Leveraging Existing Datasets to Advance Family Caregiving Research: Opportunities to Measure What Matters

Rebecca M. Goodwin, Rebecca L. Utz, Catherine E. Elmore, Katherine A. Ornstein, Djin L. Tay, Lee Ellington, Ken R. Smith & Caroline E. Stephens

To cite this article: Rebecca M. Goodwin, Rebecca L. Utz, Catherine E. Elmore, Katherine A. Ornstein, Djin L. Tay, Lee Ellington, Ken R. Smith & Caroline E. Stephens (16 Apr 2024): Leveraging Existing Datasets to Advance Family Caregiving Research: Opportunities to Measure What Matters, Journal of Aging & Social Policy, DOI: 10.1080/08959420.2024.2320043

To link to this article: https://doi.org/10.1080/08959420.2024.2320043
Leveraging Existing Datasets to Advance Family Caregiving Research: Opportunities to Measure What Matters

Rebecca M. Goodwin\textsuperscript{a,b}, Rebecca L. Utz\textsuperscript{c,d,e,f,g}, Catherine E. Elmore\textsuperscript{a,h}, Katherine A. Ornstein\textsuperscript{i}, Djin L. Tay\textsuperscript{a,c,e,i}, Lee Ellington\textsuperscript{a,d,f,g}, Ken R. Smith\textsuperscript{i,c,g,j}, and Caroline E. Stephens\textsuperscript{i,a,d,f,g}

\textsuperscript{a}College of Nursing, University of Utah, Salt Lake City, USA; \textsuperscript{b}National Library of Medicine, National Institutes of Health, Bethesda, USA; \textsuperscript{c}College of Social and Behavioral Science, University of Utah, Salt Lake City, USA; \textsuperscript{d}Consortium for Families & Health Research, University of Utah, Salt Lake City, USA; \textsuperscript{e}Family and Consumer Studies, University of Utah, Salt Lake City, USA; \textsuperscript{f}Center on Aging, University of Utah, Salt Lake City, USA; \textsuperscript{g}Family Caregiving Collaborative – Utah Caregiving Population Science, University of Utah, Salt Lake City, USA; \textsuperscript{h}Center for Equity in Aging, Johns Hopkins University, Baltimore, USA; \textsuperscript{i}Huntsman Cancer Institute, University of Utah, Salt Lake City, USA

\textbf{ABSTRACT}

More than 17.7 million people in the U.S. care for older adults. Analyzing population datasets can increase our understanding of the needs of family caregivers of older adults. We reviewed 14 U.S. population-based datasets (2003–2023) including older adults’ and caregivers’ data to assess inclusion and measurement of 8 caregiving science domains, with a focus on whether measures were validated and/or unique variables were used. Challenges exist related to survey design, sampling, and measurement. Findings highlight the need for consistent data collection by researchers, state, tribal, local, and federal programs, for improved utility of population-based datasets for caregiving and aging research.

\textbf{KEYWORDS}

Big data; caregiver; family; health equity; long-term services and supports; older adult; population data; secondary data; social determinants of health

\textbf{Introduction}

More than 17.7 million unpaid caregivers play a critical role in caring for older adults as they age and cope with chronic illness, disability, and end of life (Committee on Family Caregiving for Older Adults, Board on Health Care Services, Health and Medicine Division, & National Academies of Sciences, Engineering, and Medicine, 2016). Caregivers provide over 36 billion hours of unpaid care with an economic value of roughly $600 billion (Reinhard et al., 2023) – this is approximately three times the amount of Medicaid spending on long-term services and supports (Congressional Research Service [CRS], 2018). Given the undeniable value of family caregivers, it is vital to consider who will care for the caregiver. Addressing this question and finding ways to better understand the needs of and ways to support these individuals is
a critical public health issue requiring research and data (AARP & National Alliance for Caregiving, 2020; Centers for Disease Control and Prevention [CDC], 2018).

