National Institute on Aging (NIA)
Division of Behavioral and Social Research (BSR)

Breaking Data Barriers and Scaling Behavior Change Interventions to Benefit Older Adult Health through Public-Private Partnerships

APRIL 29-30, 2021

September 28, 2021

This meeting summary was prepared by Rose Li and Associates, Inc., under contract to the National Institute on Aging (NIA) Division of Behavioral and Social Research (BSR). The views expressed in this document reflect both individual and collective opinions of the meeting participants and not necessarily those of NIA. Contributions to this summary by the following individuals are gratefully acknowledged: Kelly Beazley, Partha Bhattacharyya, Malaz Boustani, Andrea Gilmore-Bykovskyi, Amy Kind, Rose Li, Mitesh Patel, Marcel Salive, Caroline Sferrazza, Jennifer St. Sauver, Kayte Spector-Bagdady, Nancy Tuvesson, and Gui Woolston.

Table of Contents

Executive Summary

On April 29-30, 2021, the National Institute on Aging (NIA) Behavioral and Social Research (BSR) Division convened an expert panel to discuss strategies for Breaking Data Barriers and Scaling Behavior Change Interventions to Benefit Older Adult Health through Public-Private Partnerships. The 22 invited presenters and discussants provided insight into opportunities to expand data access and integration through public–private partnerships. In addition, the panelists discussed leveraging big data assets to support the implementation of evidence-based behavioral interventions and to address health disparities in older adults, including individuals living with dementia. The panelists emphasized that richer, more complete datasets are needed to understand health disparities, particularly for older adults, including the impact of social determinants of health (SDOH) on individual health outcomes and of community-level determinants on health care policy. The infrastructure to build new datasets and support data-driven research is costly and time-consuming to maintain, and adequate sustained resourcing is required to provide broad access to richer data at a lower cost. Public–private partnerships can address the resourcing needs for data access and integration, as well as more complex issues, including social license and governance for the use of big data and can ensure the capacity for data for future usability.

Expanding Data Access and Integration through Public–Private Partnerships

The COVID-19 pandemic highlighted the need for richer data from nursing homes and the value of near real-time integration of electronic medical record (EMR) data with Medicare claims data. Integrated datasets such as these would enable public heath surveillance, pharmaco-epidemiological research, and the recruitment of nursing home facilities for cluster randomized controlled trials (RCTs). Data sharing cooperatives (e.g., the NIA IMPACT Collaboratory) employ a top-down approach, facilitating large-scale integration of data acquired from upstream sources (e.g., collecting EMR data directly from EMR providers rather than from individual nursing home facilities). Other public–private partnerships have focused on a lifecourse approach, generating complete longitudinal records from all providers across the lifespan of individuals within a very narrow geographic area. Both approaches rely on building partnerships to support the infrastructure needs of data-driven research. In addition to generating rich datasets, these partnerships may help to address systematic issues related to data harmonization, privacy preservation, and heterogeneity of data quality.

Gaps and Opportunities

1. Better coordinate interagency efforts to enable sustainable research infrastructure and reduce duplication of effort by researchers
2. Develop data infrastructures that include a strategy for self-sustainability (e.g., creating a health information exchange) rather than relying solely on a research base for funding
3. Support funding for the generation of longitudinal datasets (e.g., Framingham study) that capture lifecourse health information for data-driven research
4. Provide top-down leadership for initiatives to engage other private stakeholders (e.g., U.S. Food and Drug Administration's Sentinel Initiative)

5. Develop mentorship opportunities for new investigators working with large-scale data-driven research and tools for big data analytics
6. Create minimum standards for data collection to enable better integration of distributed data
7. Establish guidelines for investigators seeking to apply Medicare data to in silico clinical trials (i.e., computer-simulated clinical trials)
8. Engage private partners embedded with underserved and under-insured populations (e.g., safety net hospitals) to address disparities in the quality of EHR data

Leveraging Big Data to Address Health Disparities

For older adults, recent data suggest that SDOH (e.g., social isolation, lack of access to information and health literacy, and socioeconomic status) can have a significant impact on an individual's risk for developing Alzheimer's disease and related dementias (ADRD). However, the lack of data linking SDOH to health outcomes precludes the use of big data approaches to address health disparities arising from SDOH. Integrating electronic health record (EHR) data with data on engagement with community services provides an opportunity to implement screening for SDOH and assess its impacts on health outcomes. Newer digital data can be leveraged to capture older adults' engagement with community resources, both at the individual level (e.g., through digital phenotyping) and at the area level (e.g., using cellular mobility data, internet searches, and Google maps), but issues of privacy preservation and structural disparities in access to digital devices must be addressed (i.e., the digital divide).

Gaps and Opportunities

1. Engage community services to help close disparity gaps and leverage public–private partnerships to support capacity building for community services
2. Design behavioral interventions for low-resource services (e.g., safety net hospitals, rural hospitals)
3. Integrate lifetime location data from the U.S. Census with other SDOH data to understand lifecourse exposures to SDOH (e.g., timing and dose of disadvantage across an individual's lifespan)
4. Use behavioral insights from EHR data for risk stratification of patients in order to prioritize limited resources in areas or populations of high risk
5. Develop learning tools to field test behavioral decisions
6. Understand which populations are missed by digital phenotyping
7. Incorporate fairness and privacy-preservation into algorithm design
8. Develop guidelines for predictive screening of SDOH, with particular consideration for socio-ethical issues
9. Adopt evidence-based metrics to assess the impact of community services on mitigating health disparities based in part on validated measures of SDOH
10. Promote data democratization initiatives to create accessible and actionable data repositories at the community level

Social License for Data Usage and Public–Private Partnerships

The future of computational medicine includes opportunities for data-driven research to impact individual health outcomes (e.g., behavioral "nudge" interventions) as well as population-level

outcomes (e.g., optimizing care management, allocating resources, decreasing hospital readmissions, and encouraging preventive care). A fundamental barrier to all big data approaches is siloed data. Moreover, a lack of diversity in research datasets coupled with a lack of access to industry data limits independent validation and impacts the future utility of these data. Open-source clinical datasets may help to spur computational medicine. Public–private partnerships provide an avenue to build integrated datasets with rich, high-dimensional medical data (e.g., x-ray, electrocardiogram, pathology) linked to outcomes as well as the necessary research infrastructure to encourage independent validation. Partnerships focused on health equity may help to address regulatory issues and increase heterogeneity of new datasets by promoting social license to enable big data initiatives.

Gaps and Opportunities

1. Employ data-driven analytics to understand the impact of social isolation on long-term health in older adults
2. Include unstructured notes in the datasets (e.g., mentions of falls or episodic memory issues)
3. Employ cultural competency and community outreach for quality initiatives to identify outcomes that are meaningful to consumers
4. Incorporate measures of SDOH into datasets
5. Support training or mentorship programs to familiarize academic users with high-dimensional integrated datasets
6. Develop standardized regulatory and legal templates for integrated datasets
7. Enable real-time data integration for risk stratification to enable hospital capacity management
8. Raise standards for data commercialization
9. Understand governance issues regulating health data
10. Support efforts to digitize and modernize health data

Strategies for Using Big Data to Inform Clinical Interventions

EHRs provide valuable longitudinal data to inform population-level interventions, especially those aimed at maintaining function or managing symptoms of chronic conditions over time. Predictive algorithms developed using EHR data can support both clinical decision support (CDS) tools and behavioral nudge interventions, which apply the principles of behavioral economics to encourage specific behaviors that improve health outcomes. However, the utility of EHR data is limited because these data rely on medical encounters and many individuals living with dementia are undiagnosed. Incomplete or inaccurate data prompt concerns about algorithm accuracy and the ethical implications of predictive analytics. For individuals living with undiagnosed or misdiagnosed dementia, community settings provide an opportunity to capture data not based on medical encounters, such as daily activities, SDOH, mobility, and medication adherence. Moreover, because behavioral nudge interventions leverage heuristics, these interventions may be well-suited for individuals living with dementia. However, the rapid feedback cycle for refining behavioral nudges necessitates new, more agile tactics to evaluate nudge interventions (e.g., quality improvement trials and silent biomarkers). In addition, effectively implementing strategies to generate community-based data without building an

entirely new infrastructure may require building partnerships with groups already established in the community as well as an understanding of endemic biases associated with big data analytics and rapid, user-centered, iterative refinement cycles to improve agile nudge algorithms.

