Development of a Longitudinal Dataset of Persons With Dementia and Their Caregivers Through End-of-Life: A Statistical Analysis System Algorithm for Joining National Health and Aging Trends Study/National Study of Caregiving

Suzanne S. Sullivan, PhD, MBA, RN, CHPN[1] [ORCID], Chin-Shang Li, PhD[1],
Cristina de Rosa, MSN, RN, PhD[1] [ORCID], and Yu-Ping Chang, PhD, RN, FGSA, FAAN, FIAAN[1]

Abstract

Background: Alzheimer's disease and related dementias (AD/ADRD) are terminal conditions impacting families and caregivers, particularly at end-of-life. Longitudinal, secondary data analyses present opportunities for insight into dementia caregiving and decision-making over time; however, joining complex datasets and preparing them for analysis poses many challenges. **Objectives:** To describe an approach to linking national survey data of older adults with their primary caregivers to build a prospective, longitudinal dataset, and to share the Statistical Analysis System (SAS) coding statement algorithms with other researchers. **Methods:** The National Health and Aging Trends Study (NHATS) and National Study of Caregiving (NSOC) are joined using a series of algorithms based on conceptual and operational definitions of dementia, primary caregivers, and the occurrence of death. A series of SAS algorithms resulting in the final longitudinal dataset was created. **Results:** NHATS/ NSOC participants were linked using three preliminary data files (n = 12 427) and one final data join (n = 3305) over nine rounds of data collection. Presence of dementia was defined based on the indicator in the year preceding the last month-of-life (LML) interview. Primary caregivers were defined as the person providing the most frequent care over time. Additional flag variables (LML interview, dementia classification, and cohort (2011 vs 2015)) were created. The SAS algorithms are presented herein. **Discussion:** The SAS coding statement algorithms provide an opportunity to conduct longitudinal analysis of care for both members of the dyad in the context of dementia and end-of-life. Future research using the proposed dataset can further explore care and caregiving in these populations.

Keywords

Alzheimer's disease and related dementias (AD/ADRD) are a group of progressive, neurodegenerative syndromes that affect cognitive functioning and contribute to physical disability and decline, with late-stage dementia ultimately resulting in death. Globally, 50 million people are currently afflicted with AD/ADRD and its prevalence is projected to triple to an estimated 152 million people by the year 2050[1]. Thus, the treatment and care of persons living with dementia is an urgent public health concern[2]. However, AD/ADRD are poorly understood, stigmatizing conditions, which can lead to delays in diagnosis and treatment, further exacerbating the socioeconomic, physical, and psychological implications for persons with dementia (PWD), their families, and society at large.[1]

The treatment and care of PWD presents unique challenges given the prolonged trajectory of decline leading up to death, which is unlike other terminal diseases such as metastatic cancer or organ failure that follow more predictable patterns of functional decline[3,4]. Nevertheless, despite the often-lengthy downward course of illness, changes in condition or care needs for

[1]School of Nursinsg, University at Buffalo - South Campus, Buffalo, NY, USA

Corresponding Author:
Suzanne S. Sullivan, PhD, MBA, RN, CHPN, School of Nursinsg, University at Buffalo School of Nursing, 201 A Wende Hall, Buffalo, NY 14214, USA.
Email: suzanney@buffalo.edu

PWD often occur rapidly and unexpectedly, and many families find themselves unprepared to make treatment decisions because they have not discussed their wishes for end-of-life care[5]. In the absence of these discussions, PWD may experience unplanned, non-beneficial medical treatments, and burdensome care transitions at the end-of-life[6]. Thus, more research is needed to understand the needs of PWD and their caregivers over time, to better prepare PWD and their caregivers for managing care throughout the trajectory of illness and the dying process[7,8].

Secondary analysis of longitudinal, population-level data of older adults living in the community, such as the National Health and Aging Trends Study (NHATS) and its companion study, the National Study of Caregiving (NSOC)[9] offer a unique opportunity to advance research that contributes to the understanding of how the trajectory of AD/ADRD impacts PWD and their caregivers[10]. However, joining these surveys into a longitudinal dataset ready for analysis is a decidedly complex process. First, it is necessary to acquire an in-depth understanding of the study design including how the individual variables relate over time (e.g. changing caregivers). Second, it is necessary to create conceptual and operational definitions that guide decisions on how to preprocess and classify the right subpopulations for the study purpose using the data on hand.