Unfortunately, due to wide variation in the way data have been collected, survey-based findings cannot be reliably aggregated to be used as an evidence base to develop, plan, and coordinate resources and services or to measure the economic value of family caregiving (NASEM, 2016; RAISE, 2022). There have been calls to collect data more consistently, using common terminology to define caregiving constructs such as prevalence, health impacts, costs and value of care, unmet needs, and caregiver experiences (Fox-Grage, 2020; RAISE, 2022). Although longitudinal studies are needed to understand the impact of caregiving over time, it is challenging, expensive, and time-consuming to collect such data, particularly within the contexts of serious illness and end of life.

Analyzing existing datasets, particularly population-based data, holds promise to advance our understanding of the roles and needs of diverse caregivers of older adults and to inform policy, particularly when linked with data from other sources (Amjad et al., 2021; Kindratt, 2022; National Institutes of Health [NIH], 2023; Zhang et al., 2017). Our aim is to provide a landscape of the existing population-based datasets used for family caregiving research in the U.S. by: 1) identifying and describing existing population-based aging datasets with caregiving-related data; 2) assessing the inclusion and measurement of key variables relevant to caregiving science; and 3) reviewing applications and potential of secondary data analyses to enhance caregiving research and policy.

**Conducting an environmental scan of caregiving-related datasets**

We identified population-based datasets that included older adults and their caregivers in the U.S. over the last 20 years (2003–2023). We searched nine data archives using the keywords “aging,” “older adults,” and/or “caregiv*” (Appendix A) and identified 46 possible datasets. Datasets were included if they were: (1) conducted in the U.S. between 2003 and 2023; (2) publicly available; (3) focused on caregivers for older adults aged 50 and older; (4) identified respondents as a caregiver or care recipient; (5) did not focus on a single diagnosis (i.e., cancer or stroke); and (6) included self-report, survey-based data on at least one of the following eight caregiving domains: Positive Aspects of Caregiving and Resilience, Caregiver Demographics & Health Characteristics; Care Recipient Demographics & Health Characteristics, Caregiving Context, Supports and Resources for Caregiving, Healthcare Utilization, and Cost and Value of Caregiving. These caregiving domains were identified a priori by our team of gerontology and caregiving experts and domain selection was guided by the caregiving literature. Table 1 shows the caregiving domains with example constructs and related validated measures.
For each of the 14 datasets (Table 2), at least two authors independently reviewed study-related websites; dataset documentation (e.g., codebooks, questionnaires, and/or design papers); and peer reviewed manuscripts using the dataset. Data extracted for each dataset included: location, sample criteria, data collection methods, data access, example peer-reviewed manuscripts; linkages with other data sources; and caregiving-related domains.

**Caregiving content of datasets**

Of the 14 datasets reviewed, 8 used nationally representative samples, 2 were isolated to a region/single state; 10 were longitudinal; and 4 initiated data collection prior to 1995. Only one (Hispanic-EPES) specifically focused on caregiving in an ethnic, racial, or geographically minoritized population, though some datasets oversampled or included sub-studies that increased representation (HRS, NSHAP, NSOC/NHATS, SHOW). Several datasets supported linkages to administrative data, including Medicare claims (ADAMS/HRS, Hispanic EPESE, NLTCS, NHATS/NSOC, WLS); census data (NHATS/NSOC); IRS (WLS), Social Security earnings, and benefits (ADAMS/HRS, WLS-CM); and the National Security Index (HRS, NSHAP, WLS-CM). Table 3 provides an overview of the measures included in each dataset.