Gaps and Opportunities

1. Employ human-centered design approaches (e.g., qualitative interviews) into the entire design process, including end-user perspectives
2. Develop award mechanisms to establish innovation hubs to enable data-driven research
3. Include requirements for open access data in funding opportunity announcements
4. Explore ways to boost individual and community engagement using consumer technology (e.g., phone "apps") for data collection
5. Understand common biases in fit-for-purpose datasets, especially racial/ethnic, rural/urban, and bandwidth biases for technology-based approaches
6. Promote standardized formats for community-based data
7. Leverage quality improvement trials rather than traditional RCTs to enable real-time study design and evaluation of agile algorithms
8. Shift the CDS tool paradigm from push alerts to on-demand use, similar to standard medical tests (e.g., blood tests, x-rays)

Meeting Summary

Introduction and Workshop Goals

Partha Bhattacharya and Marcel Salive

The National Institute on Aging (NIA) supports translational and clinical research focused on interventions for the prevention and treatment of multiple chronic conditions in older adults, including Alzheimer's disease and related dementias (ADRD). Findings from an array of population-based studies, discovery research, and clinical trials provide insight into inherited and environmental factors that contribute to aging processes and the development of ADRD. This evidence base, coupled with predictive models, has yielded approaches to prevent or mitigate morbidity and mortality among older adults. However, challenges related to health care policy, cultural bias, delivery system inefficiencies, lack of access to quality care, and ineffective communication can prevent broad uptake of evidence-based interventions.

Advances in data analytics, predictive modeling, and interoperable software applications offer new ways to tailor behavioral interventions for both patients and clinicians. In addition, enormous potential data resources (e.g., administrative claims data from health care systems and electronic health records [EHRs]) are increasingly available to support research efforts. To maximally leverage the unprecedented availability of big data resources, and their application to the design and testing of behavior change interventions, requires a deep understanding of barriers to data access and integration (i.e., gaps) and strategies to overcome these data barriers.

The goal of this workshop is to *identify gaps and opportunities to leverage big data assets via public–private partnerships* to design and test behavior change interventions that can improve the health and wellbeing of older adults, including those with ADRD, as well as health care delivery for this vulnerable population. These gaps and opportunities fall under four broad themes: (1) expansion of data access and integration to support interventional aging studies in ADRD; (2) use of data to address health disparities across ADRD populations; (3) social license for uses of data and public–private data sharing collaborations; and (4) strategies for the use of big data for clinical interventions with older adults.

Session 1: Expanding Data Access and Integration to Support Interventional Aging Studies in ADRD

IMPACT COVID Supplement: National Nursing Home Data Sharing Cooperative

Vincent Mor

The NIA IMbedded Pragmatic ADRD Clinical Trials (IMPACT) Collaboratory conducts pragmatic cluster-randomized controlled trials (C-RCTs) to evaluate interventions within health care systems for individuals living with dementia. The COVID-19 pandemic disproportionately affected long-term care residents, highlighting a need for richer data from nursing homes to understand the impact of COVID-19 in this setting. To evaluate the potential value of real-time

access to complete electronic medical record (EMR) data from nursing homes on a large scale, the IMPACT Collaboratory conducted a pilot study with Genesis HealthCare, one of the largest long-term care providers in the United States, to monitor COVID-19 in nursing home residents and study the response to vaccination in this vulnerable population. In addition, Genesis HealthCare data were used to monitor adverse events (AEs) related to vaccine distribution programs in nursing homes. Based on these pilot studies, the IMPACT Collaboratory formalized a data sharing cooperative, working with the American HealthCare Association (AHCA), Acumen LLC, and Exponent to build a large data repository consisting of EMR data from approximately 10,000 nursing homes with matching claims data from the Centers for Medicare & Medicaid Services (CMS). AHCA leverages its relationship with nursing homes to acquire large-scale EMR data directly from EMR vendors. EMR data are transferred to Exponent, to harmonize data structure, and Acumen, to match harmonized EMR data with CMS claims data, generating a harmonized, matched dataset available to NIA investigators through a data sharing agreement. With these data, IMPACT investigators can (1) track the cumulative incidence of COVID-19 in nursing home facilities; (2) study factors that affect nursing home residents' risk of morbidity and mortality from COVID-19; (3) monitor SARS-CoV-2 vaccination of nursing home residents and characterize predictors of vaccination; and (4) examine the rates of physical and cognitive functional decline in residents with ADRD from 2020 through 2022.

The *data sharing cooperative* strategy offers several opportunities for future research, including real-time public health surveillance, pharmaco-epidemiological research (e.g., NIA de-prescribing efforts), and clinical research evaluating the impacts of new treatments and policy changes on health outcomes. In addition, this type of public–private partnership enables selective recruitment of nursing home facilities for C-RCTs and embedded RCTs, as well as real-world phase III clinical trials.

Sharing Data for Clinical Research: The Rochester Epidemiology Project

Jennifer St. Sauver

The Rochester Epidemiology Project (REP) is a research infrastructure that (1) links EMRs from multiple health care providers for a subset of Midwest residents, creating complete longitudinal records for individuals across their lifespans (i.e., a lifecourse approach); (2) provides access to these medical records for multiple institutions; and (3) archives and maintains access to historical medical records. In addition, the REP routinely serves as a sampling frame to actively invite people to participate in cohort studies or clinical trials as well as a resource to passively follow individuals through their EMR data. Founded in 1966 by Olmstead Medical Center and Mayo Clinic, the REP now has other partners (e.g., private practitioners, dental practices, Olmsted County Public Health Services, and new health care systems in the region). Similarly, the geographic scope has expanded over time to encompass a region of southern Minnesota and West Central Wisconsin.

The REP provides a complete picture of health and health care for this very specific geographic population by linking EMRs from multiple health care providers to unique individuals. Moreover, the REP infrastructure leverages dates in the EMRs to create a timeline for each unique individual in the dataset. Because the REP is geographically focused, data linkage is

critical. Another key to making the REP successful is long-term stable partnerships. These partnerships require time and trust to build as well as sustained effort and funding to maintain effective data sharing. Key components of REP data sharing agreements include (1) sharing of identifiable information, to enable data linkage; (2) insisting against data comparisons between REP partners; and (3) planning the destruction of all data if the partnership dissolves. The REP team faces significant challenges, similar to those faced by other public–private partnerships, including complex legal and regulatory issues (e.g., the Minnesota Research Authorization); maintaining data access over time (in particular due to frequently changing data storage systems); and data harmonization issues resulting from the lack of data standards.

