The National Institutes of Health
National Institute on Aging

Cost-Effective Early Detection of Cognitive Decline

October 25, 2017

Revised December 21, 2017

Table of Contents

Executive Summary

The National Institute on Aging (NIA) convened a workshop to examine the current state of the science for cost-effective early detection of cognitive decline using both passive and active approaches and to identify research opportunities in this area. The Directors of the Division of Behavioral and Social Research, John G. Haaga, and the Division of Neuroscience, Eliezer Masliah, noted that such workshops are essential for the NIA to identify gaps, areas of research interest, and promising opportunities that can inform future funding directions for the NIA. Invited speakers included researchers and experts representing academia, nonprofit organizations, and industry.

As presented by invited speakers from academia, nonprofit organizations, and industry, the state of the science offers several conclusions. Several themes emerged from discussions of future research challenges and opportunities. Participants also identified several specific research priorities and ideas for consensus work needed in the field.

State of the Science
Dementia and mild cognitive impairment (MCI) are underdiagnosed, and analyses using claims data, as well as other methods, indicate that clinical diagnoses occur late in the process of cognitive decline. Less than 50 percent of persons with dementia are diagnosed, and the rates are even lower for MCI. Applying machine learning to multiple types of administrative data, electronic health record data, or passively collected data; implementing computerized assessment and screening tools in clinical settings; and using composites and risk indices all hold promise as strategies for early detection of possible cognitive decline. Invited speakers described multiple studies that demonstrate the feasibility of these approaches for surveillance and detection.

Consensus Work
Workshop participants identified several areas where further discussion, exploration, and consensus among stakeholders are warranted.

Clarity of Language
Participants noted that terms are used interchangeably or by different communities or contexts (e.g., research versus clinical community) to denote the same idea, yet the terms used have distinct meanings. For example, the terms "validation," "cognitive decline," "functional decline," "dementia," and "cognitive impairment" have been used in different ways by different communities and at different times. It is possible to have different thresholds—and therefore definitions—for research versus clinical purposes. Consensus work should be done to define common terminology for clear and consistent use.

Ethical Guidelines
Ethical issues may be raised by passive data collection from wearables and location-based sensors, and the use of commercial financial data for early detection of cognitive decline. These issues reflect how passive monitoring and detection impact on an adults' identity, privacy, authority, and normality. Stakeholders should explicitly examine these ethical issues and, specifically, develop guidelines for providing consent or assent, ownership of data collected, principles for deciding with whom the data are shared, what actions might follow an alert or flag generated by the data, and how monitoring and detection impacts on a person's identity and normality.

Research versus Clinical Purpose

Research tools or instruments are often not ready for implementation into existing clinical workflows. It would be useful to explore the potential to delineate sensitivity and specificity thresholds for research versus clinical purposes. For example, a consistent metric such as area under the curve (AUC) is appropriate to summarize general instrument performance, but is not the best metric for making statements about individuals. The intervention design and validation methodology for new tools should be appropriate for the intended outcome and use; a single standard should not be expected to apply to all tools and tests. Stakeholders should discuss and develop standardized guidelines for how new tools, analytic methods, and data collection methods should be validated.

Establishing a Chain of Research Evidence

Participants discussed the U.S. Preventive Services Task Force (USPSTF) grade of "Insufficient" (I) for assessing the balance of benefits and harms of screening for cognitive impairment.[1] The USPSTF recognized the importance of detecting individuals with dementia and MCI, and acknowledged evidence that clinically useful screening tools exist. However, it determined there is inadequate evidence of clear benefits (e.g., treatments or interventions with significant effects) or lack of harm from screening for cognitive impairment. Workshop participants discussed the importance of crafting a research agenda that builds a chain of evidence that either (1) demonstrates a clear benefit from screening and clinically meaningful actions to take based on screening outcomes or (2) clarifies that screening does not provide benefit in some specific contexts. Cross-disciplinary stakeholders from academia, industry, nonprofit organizations, and research funding agencies would benefit from opportunities to collaborate on development of such a research agenda.

Leveraging Resources

Research presenters provided examples of screening or assessment tools that are being tested and validated in samples from ongoing nationally representative epidemiological studies. Opportunities exist to use commercially available data, or to create data consortia to enable the application of machine learning and other big data techniques, rather than simply supporting a new data collection study each time a tool is developed. Participants encouraged the NIH and other research funders to identify ways to pool data, leverage existing longitudinal studies, support the creation of crosswalks between existing measures, and connect industry's commercial data and academic research projects. Relationships among academia, industry, and the public health sector might also lead to funding for studies in low-resource areas among diverse populations. It will be important to identify entities from the federal government and industry that have the necessary data (e.g., consumer credit card data, grocery scanner data, electronic health records [EHR] data) and are willing to partner with researchers.

Research Priorities

Data Resources

Invited speakers identified a plethora of administrative and passively collected data that can be useful for early detection of cognitive decline. Further work is needed to build data resources, assess their usefulness, and apply new techniques, such as machine learning. Specifically,
1. New studies are needed to establish the validity of administrative data for approximating disease burden, including by race and ethnicity, with changes in diagnostic practices.
2. Data that are representative of multiple regions and diverse communities are needed.

---

[1] See https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/cognitive-impairment-in-older-adults-screening.

3. Other data resources, such as imaging and biomarker data and speech and language samples, with possible relationships to biomarkers of preclinical ADRD should be further explored.
4. Research is needed to explore the usefulness and feasibility of financial data in predicting cognitive decline. Research is needed to
   a. Learn more about specific kinds of transactions that are predictive of cognitive decline.
   b. Identify the longitudinal relationships among financial capacity, financial behaviors, and cognitive change, including stratification by known risk factors for cognitive decline (e.g., APOEε4, biomarkers of AD, baseline cognitive ability), and how changes in social cognition explain fraud and exploitation.
   c. Measure the size and scope of financial fraud and exploitation of older individuals.
   d. Understand uptake of financial monitoring techniques (e.g., who signs up to be monitored; who drops out, resists, or otherwise does not adopt).
   e. Incorporate multiple sources of information to understand household dynamics.

Validation Work

Validation research is needed to assess computerized or digital screening tools, automated EHR tools, and risk indices of early cognitive decline against gold standard measures. Research is needed to
1. Validate computerized test batteries, digital assessment tools, and risk indices to establish sensitivity and specificity for biomarker evidence of preclinical ADRD and to compare performance against existing measures, where appropriate.
2. Consider fit-for-purpose, community, and clinical meaningfulness.
3. Consider whether a multisite platform and coalition for validating and comparing various tools for predicting and measuring change to organize these efforts, possibly modeled on the Alzheimer's Disease Neuroimaging Initiative (ADNI) approach, could be useful to the research community.

Methods Work

Research presenters highlighted opportunities to provide robust indicators by detecting cognitive decline in multiple data sources (e.g., administrative, EHR) and using multiple methods (e.g., imaging, speech and language samples, passive collection from sensors, patient-reported assessments, DCTclock™). Machine learning techniques may offer additional opportunities to detect decline. Research is needed to
1. Make machine learning more transparent.
2. Identify sources of measurement error.
3. Assess how much data is sufficient to build reasonably accurate models.
4. Test the applications and limitations of machine learning with different types of data.
5. Develop methods to allow for longitudinal assessment that accounts for changes in diagnostic behavior.
6. Investigate and address issues of measurement error, including systematic error from clinician bias, financial incentives, etc. Sampling can induce error if correlated with the outcome variable.
7. Build, test, and apply machine learning models to multiple types of data to identify biomarkers for preclinical ADRD, including financial, sensor, and imaging data, and speech and language samples.

Development Work

The research presenters offered several lessons for future work to develop, test, and validate new assessment tools. Research is needed to

1. Develop and test technologies that enable new approaches to understanding cognition, in part by assessing process as well as product to identify subtle signs of early impairment that are overlooked because of compensatory strategies.
2. Develop and test self-administered assessments with low false-positive rates.
3. Develop and test longitudinal assessments to understand indicators of possible decline.
4. Maintain validation of existing tools on new versions of operating systems and different platforms.
5. Where possible, embed or associate new assessment methods with existing ongoing studies to leverage resources.

