U.S. Department of Health & Human Services
National Institutes of Health

National Center for Complementary and Integrative Health

Methodological Approaches for **Whole Person Research**

September 29–30, 2021

Meeting Summary

Methodological Approaches for Whole Person Research Workshop
September 29–30, 2021

National Center for Complementary and Integrative Health, National Institutes of Health

The National Center for Complementary and Integrative Health's (NCCIH's) new strategic plan defines whole person research as including three components:

- Exploring the fundamental science of interconnected systems
- Investigating multicomponent interventions or therapeutic systems
- Examining the impact of these interventions on multisystem or multiorgan outcomes

The Whole Person Research Workshop was convened to discuss examples of research studies in these three areas from diverse fields and explore methodologies potentially appropriate for whole person research. The workshop was led by NCCIH. Workshop collaborators included the National Institute on Aging, National Institute on Minority Health and Health Disparities, National Institute of Nursing Research, National Institute of Dental and Craniofacial Research, Fogarty International Center, Office of Research on Women's Health, Office of Behavioral and Social Sciences Research, Office of Disease Prevention, and the Office of Nutrition Research within the National Institutes of Health (NIH) Office of the Director.

Day 1: September 29

Welcome

Ms. Catherine Law welcomed everyone to the first day of the workshop. She noted that NIH VideoCast viewers could submit questions at any time using the VideoCast feedback form. She said the panels would not have time to answer every question during the workshop, but all questions submitted through the feedback form would be collected and shared with the workshop organizers. She noted that the workshop was being recorded and would be archived on the NIH VideoCast website and a meeting hashtag was created, #WholePersonResearch2021. She then introduced NCCIH Director Dr. Helene Langevin.

Opening Remarks and Setting the Stage

Helene Langevin, M.D., Director, NCCIH

Dr. Langevin began by thanking the workshop cochairs Drs. Wen Chen and Wendy Weber as well as the trans-NIH planning committee. She said that while whole person health and whole person research are central to NCCIH's strategic plan, many different NIH Institutes and Centers are interested in this field of endeavor. Whole person research is a counter approach to mainstream biomedical science, which focuses on analysis, precision, and breaking things down to their smallest component. Whole person research complements analysis through synthesis, that is, putting the pieces back together to understand, for example, pathogenesis and pathophysiology and their role in health and disease. It is not necessarily a new approach to studying complex interconnected systems. Integrated physiology, systems biology, and network science are established analytic and design methods that will be called on to assist in this type of complex research.

Dr. Langevin noted that while whole person research is bidirectional between analysis and synthesis, it is sometimes more challenging to synthesize then to analyze. Modern medicine and biomedicine tend to get stuck at certain levels of analysis, which can create barriers to synthesis and integration, for example, by separating the body into the organ systems. This creates silos not only in research but also in academic departments and medical specialties, resulting in gaps in knowledge across systems, for example, between gut and brain or between gut and lung. We know that there are exchanges between and among these systems that influence the functioning of each, for example, the microbiome can influence inflammatory processes that then affect the brain. Yet the scientific literature contains few reports of cross-system analyses. Understanding the different organ systems in a cross disciplinary way will expand our understanding of integrative research and eventually the whole person.

It is instructive to consider how we view disease versus health. When we think about diseases, we think about them one organ system at a time, for example, cardiovascular or neurological disease. In contrast, when we think about health, we have a more holistic view about the processes that involve the whole person. Moreover, there are bidirectional transitions between health and disease involving more than organ systems, for example, behavior, psychological stress, diet, or sedentary lifestyle. These factors can lead to poor health, such as diabetes, obesity, or depression. In the whole person paradigm, multicomponent interventions can restore health by focusing on several variables simultaneously, such as nutrition, exercise, or psychological support. Prevention of disease can begin by intervening with the whole person before health is lost.