Discussant Panel

Moderator—Julie Bynum

Panelist Comments—Daniella Meeker, Richard Platt, Judy Zhong

Dr. Meeker provided context as an academic researcher who has worked extensively with data sharing partnerships, highlighting the motivation to rapidly form partnerships as a key difference between establishing data sharing collaborations before and during COVID-19. She commended the enormous investment of effort and funding required to build relationships and overcome regulatory and privacy concerns quickly in order to track and respond to the COVID-19 pandemic. However, a significant concern moving forward is that this rapid progress will not be sustainable without solutions for common barriers that have afflicted data sharing partnerships, such as a lack of incentives for health care systems and other organizations that aggregate data to participate in research projects. Potential strategies to overcome this barrier include (1) applying lessons learned and best practices from federal organizations (e.g., the U.S. Centers for Disease Control and Prevention [CDC] or CMS) about how to engage health systems in the future and (2) identifying sources of sustained funding to maintain database resources and research infrastructures, which is not possible within the typical National Institutes of Health (NIH) funding paradigm (i.e., sequential 5-year grants). Dr. Meeker closed by inquiring about the potential for federal organizations outside of NIH to participate in developing a sustainable, unified infrastructure to leverage for both public health research (e.g., pharmaco-epidemiology) and optimizing health care using a data-driven approach.

Dr. Platt offered observations on public–private partnerships from the perspective of working with health insurance companies as part of the U.S. Food and Drug Administration (FDA) Sentinel Initiative—a multi-site, distributed database developed to monitor the safety of marketed medications. The Sentinel System collects data from five of the nation's six largest health insurers to address regulatory issues of interest to FDA. Dr. Platt highlighted key lessons learned from this collaboration, including (1) the need for shared interests between partners to help offset opportunity costs; (2) governance considerations for Medicare Advantage organizations participating in clinical trials (e.g., CMS must approve participation for its members); and (3) the value of building a distributed dataset, versus centralizing data. He emphasized that even though distributed data require greater curation and harmonization than centralized data, distributed datasets generate fewer legal issues and facilitate interpretation of unclear data. Additional considerations for a distributed data infrastructure highlighted by Dr.

Platt included (1) managing data lag when conducting clinical trials that include data from CMS or health systems and (2) engaging high-level support to facilitate linkage to external data sources in a distributed system.

Dr. Zhong offered insights on data sharing and the need for fair prediction algorithms from the perspective of a researcher at New York University. During the early stages of the COVID-19 pandemic, when New York was the center of the pandemic, she leveraged EHR data for New York City residents from two major sources—NYC Inside EHR network, funded by the Patient-Centered Outcomes Research Institute (PCORI) Network, and the New York City Health and Hospital network, the nation's largest public hospital system—to identify potential risk factors associated with COVID-19 outcomes for older, community-dwelling patients based on EHR data from ambulatory visits for the 20 months prior to their COVID-19 diagnosis. During this study, Dr. Zhong found substantial heterogeneity in EHR data quality. Importantly, this heterogeneity in EHR quality was associated with several demographic factors, including race/ethnicity. For minority patients, disparities in EHR data include (1) a greater amount of mislabeled data (e.g., COVID-19 severity), (2) smaller sample sizes, and (3) less available predictors (e.g., minority patients have fewer ambulatory encounters and fewer blood tests). Thus, for risk prediction models, the performance of predictive algorithms is driven by the non-Hispanic white patients, and accuracy may be worse for minority patients than white patients if the algorithm is not designed to account for inherent biases in the EHR data.

Gaps and Opportunities

Presenters and panelists identified common barriers to data access and integration for public–private partnerships centered on health care data and discussed strategies to overcome these data barriers as well as opportunities to leverage these integrated datasets.

Data Bias—Health care data have inherent biases based on the populations captured by different data sources and in different environments. In addition to potentially missing a population of interest, data for underserved populations are often mislabeled or incomplete. The REP works to capture more complete data for the local population by actively seeking partnerships with smaller practitioners and others (e.g., Olmstead County Public Health Services) who provide more services to underserved and underinsured populations.

Identified Data—One advantage derived from health care data is the ability to leverage identifiers to link EHR data for individuals over time and across settings to create longitudinal datasets. Data must maintain some level of identification, with appropriate protections and data use agreements, to enable data linkages. It is possible to support data linkage using complex identifiers that would not be easy to use outside of the research infrastructure by creating "hashes"; however, it is more expensive and requires more work to execute than default modes of data identification. Further, CMS data linkage provides an opportunity to create a common denominator (i.e., the total number of distinct individuals that should be represented) across different EMR datasets for adults over age 65 years.

Sustainable Resourcing—Stable funding sources are needed to maintain the databases and infrastructure to support data-driven research. Efforts to create data infrastructures should

include a strategy for self-sustainability (e.g., creating a health information exchange), rather than relying solely on a research base for funding. Other opportunities to promote sustainability include (1) better interagency collaboration, which provides the opportunity for coordination rather than duplication of efforts related to improving patient care (e.g., pharmacoepidemiology, and public health surveillance) and (2) a minimum standard for data collection to facilitate data integration across diverse data sources. The infrastructure to support longitudinal data integration (e.g., the REP and the Framingham Heart Study) requires massive ongoing investments. Regarding incentivizing interagency resourcing, the integration of real-time data may encourage investment from other agencies (e.g., CDC, CMS) interested in using big data for surveillance efforts. In addition, increasing the capacity of data for future uses may increase reusability of the research infrastructure. For longitudinal data integration efforts, both sustainable funding and stable long-term partnerships are needed. To maintain these long-term relationships, junior investigators should be paired with senior investigators to learn the lifecycle of data systems, particularly when using these data for interventional studies.

Resistance to Comparison—The same care can be delivered differently across institutions, but within data sharing partnerships (e.g., those that include multiple health care systems or providers) comparing outcomes between the individual partner organizations is often restricted, which complicates the assessment of outcomes based on delivery of care. The IMPACT Collaboratory has maintained a relatively open governance structure based on providing feedback to individual EMR providers. These EMR providers have known data insufficiencies and thus see value in receiving feedback; however, policy questions may require some level of restrictions. In general, differentiating between using these comparisons to generate more accurate machine learning algorithms versus using the data to rank providers as the purpose of the comparing these data may facilitate more open data sharing agreements.

Session 2: Leveraging Data to Address Health Disparities across ADRD Populations

Social Determinants Data in EHRs: What's Next After Closed-Loop Referrals

Michael Cantor

Social determinants of health (SDOH) have an impact across multiple populations, with social and environmental factors impacting the risk of premature death to a similar extent as genetics. In older adults, the SDOH with the highest impact include social isolation, access to information and health literacy, and socioeconomic status. Following the Factors Affecting Communities and Enabling Targeted Services (FACETS) project, which generated an open dataset with SDOH measures linked to census data at the community level, the Gravity Project created national standards for the inclusion of SDOH data in EHRs. However, evidence for the impact of screening for SDOH (e.g., connecting individuals to needed resources) on health outcomes is limited. Key socio-ethical issues related to SDOH screening include integrating screening into workflows as well as determining whether patients want to be screened, who will perform screenings, and in what settings screenings occur. Furthermore, once SDOH data are available,

it is unclear who should act on these data and the evidence base for how to act on these data is lacking.

