Using Healthcare Data in Embedded Pragmatic Clinical Trials among People Living with Dementia and Their Caregivers: State of the Art

Julie P.W. Bynum, MD, MPH, David A. Dorr, MD,† Julie Lima, MPH,‡ Ellen P. McCarthy, PHD,§¶ Ellen McCreedy, PHD,‡ Richard Platt, MD,‖ and V.G. Vinod Vydiswaran, PHD**††*

Embedded pragmatic clinical trials (ePCTs) are embedded in healthcare systems as well as their data environments. For people living with dementia (PLWD), settings of care can be different from the general population and involve additional people whose information is also important. The ePCT designs have the opportunity to leverage data that becomes available through the normal delivery of care. They may be particularly valuable in Alzheimer's disease and Alzheimer's disease-related dementia (AD/ADRD), given the complexity of case identification and the diversity of care settings. Grounded in the objectives of the Data and Technical Core of the newly established National Institute on Aging Imbedded Pragmatic Alzheimer's Disease and AD-Related Dementias Clinical Trials Collaboratory (IMPACT Collaboratory), this article summarizes the state of the art in using existing data sources (eg, Medicare claims, electronic health records) in AD/ADRD ePCTs and approaches to integrating them in real-world settings. **J Am Geriatr Soc 68:S49-S54, 2020.**

From the *Department of Internal Medicine, Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, Michigan; †Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon; ‡Center for Gerontology and Healthcare Research, School of Public Health, Brown University, Providence, Rhode Island; §Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, Massachusetts; ¶Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts; ‖Department of Population Medicine, Harvard Pilgrim Health Care Institute, Harvard Medical School, Boston, Massachusetts; **Department of Learning Health Sciences, University of Michigan, Ann Arbor, Michigan; and the ††School of Information, University of Michigan, Ann Arbor, Michigan.

Address correspondence to Julie P.W. Bynum, MD, MPH, NCRC-16, Room 446E, 2800 Plymouth Road, Ann Arbor, MI 48109-2800. E-mail: bynumju@med.umich.edu

DOI: 10.1111/jgs.16617

Keywords: pragmatic trial; dementia; measurement; Alzheimer's disease

Health system embedded pragmatic clinical trials (ePCTs) that address the needs of people living with dementia (PLWD) and their caregivers are critical for developing and disseminating evidence-based nonpharmacologic interventions. As described by Mitchell et al,[1] these trials create the opportunity to design interventions that will work in real-world patients and settings, but they require investigators to embrace new methods and partnerships. A key element of the ePCT approach is to leverage data derived from and integrated with the healthcare system workflow into the trial's design, conduct, and dissemination. This approach allows cost-effective identification of participants and outcome data ascertainment. With the rise of electronic health records (EHRs) and focused attention on ePCTs stimulated by the National Institutes of Health (NIH) Collaboratory, capabilities for using data generated by health care have been advancing. Yet unique issues facing PLWD necessitate innovative strategies in using these data across the multiple healthcare settings targeted in Alzheimer's disease and Alzheimer's disease and related dementias (AD/ADRD) ePCTs.

The unique challenges of using existing data sources to conduct ePCTs in AD/ADRD fall into several categories: (1) AD/ADRD are underdiagnosed[2] and stigmatized diseases,[3-5] (2) caregivers often need to be identified,[6] (3) data must be accessed from settings outside the traditional acute care medical system (eg, primary care, nursing homes (NHs), and assisted living), (4) patient- and caregiver-reported outcomes must be ascertained, and (5) measures are needed that span multiple settings (eg, care transitions). Despite these challenges, opportunities exist to strengthen our ability to identify PLWD and their caregivers and to measure outcomes by leveraging data sources available through administrative data or EHRs.