This orientation is especially critical because life expectancy in the United States is decreasing. A confluence of epidemics and socioeconomic factors, such as the opioid crisis, the obesity epidemic, the COVID-19 pandemic, and socioeconomic and environmental disparities, are affecting all parts of the U.S. population. A multilevel whole person health framework that looks at biological, behavioral, social, and environmental factors across individuals, families, communities, and the general population can find ways to improve health in multiple interconnected domains. This at the center of NCCIH's strategic plan. It will require studying the effects of not just a single intervention, but multiple interventions. For years, research designs and statistical analyses of studies that looked at multiple outcomes were sometimes labelled "fishing expeditions" and scored poorly in peer review. This highlights the need to refine the methodological approaches for whole person research, the topic of this workshop. In sum, whole person research involves (1) exploring the fundamental science of interconnected systems, (2) investigating multicomponent interventions or therapeutic systems, and (3) examining the impact of these interventions on multisystem or multi-organ outcomes.

Importantly, not all studies of the whole person are whole person research. Research in various individual domains can be expanded into research on the interconnectedness of domains, which is the focus of the first session of this workshop. The level of complexity in such studies calls for the application of network science, which can be conducted in observational studies, relying on the concepts of analysis and synthesis to understand, patterns, relationships, connectedness, and changes over time. We know that multiscale networks can self-organize and have the potential to grow through emergence of patterns in response to challenges. This can happen from the top down or from the bottom up. An example of a bottom-up process that could be revealed through longitudinal studies is immune responses, epigenetic changes, or microbiome alterations. An example of a top-down process is changes in behavior or social conditions that involve conscious decisions on the parts of individuals or groups of individuals, which can have profound effects on health.

The second session of the workshop focuses on how to study the impact of single component interventions or manipulation on interconnected multiple systems. For example, we know that stress reduction techniques such as mindfulness can help with sleep, glucose metabolism, weight loss, and chronic musculoskeletal pain, but we know very little about how these effects may be interrelated because that understanding requires including these various outcomes in the same study.

The third session centers on how to investigate the impact of multicomponent interventions or therapeutic systems on a single outcome. An example would be assessing the combination of buprenorphine and a support group for opioid use disorder where the outcome is abstinence. Another example is a combination of diet, exercise, and stress management for cardiac rehabilitation where the outcome is an increased ejection fraction.

The fourth session focuses on how to examine the impact of complex multicomponent interventions on multisystem or multiorgan outcomes. An example is, how does a multicomponent program of diet, exercise, and stress management influence multiple physiological functions that are likely to be interrelated, such as sleep, sympathetic activity, microbiome, aerobic capacity, or muscle strength.

Dr. Langevin closed by saying that combined these sessions and their focus on methodology will hopefully shed some light on how to not only take the whole person puzzle apart but also to put it back together in a rigorous way.

Discussion

Dr. Karyn Esser, University of Florida, said that NIH is also siloed by organ systems, making it more challenging to get funded for cross-systems research. Dr. Langevin responded that there are many noncategorical NIH Institutes and Centers and multiple NIH-wide cross disciplinary efforts, but that clearly more can be done to cross disciplines. The goal of this workshop is to identify rigorous methodological approaches that will improve the conduct and standing of whole person, or multicomponent, multiorgan research. In response to a question from the VideoCast audience, Dr. Langevin said this workshop is a step in the process of identifying gaps and research priorities to then move funding programs through the concept clearance process to funding announcements.

Dr. Janine Simmons, National Institute on Aging, reminded the audience that an NIH workshop on "The Science of Interoception and Its Roles in Nervous System Disorders" was convened in 2019 sponsored by the NIH Blueprint for Neuroscience Research. Its focus was neural circuitry underlying the dynamic interactions between the central and peripheral nervous systems. The workshop resulted in funding announcements and several awards in basic science. Dr. Langevin added that interoception is relevant to many different complementary and integrative interventions.

In response to a question about industry involvement in whole person research, Dr. Langevin said that industry might develop devices to serve as interfaces to interrogate pathways, or develop natural products, which are complex mixtures that can have effects in multiple organ systems.