FAIRE principles (i.e., findable, actionable, interested, responsible, and effective) should be applied to SDOH data. The efficacy of screening for SDOH, in particular, has been difficult to demonstrate. Current strategies to evaluate post-SDOH screening outcomes include closed-loop referrals (i.e., feedback from the service agency that a referral was received), which do not provide measures related to health outcomes. Considerations beyond closed-loop referrals to evaluate the outcomes of SDOH screening include (1) developing metrics for community service agencies and vendors based on patient-centered health outcomes; (2) ensuring efficient and equitable distribution of additional resources to support high-quality community services and enhance their capacity to meet increased demand; and (3) leveraging high-quality research on outcomes of SDOH screening to support policies outside of health care (e.g., urban planning) that address the conditions underlying disparities.

Beyond Claims: Harnessing Digital Data to Address Disparities in Health of Older Adults

Thomas Tsai

Current data on health outcomes rely predominantly on encounters within the health care sector (e.g., EMRs and CMS claims data). The COVID-19 pandemic highlighted opportunities to leverage digital data sources in order to capture data about interactions with community services and understand systemic disparities in access to health care. Digital data sources (e.g., cellular mobility data, internet search terms, and maps of health care facilities) may be able to capture area-level effects that are dynamic and include sociobehavioral context that reflects physical and social interaction within the built environment and the community.

Digital phenotyping, used to assess the effectiveness of social distancing measures during the early phase of the COVID-19 pandemic, revealed large socioeconomic disparities in COVID-19 cases. Similar digital data resources (e.g., visualization of internet searches for dementia) provide the opportunity to study chronic diseases, including ADRD. Moreover, overlaying internet search data with existing measures of social vulnerability (e.g., the Social Vulnerability Index) may provide information on underlying drivers for the increase in search terms around dementia. However, relying on digital data alone may exacerbate the digital divide—underlying structural disparities in access to digital devices (e.g., rural communities with less bandwidth, low-income individuals, older adults, and individuals living with dementia). In addition to individual digital phenotyping, community-level digital data also provide important insights on disparities by providing a richer, dynamic view of digital interactions (e.g., the association between age-related declines in health and age-related declines in the use of technology). Integrating digital health insights with community-level data on SDOH and health outcomes provides an opportunity for a public–private partnership to create a polysocial risk score.

Discussant Panel

Moderator—Julie Bynum

Panelist Comments—Andrea Gilmore-Bykovskyi, Cara James, Amy Kind, Michelle N. Meyer

Dr. Gilmore-Bykovskyi emphasized the need to consider algorithmic fairness when integrating validated digital biomarkers and digital phenotypes with other large data sources (e.g., EHRs). Standards are also needed to assess the fairness of specific digital phenotypes, taking into account the impact of various SDOH on digital biomarkers across different populations or for different outcomes of interest (i.e., the digital divide). Algorithmic fairness should ensure data sufficiency for different digital phenotypes (e.g., create another biomarker to address the digital divide); evaluate the differential performance of algorithms that generate digital phenotypes against different subpopulations based on known disparities (e.g., race/ethnicity, socioeconomic status); and provide a data quality framework for investigators in the data research field. In addition to racial disparities in the misclassification of key domains within EHR data, access to data sources differs across race/ethnicity, which suggests that metrics for SDOH may need to reflect both the causes and consequences of health disparities.

Dr. James pointed out the potential for unintended bias against those who are not captured within the digital footprint as technology-driven solutions are increasingly employed to address issues related to SDOH. The COVID-19 pandemic, and its response, have highlighted technology gaps in certain communities (e.g., rural and low-income communities). In addition, it is unclear how English proficiency and cognitive capacity (e.g., individuals living with dementia) may affect an individual's digital footprint. Another key gap is metrics for community services, in particular how an individual's digital footprint may affect their ability to access services and to get their social needs met. Dr. James highlighted an opportunity for public–private partnerships to impact how capacity for community services (e.g., financial support and other resources) is built by developing data standards for SDOH that can enable valid metrics for community services.

Dr. Kind highlighted three considerations from a social exposomal viewpoint: (1) the impact of community-level SDOH; (2) the impact of lifecourse exposures (i.e., dose and timing) of individual-level SDOH; and (3) data democratization of big data assets. Considerable evidence suggests that community-level factors impact health independently of individual-level factors, and sometimes outweigh the impact of the individual-level factors, demonstrating the value of considering both for developing actionable metrics (i.e., metrics that can be used to direct policy, social activism, or interventions). Opportunities exist to leverage data assets (e.g., census data and residential histories) to determine how dosage and timing of disadvantage exposure across the life course impact late-life health outcomes, including ADRD, in terms of both interactions with the health care system and fundamental changes in biology. Embracing data democratization and encouraging data researchers to translate data for a broad, non-technical audience will enable data repositories to become actionable at the ground level for community members and individuals impacted by disparities, as well as local leaders and policymakers who determine resource allocation to address health and other disparities.

Dr. Meyer offered insights from both data science and behavioral science (e.g., decision making) perspectives. From the data science perspective, leveraging big data assets (e.g., EHR

data) to risk stratify a patient population when there is a standard of recommended care offers opportunities to concentrate limited resources in a sub-population that can benefit the most from the intervention as well as to direct a variety of behavioral nudge interventions to encourage patients to receive and/or providers to take the time to discuss and offer care. Regarding digital phenotyping, non-invasive data sources permit the collection of more data points over time in home seeing (e.g., associated with remote patient monitoring for chronic disease patients), but disparities in access to reliable internet may exacerbate existing health disparities. From the behavioral science perspective, she emphasized that although big data may help overcome barriers to clinical implementation of evidence-based interventions, the collection and use of big data raise new behavioral challenges. These challenges stem from a lack of awareness about how health care systems collect, use, and share data as well as a lack of understanding of what SDOH are and how they influence health outcomes. This highlights the need to maximize transparency about data usage in a way that communicates effectively why the data are important, how they are used, and how they benefit the individual providing the information, as well as complicated concepts such as privacy preserving, data sharing frameworks.

Gaps and Opportunities

Panelists and presenters discussed gaps and opportunities for leveraging data to address health disparities, emphasizing the need to include SDOH data, utilize implementation science, and focus on partners in low-resource areas.

SDOH Disparities—Regarding self-report measures of SDOH and consent to use these data, minorities are less likely to consent to providing data on SDOH. This racial/ethnic disparity in data access has broad implications for the quality of the underlying data structure, in addition to disparities in access to health care and the digital divide. To address structural disparities, fairness, privacy, and use of these data must be clearly and effectively communicated.

Data Accessibility and Actionability—Two of the cornerstones to ensuring actionability are having a tool or a resource that meets the needs of the user (e.g., the policymaker or the health care system) and access to data. In many cases, the locus of control for action is not within the research community. Embracing implementation science could enable more efficient communication of actionable data to a broader audience (e.g., the policy community, nonprofit community-based organizations).

Implementation of SDOH Screening—Although the broader vision of collecting real-time patient-reported SDOH all the time and for every individual is ideal, it may be difficult to implement. Instead, researchers can consider opportunities to intervene and collect the data that may benefit patients at a very specific moment in their lifecourse (e.g., using a room-based radio sensor to track social interactions following surgery, which has been linked to early detection of complications from surgery). Moreover, SDOH screening must be aligned with both workflows, include incentives for implementation, and be targeted to areas of need. Building partnerships with low-resource partners (e.g., small nonprofit organizations, safety hospitals,

and rural providers) and adapting interventions to accommodate low-resource areas provide an opportunity to address disparities in SDOH.