Among the 14 datasets reviewed, 12 included measures of caregiver burden and/or strain, but only a few used validated measures [e.g., Self-rated Burden Scale (SHOW, NPHA), Zarit Burden Index (ADAMS/HRS), and Modified Caregiver Strain Index (ADAMS/HRS, SHOW)]. Non-validated measures were used to assess impact on relationships, interference with hobbies or leisure, and anticipation or desire to institutionalize the CR. Many datasets asked caregivers to rate their overall health using one Likert scale question, and measured depression (e.g., HINTS via PHQ-4; HRS via CES-D for depression and loneliness), anxiety (HADS-D), satisfaction with life, difficulty caring for their own health, and responsibilities for multiple CRs. Only 7 datasets (ADAMS/HRS, CG-US, HRS, MIDUS, NLTCS, NPHA, NSOC/NHATS) measured positive aspects of caregiving (e.g., gaining a sense of purpose, resilience, learning new skills, and feeling useful or closer to the care recipient), but they did not use validated measures.

All datasets captured caregiver and care recipient demographics (e.g., gender, ethnicity/race) and included measures of care recipient’s health status, including functional limitations as well as major health conditions or reasons for care needs (some comprehensively; others only one major health problem). ADAMS/HRS enabled calculation of the Charlson Comorbidity Index. Measurement of cognition varied from a single question [e.g., whether care recipient has dementia or Alzheimer’s disease (BRFSS)] to cognitive performance tests and standardized cognitive measures [e.g., AD8 Dementia
Table 1. Caregiving domains with example constructs and related validated measures.

<table>
<thead>
<tr>
<th>Domain:</th>
<th>Definition:</th>
<th>Example Constructs (and Related Standardized Measures)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caregiver Burden and Strain</td>
<td>Burden: stress and other feelings and psychological symptoms experienced by caregivers because of the care they provide. Strain: physical, mental, or financial difficulty.</td>
<td>Burden (Zarit Burden Interview&lt;sup&gt;1&lt;/sup&gt;), Self-rated Burden Scale&lt;sup&gt;2&lt;/sup&gt;, Caregiver Reaction Assessment&lt;sup&gt;3&lt;/sup&gt;, Strain (Modified Caregiver Strain Index)&lt;sup&gt;4&lt;/sup&gt;</td>
</tr>
<tr>
<td>Positive Aspects of Caregiving and Resilience</td>
<td>Self-affirmation, a sense of gain or benefit, feeling closer to the care recipient, finding personal meaning or a life purpose, fulfilling a sense of duty, enjoyment, generativity, and improving the outlook and quality of life for the caregiver and/or care recipient. Resilience: the process and outcome of adapting to challenging life experiences.</td>
<td>(Positive Aspects of Caregiving Scale (PACS)&lt;sup&gt;5&lt;/sup&gt;, Gain in Alzheimer Care Instrument, Positive Experiences Scale (PES)&lt;sup&gt;6&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Caregiver Demographics &amp; Health Characteristics</td>
<td>Demographics include age, sex, race, ethnicity, marital status, education, socioeconomic status, federal poverty status, employment, geographic area type (urban/rural/frontier). Health measures include self-rated health, physical and mental/psychological health (e.g., depression, anxiety, stress), mobility and functional health, co-morbidities, quality of life and life satisfaction, well-being, satisfaction with caregiving, and health behaviors (e.g., preventative care, substance use, physical activity).</td>
<td>Mентation-Depression (PHQ-9&lt;sup&gt;7&lt;/sup&gt;), Anxiety (HADS&lt;sup&gt;8&lt;/sup&gt;, GAD-7&lt;sup&gt;9&lt;/sup&gt;, PROMIS), Distress (PHQ-4), Mobility-Physical and Functional (ADLs, IADLs); Comorbidities (Charlson Comorbidity Index&lt;sup&gt;10&lt;/sup&gt;); Quality of Life (PROMIS Global-10, EuroQOL&lt;sup&gt;11&lt;/sup&gt;, SCRQoL-ASCOT), Satisfaction with Caregiving (EuroQOL), Health Behaviors, Social Support (PROMIS 4-item instrumental support), Advance Care Planning</td>
</tr>
<tr>
<td>Care Recipient Demographics and Health Characteristics</td>
<td>In addition to those listed above for caregivers, care recipient health measures include major health conditions or reason for care needs; mobility and functional status; types of caregiving tasks needed; whether advanced care planning has been discussed and/or documented; and receipt of support/resources (e.g., Medicaid or Medicare).</td>
<td>In addition to those for caregivers, Medicare/Medicaid Status, Level of care needed (Level of Care Index), Location of Care, Major health problem or condition (Netherlands Health Monitors Questionnaires), Dementia (Neuropsychiatric Inventory, IQCODE), Self-reported health (PROMIS).</td>
</tr>
<tr>
<td>Caregiving Context</td>
<td>Caregiver-Care Recipient Relationship: the nature of a biological, legal, or chosen relationship (e.g., parent, child, grandparent, live-in partner, friend). Physical proximity (e.g., co-residence, long-distance, institutionalized nearby); reasons for informal care provision (e.g., love and affection, no institutional care available/affordable); whether the caregiver has responsibilities caring for more than one person (e.g., two aging parents, or a parent and a child). Intensity, which is measured in terms of hours per week spent caregiving, and duration of the caregiving role in terms of weeks, months, or years. Intensity also accounts for the types of caregiving tasks provided/needed (e.g., household care, personal care, medical/nursing care, emotional support, assistance with administrative/financial issues, assistance with outside activities); as well as presence and role of other caregivers.</td>
<td>Relationship Quality (EuroQOL vertical visual analogue scale (EQ-VAS), Expectation of future caregiving; Intensity (Archangel Caregiving Intensity Index&lt;sup&gt;12&lt;/sup&gt;, Level of Care Index&lt;sup&gt;13&lt;/sup&gt;)</td>
</tr>
</tbody>
</table>