0002-8614/20/$15.00

These data are useful at multiple points in the ePCT process including the design phase, conduct of the pilot and full trial, and subsequent dissemination. But careful consideration of the "fitness for use" of a particular data strategy is critical at all stages. This report provides an overview of using healthcare data in ePCTs on which the objectives of the Data and Technical Core of the newly established National Institute on Aging Imbedded Pragmatic Alzheimer's Disease and AD-Related Dementias Clinical Trials (IMPACT) Collaboratory are based and serve as the groundwork for our future work in addressing the unique data challenges in ePCTs among PLWD.

OVERVIEW OF TYPES OF DATA AND SOURCES

Clinical trialists are well versed in data collection strategies for studies that directly recruit individual participants into intervention and control groups. These approaches use validated instruments for assessing participants and their outcomes, and they employ research staff to collect data meticulously in person, by phone, or through electronic media. The challenges of scaling this traditional approach to data collection can limit a trial's size and the settings in which it can be conducted. When conducting an ePCT, investigators have an opportunity to reduce costs and burden by using existing data that have been ascertained in the course of usual clinical care (Table 1). These data include administrative data, generated for billing or regulatory purposes, and EHR data that include structured elements (eg, laboratory results, diagnostic codes, medications) and unstructured or text fields (eg, clinical notes and imaging reports).

These existing data can be accessed from federal sources, private payers, directly from specific health systems, or in some cases through intermediaries who facilitate collaboration with multiple healthcare systems and payers, such as the Distributed Research Network (DRN) established NIH HCS Collaboratory (https://rethinkingclinicaltrials.org/nih-collaboratory-drn/). The DRN implements a common data model that facilitates use of data from both Medicare Advantage and commercially insured individuals across multiple payers. The Patient-Centered Outcomes Research network (PCORnet) is another DRN funded by the Patient-Centered Outcomes Research Institute that has implemented a related common data model based on EHR data (www.pcornet.org).

Because the age of onset of dementia is most commonly older than 65 years, data from the Centers for Medicare & Medicaid Services (CMS) are a particularly valuable asset to ePCTs for PLWD. Medicare claims from CMS include hospital, post-acute care, clinic, hospice, and skilled nursing facility billing data. Medicare claims data have the advantage of being complete because they include all Medicare beneficiaries in the healthcare setting whether the participants complete the trial or not, and they are uniform across various healthcare systems. Although these data have historically been available only for people enrolled in fee-for-service Medicare, data from Medicare Advantage are increasingly becoming available. In addition to claims, CMS also has "assessment data" from NHs captured in the Minimum Data Set (MDS) and from home health agencies captured in the Outcome and Assessment Information Set (OASIS). CMS requires regular collection of these data to calculate payments and monitor quality. MDS and OASIS also include clinical information such as cognitive and functional status and behavioral issues. For example, MDS data capture standardized assessment of all residents in more than 15,000 NHs in the United States. Assessments are administered at minimum quarterly, making the MDS a rich source of resident status over time. In addition, although the NH setting lags behind hospitals, 60% of NHs also have EHRs.[7]

Table 1. Healthcare-Generated Data Types, Their Content, Examples of Potential Uses, and the Sources Used

Data type	Information contained in data type	Examples of possible use	Example sources for data type
Claims data	Inpatient, clinic, home health, hospice, medication data used for billing	Identify participants diagnosed with dementia; measure outcomes (eg, readmissions, hospital transfers, antipsychotic use)	CMS Virtual Research Data Center (FFS Medicare), NIH Collaboratory DRN (Medicare Advantage and commercial)
Enrollment data	Demographic, geographic, and plan type for enrollees in insurance (fee for service/Medicare Advantage/ commercial)	Estimate available eligible sample Identify settings for further dissemination	CMS VRDC (FFS Medicare), NIH Collaboratory DRN (Medicare Advantage and commercial), OPTUM
Assessment files	Clinical data for quality reporting and payment in nursing homes and home care	Identify home health agencies or nursing homes with high proportion PLWD	Minimum Data Set (nursing home), OASIS (home health) from CMS or directly from healthcare setting
Electronic health records	Structured data (labs, problem lists), text fields, billing data, patient-reported outcomes	Cognitive screens and clinical notes to identify undiagnosed PLWD	Directly from participating healthcare setting, federated data intermediaries (eg, PCORnet, ACT)
Provider files	Type, size, location, ownership	Find physician practices serving high ethnic minority populations	PECOS, provider of service files, Medicare provider practice and specialty