Session 3: Social License for Uses of Data and Public–Private Data Sharing Collaborations

Nightingale Open Science and the Potential of Public–Private Data Collaborations

Katy Haynes

A growing number of computer science researchers are interested in applying their talents to the field of medicine (i.e., clinical problems), but access to large-scale clinical datasets is lacking. Open source data could help move the field of computational medicine forward; however, a key barrier is that the most useful data from a machine learning perspective (e.g., medical imaging data) are often siloed by health care systems and other private companies. This inaccessibility of data has dramatically limited the ability of computer science researchers to generate new concepts, models, and tools to move health care forward. Nightingale Open Science's goal is to unlock data that are siloed by private and public organizations, curate these data in a meaningful way, and make them broadly accessible to research communities. Current efforts aim to link high-dimensional medical imaging data (e.g., x-ray, electrocardiogram, pathology) with patient outcomes from the EHR. Nightingale has promoted public–private partnerships, and data accessibility, by creating a standardized framework to facilitate legal agreements and Institutional Review Boards (IRBs), working hand in hand with health care systems to build open source data assets. In return, public–private partnerships encourage and expand data accessibility, enabling computer science researchers to find new patterns and signals for diseases (e.g., ADRD).

Most machine learning and computer vision work in the clinical space focuses on predicting diagnosis or clinical diagnostic support (CDS). Nightingale, by contrast, is interested in algorithms that can pick up on new patterns for a disease that the human eye may not see and what those novel signals reveal about the diagnostic process as well as how physicians view patients and patient outcomes. Outcomes of these public–private partnerships include increased research capacity and data infrastructure and a subset of de-identified, open source data that are made available to researchers. Nightingale also plans to launch challenges around these open source datasets to encourage innovative research. Three datasets will be available when Nightingale launches in December 2021: (1) chest x-rays from emergency room (ER) visits linked to pulmonary deterioration outcomes, which was developed to provide diagnostic support for physicians during the COVID-19 pandemic; (2) breast cancer biopsy slides linked to specimen results, registry outcomes, and other outcomes (e.g., mortality), which is in the process of scanning and digitizing millions of biopsy slides; and (3) electrocardiogram waveforms linked to long-term outcomes, which is also in progress. In the past 2 years, Nightingale has built a network of partnerships with more than 10 health systems, providing meaningful data to researchers by offering concrete incentives (i.e., seed grants) and opportunities (e.g., increasing their data capacity and infrastructure and collaborating on interesting research) to health system partners. In addition to data access barriers, the lack of

training opportunities to work with these data remains a persistent gap. Scaling data science and computer scientists' talent requires ramping up their clinical knowledge or fostering collaboration across disciplines between physicians and researchers.

Data Justice: Ensuring Diversity and Accessibility of Research Datasets

Kayte Spector-Bagdady

Biospecimens and health data are governed by method of procurement, but what contributors care about is use. Moreover, biospecimens and health data that are procured differently end up being used similarly, and regulatory mechanisms and market forces have failed to reconcile this tension. This lack of concordance between data procurement and data use raises a key question: How can public–private partnerships achieve the desired advances in computational medicine while respecting individual contributors of these data? In terms of the limitations of our current governance structure, (1) the current regulations only apply in limited circumstances (e.g., identified data and specimens) and (2) human subjects research regulations were intended to prevent risks of physical harm and are now being applied to secondary research risks (e.g., privacy concerns or the welfare interest of how patient data are used), which are much more about dignitary harm. Furthermore, African American, Black, and Latinx patients and research participants express stronger concern about data commercialization, exacerbating other racial and ethnic disparities already present in health care data.

In addition to the regulatory divide between privacy protections for patients and for research participants, there is limited legal recourse for patients if they believe their rights have been violated. These regulatory issues are further complicated in industry, where there is less governance and individuals are less likely to read informed consent forms. Private industry databases are growing rapidly, whereas public governance structures and belabored requirements to ensure fully informed consent for research participants cause publicly developed datasets to lag. Important considerations for public–private data partnerships include (1) ensuring racial and ethnic diversity in datasets, particularly those used for secondary research and machine learning algorithm development, and (2) ensuring sufficient access to data for independent validation or derivative discoveries. Opportunities to improve the regulatory landscape include developing a new regulatory structure, which is unlikely to occur, and raising the standards for commercialization of data and specimens for both patients and participants.

Big Data, Nudges, and Rapid Learning

Gui Woolston

CVS recently acquired Aetna and is now standing up new clinical management programs aimed at driving changes in member behaviors to improve health and lower medical costs. Individual-level programs include behavioral nudge interventions—personalized, contextualized engagements with a member designed to change a specific behavior to improve their health, based on a deep understanding of their personal barriers. Population-level programs, such as clinical analytics, are used to target nudge interventions to specific members or groups of

members (e.g., using risk stratification algorithms) and to evaluate outcomes of implementing nudge interventions (e.g., hospital readmission avoidance following specific nudges).

Clinical management programs emphasize a holistic, personalized, and coordinated approach to outreach across CVS Health and Aetna channels (e.g., e-mail, web portal, text messaging, mail, MinuteClinics). For example, approximately 37 million individuals are estimated to have chronic kidney disease (CKD); most are undiagnosed, and capacity is insufficient to test every member. By stratifying member risk across two dimensions (e.g., risk of CKD and clinical benefit of diagnosis), Aetna identifies a target population for testing, employs nudge interventions to encourage testing, follows-up on clinical outcomes, and incrementally engages members in Aetna's CKD clinical management program through iterative cycles. Aetna leverages machine learning algorithms trained using claims data, EMR data, and any other data that can be gleaned about the member (e.g., zip code and SDOH) to develop nudge interventions. Members are then tracked over time to determine what types of messages are more likely to be effective and for what specific member characteristics. Because these clinical management programs leverage big data, key considerations for designing "smart" programs include (1) the ability to detect, and remedy, heterogeneity (i.e., when a program is effective on average but the effect differs substantially across diverse segments of the population); (2) the importance of a rapid feedback cycle (i.e., the turnaround time between testing and learning), particularly for nudge interventions; and (3) the ultimate need for RCTs to demonstrate an effect.

Discussant Panel

Moderator—Mitesh Patel

Panelist Comments—Rick Cohen, William Crown, Mona Siddiqui

Dr. Cohen provided insight on data usage from the health care industry perspective, noting that many people are surprised about the extensive use of medical claims data by health insurance companies to improve health outcomes. Care management as a practice began with manual sorting of claims combined with member outreach through nurses. Today, care management leverages sophisticated datasets built from internal medical claims data to improve health outcomes at both individual and population levels. Internal claims data can identify the most medically complex, high-utilizing members and connect these members with a team that works directly with providers and caregivers to help them achieve and maintain better health. On the population health side, internal data are used to identify groups of members with chronic conditions (e.g., diabetes, high blood pressure) to ensure proper allocation of resources (e.g., for diagnostics and interventions) and target diverse nudge interventions (e.g., medication adherence, recommended wellness visits, cancer screenings, and vaccinations). More recently, Aetna is leveraging external data as well as internal data to improve care management programs, including nudge campaigns. Prior to the COVID-19 pandemic, social connectedness nudge programs were built around interventions to connect members with their communities (e.g., community and religious-based groups). After the pandemic began, these care management programs pivoted to educate seniors about using technology to make virtual connections.