Translation Work

Assessment and screening tools, administrative data analysis methods, and risk indices to detect early signs of cognitive decline are often developed in research contexts. More research is needed to adapt, test, and implement such tools in clinical settings and to link them to clinically meaningful care. Research is needed to

1. Adapt existing tools and methods with input from stakeholders to fit the needs of a clinical workflow. Ideally, instruments would have utility at the individual level as both predictors and indicators of change.
2. Test existing tools and methods in a variety of diverse regions and populations, and with different modalities such as telemedicine.
3. Identify and provide evidence for the link between screening and assessment tools and clinically meaningful care recommendations.

Conduct usability studies in a variety of geographic regions and populations, and with different modalities such as telemedicine, and leverage substantial existing work from the U.S. Department of Veterans Affairs.

Meeting Summary

Introduction

The National Institute on Aging (NIA) convened a workshop to examine the current state of the science for *Cost-Effective Early Detection of Cognitive Decline* using both passive and active approaches and to identify research opportunities in this area. The Directors of the Division of Behavioral and Social Research, John G. Haaga, and the Division of Neuroscience, Eliezer Masliah, noted that such workshops are essential for the NIA to identify gaps, areas of research interest, and promising opportunities that can inform future funding directions for the NIA. (See Appendix A for agenda and Appendix B for participants list).

The workshop was chaired by Maria Carrillo, Alzheimer's Association, and Reisa Sperling, Harvard University, and featured four thematic panels that were moderated by NIA staff: (1) role of administrative data for predicting cognitive decline; (2) predicting cognitive decline; (3) low-cost cognitive assessment and screening for clinical settings; and (4) bridging the knowledge gap between industry and academia. Invited speakers included researchers and experts representing academia, nonprofit organizations, and industry. Research presentations explored several topics related to early detection of cognitive decline, including, for example, financial decision making, machine learning, dementia risk indices, computerized assessments and screening tools, and passive detection sensors.

This meeting report provides a brief overview of each research presentation, thematically organized group discussion points, and considerations about the state of the science and future research.

Role of Administrative Data for Predicting Cognitive Decline

Relationship Between Cognitive Decline and Financial Decision Making
Jason Karlawish, University of Pennsylvania

Marson et al. define financial capacity as "the capacity to manage money and financial assets in ways that meet a person's needs and which are consistent with his or her values and self-interest."[2] Decline in financial capacity, mistakes in managing money, and susceptibility to fraud targeted at older people are often signs of Alzheimer's disease and related dementias (ADRD) long before an individual presents with clinically diagnosable dementia. Data from multiple studies, including the Rush Memory and Aging Study, demonstrate that decreases in cognition are associated with decreases in financial literacy.[3] Further, confidence in one's ability to manage finances and financial knowledge do not decrease with drops in measured cognition. The gap between a patient's confidence in his or her ability to manage finances and the assessment of a family member can be an early sign of cognitive decline.
Poor financial decision making among the elderly is a public health issue because it can harm the individual and the family, it has many causes (e.g., fraud, exploitation, poor judgement), and the

---

[2] Marson, D. C. et al. (2012). Assessing civil competencies in older adults with dementia: Consent capacity, financial capacity, and testamentary capacity. In Larrabee (Ed.). *Forensic Neuropsychology: A Scientific Approach,* 2nd ed. New York: Oxford University Press.

[3] Marson, D. C. et al. (2009). Clinical interview assessment of financial capacity in older adults with mild cognitive impairment and Alzheimer's disease. *Journal of the American Geriatrics Society, 57*, 806-814.
Gamble, K. et al. (November 2013). Aging, financial literacy, and fraud. Netspar Discussion Paper. Network for Studies on Pensions, Aging, and Retirement.

situation could be improved with intervention. Education, empowerment, surveillance, and prevention are actionable at the individual, community, and societal levels. These issues have become even more problematic as retirement savings have shifted from pension plans, which are professionally managed, to 401K and other personal retirement products, which are managed directly by the client.

New rules that attempt to address one aspect of this public health issue will take effect in February 2018. The Financial Industry Regulatory Authority (FINRA) will require firms that sell securities and annuities to place a 15-day hold on transactions that seem odd, suspicious, or against the self-interest of the client. Firms also will be required to ask clients when they open an account for the name and contact information of a trustee who should be alerted in situations where the client appears to be acting against his or her self-interest.

EverSafe provides an example of a fraud and identity theft protection machine learning system—FinTech—designed to surveille and prevent financial mismanagement due to cognitive decline.[4] Machine learning systems can learn an individual's personal baseline activity and, based on the learned norms, identify changes in behavior or erratic activity. The system analyzes behavior across accounts, institutions, credit reports, and time and categories of transactions. Alerts of abnormal behavior are sent via email, text, voice, or an app to the individual and/or a previously identified trustee.

Considerations for Future Work
There are many opportunities to examine ecologically valid data on financial decision making to detect potential cognitive decline. The current state of the science indicates that cognition and financial capacity are correlated. Cognition and financial transactions are interrelated, but more information is needed about the specific transactions and their specific characteristics, and how they change over time. Methods, such as machine learning, exist that can aggregate and analyze financial transactions, but more research is needed to identify the most effective approaches.

Several research opportunities exist. Research is needed to (1) identify the longitudinal relationships among financial capacity, financial behaviors, and cognitive change, and stratify them by known risk factors for cognitive decline (e.g., APoE4, biomarkers of AD, baseline cognitive ability); (2) predict never events such as foreign transfers to selected countries; and (3) investigate how changes in social cognition explain fraud and exploitation. More research is needed to measure the true size and scope of the problem of fraud and exploitation and to better understand what types of individuals sign up to be monitored and who does and does not drop out, resist, or ignore alerts about their own behaviors.

Detecting Cognitive Impairment in Financial Data
Lauren Nicholas, The Johns Hopkins University

As was noted by Karlawish, one of the earliest signs of cognitive decline and dementia is impaired financial capacity, such as difficulties managing money and paying bills or making erratic and uncharacteristically risky financial decisions. Cognitively impaired individuals frequently overestimate their financial decision-making abilities, placing them at risk for financial fraud, inappropriate asset allocation, and credit delinquency from unpaid bills. More research is needed to determine the prevalence and magnitude of these events.

---

[4] See https://www.eversafe.com.

Consumer credit data represents a large-scale, real-time source of information about consumer events potentially tied to dementia such as late bill payments, tax liens, home foreclosures, and changes in borrowing behavior. These data are already used by commercial entities and financial institutions that, except for the new FINRA rules described by Karlawish, do not have legal responsibilities to cognitively impaired individuals.

Nicholas described a project in which Federal Reserve Bank of New York Equifax Consumer Credit Panel data are matched to Medicare claims. Regression and machine learning models will be used to assess the predictive utility of credit score data and timing of credit events relative to dementia. The presence of financial decision-making limitations alone is insufficient to predict dementia. Credit report data offer a potentially promising way of tracking a comprehensive set of financial outcomes that may be predictive of cognitive impairment, and Nicholas and colleagues are investigating whether credit scores are sufficient or whether more granular data are needed. Large sample sizes with longitudinal data are needed for such analyses, and there is currently no single dataset with all necessary variables. Several features complicate these analyses such as automatic payments, individuals without credit, and household and family-level interactions.

Considerations for Future Work
There are many opportunities for research on using consumer credit administrative data to predict future dementia. Nicholas recommended future research to (1) identify federal and industry partners with necessary data such as from credit card use, purchasing histories, and grocery scanners, tax records, and electronic health records (EHRs) and health claims data; (2) incorporate multiple sources of information to understand household dynamics; and (3) determine guidelines for passive monitoring and detection (e.g., opt in or opt out screening).

Population Surveillance: Administrative Data and the Diagnosis of Alzheimer's Disease
Julie P. W. Bynum, Dartmouth College

Many types of administrative data may be useful for population surveillance for cognitive decline including registry data, EHR data, and claims data, as well as epidemiological data from population-based studies (e.g., Health and Retirement Study, Cache County Study of Memory and Aging). Each of these types of data has comparative strengths and weaknesses in terms of representativeness, accuracy of diagnosis, cost of collection, and timeliness. For example, epidemiological data might be nationally representative (although not necessarily regionally representative) and accurate in reporting diagnoses, but the cost of data collection can be high and data release slow. Claims data have good national representativeness and low cost of collection, but only a fair accuracy of diagnosis. It is important to consider the shortcomings of different types of data to identify what gaps researchers could fill to make these data more useful for population surveillance. Using data that enables analysis of regional and subgroup variation will be important in identifying where and how to target interventions. Bynum's analyses suggest regional variability in rates of clinically detected ADRD, stage of disease at diagnosis, and survival after diagnosis.