Abbreviations: ACT, Accrual to Clinical Trials; CMS, Centers for Medicare & Medicaid Services; DRN, Distributed Research Network; FFS, fee for service; NIH, National Institutes of Health; OASIS, Outcome and Assessment Information Set; OPTUM, Optum Global Solutions; PCORnet, Patient-Centered Outcomes Research network; PECOS, Medicare Provider Enrollment, Chain, and Ownership System; PLWD, people living with dementia; POS, Medicare Place of Service File; VRDC, Virtual Research Data Center.

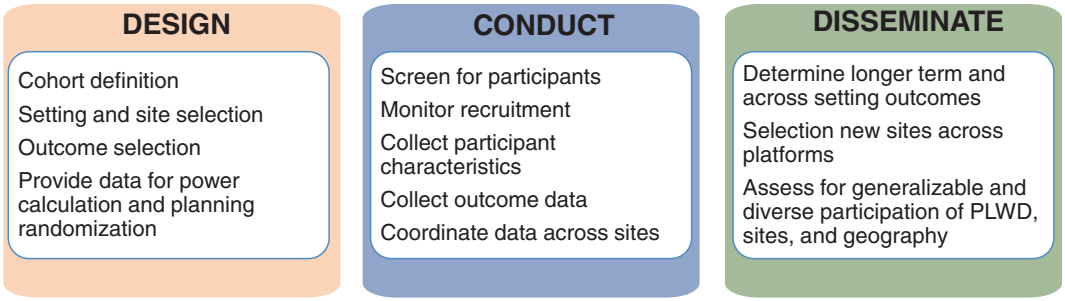

Figure 1. Overview of uses of administrative or electronic health record data in embedded pragmatic clinical trials. PLWD, people living with dementia.

USES OF HEALTHCARE-GENERATED DATA IN ePCTs

Healthcare-generated data can be used at multiple stages in the ePCT process, from the design phase, to conduct of the pilot study and full trial, to subsequent dissemination (Figure 1). In the design phase, which is the main focus of the IMPACT Collaboratory, investigators can use existing data to identify potential participants and calculate power and sample size estimates. Existing data can also help in the design phase by enabling the identification and characterization of eligible healthcare settings including considerations of representation of diverse populations. Aggregated data from healthcare systems are available from EHR data infrastructures; publicly available data sources such as provider files available through CMS (eg, Medicare Provider Enrollment, Chain, and Ownership System [PECOS],[8] Certification and Survey Provider Enhanced Reports [CASPER][9]) or through websites (eg, OHSDI Atlas,[10] Accrual to Clinical Trials [ACT],[11] Nursing HomeCompare (medicare.gov/nursing HomeCompare),[12] LTFocus (ltcfocus.org),[13] and Dartmouth Atlas[14]). These aggregated data can also be used to assure balance on key measures between clusters in each trial arm during the randomization process.

In the process of study execution, both claims and EHRs can be used to identify specific participants, evaluate adherence to protocol, and measure outcomes. Using administrative data to measure longer term outcomes, including utilization and spending, is critical because it can be done long after the trial is complete without the need for direct participant contact. Finally, administrative data are useful to identify new sites for the next phases of implementation.