(Continued)
Table 1. (Continued).

<table>
<thead>
<tr>
<th>Domain:</th>
<th>Definition:</th>
<th>Example Constructs (and Related Standardized Measures)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supports and Resources for</td>
<td>Support: tangible contributions by others to caregiving tasks, instrumental social support, the use of formal support services, and satisfaction with services and supports. Resources: financial support, workplace benefits, training, and information.</td>
<td>Social support, engagement, and capital (Berkman social network questionnaire); Perceived availability of assistance with material, cognitive, or task (PROMIS instrumental support 4-item short form); Use/cost of formal support services; Workplace benefits; Training; Use of technology; Healthcare time and costs (primary care, hospital admissions, emergency medical services, skilled nursing facility, Hospice, paid in-home care)</td>
</tr>
<tr>
<td>Caregiving</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health care Utilization</td>
<td>Health care utilization by caregiver or care recipient includes frequency and temporal measures (e.g., within 30 days of last hospital admission) of formal health system use.</td>
<td>Financial strain and impact (e.g., stopped saving, took on debt, borrowed money, put off retirement); work impacts (e.g., go in late, leave early, change work hours, take leave of absence, turn down promotion); ability to combine informal and paid care; and satisfaction with employer (EQ-VAS).</td>
</tr>
<tr>
<td>Cost and Value of Caregiving</td>
<td>The cost and value of caregiving include estimates of the financial impact (both positive and negative) on individual caregivers and care recipients, public and private employers, and local and national economies. These measures include estimates of hidden, hypothetical or opportunity costs (i.e., loss of potential gain from other alternatives).</td>
<td></td>
</tr>
</tbody>
</table>

*Definitions derived from the APA Dictionary of Psychology and by considering the constructs measured in the exemplar instruments.

1(American Psychological Association, 2011; Zarit et al., 1980); 2(van Exel et al., 2004); 3(Given et al., 1992); 4(Thornton & Travis, 2003); 5(Tarlow et al., 2004); 6(Broese van Groenou et al., 2013); 7(Kroenke et al., 2001); 8(Bjelland et al., 2002); 9(Löwe et al., 2008); 10(Charlson et al., 2022); 11(Brooks, 1996); 12(ARCHANGELS, n.d.); 13(HRS

Screening Interview (NHATS/NSOC)); short form Informant Questionnaire on Cognitive Decline in the Elderly [IQCODE; (ADAMS/HRS)].