Dr. Crown offered reflections about opportunities and challenges for improving researcher access to data from his experience with OptumLabs. In partnership with the Mayo Clinic, Optum brought together approximately 130 million lives of claims data and 85 million lives of EHR data linked to social demographics, Social Security death master files, Medicare fee-for-service data, and area health resource files. One of Optum's goals was to improve data access and incentivize innovative research as a community "give-back." Many academic researchers were discouraged about working with this more complicated, multi-dimensional data structure, but Optum provided training and access to this massive dataset, and many of these academic researchers subsequently submitted grant proposals to NIH. One interesting observation from this integrated dataset was how incomplete the EHR data were. Although 70 percent of the provider groups in the EHR data were part of integrated delivery networks, linked data revealed that EHR data often did not span different treatment settings. These incomplete EHR data raise the concern that if researchers are only working with slivers of the data (i.e., EHR data *or* claims data), they may not understand the limitations of these data (e.g., lack of clinical detail in claims data or incomplete EHR data). While Optum's give-back strategy successfully expanded researcher access to data, the long-term sustainability of this model is unclear. Standpoint, another segment of OptumLabs that sells a licensed de-identified dataset, provides scaled pricing that is lower for academic research projects, but is still expensive. This strategy is more sustainable in terms of recovering the cost of building the dataset, but it is expensive from a research perspective. Thus, although access to richer, highly dimensional datasets at lower costs may be desirable, providing access in a way that is both cost-effective and sustainable remains a challenge. One interesting possibility would be to create synthetic datasets that capture the richness and granularity of real data but can be made available to researchers at lower costs and without the identification risks associated with real data.

Dr. Siddiqui provided context around quality improvement initiatives from the perspective of a growing health care company. As Humana expands into health care services, an enormous focus has been on enabling more care in home- and community-based settings, intervening earlier in a patient's trajectory, and understanding what services are most helpful for a member at a particular point in their lifecourse. To achieve these goals, Humana has made significant investments to develop a *longitudinal human record*, bringing together data from across the Humana enterprise (e.g., pharmacy segment, behavioral and SDOH data, and administrative claims) using a rigorous process of data certification as well as a robust artificial intelligence (AI)-driven ethics program; leveraging these data to proactively design programs targeting high-risk member populations; and providing rapid iterative improvements on those programs. She highlighted two recent quality improvement efforts: (1) early in the COVID-19 pandemic, when many Medicare Advantage members (i.e., older adults) became concerned about not being able to access their usual care and medications, Humana leveraged its data and AI program investments to identify members at high risk and to target proactive outreach through care management resources, both in terms of clinical needs as well as basic needs (e.g., food and transportation) and (2) as part of a partnership focused on health equity, Humana is leveraging external data assets to evaluate the performance of its models across different sub-populations (e.g., different socioeconomic or racial and ethnic segments). Additional considerations for optimizing data-driven care management programs include (1) leveraging external data, (2)

accessing real-time data (to the extent possible), (3) connecting real-time sources of data, (4) tracking outcomes that are meaningful to consumers for making their health care decisions, and (5) continuing to foster transparency and explainability of data-driven models.

Gaps and Opportunities

Presenters and panelists discussed challenges and opportunities around building better datasets and the transparent use of these data for improving health outcomes.

Resourcing Robust Linked Datasets—Early data-driven models (e.g., to predict an individual's risk for developing ADRD) were designed using claims data and later refined by linking EMR data from fixed coded fields. However, almost 80 percent of the medical data relevant to ADRD are found in unstructured notes, and leveraging the potential of these unstructured data is a major gap from a research standpoint. From a data accessibility standpoint, significant sustained investments are needed to create meaningful datasets and data linkages across different data sources. Even if a health system is well resourced in terms of analytic capabilities, creating linked datasets remains difficult and almost always requires an additional investment. In addition to appropriate financial resources, much of the effort to build linked datasets involves digitizing and modernizing data, in particular medical imaging data. Interestingly, most imaging data are stored in third-party vendor systems. So, health systems must pay again to use their own data. Organizations such as Nightingale Open Source leverage philanthropic funding to deploy seed grants and help health systems to build these datasets. Other opportunities to reduce the cost burden of building these datasets include scaling the data creation process.

Building Representative Datasets—Racial and ethnic disparities in health care utilization may be reflected as bias in models that are built solely on administrative claims data from health systems. Understanding the reasons for lower utilization is a large gap that may require primary data collection within health systems (e.g., surveying patients in a scientifically valid way around health care access and utilization) and linkage of utilization data to administrative claims data to highlight populations with significantly lower density of data. In addition, health care utilization surveys reveal that racial disparities also exist for consent to data use by private partners (e.g., insurance companies), with minority patients much less likely to consent to these private engagements if given the chance. Public–private partnerships may provide opportunities to engage diverse populations of patients who would otherwise be missed by claims data alone, and health systems should consider aligning their outreach efforts from a racial/ethnic perspective. However, the overarching assumption that the lack of racial/ethnic diversity in datasets stems wholly from minorities being less likely to consent is incomplete, and recent studies have found that method of recruitment for RCT participants, specifically in secondary studies, builds more biases into the data structure than the lack of consent to data use. Another opportunity to encourage racially diverse research is by increasing data accessibility for researchers who are diverse, and thus are more likely to focus on questions of interest in minority communities; however, these same researchers are also less likely to be at institutions with the infrastructure (e.g., computational systems management) needed for data-driven research. In addition, health equity public–private partnerships present an opportunity to focus on sub-populations that are underrepresented or underperforming.

Scaling Nudge Interventions—In the future, nudge interventions will likely become more targeted. Given appropriate context (e.g., sensitivity and personalization), most members have been surprisingly willing to accept nudge interventions. However, as these interventions become more personalized, member data may be used in new ways. One major concern is that if members fully understood the extent to which insurers were accessing and using EMR and claims data to enables these nudges, they may be less likely to consent to the use of their data for fear of reprisal (e.g., dropped coverage or increased rates due to predictive health factors). Health systems can employ metrics (e.g., opt-out rates) and other structured feedback (e.g., surveys) to track member responses to nudge engagements. As nudges become more numerous and personalized, opportunities to further tailor these agile interventions include pinpointing the instance in which a member is most receptive to messaging, including outcomes that are meaningful and valuable to patients, and determining the best person to deliver messages. Surveys rank pharmacists and nurses as among the most trusted people to deliver health information, whereas health insurance executives are ranked at the bottom.

Session 4: Strategies for Use of Big Data for Clinical Interventions with Older Adults

EHRs for Population-Level Intervention: Promise and Pitfalls

Andrea Cheville

Patients traverse a diversity of care providers and settings in their post-operative trajectories. To feasibly engage patients throughout this process, multiple touchpoints are required. EHR data provide a potentially valuable data source for population-level interventions because they are centralized, provide the ability to track every patient within a health care system across providers, and can be leveraged to impact virtually all care processes. EHR-based interventions may be provider- or patient-directed. Examples of provider-facing interventions include CDS tools that provide hard stops (e.g., alerts if toxic medications will be administered), prompts (e.g., alerts that prioritize a patient for non-pharmacologic options to post-operative pain management), and facilitators (e.g., task lists and embedded awareness information). For patients, EHR data can be used to optimally target the delivery of educational messages or other interventions (e.g., behavioral nudges). Data elements within the EHR also provide opportunities to drive distal CDS. For example, one smoking cessation program at the Mayo Clinic highlights the utility of combining different types of EHR-based interventions: (1) patients are flagged to be asked whether they are a smoker and whether they want to quit; (2) if they answer yes then the exam room monitor auto-populates a video describing the smoking cessation programs while the patient waits for the provider; and (3) the provider receives a hard stop alert to offer the patient a referral to the smoking cessation program. Results from this EHR-based intervention strategy show a 10-fold increase in referrals to the smoking cessation program with an impressive increase in quit rates.