A medical anthropologist in the VideoCast audience asked whether culture is included in the social category, noting that the research approaches to studying culture differ from those studying social factors. Dr. Langevin said she would need to think further on it but as a first response, certainly cultural factors, such as discrimination, disparities, and stigma would be included under social factors. However, she agreed that there is also a need to consider how interventions are practiced and perceived in

different cultural contexts, for example, acupuncture in traditional Chinese medicine. Dr. Judith Arroyo, National Institute on Minority Health and Health Disparities, added that the social determinants of health often consider culture, such as traditions and belief systems that drive behavior. The PhenX Toolkit contains measures for social determinants of health. A representative of the NIH Office of Behavioral and Social Science Research added that the Office has created a handbook for conducting research in different cultural frameworks (The cultural framework for health: an integrative approach for research and program design and evaluation).

A VideoCast participant asked how the methods discussed at this workshop might be relevant in traditional medical systems. Dr. Langevin said that the Western medical system starts with a diagnostic framework and moves toward therapy, which might not be the case in other types of health care, such as Ayurvedic medicine. For example, a different health care system might consider the signs and symptoms of the patient and organize them in a completely different way than Western medicine. An intervention should be considered within the context of the system in which it is used. Nonetheless, to understand effect, the treatments have to be standardized and validated to be reliable and reproducible, despite having a different theoretical or diagnostic focus than Western medicine. Dr. Weber added that this workshop aims to identify methods that can address multicomponent interventions, including across different medical systems, which should assist not only the research community but also peer reviewers.

Session One—How to Study Interconnected Systems: Observational Studies
Moderators: Janine Simmons, M.D., Ph.D., National Institute on Aging; Qilu Yu, Ph.D., NCCIH

Dr. Yu provided a brief introduction for each of the speakers.

A Toolbox for Isolating and Studying Parts of Interconnected Systems: Almost Matching Exactly for Observational Causal Inference:
Cynthia Rudin, Ph.D., Duke University

The Almost Matching Exactly (AME) Lab at Duke University aims to perform data-driven causal analyses from complex interconnected systems. Dr. Rudin provided an example to illustrate its intent. When asking the question, "how will a drug affect a patient?" the answer to the question is affected by her age, race, medical history, and current medical condition; the rate she absorbs drugs; and the other drugs she takes. Matching allows one to isolate part of that complex system in a controlled way and study just one influence or outcome at a time. The AME approach attempts to match a current situation with almost identical situations from the past, in order to use these past situations to predict the future. It was developed because previous matching efforts (e.g., propensity score matching, prognostic scores, black box machine learning) produced uninterpretable matches and subjective analyses and were too slow to generate.

High-dimensional data cannot be processed in the human brain, so an algorithm was needed. The AME approach uses case-based reasoning for matching, a form of causal inference that is valuable because it is nonparametric and interpretable. It is a good way to study interconnected systems because it mimics a controlled experiment, it controls for other things going on in the system, and it leverages large databases of observational data when an experiment cannot be done and no match-controlled groups are available. Because individuals cannot be matched on covariances, they will differ on at least some dimensions; the aim is to select the covariance that matters, based on a series of assumptions. AME matches units almost exactly on covariates. These methods rival black box machine learning methods in

their estimation accuracy but also have the benefit of being interpretable and easier to troubleshoot. The approach then aims to learn the distance metric using machine learning on a separate training set, rather than having the human try to come up with some distance metric in their head.

Four software packages are available from the AME Lab. Dr. Rudin described an application of the Matching After Learning to Stretch (MALTS) Python Package as an example. It was used to ask whether a job training program is effective in increasing a person's salary over time. The program was used to find a match group for a cohort of individuals who underwent the training program. If one creates a match group based on prognostic scores, it is not a very cohesive group. MALTS constructs a match group that is far more cohesive, in terms of age and level of education. The program then creates an interpretable "stretch" distance metric that stretches the covariates according to how important they are. The "stretch" of each covariate is determined from machine learning. In experiments, MALTS has been as accurate in estimating treatment effects as black box machine learning techniques, and it is interpretable. Dr. Rudin provided another example of AME application to discern the effects of multiple variables contributing to seizures in critically ill patients, for example, how does the seizure activity affect the patient's outcomes along several variables.