Population surveillance from the point of view of individuals already in the health care system is limited by who presents and when. Multiple studies have demonstrated high rates of undetected dementia in

the community (~62 percent).[5] Estimates of undetected dementia in the community have remained relatively stable. However, many of the studies used for meta-analysis are old and do not include current data. Bynum and colleagues have found that 60 percent of primary care setting diagnoses are made without consultation with specialists.[6]

Administrative data can allow for observation of diagnosed disease and symptoms in health systems for the entire Medicare and Medicaid population. These data are valid diagnostically when the goal is to observe those recognized and receiving health care for dementia (i.e., *diagnosed* dementia). The variation in diagnostic practices is declining. These data are limited with respect to stage of disease and local diagnostic biases, and in populations with less certain expected incidence and prevalence.

Considerations for Future Work
Population estimates would be strengthened by obtaining estimates by race for assessment of expected rates. Many of the studies to assess whether administrative data provide valid approximations of disease burden are dated and need to be conducted with new data. Methods should be developed to allow for longitudinal assessment that accounts for changes in diagnostic behavior. Objective measures of cognitive performance should be used in studies with representative populations large enough to allow for regional and subgroup analyses.

Discussion
Participants focused on two concerns about the use of financial and credit data to predict cognitive decline: (1) the sensitivity and specificity of financial decision-making data as a signal of cognitive decline is unclear; these data may not reveal situations in which a spouse or family member is the one managing accounts, or if there is a short-term situation leading to decisions not necessarily in the person's long-term interest (e.g., vacation, medical emergency, homelessness), and (2) ethical issues raised by financial institutions asking for consent from a potentially cognitively impaired person to alert a trustee. Ecologically valid data that reflect a personal baseline from which deviations might trigger alerts could address some of these concerns. An alert generated by a financial institution and sent to the client and his or her trustee does not necessarily trigger action or a diagnosis, but rather, could indicate a potential functional decline that should be further examined.

Financial data are changing with advances in technology, and firms will develop business models in their best interests. The challenge for researchers is to identify ways to leverage these developments and data for public health purposes. There is an opportunity to investigate the relationship between financial capacity, social environment, and cognitive function. A model that triangulates financial data with other types of information, such as pharmacy behavior and driving behavior, would be more comprehensive. There may be early signs of cognitive decline in one, but not all, possible areas (e.g., money, medications, transportation, telephone).

---

[5] Lang, L. et al. (2017). Prevalence and determinants of undetected dementia in the community: A systematic literature review and meta-analysis. *BMJ Open, 7*: e011146. doi:10.1136/bmjopen-2016- 011146.
[6] Koller, D., Hua, T., & Bynum, J. P.W. (2016). Treatment patterns with antidementia drugs in the United States: Medicare cohort study. *Journal of the American Geriatrics Society, 64*, 1540-1548.

Predicting Cognitive Decline

Uses and Misuses of Machine Learning in Health
Ziad Obermeyer, Harvard University

There has been rapid progress in turning certain physician tasks into machine learning problems, and there is enormous potential for early detection, improved safety, and lower costs in using these methods. However, using machine learning with health data raises statistical issues that can distort research and, if applied, cause patient harm. Mismeasurement is everywhere and has consequences, including (1) event-based measurement that misses people who do not present; (2) selective measurement; and (3) subjective measurement that can be impacted by cognitive bias and misdiagnosis, or financial incentives for over diagnosis.

One understudied problem is mismeasurement of the outcome/dependent (Y) variable and how it can distort machine learning predictions. There can be multiple predictor/independent (X) variables, and any X can be correlated with mismeasurement in the Y variable, which will impact predictions. The prediction is the sum of the predicted truth and the predicted difference between the truth and observed variable. It is difficult to undo errors built into the Y variable, and good predictors could be fitting to errors. X variables should be chosen carefully to differentially reduce error.

Considerations for Future Work
Obermeyer offered three overarching considerations for future work. First, researchers should think carefully about error in Y. There is systematic error from clinician bias and financial incentives, for example, and sampling can induce error if correlated with Y (e.g., test results from phone data). Second, investments are recommended in obtaining and evaluating new sources of X variables, such as linkages to phones, wearables, financial, and retail data. However, it is important to remember that the X variables collected are only as good as the Y variable they predict. Finally, research funding agencies should support work to improve closer measurements of Y variables. For example, research could build on cohort studies of well-defined populations (or conduct new cohort studies) to validate existing Y variables by measuring Y in a subsample to see how closely imperfect measures are relative to the truth.

Algorithms for Disease Detection: Lessons from Deep Learning and Retinal Imaging
Lily Peng, Google Inc. (via WebEx)

Rather than trying to make systems intelligent by telling them all they need to know up front, machine learning makes systems intelligent by teaching them to teach themselves. Algorithms can be developed that can learn simply from observing the surrounding world, similar to how humans and animals learn. Deep learning is a particular kind of machine learning that has been shown to be remarkably effective in the past few years, although many of the techniques used in deep learning are much older. Peng provided an overview of an example of a Google Inc. project that applied deep learning to diabetic retinopathy.[7] Image recognition is one of the most robust types of machine learning, and Google Inc. is actively looking for new applications in medicine. Peng explored the possibility that imaging could be used as a biomarker for aging and/or cognitive decline.

---

[7] Gushan, V., Peng, L., Coram, M., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA, 316*, 2402-2410.

There are several benefits to deep learning, including (1) highly accurate predictions without feature engineering; (2) quantitative, reproducible measurement; and (3) unbiased hypothesis generation. Other research suggests that, in addition to cardiovascular and eye disease, retinal imaging may contain signals about other diseases including Alzheimer's plaques visualized after ingesting curcumin, multiple sclerosis, and cerebrovascular disease.

Considerations for Future Work
Machine learning requires a great deal of data. Peng recommended that funding agencies encourage sharing and pooling of retrospective data, especially clinical trial data. In addition, supporting work that gathers data prospectively and longitudinally would be valuable.

Predicting Dementia Using Routinely Collected Clinical Data
Deborah Barnes, University of California, San Francisco

Many tools are available to identify patients with an increased risk of developing dementia. EHR data contain clues for detecting undiagnosed dementia, which is ideal for identifying patients who would benefit from additional screening. Barnes provided overviews of ongoing studies using clinical data to predict future dementia and to detect undiagnosed dementia. Barnes discussed two methods developed to predict future dementia: (1) late-life dementia risk indices and (2) a dementia screening indicator.

Prognostic risk indices are typically used to predict the likelihood of an event in a particular time frame using combined individual risk factors. Accuracy of risk indices is evaluated by measuring the area under receiver operating characteristic (ROC) curve, or area under the curve (AUC) or the c-statistic. Barnes presented work to develop the late-life dementia risk index and the brief dementia risk index using the Cardiovascular Health Cognition Study[8] as well as research that supported the development of the dementia screening indicator.[9]

Barnes and colleagues explored whether EHR data could be used to detect patients with undiagnosed dementia and are developing and validating the EHR Risk of Alzheimer's and Dementia Assessment Rule (eRADAR). They will determine whether eRADAR can improve dementia detection rates and clinical outcomes.

Considerations for Future Work
Barnes recommended more research to study the implementation of existing dementia risk prediction tools as well to develop, validate, and test automated EHR-based tools to detect undiagnosed dementia.

Inferring Clinical Outcomes Using Speech and Language Processing Algorithms
Izhak Shafran, Google Inc.

Shafran provided examples of clinical inference derived from speech and language processing in the areas of clinical depression, autism spectrum disorder, Parkinson's disease, and MCI. In the examples provided, researchers detected depression with an accuracy of 74.8 percent, measured severity of Parkinson's disease with a mean absolute error of 5.5 percent, and detected MCI with an AUC of 0.81.

---

[8] Barnes, D. E., et al. (2010). Dementia risk indices: A framework for identifying individuals with a high dementia risk. *Alzheimer's and Dementia, 6*, 138-141.
[9] Barnes, D. E., et al. (2014). Development and validation of a brief dementia screening indicator for primary care. *Alzheimer's and Dementia, 10*, 656-665e1.