Other opportunities for population-level use of EHR data include insight into care delivery (i.e., what does and does not occur) as well as mediator analysis to identify which components of a multi-pronged care process are impactful. For example, customizable interfaces and dashboard

analytics (e.g., duration of provider use) can effectively highlight whether, and to what degree, an intervention limits a workflow or renders it more efficient. However, there are significant barriers to efficiently leveraging EHR data, including provider-level workflow variations, the need for robust implementation and engagement strategies, and onerous oversight and governance structures.

Machine Learning and Behavioral Nudges in Advanced Illness Care Delivery: A Case Study in Serious Illness Communication

Ravi Parikh

Early conversations about patient goals and end-of-life wishes are appropriate at any stage and for any type of serious illness. Serious illness conversations are a priority in cancer care because end-of-life spending is responsible for over one-quarter of cancer care spending and these conversations lead to increased utilization of hospice in lieu of aggressive end-of-life care, which may reduce end-of-life spending by up to 40 percent. The University of Pennsylvania adopted a checklist strategy to standardize serious illness conversations, but after3 years, serious illness conversations were only documented for approximately 10 percent of patients. Two main reasons for this failed implementation were proposed: (1) poor patient identification using only patient prognosis to prioritize which patients needed conversations and (2) difficulties changing patient behaviors using advanced care planning education initiatives.

To address the patient identification issue, machine learning and behavioral nudges were leveraged to develop a prognostic tool for serious illness communication. The machine learning algorithm generated individual real-time estimates of 6-month mortality using more than 500 routinely collected variables embedded in the EHR. The predictions were generated immediately prior to a clinical encounter, prompting an alert to the provider that a serious illness conversation was warranted. The prognostic tool was developed using a six-step process, leveraging the principles of behavioral economics from design to impact, which may be generalized to developing other predictive analytics tools: (1) conduct stakeholder interviews with end users (e.g., clinicians who could be using the prognostic tool) to learn their perceptions on using predictive algorithms in their practice; (2) design and train the algorithm (i.e., gradient-boosting classifier for predicting 6-month mortality based EHR data) using retrospective data; (3) conduct stakeholder surveys to determine perceptions of fit and usability (i.e., did the tool meet their initial expectations); (4) prospectively validate the tool (e.g., integrate the tool into the EHR and generate silent predictions of mortality risk); (5) conduct a feasibility study for implementation of the tool; and (6) conduct an RCT for the intervention. At the time of clinical encounter, the tool could discriminate between high- versus low-risk individuals with an almost 16-fold difference between risk of mortality in the high- versus low-risk groups. Moreover, the machine learning nudge quadrupled the rates of serious illness conversations for patients with cancer. Important considerations for implementation include recognizing capacity constraints and provider fatigue from alerts.

Discussant Panel

Moderator—Mitesh Patel

Panelist Comments—Malaz Boustani, Niteesh Choudhry, Mark Friedberg

Dr. Boustani offered reflections on agile science, which integrates insights from behavioral economics, complexity science, and network science to create more rapid turnaround processes. The coupling of data science and machine learning to changing human behavior as well as overall implementation processes provides the opportunity to address the myriad of negative results seen with CDS tools in the past. CDS does not need to be limited to pushing alerts; instead, "on-demand" CDS tools can be used in a similar manner to clinicians ordering diagnostic tests (e.g., blood tests). In addition, the CDS paradigm shift can be coupled with modified machine learning tactics. For example, NIA is supporting a new pragmatic trial using machine learning to create a passive digital marker for the early detection ADRD and developing agile nudge processes to improve detection of this marker by comparing this marker with patient-reported outcomes. Because the purpose of nudge interventions is to change human behaviors, a key step in the process is human-centered design. Simply having a demand signal from executives may be insufficient to change patient behaviors. In addition to "C-suite confirmation," buy-in for proposed tools and interventions is needed from providers, patients, and caregivers. Dr. Boustani also emphasized the need to consider new, more agile tactics for the evaluation of these agile interventions, as opposed to pragmatic trials.

Dr. Choudhry provided insights on changing patient and physician behaviors from the implementation science perspective, emphasizing the need to consider interventions for both health care-oriented and community-based settings. The majority of machine learning research focuses on leveraging traditional big datasets (e.g., EHR and claims data) to develop algorithms that can be applied to observational and interventional efforts in the health care setting. However, for more interactive interventions that measure and act upon outcomes in real time, as well as interventions targeting older adults and individuals living with dementia in community-based settings, the types of data and data integration needed may differ. Data related to daily life activities (e.g., walking, medication taking, cognitive function, mobility symptoms) and SDOH may be more useful predicters of age-related cognitive decline than data based solely on an individual's interactions with the health care system (e.g., claims and EHR data or retail transaction data from pharmacies). Moreover, for real-time interventions, claims data are not timely and EHR data require sophisticated processing of unstructured fields to become usable in real-time. Technical considerations for community-based interventions include (1) designing ways to gather data through connected ambient monitoring devices in home settings that do not require patients to charge devices and (2) defining ideal systems to deliver interventions or SDOH solutions (e.g., texting platforms). While third-party commercial systems are working on these solutions and health care institutions appear committed to the integration of new data sources, issues around format, legality, data security, ethics, and IRB approvals must be overcome in order to move beyond only claims and EHR data sources.

Dr. Friedberg commented on nudge interventions from the perspective of a clinical researcher transitioned to private industry. Regarding resourcing, the cost of data can be significant. In

thinking about what research to fund, an industry perspective may help inform researchers about the actual value of any tools from an industry perspective. Regarding the limits of administrative data, Dr. Friedberg noted that vendors are beginning to collect live data from patients, including remote symptom monitoring programs that are available as phone applications. Consumer technology is good at active user engagement, and leveraging user engagement and retention techniques from the application development space provides an opportunity to enhance data collection for remote monitoring programs. Moreover, for individuals living with dementia, heuristic responses persist for many patients who have lost executive and deliberative cognitive functions, providing the opportunity to design nudge interventions specifically for patients with ADRD that may improve both their safety and their quality of life.

Gaps and Opportunities

Presenters and panelists outlined opportunities and challenges for leveraging machine learning based solutions (e.g., CDS tools and nudge interventions).

Biases in Data Acquisition—Biases from incomplete or inaccurate data will be reflected as bias in machine learning algorithms. Data quality should be considered during dataset construction because hidden structures in the data may not be obvious. When considering interventions using big data in the context of health systems, recognition that fit-for-purpose datasets have endemic biases (e.g., racial/ethnic bias, rural versus urban, and the technology gap) is crucial. Moreover, building fit-for-purpose datasets is time consuming and expensive. Opportunities to overcome data collection biases for a specific population include (1) building an entirely new data infrastructure in a target community of interest or (2) engaging in public–private partnerships to improve outreach in these communities by leveraging the relationship of an existing organization within the community, in particular for underserved populations. Data collection from commercial technology (e.g., Fitbits) exacerbate the potential for unintentional bias because issues of algorithmic bias observed in other contexts may be complemented by new biases present for those data streams.

Rapid Cycle Innovation—From an implementation perspective, a priori assumptions are a poor substitute for innovation. Assumptions that appear sound may in fact not be. Rapid user-centered iterative refinement cycles offer a level of agility that has not been accessible to trialists in the past. Building this process into design approaches enables leveraging of unanticipated signals for learning. However, potential regulatory issues (e.g., IRB approvals) may be unable to keep pace with these rapid cycles. Other considerations for this study design and evaluation paradigm shift include (1) addressing measuring outcomes for no harm; (2) engaging website metrics; (3) leveraging adaptive trial designs with post-hoc validation; and (4) scaling replication to create multi-site, multicenter, and multi system relationships.