In sum, this framework has several important elements. First, its algorithms create matched groups that are interpretable. The goal is to match treatment and control units as closely as possible, or "almost exactly." Second, its algorithms create accurate estimates of individual treatment effects. This is because it uses machine learning on a separate training set to learn which features are important for matching. Variables that are important are "stretched" so that the matched groups agree closely on these variables. Third, the methods are fast and scalable. This approach has proven invaluable in the study of complex systems where causal effects can easily be confused with correlations. It leverages machine learning and database tools, and is scalable, fast, and accurate. Importantly, it provides a toolkit for granular insights into data and strengthens the causal claims that can be made.

Machine Learning Methods for Studying Dynamic, Interconnected Multisystems
Ziv Bar-Joseph, Ph.D., M.Sc., Carnegie Mellon University

Molecular interconnected systems at the cell, tissue, and organ levels are composed of several interacting entities that, together, play a critical role in all biological and biomedical processes. Thus, our body is a complex multisystem comprised of several interacting components at various levels. Dr. Bar-Joseph said that he and his colleagues are focused on the molecular level and on how systems work within and between cells. They then model these systems using networks, and in reconstructing these networks using machine learning methods, analyze them to derive insights and actions. The aim is to develop tools to create an open, global atlas of the human body at the cellular level leading us to a better understanding of how the relationships among our cells affect our health.

Dr. Bar-Joseph provided an overview of machine learning methods, both supervised and unsupervised, that have been used to study and model various dynamic interconnected networks within and between cells.

Unsupervised learning uses machine learning algorithms to analyze and cluster unlabeled datasets. This is useful when you have collected a lot of data but do not have a good understanding of each of the components in the data or even the data alone. These algorithms discover hidden patterns or data groupings without the need for human intervention. Data are usually very high dimensional (e.g., genes, cells, images), and several methods can be used to reduce the dimension for both downstream analysis

and visualization. The next step is clustering the reduced dimensional data, that is, grouping of entities (e.g., genes, cells, tissues) based on their similarity or distance. The next step is to look at joint interactions through modeling using probabilistic methods. These probabilistic methods can provide insight into not only interactions but also information on and confidence in the probability of events occurring. One of the well-known approaches is use of Bayesian networks that allows one to assume independence between some objects in the sample and identify factors that are dependent or independent. This method is useful for static, snapshot data and can provide information on causation and directionality. Another method is Markov random fields, which is useful for static, undirected interactions (e.g., protein complexes, spatial organization). In contrast, Hidden Markov Models (HMMs) are useful to model dynamic networks using time series data. This method can be used to integrate static and time series data to learn causal models. Because it provides a discrete set of states, it is primarily useful for data sampled at low frequencies. Continuous states HMMs extend the methods of HMMs. They can be used to integrate signaling, regulatory, and transcription data and to model highly sampled data, including single cell data. These analyses can lead to identifying which specific intervention would lead to a better outcome.

Supervised machine learning uses labeled datasets to train algorithms to classify data or predict outcomes. In short, as input data are entered into the model, it adjusts its weights until the model has been fitted appropriately. Deep neural networks (a supervised method) have revolutionized machine learning over the past decade. A deep neural network is a neural network with a certain level of complexity, requiring sophisticated mathematical modeling to process data in complex ways. This approach works well for data with some locality properties, for example, image or sequence data. It can be used to learn things that might not have been immediately known or recognized, for example, some type of relationship occurring at different locations or interactions in spatial data. Other types of data will require either transformation or extensions. This approach is considered supervised in that it requires positive and negative examples.