All datasets measured some aspect of caregiving context but did not use validated measures or consistent definitions. Constructs included the type of relationship between caregiver and care recipient (e.g., spouse, child, sibling; some included choices for live-in partners, in-laws, and non-family friends); physical proximity (e.g., co-residence, long-distance); duration of caregiving relationship (months, years); time spent providing care (hours/week, days/months), frequency and types of caregiving tasks provided/needed (e.g., household, nursing, emotional, financial, coordination); relationship quality; expectation of future caregiving; and choice in taking on caregiver role. Some datasets inquired if the caregiver cares for multiple care recipients (e.g., parent and child, two parents).

Twelve datasets measured supports and resources for caregivers including social support (e.g., other unpaid caregivers, paid help); social connection; use, cost, and insurance coverage for formal support services (e.g., transport, respite, hospice); workplace benefits (e.g., paid leave); training; and unmet needs. Although validated measures exist for some of these supports and resources constructs, we only identified use of non-validated measures. Only nine datasets assessed health care utilization (e.g., hospital, ER, Hospice). Cost and value of caregiving was captured in seven datasets and included financial
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Time Period, Design, Data Collection</th>
<th>Target Population &amp; Sampling</th>
<th>Data Access &amp; Use</th>
<th>Example Secondary Data Analysis Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aging, Demographics, and Memory Study (ADAMS)/Health and Retirement Study (HRS)</td>
<td>2001-ongoing; Longitudinal Sub-study of and follow-up via HRS. Caregivers completed self-administered paper-and-pencil caregiver questionnaire. Also followed through HRS.</td>
<td>n = 856 U.S. Adults 70+ care recipients as subjects &amp; n = 856 matched caregivers informants. Population-based using HRS sample, stratified based on cognitive function. Protocol required participation of both the HRS respondent (subject) and an informant who was familiar with the subject's daily activities and medical history (usually a close family member e.g., spouse or child).</td>
<td>Restricted datasets and public release datasets (omit day of birth, day of assessment, and information on geographical location. <a href="https://hrsdata.isr.umich.edu/data-products/aging-demographics-and-memory-study-adams-wave-c">https://hrsdata.isr.umich.edu/data-products/aging-demographics-and-memory-study-adams-wave-c</a></td>
<td>Stites et al. (2023). PMID: 36790294</td>
</tr>
</tbody>
</table>

(Continued)
Table 2. (Continued).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Time Period, Design, Data Collection</th>
<th>Target Population &amp; Sampling</th>
<th>Data Access &amp; Use</th>
<th>Example Secondary Data Analysis Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health and Retirement Study (HRS)</td>
<td>1992-ongoing; Longitudinal. Cohort sample; Biennial interviews.</td>
<td>n = 15,723 respondents in the 2020 wave. US adults ages 50+ who reside in private households. CG &amp; CR = respondents; Spouses or partners recruited regardless of age. Biennial interviews and post-death interview with a knowledgeable proxy. The 2010 new cohort enrollment included a large expansion of the minority (Black and Hispanic) oversample. Panel response rates for this group have been comparable to the full sample.</td>
<td>1) Public (anonymized; download with registration), 2) sensitive (requires additional approval), 3) restricted (secure remote access or under a license). <a href="https://hrs.isr.umich.edu/data-products">https://hrs.isr.umich.edu/data-products</a></td>
<td>Miller, et al. (2023). PMID: 37081828</td>
</tr>
</tbody>
</table>