Our program of research, funded by the National Institute of Aging (NIA), focuses on understanding the needs of PWD and their caregivers in planning and preparing for end-of-life care (R03AG067159). The purpose of this paper is to describe our approach in building a prospective, longitudinal dataset of NHATS participants linked to their primary caregiver (NSOC), and to publish our Statistical Analysis System (SAS) coding statement algorithms for use by other researchers in hopes of furthering research addressing the needs of PWD and their primary caregivers through the end-of-life experience.

Methods

National Health and Aging Trends Study and National Study of Caregiving Datasets

The NHATS and NSOC are longitudinal surveys of Medicare beneficiaries ages 65+ and their caregivers that may be linked together to further understanding of relationships and aging and disability trends through the end-of-life. NHATS/NSOC were developed by a consortium of multidisciplinary researchers at the Johns Hopkins University and are sponsored by the NIA (U01AG32947). Commencing in 2011, NHATS is an annual survey of 12,427 older adults, and the NSOC periodically surveys caregivers of a select group of NHATS study participants (n = 3305). The studies are guided by a conceptual framework that blends the language of the World Health Organization (WHO) International Classification of Functioning with the Nagi Model of Disablement[11–13].

National Health and Aging Trends Study

NHATS utilizes a prospective, complex, stratified, three-stage design which oversamples the oldest age groups and Black non-Hispanic persons in the coterminous United States (excluding Alaska and Hawaii). There are two cohorts of participants (2011 with replenishment of the sample in 2015). NHATS data are nationally-representative (when the data are weighted). The key content areas of the NHATS survey include health conditions, impairments and symptoms, physical and cognitive capacity, self-care and mobility, participation in valued activities, and the physical, social, technological, and service environment[14]. Within NHATS are periodic supplemental questions such as those pertaining to COVID-19 and advance care planning. NHATS contains structured and unstructured data elements such as clock drawings and dried blood spots[14]. For NHATS participants who die, there is a last month of life interview (LML) available for understanding the end-of-life experience. The LML interview is completed by a proxy that provides insights into the place and quality of end-of-life care, daily activities, and additional questions about interacting with the healthcare system.

National Study of Caregiving

The NSOC is a periodic survey of select caregivers of NHATS participants who help with self-care, mobility, medical, or household activities[15]. The NSOC gathers demographic and other data on the caregiver's situation and caregiving experience in up to five family or unpaid caregivers per NHATS participant. Caregivers were interviewed at three timepoints (NSOC-I-III 2011, 2015, and 2017). There are ten key content areas in the NSOC, including activities, intensity and duration of care, positive and negative aspects of caregiving, caregiver health and well-being, and socioeconomic variables, among others items such as 24-hour time diaries. The NSOC I (2011) and II (2015) provide cross-sectional data and the NSOC III (2017) provides both cross-sectional and longitudinal data by re-interviewing caregivers who participated in NSOC II.

Prior National Health and Aging Trends Study/ National Study of Caregiving Research

NHATS and NSOC data have been used to understand a multitude of aspects of aging, including issues related to dementia, caregiving, and the quality of end-of-life care. For example, Amjad and colleagues[16] determined that PWD experience higher symptom burden and more limitations in their social activities near the end-of-life. Kasper and colleagues[17] identified that more than 90% of those in community settings rely on family or unpaid caregivers, particularly spouses and daughters to meet the needs of PWD. In another example, Vick and colleagues[18] determined that the odds of experiencing caregiving strain were nearly twice as high among dementia caregivers at the end-of-life compared to non-dementia caregivers.

Accessing National Health and Aging Trends Study/National Study of Caregiving Data

Publicly-available, sensitive, and restricted data are available to researchers without a fee. Publicly-available data are broad and sufficient for a multitude of analyses. Sensitive data includes protected information such as age-in-years, month and year of birth and death, and restricted data includes items such as geographic location (e.g. 5-digit zip code). Access to restricted data elements occurs with permission of the NHATS principal investigators under a data use agreement, and requires evidence of funding and ethics approval of human subject research by an institutional review board (IRB). With additional permission, NHATS/NSOC studies may be linked to external sources (e.g. Centers for Medicare and Medicaid Services (CMS) data).

Algorithm Development Process

The purpose of developing this SAS algorithm is to develop a dataset for investigating aspects of caregiving and other factors associated with quality-of-life and end-of-life care transitions of PWD. Key conceptual and operational definitions of the following items were developed: (a) dementia classification, (b) primary caregiver, and (c) death. Data preprocessing included descriptive statistics, data visualization, and content expertise from a multidisciplinary team of experts representing palliative care nursing, gerontology, and biostatistics.