IDENTIFICATION OF PLWD FROM HEALTHCARE-GENERATED DATA

One of the greatest potential advantages of using healthcare data in AD/ADRD ePCTs is the ability to identify eligible participants without directly assessing an individual participant's cognitive status. Instead, diagnoses in administrative data required for clinician billing can be used to identify a PLWD. Similarly, EHRs contain structured data elements with diagnoses populating problem and medical history lists. These diagnoses can be used to identify participants for ePCTs, with the major caveat that underdiagnosis limits this approach. A recent meta-analysis estimated 60% of AD/ADRD cases are undiagnosed in the community.[2] Studies evaluating the accuracy of claims diagnoses have shown good performance but with under-ascertainment of mild disease in particular.[15-19] Moreover, none of the algorithms for the identification of AD/ADRD in Medicare claims has yet been validated using International Classification of Diseases, Tenth Revision, diagnostic coding that began in 2015.

Access to EHRs that can be searched for symptoms, clinician comments, results of annual wellness screening examination, and other data elements presents an opportunity to rely less on confirmed diagnoses from purely administrative sources. The ability to use documented clinical data may be particularly important for identification of racial/ethnic minorities in whom differences in stigma attached to AD/ADRD and beliefs about cognitive loss as a normal part of aging may contribute to lower rates of formal diagnosis.[4,20-23] Despite major gaps in consistent assessment of AD/ADRD, significant advancements have been made by changing data collection strategies from reliance on one data source to using combinations of EHR, claims, survey, and other data.

By combining different data types and sources, validated data combinations, called "computable phenotypes" in informatics, can be created for more sensitive and reliable assessments of AD/ADRD status. Barnes et al[24] describe eRADAR, a high-performing algorithm that uses common EHR data to identify patients with undiagnosed dementia. The eMERGE consortium has a public computable phenotype to identify people visits for dementia-related diagnosis or prescriptions for dementia-related drugs.[25] In other recent work, McCoy et al applied a validated natural language processing (NLP) tool to examine the association of cognitive symptoms with incident dementia diagnosis using longitudinal EHRs,[26] and Beltrami et al[27] used NLP to identify early linguistic signs of cognitive decline, not necessarily dementia itself, in a population of older adults.[27] In addition, a large database of multimedia interactions and transcripts, DementiaBank,[28] is available for the study of communication in dementia patients, and it was used to study NLP techniques to classify and analyze the linguistic characteristics of AD patients.[29-31]

Despite these exciting developments in the use of EHR data for identifying PLWD, a number of critical issues are left to address. It is imperative that investigators using their own "computable phenotype" to identify PLWD in a healthcare system validate and share their approaches. Investigators who opt to use an existing validated computable phenotype can both shorten development time and provide measures of accuracy for further validation in this growing field. Several online resources are available

including the Phenotype Knowledge Base (PheKB [phekb.org]); via PhenX, a curated resource for research-specific definitions; and via the NIH Clinical Data Elements database, among others. Sharing the definitions used in ePCTs via open source online resources can help continue to improve the quality and transferability of future trials.

Several cautionary issues pertaining to use of EHR data in conducting ePCTs in PLWD are worth noting. It is important for investigators to be aware of poor or uneven data quality that continues to exist in EHRs. Even previously validated definitions and algorithms for identifying AD/ADRD populations and assessing their outcomes must be validated locally in each healthcare system to account for variations across settings and purposes. When data quality issues are discovered, it may be possible to mitigate them by combining with other data, using advanced statistical approaches such as imputation or by performing sensitivity analyses. In addition, there are trade-offs to consider when choosing which healthcare-generated data source to use. Although EHR data are timely, their completeness and accuracy are variable. In addition, algorithms identifying PLWD suffer from potential biases, such as lower accuracy in minority populations and those with lower healthcare access. However, Medicare claims data may be less timely and have lower sensitivity for early disease but are simpler to implement. They have a reasonable evidence base in terms of validation and minimal missing data.