Deep learning has been used for inferring gene relationships from single-cell expression data. Usual methods for inferring gene-gene interactions from expression data have focused on intracellular interactions. High-throughput spatial expression data has facilitated methods to deduce such interactions both within and between cells. It may be hard to organize genes and proteins in an order or way that preserves locality relationships when using vectors or metrices. An alternative is to feed graphs into the neural networks. This can be based on known interactions, known causal relationships, or anything else. In this approach some of the graph (interactions) aspects are known and the inference is related to other attributes (e.g., causality or gene interaction based on spatial data).

In sum, unsupervised and supervised machine learning methods have been developed and used to model multisystems from large datasets. Each approach has its own assumptions and requirements, which are important to understand before choosing a method. Some methods are more appropriate for some data types depending on the information available. Nonetheless, data integration is often possible using many of these methods.

A Person-Oriented Approach to the Analysis of Interconnected, Multicomponent Systems: Using Latent Class/Profile Analysis to Identify Prototypical Profiles of Risk
Daniel Bauer, Ph.D., University of North Carolina at Chapel Hill

Dr. Bauer began by stating that the person-oriented approach to the analysis of interconnected, multicomponent systems is motivated from a developmental systems perspective. A developmental

system is comprised of multiple levels (e.g., biological, psychological, sociological, and cultural) that are "inextricably fused" to create a functioning holism. This fusion reflects high levels of interactions both within and between levels of the system.

In developmental psychology, a distinction has long been made between variable-oriented and person-oriented approaches to research. Variable-oriented methodology, reflected in many contemporary statistical methods, is characterized by the estimation of unique effects for specific variables, such as examining the predictive relationship between blood pressure and heart disease when controlling for other risk factors. However, in the real world everything else does not stay constant, and high blood pressure is probably related to many other factors that are also moving. This methodology takes an atomistic rather than holistic perspective. Moreover, when including interactions, these tend to be low order (e.g., one variable interacting with another), although some machine learning techniques can embed complex interaction patterns.

In contrast, the person-oriented approach eschews this atomistic focus on the (often additive) effects of specific variables in favor of a more holistic representation of the individual. Motivated from the perspective of systems theory, person-oriented research typically seeks to identify prototypical individual profiles across a set of variables characterizing the process under study; for example, configurations that reflect patterns of individual functioning or dense areas within multidimensional space.

Often, these profiles are obtained using heuristic clustering algorithms like K-Means or, increasingly, model-based approaches like latent class/profile analysis and other finite mixture models. What these unsupervised learning techniques share in common is the ability to identify configurations, or points in multivariate space, that reflect representative patterns of individual functioning across multiple domains, and that can be used as predictors or outcomes. As an illustration, assume classes of individuals are mixed together in a population. The classes differ in their values on a set of observed variables, for example, categorical indicators (latent class analysis) or continuous indicators (latent profile analysis) to predict which class a person is in. After obtaining classes, they can then be related to other variables to predict latent class membership, distal outcomes, or moderators. In other words, can certain classes predict higher risks? Dr. Bauer used an example of predicting school dropout based on a social competence profile with four variables.

Whole person research is concerned with interconnected systems of predictors or outcomes. Therefore, latent class/profile models provide one way of getting traction on complex systems by identifying prototypical profiles that represent individuals as wholes. This provides a potentially beneficial match between objectives and research methodology. For instance, a person-oriented approach would be ideal for evaluating hypotheses regarding metabolic syndrome, defined as a constellation of risk factors (high blood pressure, high blood glucose, low high-density lipoprotein (HDL), high triglycerides, large waist circumference), and its relation to heart disease and other health problems. Latent profile analysis could be used to identify representative profiles for these risks risk factors, for example, does metabolic syndrome emerge as a distinct profile and what is its prevalence, or what are the relative risks of different patterns for heart disease or diabetes?