The accuracy could be improved with more data, continual improvements in newer deep learning algorithms, and integration of modalities in addition to speech and language (e.g., motor movements, writing, cognitive tasks).[10]

Attention mechanism for modeling sequences is a new breakthrough tool in model interpretability. It can shed light on how the model infers, specifically inputs on which the model relies. It does not make any independence assumption and does not need manual training data.

Speech and language data can play a role in clinical inference, particularly with tools available for real-world acoustic speech. Shafran provided overviews of feasibility studies intended to infer social life from everyday conversations and ambient sounds.

Considerations for Future Work
Shafran presented several ideas for developing new models and tools, creating data resources, and providing opportunities for clinical researchers and machine learning scientists to work together.
1. Develop standard models and tools to measure effective communication, reciprocal social interactions, and quality of emotional interactions
2. Develop standard task definition and evaluation criteria for the above measures with the help of consortia resources (e.g., Speech and Language Group at National Institute of Technology, Linguistic Data Consortium, European Language Resource Association).
3. Create the data (corpora) to evaluate whether models learned in one study design can be effectively used in another (e.g., social engagement versus depression).
4. Create data to evaluate how much data is sufficient/necessary to build reasonably accurate models.
5. Organize a mini summer course and hands-on workshop to bring together clinical researchers, machine learning experts, and graduate students to solve selected problems

Detecting Presymptomatic Cognitive Decline with DCTclock™: Understanding Cognition
Randall Davis, Massachusetts Institute of Technology
Dana L. Penney, Lahey Hospital and Medical Center

Subtle impairment precedes error, but such impairment is often obscured by compensatory strategies. New technologies enable entirely new approaches to understanding cognition by seeing through compensatory strategies and revealing previously undetected behaviors. These technologies also enable new approaches to assessment that can occur before an individual "looks" impaired, and instead offer the possibility of identifying individuals who are working harder or longer to compensate for subtle impairments. Gold standard approaches to cognitive assessment, such as the Mini-Mental State Examination (MMSE), are time consuming, require a clinician to administer, are based in part on subjective assessment, and are relatively insensitive to mild impairment.

---

[10] Bayestehtashk, A. et al. (2013). Fully automated assessment of the severity of Parkinson's disease from speech. *Computer Speech and Language, 29*, 172-185.
Asgari, M., Shafran, I., & Bayestehtashk, A. (2014). Inferring social contexts from audio recordings using deep neural networks. Proceedings of *Machine Learning for Signal Processing*, IEEE International Workshop.
Asgari, M., Shafran, I., & Sheeber, L. B. (2014). Inferring clinical depression from speech and spoken utterances. Proceedings of *Machine Learning for Signal Processing*, IEEE International Workshop.
Lehr, M., Shafran, I., Prud'hommeaux, E., & Roark, B. (2013). Discriminative joint modeling of lexical variation and acoustic confusion for automated narrative retelling assessment. Proceedings of *NAACL-HLT*, Association for Computational Linguistics.

Davis presented an overview of a new digital platform using machine learning and signal processing informed by clinical expertise. The first test using this platform is called DCTclock—a radical revision of the traditional clock drawing test that uses a digitizing ballpoint pen for test administration. This approach yields detailed information about clock drawing behavior, as well as the final product. A research version of the DCTclock test has been in use at Lahey Hospital and Medical Center since 2005, and by the Framingham Heart Study since 2011, producing large numbers of clinical and normative samples. Results indicate that the DCTclock test detects mild impairment where the MMSE misses it. Its sensitivity is 73 percent versus 66 percent for the MMSE and yields the same specificity of 90 percent. The DCTclock test also yields scores that differ significantly across disease stages (healthy, amnestic MCI, and AD). Davis also discussed a newer test called the digital maze; research on this tool suggests that total maze completion time distinguishes AD from healthy controls.

Davis noted that re-conceptualized assessments enable measures of both process and product, and that measuring process provides insights not available from counting errors. Machine learning techniques can be significantly more accurate than existing scoring systems even when designed to be more transparent. In general, subtle behaviors can reveal cognitive state, which can inform alternative assessment strategies.

Considerations for Future Work
Davis suggested that cognition research be reconceived to acknowledge that subtle impairment precedes error and that impairment is often obscured by compensatory strategies. He also noted that more work needs to be done to make machine learning models more understandable. Finally, these research efforts would benefit from the involvement of practicing clinicians early in the process to support translation from laboratory development to implementation in clinical settings.

Discussion
Participants discussed the challenge of classifying dementia as a discrete event versus the idea that cognitive decline occurs along a continuum pathway over a longer period. Conceptually understanding dementia as a process is difficult to translate into a quantitative measure. Machine learning techniques might be able to detect subtle impairments that precede a diagnosis of dementia. The clinical community defines dementia as a discrete event, which is consistent with coding for health care services, whereas intervention development researchers are beginning to include the rate of change in an outcome variable in recognition of dementia as a gradual process.

One participant raised a question about how the collected speech and language data that Shafran presented are evaluated compared to speech among a normal aging population. For example, individuals experiencing normal aging could demonstrate non-dementia–related speech processing deficits. Shafran clarified that the machine learning algorithms are powerful sequence pattern recognizers. Normal aging is controlled for if the training data used to develop the machine learning model is normalized for age. In the case of the MCI research Shafran presented, the training data included an age-matched population.

The workshop chairs and other participants stressed the importance of using clear and consistent language to discuss constructs for which words might be used interchangeably, but have different intended meanings depending on context, purpose, and discipline, such as "dementia," "cognitive decline," "functional decline," and "cognitive impairment."

Ethics of Early Detection
Jason Karlawish, University of Pennsylvania

Remote monitoring and passive detection of health and behavior data raise ethical issues in addition to the known issues regarding how medical test results can impact a person's self-identity and experiences of stigma. Consenting to remote passive monitoring is submitting to being your own informant. What issues arise when consent is given at an earlier date, but collected from a person later experiencing cognitive decline? Who owns the information generated? What events would trigger alerts, and to whom, and what actions can be taken next?

Agency is the ability to act with intention, and it is entangled with identity, responsibility, privacy, authority, and ideas of normality. How will passive detection impact individual agency? Structural agency is the idea that actions are determined by how the world is organized, and less about individual agency. Ethical challenges in passive detection arise in the tension between individual and structural agency and balancing the rights to individual agency (e.g., ability for a person to drive) and a public health imperative (e.g., ability to deny a driver's license because a person is deemed a danger to others).

Low-Cost Cognitive Assessment and Screening for Clinical Settings

Online Assessment of Risk for Cognitive Decline
Michael Weiner, University of California, San Francisco

It is difficult to detect longitudinal decline in cognitively normal individuals, especially because of within-subject variance. The ability to assess risk for cognitive decline is important both for enrollment in clinical trials to assess group effects and for determination of clinical treatment needs at the individual level. The desired test for detecting risk of cognitive decline would have a high signal-to-noise ratio, where noise is within-subject variance.

Weiner reported on efforts using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Brain Health Registry to conduct naturalistic studies of AD progression. These data demonstrate that cognitively normal individuals decline very slowly, and it is difficult to detect a group effect in only 1 to 3 years. It is even more difficult to detect decline in one person. Composites can mitigate this problem by reducing within-subject variance. Weiner discussed the utility and feasibility of unsupervised online assessments, such as Cogstate, and the value added by collecting data from a study partner. The research presented suggests that it is feasible to use unsupervised online tests and to administer verbal recall assessments online. It is also important to use multiple types of data. Validation is critical, and very difficult and expensive. Other information that can be used to help predict cognitive decline include genetic data (e.g., ApoE4 plus polygenic risk), blood tests, eye exams, and passive monitoring of activity and sleep. Issues with passive monitoring data include sensitivity, specificity, and the need for validation.

Considerations for Future Work
Future research needs to address the development and validation of new tools using active and passive methods to measure cognition with attention to comparing different methods on different samples.

Weiner recommended the establishment of a multisite platform for validating and comparing various tools for predicting and measuring change (e.g., a coalition for online assessment). Multiple entities could be involved including the NIA, academic institutions, foundations, pharmacological industry, and

technology industry. The ADNI Private Partner Scientific Board (PPSB) might be interested in supporting such an effort.