CASE STUDIES: THE METRIcAL AND PROVEN TRIALS

We provide two case studies of ePCTs in NHs for patients with advanced dementia that use administrative data and EHR data to highlight the advantages and potential limits of using the pragmatic healthcare data approach. Although using similar data sources, the differing aims of each study highlight how the degree of pragmatism that can be achieved varies and the importance of piloting the planned data strategy to assess its fitness for the intended use in the trial.

MUSIC AND MEMORY: A PRAGMATIC TRIAL FOR NURSING HOME RESIDENTS WITH ALZHEIMER'S DISEASE

Music and Memory: A Pragmatic Trial for Nursing Home Residents with Alzheimer's Disease (METRIcAL) is an ePCT of a personalized music intervention, is one of several sensory and reminiscence therapies being explored as low-risk alternatives to pharmaceutical approaches in managing behavioral and psychological symptoms of dementia.[32] In METRIcAL, nursing home staff identify music a resident preferred when they were younger and deliver the music at early signs of agitation. The primary aim of METRIcAL is to determine whether or not personalized music reduces agitation among residents with advanced dementia compared with usual care. The pilot phase of METRIcAL was completed in 2018. The ePCT is currently underway; 81 NHs from four corporations are enrolled, 27 NHs receive the intervention in each study year (2019, 2020, and 2021).

PRAGMATIC TRIAL OF VIDEO EDUCATION IN NURSING HOMES

The Pragmatic Trial of Video Education in Nursing Homes (PROVEN) was an ePCT of a video to assist with advance care planning for NH residents with advanced dementia or advanced cardiopulmonary disease.[33] This population was chosen because it is likely to experience unnecessary and nonbeneficial care at the end of life including multiple hospital transfers. The primary outcome of interest was hospital transfers per person-days alive. A total of 360 NHs (intervention arm n = 119; control arm n = 241) within two NH healthcare systems were enrolled in the trial. Early results suggest the videos helped residents and their surrogates think differently about their medical choices and prompted conversations with a provider.[34]

DATA USE AND LESSONS LEARNED

Both ePCTs benefited from the routine collection of MDS assessments that contain diagnosis and cognitive and physical function for all NH residents. MDS was used to identify residents who had been in the NH at least 90 of the last 100 days and who had a dementia diagnosis. The PROVEN intervention was delivered as a quality improvement intervention to everyone in the NH during the study period, so NH staff did not need to be aware of which residents were targeted for outcome analyses. This highly pragmatic approach was not possible in METRIcAL in which a subset of eligible residents were targeted to receive the intervention because of its resource-intensive nature, requiring equipment (mp3 players, headphones, etc) and staff effort in personalizing music selection. An on-site formal process for selecting study targets from potentially eligible residents identified in MDS was necessary in treatment and control NHs.

A second important use of healthcare data in both ePCTs was to measure the main outcome, hospital transfers using Medicare claims data for PROVEN and occurrence of agitated behaviors assessed in MDS for METRIcAL. The PROVEN trial approach was straightforward; claims data allowed for complete case ascertainment because participants were all in Medicare, and hospitals uniformly submit bills for admission. For METRIcAL, the main outcome of agitated behavior collected through MDS depends on recognition and documentation of the behaviors by nursing staff. During the pilot phase of METRIcAL, investigators discovered that agitated behaviors were underdetected in MDS, likely due to staff normalization of those behaviors over time.[35] The measurement strategy for the full-scale ePCT was adjusted to include on-site data collection for a randomly selected subset of participants.[36]

In the design and pilot phase, both trials used what they observed in administrative data about their main outcome measures to address imbalances across sites that could be addressed by altering their randomization protocols. In METRIcAL, based on observed variation in documentation of agitated behaviors across NHs and the process for selecting eligible residents, study arms were balanced on behaviors and number of potentially eligible residents before randomization. Similarly, in PROVEN, trial arms were purposefully balanced at baseline on their historical

rate of hospital transfers (primary outcome) to address the high underlying variation in the rate of hospitalization across NHs.