Dementia Classification

The NHATS principal investigators have published a technical paper, Classification of Persons by Dementia Status in the National Health and Aging Trends Study[19] and NHATS Dementia Classification Addendum[20] for follow-up rounds along with SAS programming statements for identifying individuals with either possible, probable, or no dementia[21]. NHATS participants are identified using this approach based on either a report of dementia or Alzheimer's disease diagnosis by a physician (hc#diescn9, "has dementia or Alzheimer's" subsequently labeled as "previously reported"), cognitive tests of memory, orientation, and executive function to form a derived variable (cp#dad8dem) that is calculated using the Eight-item Interview to Differentiate Aging and Dementia (AD8)[22,23] instrument, with the # sign indicating the round of data collection. However, the NHATS dataset does not differentiate between Alzheimer's disease or other forms of dementia.[19–23]

Primary Caregiver Classification

Individuals assisting NHATS participants are screened for eligibility for participation in the NSOC. During the yearly data collection interview process, NHATS participants are asked whether they are receiving assistance for mobility, self-care, or household activities for health or functioning in the last month, or if they are living in a residential care facility. Identified "helpers" who are either, 1) related to the NHATS participant (paid or not), or 2) unrelated but not paid to help, are eligible to participate in the NSOC study. The NSOC surveys up to five eligible caregivers for each NHATS participant. It is important to note that not all NHATS participants have caregivers.

Death Classification

There are two primary ways to identify whether an NHATS participant has died during the NHATS study using either the LML interview derived variable (r#dlmlint), or by using the month (pd#mthdied) (1 = January through 12 = December) (pd#mthdied) and four-digit year (pd#yrdied) of death. The LML interview item is a dichotomous variable available in the sensitive data file. Although unrelated to the algorithm development discussed herein, it may be of interest to some readers to note that is also possible to identify those who died prior to each round (fl#spdied), the place of death (pd#placedied), as well as the death of a spouse (op#dspoudec) or others (e.g. child).

Joining the National Health and Aging Trends Study and National Study of Caregiving Studies

The first step in joining the NHATS and NSOC studies is to access the data from the NHATS and NSOC website with permission under a data use agreement with the principal investigators of NHATS. Individual participants in NHATS are assigned a unique Sample Person ID (SPID) which can be matched year-to-year across all survey rounds resulting in a row of data for each individual NHATS participant annually until death. NSOC participants are assigned an Other Person ID (OPID) that can be linked to the SPID in a one-to-many join. We elected to focus on the primary caregiver only for this dataset, creating a one-to-one data join.

Results

In this study, we developed four SAS algorithms for joining the NHATS/NSOC annual surveys (2011–2019) into a longitudinal dataset using individual and caregiver data as well as additional "flag" variables to enable data analysis (Supplemental Materials 1–4).

Demographics

National Health and Aging Trends Study Participants

The final file consists of 3305 NHATS participants from both cohorts (2011 and 2015) linked to their primary caregiver (Table 1). There are 1466 (44.4%) NHATS participants without dementia, 1366 (41.3%) with probable dementia, 380 (11.5%), possible dementia, and 93 (2.8%) missing data (Table 2). The prevalence of possible or probable dementia in

Table 1. Dementia and death demographics by NHATS rounds 1–9 (2011–2019) joined to NSOC by cohort.