Using Computerized Batteries to Measure Cognitive Function
Richard Gershon, Northwestern University

Funded through the trans-NIH Blueprint for Neuroscience Research,[11] the NIH Toolbox provides a rich battery of computerized assessments for the clinic that can measure cognitive function.[12] Several measures in the toolbox are useful for detecting cognitive decline. The NIH Toolbox includes measures from the NIH-funded Patient-Reported Outcomes Measurement Information System (PROMIS)[13] and is available in English and Spanish. It is often difficult to ensure measures developed using NIH funding are continually updated and used in the real world. The NIH Toolbox has succeeded in maintaining its usefulness by providing and maintaining the measures for the iPad via an Apple App Store.

The NIH Toolbox instruments are computer driven using Item Response Theory and Computerized Adaptive Testing (CAT). Hundreds of items underlie each instrument; the computer selects a unique subset of items based on demonstrated ability. The assessments are appropriate for individuals aged 3 to 85. The NIH Toolbox includes measures from four domains: cognition, emotion, motor, and sensation.

Considerations for Future Work
Gershon highlighted three forthcoming or proposed projects:
1. Advancing Reliable Measurement in Alzheimer's Disease and Cognitive Aging—This is a validation study using NIH Toolbox measures to identify additional measures that would be helpful for the oldest old population.
2. MyCog: Rapid Detection of Cognitive Impairment in Everyday Clinical Settings—This proposed study will use measures in clinical settings to flag individuals for further testing of cognitive impairment and highlight discrepancies between self-report and study partner report.
3. Mobile Monitoring of Cognitive Change—This project proposes development of Android and iOS apps, validation of tests, and norming of successfully validated measures to a nationally representative U.S. sample population to enable the mobile monitoring of cognitive abilities by age, state, context, or health condition–related changes in cognitive abilities.

Research accuracy and clinical accuracy are not necessarily synonymous. Less burden is required for research accuracy, and assessments can be modified for clinical accuracy. The two purposes need to be identified when designing and validating assessments.

The UCSF Brain Health Assessment: A Tablet-based Screen and Assessment for Clinical Settings
Katherine Possin, University of California, San Francisco

Dementia and MCI are underdiagnosed. Estimates indicate less than 50 percent of persons with dementia have been diagnosed, and the rates are even lower for MCI. Widely used brief screening tools have many limitations including poor sensitivity and specificity for mild presentations, difficulty integrating into primary care workflows, and lack of automated results-based care recommendations.

---

[11] See https://neuroscienceblueprint.nih.gov.
[12] See http://www.healthmeasures.net/explore-measurement-systems/nih-toolbox.
[13] See http://www.healthmeasures.net/explore-measurement-systems/promis.

Possin's work has yielded several lessons to guide development of primary care-based screening tools: (1) the assessment must take 10 minutes or less; (2) the reporting should be instant and automatically integrated with EHR; (3) results interpretation should be simple; and (4) result-specific educational materials should be immediately available to guide post-screen interpretation and care management. Primary care physicians often lack confidence in their abilities to provide cognitive care, which is predictive of low screening, referral, and diagnosis rates. The UCSF Brain Health Assessment (BHA), designed for use on tablets, was developed and tested with these lessons in mind.

Tests of the BHA indicate that it is more accurate than the Montreal Cognitive Assessment (MoCA) at detecting dementia and MCI; the higher accuracy is particularly dramatic for non-AD MCI. The BHA subtests demonstrate expected correlations with established neurocognitive tests and with regional gray matter volumes. The provider interpretive report for the BHA provides guidance on the likelihood of MCI or dementia, a cognitive impairments profile important for diagnostic criteria, and next steps for care. Next steps in this work will include integrating the interpretive report with Epic EHR.

Considerations for Future Work
The UCSF BHA should be validated in less educated and ethnically diverse patients and implemented in seven primary care clinics in the San Francisco area. Validation projects are under way in Cuba and Greece. In addition to the BHA, the research team plans to provide effective and user-friendly cognitive measures to improve a diverse range of research and clinical investigations. Possin and Gershon are also part of a National Institute on Neurological Disorders and Stroke (NINDS)-funded Consortium for Detecting Cognitive Impairment whose goal is to address the unmet need to detect cognitive impairment, including dementia, in large and diverse populations seen in primary care in the United States, including populations experiencing health disparities.

Using Sensors for Detecting Cognitive Decline
Jeffrey Kaye, Oregon Health and Science University

Cognitive detection methodology must match the use case. Methods to detect relevant cognitive change (i.e., change that predicts or is associated with a loss of functional ability) depends on the purpose: epidemiology (surveillance), health system (care delivery), clinical research studies, or clinical trials. The definitions of cognitive decline (relative to what), detection (sensitivity, specificity, meaningfulness), early (before what), cost-effective (relative to what), and sensors might vary depending on the purpose and use case.

Cognition has historically been measured with contrived tasks that aim to detect functional problems that cannot be measured directly. An alternative is to measure cognitive function directly in everyday life with sensors and passive detection (e.g., medication management, driving ability, computer use, sleep, body-mass index, and mobility). Kaye provided overviews of three use case examples including an ongoing national epidemiology study to identify cognitive decline among those living alone, in seemingly normal individuals for enrollment in a dementia prevention clinical trial, and a study to detect decline in a nonpharmacologic intervention focused on increasing social engagement in the context of a dementia prevention clinical trial.

Considerations for Future Work
Kaye provided general recommendations for designing and executing future research studies on detecting cognitive decline:

1. Clearly identify and specify the major use cases.
2. Use technologies and methods that are maximally unobtrusive and appropriate for the use case.
3. Make research transparent and replicable.
4. Consider whether more frequent monitoring of a smaller sample—rather than less frequent monitoring of a large sample—can be appropriate depending on the use case.
5. Conduct careful usability studies for all launches.
6. Where possible, embed or associate new methods with existing studies or studies in the development phase; this can be a cost-effective approach to help grow the research workforce.

Use of Digital Technology for Clinical Research
Dorene Rentz, Harvard Medical School

For this discussion, the use of digital technology for clinical research means the software development that enables use of computerized devices to capture clinically relevant behavior. Portable electronic devices such as smart phones and tablets can be accessible to older adults. The potential to use these types of devices to capture cognitive assessment performance in clinical trials is appealing. However, questions of feasibility, reliability, and validity are critical. Do computerized tests (1) substitute for standardized testing, (2) provide complementary or unique information, (3) and provide diagnostic sensitivity for discriminating groups or detecting biomarker evidence of preclinical AD?

Rentz presented work on the initial development of the iPad-based Computerized Cognitive Composite (C3) tool in the Anti-Amyloid Treatment in Asymptomatic Alzheimer's (A4) Study. The team has assessed feasibility and validity of at-home remote iPad assessments compared to in-clinic measurements, and validated the measures against the NIH Toolbox, the A4 primary outcome measure (ACDS Preclinical Alzheimer Cognitive Composite [PACC]), and neuropsychological tests. The study has also involved use of the REDCap™ for remote assessment of patient-reported outcomes, and the DCTclock™ digital pen.

The technologies included in these studies can all be cost-effective and have the potential to detect early cognitive decline. Results indicate that computerized tests can substitute for conventional testing once they are validated. Validated iPad-based assessments (e.g., Cogstate, NIHTB, and C3) also provide complementary information to standard cognitive measures. The potential for use of tablet-based assessments at home with repeated administrations on a participant's own device is a particularly appealing feature. Some computerized applications can provide unique information, such as the digital pen capturing diagnostically relevant underlying cognitive processes (e.g., time spent completing the task, hesitation at decision points). The digital pen can be used to distinguish diagnostic groups and shows promise in detecting biomarker evidence of preclinical AD. It is unclear whether tablet-based assessments can be used in this manner.

Considerations for Future Work
Lessons learned from Rentz's work yield several considerations for researchers when designing and testing digital technology for clinical research:
1. Computerized tests should be compatible with multiple platforms.
2. Participants should be able to independently use the test platform at home and on their own device.
3. Directions should be easy to read so participants can correctly perform tests.
4. Data should be easy to upload and synced with larger Health Insurance Portability and Accountability Act (HIPAA)-compliant databases.
5. Researchers should have real-time access to data to monitor for errors.

6. Computerized test batteries need to be validated and have good sensitivity and specificity for biomarker evidence of preclinical AD.