(Continued)
Table 2. (Continued).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Time Period, Design, Data Collection</th>
<th>Target Population &amp; Sampling</th>
<th>Data Access &amp; Use</th>
<th>Example Secondary Data Analysis Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey of the Health of Wisconsin (SHOW)</td>
<td>2008-ongoing; Longitudinal. Modeled after the CDC’s National Health and Nutrition Examination Survey (NHANES). Annual statewide survey (mailing followed by field staff in-home visits). computer-assisted personal interview (CAPI) program. Includes Caregiver Module.</td>
<td>n = 6,000+ U.S. (Wisconsin residents) Aged 18+: more than 1/3 rural participants, 34% of whom are living at or below 200% of the Federal Poverty Level (FPL) and 13% without stable health insurance.</td>
<td>Publicly available and restricted use datasets.</td>
<td>Litzelman, et al (2014). PMID: 24780842</td>
</tr>
<tr>
<td>Dataset</td>
<td>Time Period, Design, Data Collection</td>
<td>Target Population &amp; Sampling</td>
<td>Data Access &amp; Use</td>
<td>Example</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------</td>
<td>-----------------------------</td>
<td>-----------------</td>
<td>---------</td>
</tr>
</tbody>
</table>

Acronyms: CDC = Centers for Disease Control and Prevention; CMS = Centers for Medicare and Medicaid Services; NAC = National Alliance for Caregiving; NCI = National Cancer Institute; NDI = National Death Index; NIA = National Institute on Aging; NIH = National Institutes of Health.
strain (e.g., debt, bills unpaid/late, evicted/foreclosed on); and impacts of caregiving on employment (e.g., arrive late/leave early, unpaid leave, no promotion, lose job, leave workforce).

Discussion

It is critical to establish a robust, reliable, and standard evidence base to facilitate needed research about family caregivers. This review identifies existing publicly available survey-based population datasets, describes features that may encourage broader use, and allows us to identify deficiencies in the existing data infrastructure. These datasets offer several strengths for accelerating caregiving science, such as large, representative sample sizes and detailed data provided by caregivers about their health, the context of caregiving, and their unmet needs. Five are linked to administrative data sources (e.g., Medicare claims, Social Security, and the National Death Index) to enable additional information about survey participants. Resources such as the Federal Statistical Research Data Centers (U.S. Census Bureau) offer new opportunities to find and link administrative data to these caregiver-focused datasets.

Most existing datasets focused on the negative aspects of caregiving (e.g., burden, strain), with only half capturing measures of positive aspects (e.g., greater sense of purpose and satisfaction) (Berg et al., 2021; Marino et al., 2017; Pysklywec et al., 2020; Schulz et al., 2020; Smaling et al., 2021). More research is needed to better understand the positive aspects of caregiving, however collecting the data needed for this purpose relies on testing applicability of existing measures, developing new measures, and including them in survey-based studies (Galvin et al., 2020; Martin et al., 2021; Smaling et al., 2021).

Data are needed to address the many health disparities that impact family caregivers, because “health equity requires data equity” (Boyd et al., 2020; Ponce et al., 2023; Rote et al., 2019). Although large datasets provide sufficient power for analysis, most still do not adequately represent diverse communities (Marani, 2021; Pruchno et al., 2008). Several datasets address this by oversampling ethnic and racial populations; others conduct ancillary studies to increase underrepresented populations. Regardless, volunteer bias remains an issue (Brayne & Moffitt, 2022). Use of consistent measures across survey-based studies would enable analysis of data across multiple studies to gain better insight about caregivers in diverse contexts (Giovannetti & Wolff, 2010).