	R1	R2	R3	R4	R5	R6	R7	R8	R9
	2011	2012	2013	2014	2015	2016	2017	2018	2019
Cohort 1 - joined to NSOC (2011)	N = 2281	N = 2150	N = 1901	N = 1685	N = 1537	N = 1406	N = 1242	N = 1029	N = 859
Has dementia[a](%)									
Yes	307 (13.5)	62 (2.9)	63 (3.3)	61 (3.6)	55 (3.6)	40 (2.8)	46 (3.7)	28 (2.7)	30 (3.5)
No	1967 (86.2)	1682 (78.2)	1427 (75.1)	1257 (74.6)	1149 (74.8)	1004 (71.4)	818 (65.9)	675 (65.6)	553 (64.4)
Previously reported	—	225 (10.5)	217 (11.4)	200 (11.9)	216 (14.1)	204 (14.5)	180 (14.5)	157 (15.3)	145 (16.9)
[b,c]Missing, don't know, inapplicable	—	181 (8.4)	194 (10.3)	167 (9.9)	117 (7.6)	158 (11.3)	198 (16.0)	169 (16.4)	131 (15.2)
[d]Last month of life interview (%)									
Yes	–	155 (7.2)	155 (8.2)	133 (7.9)	87 (5.7)	127 (9.0)	175 (14.1)	148 (14.4)	117 (13.6)
Cohort 2 - joined to NSOC (2015)					N = 1024	N = 962	N = 864	N = 720	N = 615
Has dementiaa (%)									
Yes	–	–	–	–	139 (13.6)	58 (6.0)	38 (4.4)	27 (3.8)	
No	–	–	–	–	880 (85.9)	730 (75.9)	604 (69.9)	486 (67.5)	429 (69.8)
Previously reported	–	–	–	–	—	97 (10.1)	116 (13.4)	110 (15.3)	91 (14.8)
[b,c]Missing, don't know, inapplicable	–	–	–	–		77 (8.0)	106 (12.2)	97 (13.5)	86 (14.0)
[d]Last month of life interview (%)									
Yes	–	–	–	–	–	62 (6.4)	98 (11.3)	88 (12.2)	72 (11.7)
Dementia class change[e] (%)		159 (1.28)	80 (.64)	65 (.52)	46 (.37)	73 (.59)	76 (.61)	50 (.40)	41 (.33)

Abbreviation. NHATS, National Health Aging and Trends Study. NSOC, National Study of Caregiving.
[a]NHATS Dementia variable (hc#disescn9).
[b]NHATS coded Missing (−9), do not know (−8), RF (−7), inapplicable (−1).
[c]Inapplicable refers to individuals who have died in a previous round.
[d]Last month of life interview (r#dlmlint).
[e]Dementia class change from probable dementia to either possible dementia or no dementia (n = 12 427).

cohort 1 (n = 2281) is 55%. Cohort 2 includes 1024 individuals with probable (35.7%), possible (12.2%), or no dementia (49.5%). Nearly half of the NHATS participants have an LML interview (42.9%). The majority of NHATS participants are female (65.3%) and white (cohort 1 = 69.8%; cohort 2 = 66.4%). Ages range from 65-90+ years old.

National Study of Caregiving Study Participants

There are 3305 primary caregivers of NHATS participants, with 2281 in cohort 1 and 1024 in cohort 2. The mean caregiver age ranges from 60-65 years, the majority are female, non-Hispanic white, and are most often primary relatives to the NHATS participant such as a spouse/partner, daughter, or son (Table 3). For example, NSOC III caregivers (n = 1643) are most frequently, daughters (36.4%), female spouse/partners (16.5%), granddaughters (3.3%), friends (2.9%), sisters (2.3%), or nieces (2.1%). Male caregivers primarily consist of sons (14.5%), spouse/partners (9.6%), grandsons (1.2%), or friends (1.2%).

Dataset Characteristics

Dementia Classification

The final joined file represents all NHATS participants with a linked NSOC survey (n = 3305) and determination of dementia or Alzheimer's disease status (1 = possible, 2 = probable, or 3 = no dementia) as classified by the dementia SAS coding statements provided by NHATS[21]. During data preprocessing it was noted that dementia status was not always classified consistently over the nine rounds of data collection. That is, some individuals were initially coded as having 1 = possible dementia or 2 = probable dementia, followed by subsequent years where 3 = no dementia was noted. The frequency of dementia class change occurred in the full dataset (n= 12,427) from 2 = probable dementia to either 1 = possible dementia or 3 = no dementia once 478 (3.81), twice 55 (.44), or 3 times 2 (.02) across the 9 rounds of data. Therefore, we elected to classify dementia status with a "flag" item using the variable indicating the dementia status provided in the round immediately preceding the LML interview (e.g. round 3 dementia status for a

Table 2. Demographic variables of NHATS participants final join (Longitudinal Dataset).