Accelerating Research—From the academic research perspective, true innovation may require radical changes to the current system (e.g., funding and regulatory structures). There are also opportunities to build de-centralized clinical trial infrastructure to accelerate paradigm-shifting research include (1) focusing on the implementation of evidence-based guidelines, which

enables regulation as quality improvement; (2) using an agent-based modeling approach to gain insights into implementation; (3) considering new regulatory tiers to compare interventions that will be independently used as standard practice; and (4) creating new contract mechanisms and innovation hubs with internal governance structures. Potential challenges associated with these paradigm shifts include data privacy, data security, informatics infrastructure, Medicare payment infrastructure, and research management infrastructure, all of which exist for legitimate reasons but hamper innovation.

Appendix A: Meeting Agenda

Workshop Organizers: Partha Bhattacharyya, Marcel Salive, and Nina Silverberg, NIA

Workshop Co-Chairs: Julie Bynum, University of Michigan and Mitesh Patel, University of Pennsylvania

Findings from an array of population-based studies, discovery research, and clinical trials provide insight into the inherited and environmental factors that contribute to aging processes and the development of Alzheimer's Disease and related dementias. This evidence base, coupled with predictive models, has yielded strategies to prevent or mitigate a significant portion of morbidity and mortality among older adults. Efforts to address challenges related to health care policy, behavioral economics, cultural bias, delivery system inefficiencies, lack of access to quality care, and ineffective communication would provide an opportunity to dramatically improve uptake of these evidence-based interventions.

These barriers can be addressed by leveraging data from electronic health records, administrative claims, and other real-world clinical settings. Advances in data analytics, interoperable software applications, and decision science tools offer new ways to tailor behavioral interventions for individual patients and/or clinicians. This workshop will identify gaps and opportunities for integrating big data assets via public private partnerships, overcoming barriers to clinical adoption of effective interventions, and addressing the health and wellbeing of older adults, including people living with dementia.

Day 1: April 29, 2021

Welcome Remarks

1:00 pm Partha Bhattacharyya and Marcel Salive, NIA

1:05 pm Richard Hodes, NIA

1:10 pm Julie Bynum, University of Michigan and Mitesh Patel, University of Pennsylvania

Theme 1: Expanding Data Access and Integration to Support Interventional Aging Studies in ADRD

1:15 pm IMPACT COVID Supplement: National Nursing Home Data Sharing Cooperative– Vincent Mor, Brown University

1:25 pm Sharing Data for Clinical Research: The Rochester Epidemiology Project – Jennifer St. Sauver, Mayo Clinic

1:35 pm Discussant Panel – Daniella Meeker, University of Southern California; Richard Platt, Harvard University; and Judy Zhong, New York University

2:05 pm Broad Group Comment and Questions

Theme 1 Discussion Topics:

- What are the major barriers limiting the linkage of databases (clinical data, claims data, public health data, patient-initiated data) to capture the expanse of the patient experience?
- How have research programs modified their cohort development processes in response to the COVID-19 public health emergencies?
- Where are the major gaps in access to clinical and non-clinical data assets that limit design of interventional studies?
- Are there exemplary data sharing agreements and partnerships that are enabling innovative cohort designs for longitudinal population studies?
- What are the opportunities for integrating data from social services, social determinants of health, and behavioral health interventions with clinical and population health data?
- Are new tools and technologies needed to support the design and analysis of interventional aging studies?
- How do we ensure inclusion of people with limited access to both internet and healthcare?

2:15 pm Break

Theme 2: Leveraging Data to Address Health Disparities Across ADRD Populations

2:20 pm Social Determinants Data in EHRs: What's Next After Closed-Loop Referrals? – Michael Cantor, Regeneron

2:30 pm Beyond Claims: Harnessing Digital Data to Address Disparities in Health of Older Adults – Thomas Tsai, Harvard University

2:40 pm Discussant Panel – Andrea Gilmore-Bykovskyi, University of Wisconsin; Cara James, Grantmakers in Health; Amy Kind, University of Wisconsin; and Michelle N. Meyer, Geisinger Health System

3:10 pm Broad Group Comment and Questions

Theme 2 Discussion Topics:

- How might novel data sources enable integration of life course approaches into health disparities interventions for older adults?
- What are the sources of sampling bias and lack of access to data from underserved populations and how can the gaps be overcome?
- Are there innovative models for engaging underrepresented populations of older adults in interventional aging studies through the use of technology and integrated data networks?
- What are common challenges to data sharing and engagement of minority populations in addressing high priority areas of aging research for older adults?

- What approaches can be taken in the integration of data sets in aging research to overcome common cultural and communication barriers among underserved and underrepresented populations of older adults?

3:20 pm Summary of the Day – Julie Bynum and Mitesh Patel

3:30 pm Adjourn for Day

Day 2: April 30, 2021

1:00 pm Welcome – Julie Bynum and Mitesh Patel

Theme 3: Social License for Uses of Data and Public-Private Data Sharing Collaborations

1:05 pm Nightingale Open Science and the Potential of Public-Private Data Collaborations – Katy Haynes, Nightingale Open Science

1:15 pm Data Justice: Ensuring Diversity and Accessibility of Research Datasets – Kayte Spector-Bagdady, University of Michigan

1:25 pm Big Data, Nudges, and Rapid Learning – Gui Woolston, CVS

1:35 pm Discussant Panel – Rick Cohen, CVS; William Crown, Brandeis University; and Mona Siddiqui, Humana

2:05 pm Broad Group Comment and Questions

Theme 3 Discussion Topics:

- What are common approaches to address whether big data resources accurately reflect population health effects?
- What are meaningful approaches to enable communities to speak to the use of personal health data for the benefit of the community?
- What are the consequences of under-represented populations in large data studies for introducing bias?
- What are the principles of some of the more successful data sharing models among academic institutions and private sector organizations to support longitudinal aging studies and intervention studies?
- How might community members be best engaged in consenting processes for data use in aging populations?

2:15 pm Break

Theme 4: Strategies for Use of Big Data for Clinical Interventions with Older Adults

2:20 pm EHRs for Population-Level Intervention: Promise and Pitfalls – Andrea Cheville, Mayo Clinic

2:30 pm Machine Learning and Behavioral Nudges in Advanced Illness Care Delivery: A Case Study in Serious Illness Communication – Ravi Parikh, University of Pennsylvania

2:40 pm Discussant Panel – Malaz Boustani, Indiana University/Regenstrief; Niteesh Choudhry, Harvard University; and Mark Friedberg, Blue Cross Blue Shield/Harvard University

3:10 pm Broad Group Comment and Questions

Theme 4 Discussion Topics:

- What are successful approaches to integrate data resources to support studies of evidence-based screening, interventions. and preventive care for ADRD?
- How might data hubs be established to enable integration of real-world data and other sources to enable assessment of clinical interventions (e.g., nudges)?
- What tools are needed to enable nimble and accurate data extraction from unstructured data resources to inform appropriate care for older adults?
- How can analytic methods (ML/AI) be adapted to avoid integration of bias into algorithms addressing clinical and population data?
- Can existing EHR data resources be used for accurate patient-reported outcomes of investigations of ADRD clinical interventions?

3:20 pm Summary Comments – Julie Bynum and Mitesh Patel

3:35 pm Closing Remarks – Partha Bhattacharyya and Marcel Salive

3:40 pm Adjourn