Finally, both PROVEN and the pilot phase of METRICAL inserted new fields into the EHR to capture implementation adherence. The customized report integrated into the EHR to capture implementation adherence with the ePCT was underused and disliked by frontline providers, for whom the report had no relevance to clinical care.[37] When planning an ePCT, researchers should use caution when inserting new elements into a workflow that do not serve a clinical purpose evident to frontline providers.

These two recent ePCTs conducted in NHs illustrate how routinely collected administrative data can be leveraged to promote balanced randomized clusters, streamline NH recruitment, facilitate patient selection, and enable an efficient, pragmatic approach to outcome ascertainment. However, the data strategy can introduce new challenges, and like many aspects of conducting a clinical trial, it can benefit from pilot testing.

ACCESS, PROTECTION, AND SHARING OF DATA

Investigators need to initiate plans for accessing healthcare data early and plan for a lengthy process of gaining approvals and developing partnerships. Obtaining administrative data from federal sources, such as CMS, has a well-defined but lengthy process managed by ResDAC (https://www.resdac.org). Obtaining data directly from health systems is attractive, but healthcare systems may not be familiar with the Health Insurance Portability and Accountability Act Privacy Rule as it pertains to research and may find the regulatory process daunting. Even when willing, it can take months or years to enter into the appropriate arrangements if healthcare systems are unfamiliar with the process.

The use of EHR and other administrative data within ePCT designs, coupled with a focus on PLWD, require researchers to consider their data privacy and sharing options very carefully. Traditional efficacy trials typically obtain informed consent that includes explicit assurances to protect privacy and also authorizes plans for data sharing covering future use. In fact, the NIH has proposed an expansion of data-sharing rules for funded research.[38] When ePCTs use data generated through the delivery of health care that is in the possession of providers, delivery systems, and payers, there are additional considerations: (1) data obtained with waiver of consent, especially for system-level interventions or cluster randomized trials, precludes specific consent for data sharing; (2) data volume and content include a large number of data points, making deidentification of individuals difficult, perhaps impossible[39]; and (3) providers, delivery systems, and payers may not agree to participate if data about their organization can be used for unspecified secondary purposes.[39] In many cases, the ability to share individual data may ultimately be limited, and researchers should be able to provide a detailed description of the steps they took in obtaining their data that provides a practical guide for researchers looking to do similar research in the future.

In summary, data available through normal delivery of care to PLWD within healthcare systems present tremendous opportunities to strengthen the design and conduct of AD/ADRD ePCTs. However, because this field is both complex and relatively nascent, novel methodologies and approaches must proceed thoughtfully and rigorously. Under the leadership of the Technical Data Core, the IMPACT Collaboratory will help advance our ability to conduct successful ePCTs that can improve care for PLWD by supporting investigators' efforts to use healthcare-generated data. The core will help devise approaches to overcoming some of the barriers associated with using data obtained in the course of care by (1) connecting investigators to validated algorithms for identifying PLWD and contributing to creating them where they do not exist; (2) finding and developing measures for outcomes important to stakeholders including PLWD, caregivers, and health systems; and (3) generating information to help investigators find settings and healthcare system partners whose characteristics and populations served are well aligned with the study's aims.

ACKNOWLEDGMENTS

Financial Disclosure: This work was supported by the National Institute on Aging (NIA) of the National Institutes of Health under Award Number U54AG063546 that funds the NIA Imbedded Pragmatic Alzheimer's Disease and AD-Related Dementias Clinical Trials Collaboratory (NIA IMPACT Collaboratory).

Conflict of Interest: All authors are members of the Technical Data Core, and none have declared any conflicts of interest for this article.

Author Contributions: All authors are members of the Technical Data Core and contributed to all phases of writing this manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Sponsor's Role: None.