Technologies for Improved Patient Assessment
J. Jeremy Rice, IBM

Rice provided an overview of IBM-conducted research studies on (1) speech-based assessment of cognitive state and neurodegenerative diseases; (2) a project to collect data on movements in activities of daily living with wearable sensors in Parkinson's disease clinical trials; (3) sensor-based environmental systems for elder care (Rice specifically mentioned a pilot project with Avamere Health Services[14]); (4) plasma-based biomarker for AD using machine learning; and (5) computational modeling of potential pathophysiological mechanisms in AD. The speech-based assessment research, for example, demonstrated that syntactic analysis of 1-minute samples of natural speech could detect group differences and inference of scales in Parkinson's disease.

Considerations for Future Work
Rice offered several considerations for future research:
1. Combine heterogeneous data sources, and look for passive methods of data collection.
2. Create a national repository of speech samples to foster development of better analytics.
3. Define new endpoints and standards for continuously and passively collected data.
4. Conduct research that improves understanding of disease phenotype, which may be useful for defining progression biomarkers of cognitive decline and ultimately inform targeted therapies.

Low-Cost Cognitive Assessment Among Low SES and Diverse Populations
Deborah Gustafson, State University of New York, Downstate Medical Center

Gustafson provided examples of her own work to illustrate her recommendations for reducing the costs of cognitive assessments in low socio-economic status (SES) and diverse communities. She is involved in three epidemiological projects being conducted in diverse populations on three continents. Researchers can bring collaborations with the pharmaceutical industry and others to low-resource communities and in turn increase research participation diversity and better serve the public health and care needs of these communities. The cumulative load of factors affecting low-resource communities should be considered, including multimorbidity, access to and appropriate use of health care, medication adherence, lifestyle challenges (substance abuse, cigarette smoking, transportation, etc.), and traumatic life events.

Gustafson presented an overview of the Cognigram™ tool, which is a digital cognitive assessment system. The Cognigram is a 10- to 15-minute computerized cognitive assessment tool developed for clinical practice, which provides a simple assessment of changes in cognitive impairment over time. The tool has been reviewed by the U.S. Food and Drug Administration and is HIPAA-compliant. The tool design leverages the scientifically validated Cogstate Brief Battery, which has been widely researched and used in major public-private partnership studies in AD.

Considerations for Future Work
Gustafson contributed several suggestions for reducing the costs of cognitive assessments in low-SES and diverse communities. She recommended that researchers work to (1) improve definitions and descriptions of terminology such as low SES, diversity, and health disparities on regional, national, and

---

[14] See https://www-03.ibm.com/press/us/en/pressrelease/51672.wss.

global levels; (2) advance identification and characterization of populations at risk, and actively recognize and use the characteristics in study design and data interpretation; (3) engage innovative, language invariant e-cognitive methods to measure cognition and detect cognitive impairments; and (4) broaden measurement and detection environments through active and integrated partnership between industry, health care systems, public health, government, and academic entities—each of which might require unique approaches.

Discussion

The discussion centered on the concept of **validation**, which was addressed in multiple presentations. The term "validation" is being used in different ways, and it is important to be clear about the intended meaning. Validity refers generally to the extent to which a test measurement or other device measures what it is intended to measure, and there are many types of validity including construct validity, content validity, face validity, and predictive validity. In some of the presentations, the term validation was used more specifically to describe the research conducted to establish how well a new computerized tool correlated with an established gold standard method for measuring constructs related to cognitive decline. The latter type of validation work needs to occur for new instruments being developed, and researchers should be transparent and consistent when discussing and evaluating how a tool is fit for its purpose.

Participants discussed the concept of a **gold standard** in the age of rapidly developing technologies, some of which may enable measurement of constructs that is more accurate and efficient than the 10-year-old gold standard. It is unclear how the concept of a gold standard might shift with the advent of computerized-based assessments and machine learning algorithms that could be used for surveillance. Validation strategies may need to be adjusted or repeated with each hardware and software update for the platforms being used (e.g., the NIH Toolbox measures are re-validated with every iOS version update).

Sperling noted another distinction—validating a tool for **clinical meaningfulness**, which necessitates a sensitive measure, and validating a tool for research purposes. Kaye added that the concept of "fit for purpose" might be useful because it encompasses validation, but includes consideration of how well the instrument fits the intended purpose and community. The intervention design and validation methodology should be appropriate for the intended outcome and use of the tool, rather than expect a single standard to apply to all tools and tests. Sometimes technology developed for research or consumer purposes is adopted in the clinical setting without adaptation, which can be unhelpful and confusing. Some fields have a research definition and a clinical definition with different thresholds of certainty. Perhaps using specific research and clinical definitions should be considered for cognitive decline.

Weiner proposed the establishment of a consensus conference to discuss and develop a set of definitions, standards, and expectations for the field to guide consistent, transparent, and replicable progress. Shafran noted that in the field of speech and language technologies, the National Institute of Standards and Technology (NIST) created benchmarks and standard tests to guide the field.

Roundtable: Bridging the Knowledge Gap from Industry to Academia

Commentary from Nancy Smider, Epic; Izhak Shafran, Google Inc.; J. Jeremy Rice, IBM; Richard Gershon, Northwestern Medicine; and Dana Penney, Digital Cognitive Technologies

Smider discussed processes for translating tools developed for research purposes into the workflow of a clinical setting. As an example, she discussed how Epic, an EHR platform company, has had support for years for the PRO collection within the patient portal and clinical care workflow, integrated with the health record in clinical settings. More recently, Epic has collaborated with PROMIS to enable the use of CAT-based measures in the same, integrated manner. The stakeholders involved—patients, clinicians, and researchers—have different concerns that need to be considered. Clinicians need to see value in what a new tool or workflow offers; the clinician and patient need to see evidence that the tool adds clinical value. The information provided to clinicians needs to be clear and actionable; the clinician should receive the next steps for clinical care given the result of the tool. The recommendation to provide clinicians with actionable next steps is in concert with findings from other research presented during the workshop (e.g., Possin, Penney, Gershon). The Epic EHR platform also provides potential opportunities, using industry standard interoperability mechanisms, for use of external applications to capture data and to return visualizations of results or even the results themselves to the EHR for clinical review/care purposes.

Shafran described a successful effort in the speech and language field as a potential model for providing opportunities for cross-disciplinary work to solve problems. For the past 15 summers, The Johns Hopkins University has convened a cross-disciplinary group of senior researchers and graduate students to work collaboratively on selected problems in a paid 6-week workshop. During the 6 weeks, the teams work on their projects, provide team updates to the other participants, and attend guest speaker lectures. At the end of the 6 weeks, the top three projects receive funding from an entity such as the National Science Foundation (NSF) to continue. The experience creates a community that has yielded substantial advances beyond the length of the program. For example, 8 years ago, Shafran's team worked on computerized language translation. He was later hired by Google Inc. and developed Google Translation. The workshop provided a space and real-world context for conversations and collaborations to occurred that would not have occurred otherwise. Shafran also noted the utility of having data consortia to collect and maintain big data that can then be made available freely to researchers. Data consortia free up researchers to address higher level problems without having the burden of also collecting data.

The IBM Chief Executive Officer often says that data are the new natural resource.[15] Rice noted that IBM is looking for ways to be complementary to academia and that big data are valuable to both industry and research organizations. Industry and academia should work together to send a clear message to funding agencies that diverse big data can serve multiple purposes.

Penney shared lessons learned from commercializing the DCTclock™ test, a tool originally developed for a research purpose. Tools developed for research purposes are not often commercial ready, and researchers are unprepared for the plethora of considerations needed to make a business case for an instrument or device. She recommended that persons in academia considering such a process consult early with industry experts to help inform the effort from the beginning, and she explained that testing the tool in established longitudinal study samples, such as how the DCTclock test was tested in the Framingham Heart Study, can be invaluable.

---

[15] See https://www.forbes.com/sites/robertlenzner/2013/03/07/ibm-ceo-ginni-rometty-crowns-data-as-the-globes-new-natural-resource/.

Gershon reiterated comments already discussed about the importance of creating tools that are clinically useful. Clinicians need tools that take minimal time, produce clinically meaningful results, and quickly provide actionable next steps for patient care. In the clinical setting, assessments should have low false-positive rates that are as close as possible to self-administered. Gershon asserted that passive detection models hold phenomenal promise for the future, but a great deal more research is needed.

Haaga invited participants to provide feedback to NIA staff members on their experiences, if relevant, in applying for Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) awards to create commercially relevant tools and assessments.