Despite decades of research documenting the impact of caregiving on families (Albright et al., 2016; Keita Fakeye et al., 2023; Koehly et al., 2015; Mellins, 1993), most data collection focuses on a single primary caregiver. Experts argue for a more inclusive and consistent definition of family caregivers (RAISE Family Caregiving Advisory Council & The Advisory Council to Support Grandparents Raising Grandchildren, 2022), with greater focus on
<table>
<thead>
<tr>
<th>Caregiving domains examined in survey-based population datasets.</th>
<th>ADAMS/HRS</th>
<th>BRFSS-CM</th>
<th>CG-US</th>
<th>HINTS</th>
<th>Hispanic-EPESE</th>
<th>HRS</th>
<th>MIDUS</th>
<th>NLTCS</th>
<th>NPHA</th>
<th>NSHAP</th>
<th>NSOC/NHATS</th>
<th>SHOW</th>
<th>WLS-CM</th>
<th>WLS-SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caregiver Burden and Strain</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Positive Aspects of Caregiving and Resilience</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caregiver (CG) Demographics &amp; Health</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Care Recipient (CR) Demographics and Health</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caregiving Context</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supports and Resources for Caregiving</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthcare Utilization</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(by CG or CR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost and Value of Caregiving</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

social networks that may be sharing care responsibilities (Kristensen et al., 2021; Lowers et al., 2023). Findings from this analysis highlight how the nature and needs of these family networks are masked by survey design. Even when studies ask about how many caregivers are involved in caring for a person, respondents are typically forced to pick the “primary” caregiver. As a result, data are only collected about that primary caregiver, thereby limiting our ability to consider how multiple caregivers share roles and how roles differ (Douglas et al., 2016; Griffin et al., 2022).

Analyzing representative population-based datasets, especially those with linked administrative, health, and vital records, may hold promise for advancing our understanding of the role families play throughout the caregiving and bereavement continuum. Such linkages also enable much needed research about healthcare utilization and the cost and value of care, which typically cannot be evaluated based on survey-based data alone (Giovannetti & Wolff, 2010). Work being conducted outside the U.S. may provide useful guidance for researchers and policymakers alike. Population datasets such as in Denmark, Sweden, and the Utah Population Database, capture data from the entire country or state population, instead of attempting to survey a representative sample (Hollingshaus, et al., 2024; Kristensen et al., 2021; Smith et al., 2022; Stephens et al., 2023; Stephens, et al., 2024; Tay et al., 2022). By linking administrative, genealogical, environmental, and detailed verified health records these datasets offer the opportunity to characterize health information across family networks and changes over time to provide a different source of understanding caregiver outcomes.

Implications for policy development, implementation, and research

This review highlights continued gaps in the consistent use of validated measures and representative populations to generate evidence needed to assess the experiences, needs, and economic value of family caregivers (Freedman et al., 2004; RAISE, 2022. Linking survey-based datasets with population datasets and administrative records data has the potential to inform policymakers about the economic impact of caregiving including work-related opportunity costs for the caregivers, and costs borne by U.S. businesses. Current estimates suggest that U.S. businesses are losing more than $33 billion per year due to employees’ caregiving responsibilities, but estimates vary widely and are based on old data (Bauer & Sousa-Poza, 2015; Committee on Family Caregiving for Older Adults et al., 2016; Meyer et al., 2023; Mudrazija, 2019; RAISE, 2022). Better data about these economic impacts could help to overcome the policy drift that has inhibited caregiver policy development and implementation (Levitsky, 2014; Rocco, 2017).

Implementation of policies like the Caregiver, Advise, Record, and Enable (CARE) Act in the US might provide opportunities to expand
claims-based data to directly address caregiver education or use proxy measures such as the “Patient and Caregiver-Centered Experience of Care/ Care Coordination” domain added in 2018 to Medicare’s Hospital Value-Based Purchasing Program (Leighton et al., 2020); however, these changes may amplify disparities in rural and medically underserved areas where implementation involves numerous additional challenges (Griffin et al., 2022). Nevertheless, even with administrative data, rigorous measures, and all possible dataset linkages, it is still challenging to gain needed insight about the structure, impact, and needs of nontraditional or not readily identifiable “chosen” family networks that include neighbors, friends, people providing care from a distance, and others with whom relationships are not readily ascertainable from datasets. This is particularly vital for older adults who are “aging solo” without traditional family support (Lowers et al., 2023). Mixed methods research and other approaches to complement data science methods will be critical to understand and respond to those needs.