REFERENCES

1. Mitchell SL, Mor V, Harrison J, McCarthy EP. Embedded Pragmatic Trials in Dementia Care: Realizing the Vision of the NIA IMPACT Collaboratory. J Am Geriatr Soc. 2020;68(S2):S1-S7.
2. Iliffe S, Robinson L, Brayne C, et al. Primary care and dementia: 1. Diagnosis, screening and disclosure. Int J Geriatr Psychiatry. 2009;24:895-901.
3. Holston EC. Stigmatization in Alzheimer's disease research on African American elders. Issues Ment Health Nurs. 2005;26:1103-1127.
4. Rovner BW, Casten RJ, Harris LF. Cultural diversity and views on Alzheimer disease in older African Americans. Alzheimer Dis Assoc Disord. 2013;27:133-137.
5. Justiss MD, Boustani M, Fox C, et al. Patients' attitudes of dementia screening across the Atlantic. Int J Geriatr Psychiatry. 2009;24:632-637.
6. Langa KM, Chernew ME, Kabeto MU, et al. National estimates of the quantity and cost of informal caregiving for the elderly with dementia. J Gen Intern Med. 2001;16:770-778.
7. Alvarado CS, Zook K, Henry J. Electronic health record adoption and interoperability among U.S. nursing facilities in 2016. *ONC Data Brief 39*. 2017. https://dashboard.healthit.gov/evaluations/data-briefs/electronic-health-record-adoption-interoperability-nursing-facilities-2016.php. Accessed June 5, 2020.
8. Medicare Provider Enrollment, Chain, and Ownership System (PECOS). https://pecos.cms.hhs.gov/pecos/login.do#headingLv1. Accessed February 4, 2020.
9. Certification and Survey Provider Enhanced Reports (CASPER). https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Skilled-Nursing-Facility-Quality-Reporting-Program/SNF-Quality-Reporting-Program-Public-Reporting. Accessed February 4, 2020.
10. Observational Health Data Sciences and Informatics (OHSDI) ATLAS. https://www.ohdsi.org/data-standardization/the-common-data-model/. Accessed February 4, 2020.