NHATS totals	N (%)	Cohort 1 (2011) N (%)	Cohort 2 (2015) N (%)
Sample person ID (SPID)	3305	2281	1024
Dementia status (%)			
Probable	1366 (41.3)	1000 (43.8)	366 (35.7)
Possible	380 (11.5)	255 (11.2)	125 (12.2)
No dementia	1466 (44.4)	959 (42.0)	507 (49.5)
Missing or classified otherwise	93 (2.8)	67 (2.9)	26 (2.5)
Last month of life interview (%)			
Yes	1417 (42.9)	1097 (48.1)	320 (31.3)
No	1888 (57.1)	1184 (51.9)	704 (68.8)
Sex (%)			
Female	–	1509 (66.2)	649 (63.4)
Male	–	772 (33.8)	375 (36.6)
Age range (%)			
65–69	–	239 (10.5)	143 (14.0)
70–74	–	309 (13.5)	146 (14.3)
75–79	–	444 (19.5)	192 (18.8)
80–84	–	535 (23.5)	189 (18.5)
85–89	–	414 (18.1)	164 (16.0)
90+	–	340 (14.9)	190 (18.6)
Inapplicable	–	–	
Race and ethnicity (%)			
White	–	1592 (69.8)	680 (66.4)
Black/African American	–	603 (26.4)	272 (26.6)
American Indian	–	73 (3.2)	48 (4.7)
Asian	–	26 (1.1)	13 (1.2)
Hispanic/Latino	–	132 (5.8)	54 (5.3)

Abbreviation. NHATS, National Health and Aging Trends Study

person with an LML interview in round 4). See Table 1 for a detailed description of dementia class change by round.

Primary Caregiver Classification

The primary caregiver was defined as the individual with the longest caregiving relationship with the NHATS participant, that is, the caregiver who had the greatest number of interviews over the three cross-sectional NSOC interviews (NSOC I, II, and III). We developed this definition because the longest caregiving relationship would provide the most accurate longitudinal information regarding caregiving perceptions and care needs for the PWD. For an individual NHATS participant, if more than one caregiver was interviewed on multiple occasions (i.e. a tie), the primary caregiver was further defined as the person who provided the greatest number of hours of care (cdc1hlphrsdy, cdc5hlphrsdy, and cdc7hlphrsdy) in the prior month, which is consistent with previously published research of caregiver care-recipient dyads using the NHATS/NSOC study[24,25]. This decision was two-fold. First, after limiting the NHATS participants to PWD, there were very few individuals

with a longitudinal NSOC III survey and an LML interview. Secondarily, the option to conduct a one-to-many data join of a single NHATS participant with up to five caregivers added a level of conceptual and operational complexity that was determined to be beyond the scope of this study.

Death Classification

The LML interview was chosen to determine the occurrence of death among NHATS participants for its ease of use because the LML interview is conducted after confirmation that the NHATS participant is deceased by an individual, who is usually a close family member, with knowledge of the end-of-life experience[7,24,26]. The month and year of death variables were also retained in the final dataset. It is important to keep in mind that the LML interview always follows the NHATS interview from the previous round and that any given individual may die at any time along the trajectory of the NHATS study. Therefore, it is essential to first identify in which round the NHATS participant died to determine an accurate dementia classification and to pull predictor variables from the rounds

Table 3. Primary caregiver demographic data by NSOC I-III.

	NSOC I	NSOC II	NSOC III
	2011	2015	2017
Cohort I (2011)	N = 1320	N = 802	N = 929
Age (SD)	60.53 (13.8)	63.44 (13.78)	64.58 (12.96)
Sex (%)			
Female	907 (68.7)	568 (70.8)	640 (68.9)
Male	413 (31.3)	228 (28.4)	289 (31.1)
Race and ethnicity (%)[a]			
White, non-Hispanic	–	615 (66.2)	615 (66.2)
Black, non-Hispanic	–	212 (22.8)	212 (22.8)
Other	–	21 (2.3)	21 (2.3)
Hispanic	–	57 (6.1)	57 (6.1)
Don't know/Refuse	–	24 (2.6)	25 (3.1)
Relationship (%)			
Spouse/Partner	351 (26.6)	211 (26.3)	233 (25.1)
Daughter	474 (35.9)	280 (34.9)	333 (35.8)
Son	188 (8.2)	111 (13.8)	151 (16.3)
Other (combined)	307 (29.3)	200 (25.0)	212 (22.8)
Cohort 2 (2015)		N = 594	N = 714
Age (SD)	–	61.91 (13.99)	62.29 (13.90)
Sex (%)			
Female		190 (32.0)	504 (70.6)
Male		404 (68.0)	210 (29.4)
Race and ethnicity (%)	–		
White, non-Hispanic	–	352 (59.3)	450 (63.0)
Black, non-Hispanic	–	168 (28.3)	174 (24.4)
Other	–	15 (2.5)	15 (2.1)
Hispanic	–	36 (6.1)	52 (7.3)
Don't know/Refuse	–	23 (3.9)	23 (3.2)
Relationship (%)			
Spouse/Partner	–	191 (32.2)	195 (27.3)
Daughter	–	216 (36.4)	265 (37.1)
Son	–	75 (12.6)	87 (12.2)
Other (combined)	–	112 (18.8)	167 (23.4)

Abbreviation. NSOC, National Study of Caregiving.
[a]Caregiver race and ethnicity were not reported in NSOC I.

preceding death (e.g. LML interview variables). In this case, our study used two cohorts of NHATS participants and three cohorts of NSOC participants.