We identified lesser-known datasets that are single diagnosis-focused or focused on specific minoritized populations (Hispanic-EPESE, MIDUS ancillary Milwaukee African American sample, SHOW). Such datasets may provide useful insights about culturally competent measures and study designs especially suited for diverse communities. The absence of validated measures, in general, may reflect a need for the development of more culturally and linguistically representative measures that account for the distinct care experiences and specific family caregiving support needs of LGBTQ+, rural, impoverished, Black/African American, Hispanic/Latino, Asian American, Native Hawaiian and Pacific Islander, and American Indian and Alaska Native families (Fabius & Parker, 2023; Reinhard et al., 2023; Rote et al., 2019).

As we strive for data harmonization in the field of caregiving science, future research should examine cross-cutting domains, constructs, and measures. Much can be learned from international colleagues conducting caregiving research. For example, the Midlife in Japan (MIDJA) study’s survey and biomarker measures parallel those used in MIDUS, enabling testing of hypotheses about psychosocial factors, including cultural influences, in the health of mid- and later-life adults in both the U.S. and Japan. HRS sister studies collect data from more than 30 countries by using consistent measures to enable data harmonization for multinational studies.

**Conclusion**

While leveraging population datasets has potential to accelerate caregiving science and health policy, this review finds a lack of standardization or
harmonization across data sets for data collection, sampling, and measurement of key caregiving constructs. Aligned with the RAISE Family Caregiving Advisory Council recommendations (2022), findings highlight the need to collect data more consistently, using common terminology. This requires assessing the applicability and equity of existing measures, developing new measures, and including them in survey-based studies (Galvin et al., 2020; Martin et al., 2021). By improving data infrastructure and harmonization efforts, we can better maximize the utility of population-based datasets for caregiving research and provide more effective policy support.

**Key points**

- Despite growing reliance on family caregivers for older adults, much is unknown.
- Longitudinal family caregiver studies are challenging, costly and time-consuming.
- We identified 14 survey-based datasets with family caregiving data.
- Lack of validated measures and underrepresented populations is common.
- Population datasets may accelerate caregiving science and health policy.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This work was supported by the National Institute on Aging at the National Institutes of Health under Paul B. Beeson Award 5K76AG054862-05 and the Academic Leadership Career Award K07AG068185; the University of Utah Family Caregiving Collaborative [FCC]; University of Utah Center on Aging; University of Utah Consortium-Families and Health Research [C-FAHR]; and the Mt. Sinai Pepper Center. Dr. Elmore was supported by the National Institute of Nursing Research at the National Institutes of Health under grant T32NR013456. Ms. Goodwin was supported by the National Library of Medicine at the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Disclosure statement: Authors have no conflicts of interest to disclose.

**ORCID**

Rebecca M. Goodwin [http://orcid.org/0000-0001-8792-8845](http://orcid.org/0000-0001-8792-8845)

Rebecca L. Utz [http://orcid.org/0000-0001-7734-7676](http://orcid.org/0000-0001-7734-7676)

Catherine E. Elmore [http://orcid.org/0000-0002-5908-8727](http://orcid.org/0000-0002-5908-8727)

Katherine A. Ornstein [http://orcid.org/0000-0001-6270-4423](http://orcid.org/0000-0001-6270-4423)
Data availability statement

Data sharing is not applicable to this article as no new data was created or analyzed in this review.

References


ARCHANGELS. (n.d.). Caregiver intensity index. https://www.archangels.me/


Pharmacoeconomics and Outcomes Research, 26(5), 712–720. [https://doi.org/10.1016/j.jval.2022.06.014](https://doi.org/10.1016/j.jval.2022.06.014)


Smith, K. R., Fraser, A., Reed, D. L., Barlow, J., Hanson, H. A., West, J., Knight, S., Forsythe, N., & Mineau, G. P. (2022). The Utah population database: A model for linking medical and