11. Accrual to Clinical Trials (ACT) Network. https://www.ctsi.umn.edu/consultations-and-services/multi-site-study-support/accrual-clinical-trials-act-network. Accessed February 4, 2020.
12. Nursing Home Compare. https://www.medicare.gov/nursinghomecompare/search.html? Accessed February 4, 2020.
13. Long-term care: Facts on Care in the US. http://ltcfocus.org/. Accessed February 4, 2020.
14. The Dartmouth Atlas of Health Care. https://www.dartmouthatlas.org/. Accessed February 4, 2020.
15. Taylor DH Jr, Ostbye T, Langa KM, Weir D, Plassman BL. The accuracy of Medicare claims as an epidemiological tool: the case of dementia revisited. J Alzheimers Dis. 2009;17:807-815.
16. Newcomer R, Clay T, Luxenberg JS, Miller RH. Misclassification and selection bias when identifying Alzheimer's disease solely from Medicare claims records. J Am Geriatr Soc. 1999;47:215-219.
17. Ostbye T, Taylor DH Jr, Clipp EC, Scoyoc LV, Plassman BL. Identification of dementia: agreement among national survey data, Medicare claims, and death certificates. Health Serv Res. 2008;43:313-326.
18. Pressley JC, Trott C, Tang M, Durkin M, Stern Y. Dementia in community-dwelling elderly patients: a comparison of survey data, Medicare claims, cognitive screening, reported symptoms, and activity limitations. J Clin Epidemiol. 2003;56:896-905.
19. Taylor DH Jr, Fillenbaum GG, Ezell ME. The accuracy of Medicare claims data in identifying Alzheimer's disease. J Clin Epidemiol. 2002;55:929-937.
20. Chui HC, Gatz M. Cultural diversity in Alzheimer disease: the interface between biology, belief, and behavior. Alzheimer Dis Assoc Disord. 2005;19:250-255.
21. Chin AL, Negash S, Hamilton R. Diversity and disparity in dementia: the impact of ethnoracial differences in Alzheimer disease. Alzheimer Dis Assoc Disord. 2011;25:187-195.
22. Connell CM, Scott Roberts J, McLaughlin SJ. Public opinion about Alzheimer disease among blacks, Hispanics, and whites: results from a national survey. Alzheimer Dis Assoc Disord. 2007;21:232-240.
23. Connell CM, Scott Roberts J, McLaughlin SJ, Akinleye D. Racial differences in knowledge and beliefs about Alzheimer disease. Alzheimer Dis Assoc Disord. 2009;23:110-116.
24. Barnes DE, Zhou J, Walker RL, et al. Development and validation of eRADAR: a tool using EHR data to detect unrecognized dementia. J Am Geriatr Soc. 2020;68:103-111.
25. Group Health Cooperative. Dementia. https://phekb.org/phenotype/dementia. Accessed February 4, 2020.
26. McCoy TH Jr, Han L, Pellegrini AM, Tanzi RE, Berretta S, Perlis RH. Stratifying risk for dementia onset using large-scale electronic health record data: a retrospective cohort study. Alzheimers Dement. 2019;16:531-540.
27. Beltrami D, Gagliardi G, Rossini Favretti R, Ghidoni E, Tamburini F, Calza L. Speech analysis by natural language processing techniques: a possible tool for very early detection of cognitive decline? Front Aging Neurosci. 2018;10:369.
28. DementiaBank. https://dementia.talkbank.org/. Accessed January 20, 2020.
29. Chinaei H, Currie LC, Danks A, Lin H, Mehta T, Rudzicz F. Identifying and avoiding confusion in dialogue with people with Alzheimer's disease. Comput Linguist. 2017;43(2):377-406. https://doi.org/10.1162/COLI_a_00290
30. Orimaye SO, Wong JS, Golden KJ, Wong CP, Soyiri IN. Predicting probable Alzheimer's disease using linguistic deficits and biomarkers. BMC Bioinformatics. 2017;18:34.
31. Karlekar S, Niu T, Bansal M. Detecting linguistic characteristics of Alzheimer's dementia by interpreting neural models. 2018. https://arxiv.org/pdf/1804.06440.pdf. Accessed June 2, 2020.
32. Scales K, Zimmerman S, Miller SJ. Evidence-based nonpharmacological practices to address behavioral and psychological symptoms of dementia. Gerontologist. 2018;58:S88-S102.
33. Mor V, Volandes AE, Gutman R, Gatsonis C, Mitchell SL. Pragmatic trial of video education in nursing homes: the design and rationale for a pragmatic cluster randomized trial in the nursing home setting. Clin Trials. 2017;14:140-151.
34. Loomer L, McCreedy E, Belanger E, et al. Proxies viewing decision support video in nursing home report higher advance care planning engagement. J Am Med Dir Assoc. 2019;20:1467-1469.
35. McCreedy E, Ogarek JA, Thomas KS, Mor V. The Minimum Data Set Agitated and Reactive Behavior Scale: measuring behaviors in nursing home residents with dementia. J Am Med Dir Assoc. 2019;20:1548-1552.
36. McCreedy EM, Yang X, Baier RR, Rudolph JL, Thomas KS, Mor V. Measuring effects of nondrug interventions on behaviors: music & memory pilot study. J Am Geriatr Soc. 2019;67:2134-2138.
37. Loomer L, McCreedy E, Belanger E, et al. Nursing home characteristics associated with implementation of an advance care planning video intervention. J Am Med Dir Assoc. 2019;20:804-809.e801.
38. National Institutes of Health (NIH). Request for Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-013.html. Accessed February 4, 2020.
39. Simon GE, Coronado G, DeBar LL, et al. Data sharing and embedded research. Ann Intern Med. 2017;167:668-670.