medicaid.gov/	Gorges RJ, Sanghavi P, Konetzka RT. A National Examination Of Long-Term Care Setting, Outcomes, And Disparities Among Elderly Dual Eligibles. Health Aff (Millwood). 2019 Jul;38(7):1110-1118. doi: 10.1377/hlthaff.2018.05409. PMID: 31260370; PMCID: PMC7147241. URL: https://pubmed.ncbi.nlm.nih.gov/31260370/
Surveillance, Epidemiology, and End Results-Medicare (SEER-Medicare)	The SEER-Medicare linked dataset combines high-quality cancer registry data from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program with Medicare enrollment and claims data. SEER captures detailed information on cancer incidence, stage at diagnosis, tumor characteristics, and initial treatment across participating regions that represent about 28% of the U.S. population. When linked with Medicare claims, researchers can follow cancer patients longitudinally to assess patterns of care, comorbidities, costs, and outcomes. This linkage is particularly valuable for studying older adults with cancer, including those residing in nursing homes, where cancer treatment decisions often intersect with frailty, multimorbidity, and goals of care. Limitations include the restriction to Medicare beneficiaries in SEER regions and the lag time in data availability.	SEER-Medicare data are a powerful resource to examine how cancer is managed in the nursing homepopulation, disparities in access to treatment, and downstream outcomes such as hospitalizations, transitions of care, and mortality.	https://healthcaredelivery.cancer.gov/seermedicare/	Liu MA, Keeney T, Papaila A, Ogarek J, Khurshid H, Wulff-Burchfield E, Olszewski A, Bélanger E, Panagiotou OA. Functional Status and Survival in Older Nursing Home Residents With Advanced Non-Small-Cell Lung Cancer: A SEER-Medicare Analysis. JCO Oncol Pract. 2022 Jun;18(6):e886-e895. doi: 10.1200/OP.21.00460. Epub 2022 Feb 7. PMID: 35130040; PMCID: PMC9191367. URL: https://pubmed.ncbi.nlm.nih.gov/35130040/

Data Source	Description	Nursing Home Relevance	Link for Additional Info	Example Manuscript
U.S. Census Bureau Data	The U.S. Census Bureau provides several major data sources that are widely used in health services and policy research, including the Decennial Census and the American Community Survey (ACS). The Decennial Census offers a complete population count every 10 years, including demographic, housing, and geographic characteristics at fine levels of detail. The ACS complements this by providing annual estimates on a broad range of socioeconomic indicators such as income, education, disability status, and housing conditions, available down to the census tract and block group level. While rich and publicly available, one limitation is that ACS estimates are based on samples and can have reduced reliability for very small geographies or subpopulations.	These datasets are invaluable for capturing the sociodemographic context in which facilities operate, examining neighborhood-level determinants of resident health, and linking facility or resident outcomes to community characteristics. They are also commonly used for stratifying analyses by urban–rural status or assessing disparities by race, ethnicity, and socioeconomic status.	https://data.census.gov/	Park C, Kim D, Briesacher BA. Association of Social Isolation of Long-term Care Facilities in the United States With 30-Day Mortality. JAMA Netw Open. 2021 Jun 1;4(6):e2113361. doi: 10.1001/jamanetworkopen.2021.13361. PMID: 34132793; PMCID: PMC8209586. URL: https://pubmed.ncbi.nlm.nih.gov/34132793/
United States Department of Agriculture (USDA) Rural-Urban Continuum Codes	The USDA Rural-Urban Continuum Codes (RUCCs) classify all U.S. counties into nine categories based on population size, degree of urbanization, and adjacency to metropolitan areas. These codes allow researchers to distinguish gradations of rural and urban settings beyond the simple metropolitan versus nonmetropolitan classification used by the Office of Management and Budget. The RUCCs are updated every 10 years following the decennial census and are widely used in health services research to capture geographic variation in health care access, provider availability, and population health. Their strengths include nationwide coverage and consistent categorization, though a limitation is that they are assigned at the county level, which may mask more localized differences in community resources or access.	RUCCs are particularly useful for examining rural–urban disparities in nursing home facility distribution, staffing, resident outcomes, and access to post-acute or specialty care.	https://www.ers.usda.gov/data-products/rural-urban-continuum-codes	Kosar CM, Loomer L, Ferdows NB, Trivedi AN, Panagiotou OA, Rahman M. Assessment of Rural-Urban Differences in Postacute Care Utilization and Outcomes Among Older US Adults. JAMA Netw Open. 2020 Jan 3;3(1):e1918738. doi: 10.1001/jamanetworkopen.2019.18738. PMID: 31913495; PMCID: PMC6991315. URL: https://pubmed.ncbi.nlm.nih.gov/31913495/

Funded by U24AG087939, NIA