Conclusions

As presented by invited speakers spanning academia, nonprofit organizations, and industry, the state of the science offers several conclusions. Several themes emerged from discussions of future research challenges and opportunities. Participants also identified several specific research priorities and ideas for consensus work needed in the field.

State of the Science
Dementia and MCI are underdiagnosed, and analyses using claims data indicate that clinical diagnoses occur late in the process of cognitive decline. Less than 50 percent of persons with dementia are diagnosed, and the rates are even lower for MCI. Applying machine learning to multiple types of administrative data, EHR data, or passively collected data; implementing computerized assessment and screening tools in clinical settings; and using composites and risk indices all hold promise as strategies for early detection of cognitive decline. Invited speakers described multiple studies that demonstrate the feasibility of these approaches for surveillance and detection of early cognitive decline.

For example:
1. Data demonstrate that decreases in cognition are associated with decreases in financial literacy; yet, confidence in one's ability to manage finances does not decrease with drops in measured cognition. Using machine learning to surveille financial data can identify signs of potential risk for cognitive decline based on financial behaviors that deviate from a personal norm.
2. Image recognition is one of the most robust types of data to which machine learning can be applied, and more work is needed to explore the possibility that imaging could be used as biomarkers for aging and/or cognitive decline.
3. Research has demonstrated the possibilities for clinical inference derived from speech and language processing using acoustic speech and ambient noise.
4. Multiple computerized screening and assessment tools have been or are being developed and validated against gold standard cognition assessments and in multiple diverse populations. Several tools demonstrate comparable sensitivity and specificity and have been feasibly implemented in the real world. These tools—such as the DCTclock™ pen, Cogstate (both the research tool and the Cognigram clinical practice tool), NIH Toolbox, UCSF Brain Health Assessment, and Computerized Cognitive Composite—can complement or add unique value, and/or replace older, more time-consuming gold standard assessments. For example, using the DCTclock™ digital pen for the clock drawing task yields results that shed light on process as well as product, which provides insights about subtle impairment that precedes dementia diagnosis. In addition, both clinical practice tool (Cognigram) and clinical research tool from Cogstate were

able to detect subtle changes in cognition, free from any cultural/education/language/practice biases.

Despite these promising early findings, much more work is needed. Participants agreed that big data, new analytic methods, new data collection methods, new sources and types of administrative data, and the advent of computer-based assessments and screening tools offer great promise for working to achieve the National Alzheimer's Plan Act (NAPA) goal of accelerating efforts to identify early and presymptomatic stages of ADRD. However, all agreed that these exciting new opportunities also require a great deal more research and careful ethical and guidelines considerations.

Administrative data suitable for surveillance and early detection of ADRD—with varying strengths and weaknesses in representativeness, accuracy, cost, and timeliness—include claims data, EHR data, epidemiological data, and registry data. However, other data sources may also prove useful for detecting early signs of cognitive and functional decline, such as financial data, consumer credit data, tax records, and grocery store data.

Passive detection data collection through a plethora of commercially developed sensors—both wearables and location-based—generates vast amounts of data and holds great promise for early detection of ADRD. More research is needed to determine what types of data in what contexts for what types of individuals are most useful, and what systems should be used to best analyze these data. Passive detection raises ethical considerations that need to be better explored and defined with all stakeholders.

Machine learning models hold great promise for expanding the possibilities of applying self-learning systems to various types of big data to identify biomarkers of presymptomatic ADRD. Machine learning systems need to be made more transparent and require a large amount of data. New pooled big data resources are needed to develop, test, and apply machine learning systems to the problem of detecting early cognitive decline. Using machine learning with health data raises understudied statistical issues that can distort research, and possibly even cause patient harm.

Consensus Work
Workshop participants identified several areas where further discussion, exploration, and consensus among stakeholders are warranted.

Clarity of Language
Participants noted that terms are used interchangeably or by different communities or contexts (e.g., research versus clinical community) to denote the same idea, yet the terms used have distinct meanings. For example, the terms "validation," "cognitive decline," "functional decline," "dementia," and "cognitive impairment" have been used in different ways by different communities and at different times. It is possible to have different thresholds—and therefore definitions—for research versus clinical purposes. Consensus work should be done to define common terminology for clear and consistent use.

Ethical Guidelines
Ethical issues may be raised by passive data collection from wearables and location-based sensors, and the use of commercial financial data for early detection of cognitive decline. These issues reflect how passive monitoring and detection impact on an adults' identity, privacy, authority, and normality. Stakeholders should explicitly examine these ethical issues and, specifically, develop guidelines for providing consent or assent, ownership of data collected, principles for deciding with whom the data are

shared, what actions might follow an alert or flag generated by the data, and how monitoring and detection impacts on a person's identity and normality.

Research versus Clinical Purpose

Research tools or instruments are often not ready for implementation into existing clinical workflows. It would be useful to explore the potential to delineate sensitivity and specificity thresholds for research versus clinical purposes. For example, a consistent metric such as area under the curve (AUC) is appropriate to summarize general instrument performance, but is not the best metric for making statements about individuals. The intervention design and validation methodology for new tools should be appropriate for the intended outcome and use; a single standard should not be expected to apply to all tools and tests. Stakeholders should discuss and develop standardized guidelines for how new tools, analytic methods, and data collection methods should be validated.

Establishing a Chain of Research Evidence

Participants discussed the U.S. Preventive Services Task Force (USPSTF) grade of "Insufficient" (I) for assessing the balance of benefits and harms of screening for cognitive impairment.[16] The USPSTF recognized the importance of detecting individuals with dementia and MCI, and acknowledged evidence that clinically useful screening tools exist. However, it determined there is inadequate evidence of clear benefits (e.g., treatments or interventions with significant effects) or lack of harm from screening for cognitive impairment. Workshop participants discussed the importance of crafting a research agenda that builds a chain of evidence that either (1) demonstrates a clear benefit from screening and clinically meaningful actions to take based on screening outcomes or (2) clarifies that screening does not provide benefit in some specific contexts. Cross-disciplinary stakeholders from academia, industry, nonprofit organizations, and research funding agencies would benefit from opportunities to collaborate on development of such a research agenda.

Leveraging Resources

Research presenters provided examples of screening or assessment tools that are being tested and validated in samples from ongoing nationally representative epidemiological studies. Opportunities exist to use commercially available data, or to create data consortia to enable the application of machine learning and other big data techniques, rather than simply supporting a new data collection study each time a tool is developed. Participants encouraged the NIH and other research funders to identify ways to pool data, leverage existing longitudinal studies, support the creation of crosswalks between existing measures, and connect industry's commercial data and academic research projects. Relationships among academia, industry, and the public health sector might also lead to funding for studies in low-resource areas among diverse populations. It will be important to identify entities from the federal government and industry that have the necessary data (e.g., consumer credit card data, grocery scanner data, electronic health records [EHR] data) and are willing to partner with researchers.

Research Priorities

Data Resources

Invited speakers identified a plethora of administrative and passively collected data that can be useful for early detection of cognitive decline. Further work is needed to build data resources, assess their usefulness, and apply new techniques, such as machine learning. Specifically,

---

[16] See https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/cognitive-impairment-in-older-adults-screening.

1. New studies are needed to establish the validity of administrative data for approximating disease burden, including by race and ethnicity, with changes in diagnostic practices.
2. Data that are representative of multiple regions and diverse communities are needed.
3. Other data resources, such as imaging and biomarker data and speech and language samples, with possible relationships to biomarkers of preclinical ADRD should be further explored.
4. Research is needed to explore the usefulness and feasibility of financial data in predicting cognitive decline. Research is needed to
   a. Learn more about specific kinds of transactions that are predictive of cognitive decline.
   b. Identify the longitudinal relationships among financial capacity, financial behaviors, and cognitive change, including stratification by known risk factors for cognitive decline (e.g., APOEε4, biomarkers of AD, baseline cognitive ability), and how changes in social cognition explain fraud and exploitation.
   c. Measure the size and scope of financial fraud and exploitation of older individuals.
   d. Understand uptake of financial monitoring techniques (e.g., who signs up to be monitored; who drops out, resists, or otherwise does not adopt).
   e. Incorporate multiple sources of information to understand household dynamics.