The advantage of this type of approach is that prototypical patterns are easy to interpret and communicate, there is a holistic focus on individual rather than an atomistic approach, and one can capture highly interactive nonlinear relationships among risk factors and outcomes. There are some disadvantages, however, such as the number of patterns assumed finite and the possibility that

reduction of complexity can be excessive. Further, the patterns obtained can be influenced by analytic choices and assumptions, which might be wrong. Finally, optimization can be challenging and rare patterns can be difficult to find without large samples. Nevertheless, the person-oriented approach and its attendant research methods are well suited for studying interconnected systems in whole person health research.

Towards a Precision Medicine Based on Interpretable Machine Learning
Trey Ideker, Ph.D., University of California, San Diego

Most drugs entering clinical trials fail, often related to an incomplete understanding of the mechanisms governing drug response. Machine learning techniques hold immense promise for better drug response predictions, but most have not reached clinical practice due to their lack of interpretability and their focus on monotherapies. Deep learning offers the potential for integrating data from the level of the molecular profile to the predicted response from a potential therapy. In contrast to black box learning systems, deep learning allows one to integrate multiple layers of data, for example, transferring learning from cell lines to clinical samples or developing interpretable predictions using knowledge maps of human biology. In the life sciences, extensive knowledge of cell biology provides an opportunity to design visible neural networks that couple the model's inner workings to those of real systems, facilitating genotype to phenotype translation.

Dr. Ideker said that a conventional neural network translates input to output as a black box without knowledge of system structure. In a visible neural network, however, input/output translation is based on prior structural knowledge. The DrugCell model is an interpretable deep learning model of human cancer cells trained on the responses of thousands of tumor cell lines to thousands of approved or exploratory therapeutic agents. In this visible neural network, gene disruption genotypes are translated to cell growth predictions through a hierarchy of cell subsystems. The structure of the model is built from a knowledge base of molecular pathways important for cancer, which can be drawn from literature or formulated directly from integration of data from genomics, proteomics, and imaging. Based on this structure, alterations to the tumor genome induce states on specific pathways, which combine with drug structure to yield a predicted response to therapy. High-throughput screens on thousands of potential therapeutic compounds can be done to predict responses. The key pathways in capturing a drug response lead directly to design of synergistic drug combinations, which are validated systematically by combinatorial clustered regularly interspaced short palindromic repeats (CRISPR), drug-drug screening in vitro, and patient-derived xenografts.

This method has been used to predict tumor response to taxol, for example, by revealing the pathways in which mutations moderate response. It has also been used in clinical trials of agents for metastatic breast cancer to expose other genes in the pathway that play a role in response, which would not have been visible without the neural network, and that could explain why some patients do or do not respond well to therapy. The model is also being used to look at epigenetic changes, that is, DNA methylation changes that occur that could be an indicator of the overall health or state of aging of an individual (e.g., slow or fast aging).

Dr. Ideker noted that work in this area is just beginning. More needs to be done to better represent and expand input and output data, include expression to better capture plasticity, promote joint learning of model structure and function, and develop more systematic experimental designs.

Measuring Patients' Pace of Biological Aging with Longitudinal Data, Growth Curves, and Elastic Net Regression of DNA Methylation
Terrie Moffitt, Ph.D., Duke University

Dr. Moffitt said her team has developed a new measure of an individual's personal pace of biological aging. It is designed for use in clinical trial research and in prevention research aiming to extend years of healthy life. Dr. Moffitt began by displaying the lifetime trajectory for development of chronic diseases to demonstrate that the optimal timing of interventions to slow aging before disease onset is between the second and fifth decades of life.

The Dunedin Multidisciplinary Health and Development Study has been tracking a birth cohort of 1,000 individuals born in 1972. Because the study members have now reached their late 40s, Dr. Moffitt's team believes it is well situated to contribute knowledge about those key second to fifth decades of life. They have found variation on measurements relevant to aging. Regular assessments of participants included postural hypotension, bone density, physical function, GAITrite Walk Speed Assessment, vision assessment (e.g., optic nerve scan), dental exam, respiratory lung function, and clinical interviews such as life history calendar, psychiatric and substance abuse assessment, and financial knowledge. A cognitive neuropsych test battery is performed as is brain scanning starting at age 45.