Flag Variables

In addition to the key variables described above, we created several columns to "flag" items of interest and further our analysis. These included (1) a dichotomous column indicating whether the NHATS participant had an LML interview (1 = yes, 0 = no) at any time to facilitate data preprocessing, cross-sectional analysis, or to split the file, (2) final dementia classification (1 = possible, 2 = probable, or 3 = no dementia), and (3) to which cohort the individual NHATS participant belongs (1 = 2011, 2 = 2015).

Final Data Files

Three subset data files were created prior to the final data join.

File 1.

The first file, named NHATS_SP_R1_R9_Dementia and Status Simple_061,521.sav (n = 12,427) was created by joining NHATS files round 1 through 9 to determine the dementia classifier and if there was an LML interview (status = 0 = alive, 1 = not alive) (Supplemental Material 1). This file enabled us to run descriptive statistics and manually verify (by randomly selecting cases) that the correct final dementia class status corresponded accurately with the dementia classification provided in the year prior to death.

File 2.

The second file, NSOC_R1_R5_R7_simple_061,521.sav was created from the NSOC I, II, and III (n = 12,427) and was used to identify the primary caregiver based on participation in the greatest number of interviews, or in the case of a tie, the greatest number of hours of care provided to the NHATS participant (Supplemental Material 2).

File 3.

The third file, NHATSNSOC_Dementia Status and interview_simple_061,521.sav joined the NHATS_SP_R1_R9_Dementia and Status Simple_061,521.sav with the NSOC_R1_R5_R7_simple_061,521.sav, creating a joined cohort file (n = 12,427) (Supplemental Material 3).

Final data join.

The final data file, NHATSNSOC_clean_06.15.21.sav (n = 3305) was created by merging all datasets based on the selected primary caregivers (Supplemental Material 4).

Discussion

Population-level longitudinal datasets are promising resources for furthering research of PWD and their caregivers; however, the process of joining datasets is complex and requires thoughtful consideration. In this study, we defined key variables including dementia classification, primary caregiver, and death, which enabled the development of a final longitudinal data file of the NHATS survey for nine NHATS rounds (2011–2019) joined to the NSOC I-III surveys. Further, we provide our SAS coding statement algorithms for use by other researchers who would like to acquire a better understanding of how care needs of PWD might develop over time, to help identify caregivers at risk for burnout, fatigue, or physical injury associated with caregiving, and inform caregivers' expectations and planning for the future. Moreover, the known, disparate impact of dementia caregiving on women, racial or ethnic minorities, and low socioeconomic status groups[17,27–29] can also be examined as it exists over time for determining ways to optimize support using our version of the NHATS/NSOC dataset.

This study has some limitations. First, using the NHATS-provided SAS coding statements, dementia status was not always consistent along the nine-year trajectory of the NHATS study. To address this issue, we elected to use the dementia classification indicated in the year immediately preceding the LML interview. Given the inability to differentiate between individuals with a diagnosis of Alzheimer's disease from those with non-specific dementia types, researchers should note this limitation when using our algorithm. Second, the process for identifying the primary caregiver was complex and required multiple assumptions. Although we preferred to use the NSOC III longitudinal file, we were not able to as its use would limit

the sample size significantly. Moreover, our conceptual definition was further limited by NSOC being collected over only three time points, rendering it difficult to determine the amount of caregiving provided over time. Therefore, we defined the primary caregiver as the person who was interviewed the greatest number of times, followed by the most hours of care in the previous month in the event of a tie, which is consistent with previously published research[24,25]. Third, we identified individuals who died in the study using the LML interview. Without an actual death date, it is impossible to know with certainty when and if the NHATS participant actually died. However, use of the LML interview to classify mortality in the NHATS dataset is well established[7,26]. Finally, as with all secondary data analysis, the variables included in the final data files are limited to the scope of the original NHATS and NSOC studies. However, both studies were rigorously designed for the purpose of population-level analysis of older adults who reside in the community.