Validation Work
Validation research is needed to assess computerized or digital screening tools, automated EHR tools, and risk indices of early cognitive decline against gold standard measures. Research is needed to
1. Validate computerized test batteries, digital assessment tools, and risk indices to establish sensitivity and specificity for biomarker evidence of preclinical ADRD and to compare performance against existing measures, where appropriate.
2. Consider fit-for-purpose, community, and clinical meaningfulness.
3. Consider whether a multisite platform and coalition for validating and comparing various tools for predicting and measuring change to organize these efforts, possibly modeled on the Alzheimer's Disease Neuroimaging Initiative (ADNI) approach, could be useful to the research community.

Methods Work
Research presenters highlighted opportunities to provide robust indicators by detecting cognitive decline in multiple data sources (e.g., administrative, EHR) and using multiple methods (e.g., imaging, speech and language samples, passive collection from sensors, patient-reported assessments, DCTclock™). Machine learning techniques may offer additional opportunities to detect decline. Research is needed to
1. Make machine learning more transparent.
2. Identify sources of measurement error.
3. Assess how much data is sufficient to build reasonably accurate models.
4. Test the applications and limitations of machine learning with different types of data.
5. Develop methods to allow for longitudinal assessment that accounts for changes in diagnostic behavior.
6. Investigate and address issues of measurement error, including systematic error from clinician bias, financial incentives, etc. Sampling can induce error if correlated with the outcome variable.
7. Build, test, and apply machine learning models to multiple types of data to identify biomarkers for preclinical ADRD, including financial, sensor, and imaging data, and speech and language samples.

Development Work

The research presenters offered several lessons for future work to develop, test, and validate new assessment tools. Research is needed to

1. Develop and test technologies that enable new approaches to understanding cognition, in part by assessing process as well as product to identify subtle signs of early impairment that are overlooked because of compensatory strategies.
2. Develop and test self-administered assessments with low false-positive rates.
3. Develop and test longitudinal assessments to understand indicators of possible decline.
4. Maintain validation of existing tools on new versions of operating systems and different platforms.
5. Where possible, embed or associate new assessment methods with existing ongoing studies to leverage resources.

Translation Work

Assessment and screening tools, administrative data analysis methods, and risk indices to detect early signs of cognitive decline are often developed in research contexts. More research is needed to adapt, test, and implement such tools in clinical settings and to link them to clinically meaningful care. Research is needed to

1. Adapt existing tools and methods with input from stakeholders to fit the needs of a clinical workflow. Ideally, instruments would have utility at the individual level as both predictors and indicators of change.
2. Test existing tools and methods in a variety of diverse regions and populations, and with different modalities such as telemedicine.
3. Identify and provide evidence for the link between screening and assessment tools and clinically meaningful care recommendations.
4. Conduct usability studies in a variety of geographic regions and populations, and with different modalities such as telemedicine, and leverage substantial existing work from the U.S. Department of Veterans Affairs.

Appendix 1: Agenda
Revised October 23, 2017

Meeting Goals: Describe the current state of the science for the early and cost-effective detection of cognitive decline using both passive and active approaches, and identify research opportunities in this area.

9:00 a.m.	Welcome, Introductions, and Purpose	John Haaga and Eliezer Masliah
	Co-chairs:	Reisa Sperling
		Maria Carrillo

Role of Administrative Data for Predicting Cognitive Decline
Moderator: Partha Bhattacharyya

9:10	Relationship Between Cognitive Decline and Financial Decision Making	*Jason Karlawish*
9:30	Potential of Using Credit Reports and Other Financial Data for Detecting Cognitive Impairment	*Lauren Nicholas*
9:50	Using Medicare Administrative Data to Predict Cognitive Decline	*Julie Bynum*
10:10	Q&A	
10:20	**BREAK**	

Predicting Cognitive Decline
Moderator: Jonathan W. King

10:30	Uses and Misuses of Machine Learning in Health	*Ziad Obermeyer*
10:50	Using Algorithms for Disease Detection	*Lily Peng (via WebEx)*
11:10	Predicting Dementia Risk: Development and Validation of the Dementia Risk Score Using Routinely Collected Clinical Data	*Deborah Barnes*
11:30	Inferring Clinical Outcomes from Speech and Language Processing: Role of Algorithms	*Izhak Shafran*
11:50	Non-invasive Tests to Detect Presymptomatic Cognitive Impairment	*Randall Davis Dana Penney*
12:10 p.m.	Q&A	
12:20	**LUNCH**	
12:30	Ethics of Early Detection	*Jason Karlawish*

Low-Cost Cognitive Assessment and Screening for Clinical Settings
Moderator: Nina Silverberg

1:00	Assessment of Cognition in Early Dementia: The Role of Screening and Detection of Cognitive Impairment	*Michael Weiner*
1:20	Using Computerized Batteries to Measure Cognitive Function	*Richard Gershon*
1:40	The UCSF Brain Health Assessment: A Tablet-based Screen and Assessment for Clinical Settings	*Katherine Possin (via WebEx)*
2:00	Using Sensors for Detecting Cognitive Decline	*Jeffrey Kaye*
2:20	Use of Digital Technology for Clinical Research	*Dorene Rentz*
2:40	Technologies for Improved Patient Assessment	*J. Jeremy Rice*
3:00	Low-Cost Cognitive Assessment Among Low SES and Diverse Populations	*Deborah Gustafson*
3:20	Q&A	
3:30	**BREAK**	

Bridging the Knowledge Gap Between Industry and Academia
Moderator: Partha Bhattacharyya

3:40	Views and Recommendations from Representatives from Industry and Academia
	Nancy Smider (Epic)
	Lily Peng/Izhak Shafran (Google)
	J. Jeremy Rice (IBM)
	Richard Gershon (Northwestern Medicine)
	Dana Penney (Digital Cognitive Technologies)
4:30	**General Discussion to Identify Future Research Priorities and Recommendations**
	Moderators: Reisa Sperling and Maria Carrillo
5:00	**ADJOURN**

Appendix 2: Participant List
Revised October 26, 2017

Meeting Co-Chairs
Maria Carrillo, Alzheimer's Association
Reisa Sperling, Harvard University and Brigham and Women's Hospital

Invited Speakers
Deborah Barnes, University of California, San Francisco
Julie P. W. Bynum, Dartmouth College
Randall Davis, Massachusetts Institute of Technology
Richard Gershon, Northwestern Medicine
Deborah Gustafson, State University of New York, Downstate Medical Center
Jason Karlawish, University of Pennsylvania
Jeffrey Kaye, Oregon Health & Science University
Lauren Nicholas, The Johns Hopkins University
Ziad Obermeyer, Harvard University
Lily Peng, Google Inc. (*via WebEx*)
Dana Penney, Lahey Hospital and Medical Center and Digital Cognition Technologies
Katherine Possin, University of California, San Francisco (*via WebEx*)
Dorene Rentz, Harvard University and Brigham and Women's Hospital
J. Jermey Rice, IBM
Izhak Shafran, Google Inc.
Nancy Smider, Epic
Michael Weiner, University of California, San Francisco

National Institute on Aging
Dallas Anderson, Division of Neuroscience (DN)
Nicole Armstrong, Intramural Research Program (IRP)
Partha Bhattacharyya, Division of Behavioral and Social Research (BSR)
Prisca Fall, BSR
Melissa Gerald, BSR (*via WebEx*)
John G. Haaga, BSR
John Hsiao, DN (*via WebEx*)
Chandra Keller-Allen, BSR [C] and Rose Li and Associates, Inc.
Jonathan W. King, BSR
Melissa Kitner Triolo, IRP (*via WebEx*)
Laura Major, BSR
Eliezer Masliah, DN
Kristina McLinden, DN (*via WebEx*)
Evelyn Neil, BSR
Emerald Nguyen, BSR
Lisbeth Nielsen, BSR
Georgeanne Patmios, BSR

Dana Plude, BSR
Luci Roberts, DN
Laurie Ryan, DN (*via WebEx*)
Marcel Salive, Division of Geriatrics and Gerontology
David Schlessinger, IRP (*via WebEx*)
Charryse Shell, BSR
Nina Silverberg, DN
Molly Wagster, DN
Courtney Wallin, BSR

Other Participants

Nadia Biassou, NIH Clinical Center (*via WebEx*)
Regina Bures, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development
Roderick Corriveau, National Institute of Neurological Disorders and Stroke
Zane Martin, Office of the Director, NIH (*via WebEx*)