In conclusion, while secondary data analysis of prospective, longitudinal studies holds potential for examining experiences at the end-of-life for PWD and their caregivers, developing the right data subset requires deliberate consideration of the concepts underlying the variables. Creating the SAS coding statement algorithms using NHATS and NSOC opens many possibilities for future research.

ORCID iD

Suzanne S. Sullivan https://orcid.org/0000-0003-3064-6884
Cristina de Rosa https://orcid.org/0000-0002-5261-2514

Supplemental Material

Supplemental material for this article is available online.

References

1. World Health Organization. *Dementia.* https://www.who.int/news-room/fact-sheets/detail/dementia. Updated September 21, 2020. Accessed June 28, 2021.

2. World Health Organization. *Global Action Plan on the Public Health Response to Dementia 2017-2025.* 2017. https://apps.who.int/iris/bitstream/handle/10665/259615/9789241513487-eng.pdf. Accessed June 28, 2021.

3. Lunney JR, Lynn J, Foley DJ, Lipson S, Guralnik JM. Patterns of functional decline at the end of life. *JAMA.* 2003;289(18):2387-2392. doi:10.1001/jama.289.18.2387.

4. Stallard E, Kinosian B, Stern Y. Personalized predictive modeling for patients with Alzheimer's disease using an extension of Sullivan's life table model. *Alz Res Therapy.* 2017;9:75. doi:10.1186/s13195-017-0302-6.

5. Yadav KN, Gabler NB, Cooney E, et al.. Approximately one in three US adults completes any type of advance directive for end-of-life care. *Health Aff.* 2017;36(7):1244-1251. doi:10.1377/hlthaff.2017.0175.

6. Dixon J, Karagiannidou M, Knapp M. The effectiveness of advance care planning in improving end-of-life outcomes for people with dementia and their carers: A systematic review and critical discussion. *J Pain Symptom Manag.* 2018;55(1):132-150. doi:10.1016/j.jpainsymman.2017.04.009.

7. Ornstein KA, Kelley AS, Bollens-Lund E, Wolff JL. A national profile of end-of-life caregiving in the United States. *Health Aff.* 2017;36(7):1184-1192. doi:10.1377/hlthaff.2017.0134.

8. U.S. Department of Health and Human Services. *National Plan to Address Alzheimer's Disease: 2018 Update.* 2018. https://aspe.hhs.gov/system/files/pdf/259581/NatPlan2018.pdf. Accessed June 28, 2021.

9. NHATS. About NHATS & NSOC. https://www.nhats.org/researcher/about. https://www.nhats.org/researcher/about. Accessed June 28, 2021

10. Hunt LJ, Lee SJ, Harrison KL, Smith AK. Secondary analysis of existing datasets for dementia and palliative care research: high-value applications and key considerations. *J Palliat Med.* 2018;21(2):130-142. doi:10.1089/jpm.2017.0309.

11. Freedman VA. Adopting the ICF language for studying late-life disability: A field of dreams?. *J Gerontol A Biol Sci Med Sci.* 2009;64A(11):1172-1174. doi:10.1093/gerona/glp095.

12. Montaquila J, Freedman VA, Edwards B, Kasper JD. *National Health and Aging Trends Study Round 1 Sample Design and Selection.* NHATS Technical Paper #1; 2012. https://www.nhats.org/sites/default/files/2021-01/NHATS%20Round%201%20Sample%20Design%2005_10_12_2.pdf. Accessed May 7, 2021.

13. Montaquila J, Freedman VA, Spillman B, Kasper JD. *National Health and Aging Trends Study Development of Round 1 Survey Weights.* NHATS Technical Paper #2; 2012. https://www.nhats.org/sites/default/files/2021-01/NHATS%20Round%201%20Weighting%20Description_Nov2012_3.pdf. Accessed August 4, 2021.

14. Kasper JD, Freedman VA. *National Health and Aging Trends Study (NHATS) User Guide: Rounds 1-9 Final Release;* 2020. https://www.nhats.org/sites/default/files/2021-04/NHATS_User_Guide_R9_Final_Release_0.pdf. Accessed June 28, 2021.

15. Freedman VA, Skehan ME, Hu M, Wolff J, Kasper JD. *National Study of Caregiving I-III User Guide.Version 5.0;* 2020. https://www.nhats.org/sites/default/files/2021-04/NSOC%20I-III%20USER%20GUIDE%20Version%205_Final.pdf. Accessed June 28, 2021

16. Amjad H, Snyder SH, Wolff JL, Oh E, Samus QM. Before hospice: Symptom burden, dementia, and social participation in the last year of life. *J Palliat Med.* 2019;22(9):1106-1114. doi:10.1089/jpm.2018.0479.

17. Kasper JD, Freedman VA, Spillman BC, Wolff JL. The disproportionate impact of dementia on family and unpaid caregiving to older adults. *Health Aff.* 2015;34(10):1642-1649. doi:10.1377/hlthaff.2015.0536.

18. Vick JB, Ornstein KA, Szanton SL, Dy SM, Wolff JL. Does caregiving strain increase as patients with and without dementia approach the end of life? *J Pain Symptom Manag.* 2019;57(2):199-208. e2. doi:10.1016/j.jpainsymman.2018.11.00419.

19. Kasper JD, Freedman VA, Spillman BC. *Classification of persons by dementia status in the National Health and Aging Trends Study.* 2013. Technical Paper #5. https://nhatspubdemo.westat.com/scripts/documents/NHATS_Dementia_Technical_Paper_5_Jul2013.pdf. Accessed June 28, 2021

20. Kasper JD, Freedman VA, Spillman BC, Skehan ME. *Addendum to Classification of Persons by Dementia Status in the National Health and Aging Trends Study for Follow-Up Rounds;* 2020. https://www.nhats.org/researcher/nhats/methods-documentation?id=technical-papers. Accessed June 28, 2021.

21. Spillman BC, Skehan ME. *SAS Programming Statements for Construction of Dementia Classification in the National Health and Aging Trends Study;* 2013. Addendum to NHATS Technical Paper #5. https://nhatspubdemo.westat.com/Scripts/documents/NHATS_Addendum_to_Technical_Paper_5_SAS_Programming_Statements_Jul2013.pdf. Accessed June 28, 2021.

22. Galvin JE, Roe CM, Powlishta KK, Coats MA, Muich SJ, Grant E, et al. The AD8: A brief informant interview to detect dementia. *Neurology.* 2005;65(4):559-564. doi:10.1212/01.wnl.0000172958.95282.2a.

23. Galvin JE, Roe CM, Xiong C, Morris JC. Validity and reliability of the AD8 informant interview in dementia. *Neurology.* 2006;67(11):1942-1948. doi:10.1212/01.wnl.0000247042.15547.eb.

24. Pristavec T, Luth EA. Informal caregiver burden, benefits, and older adult mortality: A survival analysis. *J Gerontol B Psychol Sci Soc Sci.* 2020;75(10):2193-2206. doi:10.1093/geronb/gbaa001T.

25. Spillman BC, Wolff JL, Freedman VA, Kasper JD. *Informal caregiving for older Americans: An analysis of the 2011 National Survey of Caregiving.* Washington, DC: Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. https://aspe.hhs.gov/system/files/pdf/77146/NHATS-IC.pdf. Accessed October 11, 2021.

26. Freedman VA, Kasper JD. Cohort Profile: The National Health and Aging Trends Study (NHATS). *Int J Epidemiol*. 2019;48(4): 1044-1045g. doi:10.1093/ije/dyz109.

27. Alzheimer's Association. *Facts and Figures*; 2021. https://www.alz.org/alzheimers-dementia/facts-figures. Accessed June 28, 2021.

28. Fetherstonhaugh D, McAuliffe L, Bauer M, Shanley C. Decision-making on behalf of people living with dementia: How do surrogate decision-makers decide? *J Med Ethics*. 2017;43(1): 35-40. doi:10.1136/medethics-2015-103301.

29. Melis RJF, Haaksma ML, Muniz-Terrera G. Understanding and predicting the longitudinal course of dementia. *Curr Opin Psychiatry*. 2019;32(2):123-129. doi:10.1097/YCO.0000000000000482.

Author Biographies

Suzanne S. Sullivan, PhD, MBA, RN, CHPN is an assistant professor, in the School of Nursing at the University at Buffalo.

Chin-Shang Li, PhD is a professor of biostatistics in the School of Nursing, University at Buffalo.

Cristina de Rosa, MSN, RN is a PhD student and research assistant in the School of Nursing, University at Buffalo.

Yu-Ping Chang, PhD, RN, FGSA, FAAN, FIAAN is a professor and the Senior Associate Dean, in the University at Buffalo, School of Nursing.