Dr. Moffitt said aging should be thought of as a life course process. We know that exposures accumulate from early life, and that changes to physiology occur years before disease diagnosis. Organ damage is difficult to reverse fully so preventive interventions must begin early. As such, research is needed to understand aging in younger people. A key question is how to measure how fast a young person is aging. One challenge is that there is no accepted measure of aging, although geroscience has an operational definition of aging, which is "the gradual progressive coordinated deterioration of physiological integrity across multiple bodily systems," that is, an interconnected system of change.

To develop a measure of aging, the team tracked decline in seven organ systems by repeatedly assessing 19 biomarkers at ages 26, 32, 38, and 45. Virtually all biomarkers showed gradual, progressive, coordinated worsening of physiological integrity over time and across multiple organ systems. However, some cohort members declined faster than others and some hardly declined at all so the next step was to capture the differences among these individuals. The team combined the 19 change slopes across the 19 biomarkers to create a measure of coordinated decline, or the pace of aging. Each individual in the cohort receives a single score that represents the gradual change that is consistent across all of the 19 biomarkers. They then validated the pace of agents scored by testing if it predicted cohort members' outcomes on all the cognitive perceptual and sensory motor capacity measures. The slow agers scored better than the fast agers. The faster pace of aging score was associated with a thinner cortex and smaller surface area of the brain as assessed by MRI at age 45. In summary, the pace of biological aging can be measured as coordinated decline across organ systems. The pace of aging score validly predicts physical limitations, cognitive decline, sensory problems, subjective facial aging, and brain aging.

The resulting measure, DunedinPACE, is implementable in whole blood and has strong test-retest reliability. Because it was developed in a single-year birth cohort, DunedinPACE is unconfounded by generational differences in exposures to factors that alter DNA methylation. Because it was developed from analysis of longitudinal change, DunedinPACE measures recent ongoing aging-related changes, not long-standing differences in health from early life. Several epigenetic age clocks have already been developed to assess how much aging has occurred since conception. But they do not pin down when in life aging accelerates, whereas DunedinPACE measures how fast you are aging now. The tool has been

used to predict mortality in the Framingham Heart Study, to capture Black-White disparities in mortality, and to capture accelerated aging in Texas adolescents who had early life adversity. Studies have also shown that a faster DunedinPACE score predicts Alzheimer's disease and mild cognitive impairment. The challenge now is to create an exportable version that other researchers can use in different populations.

Panel Discussion

Drs. Simmons and Yu led the panel in a discussion of three questions:

- What methodologies can be used in preclinical models, human subjects, or both for studying interconnected systems, and what are the advantages/strengths/pros (e.g., ability to assess temporal dynamic range and responses) and limitations/cons (e.g., some approaches are considered fishing expeditions; need to adjustment for multiple comparisons, lack of causality analysis of machine learning) of these methodologies?
- What methodologies can be commonly used in both preclinical and clinical studies to analyze interconnected systems, and which types of data can be captured in both preclinical and human subject studies?
- Which computational and analytic methods are unique to preclinical studies vs. human subject research to study interconnected systems, and what types of data can be captured in preclinical models or human subjects uniquely?

Dr. Rudin started the discussion by noting that there are basically two types of models. Data-driven models include machine learning, and mechanistic models include, for example, differential equations. These models trade off on each other depending how much data and domain knowledge are available. Choice of methods also depends on what is being modelled, for example, time series. Dr. Bar-Joseph added that there are some hybrids of the two approaches, for example, where one starts with some knowledge and then incorporates data into it. The methods and modelling should be based on the type of intervention being tested. Dr. Bauer said that the choice depends on the types of questions being asked, the kinds of variables being measured, and the number of observations available. In animal research, the sample sizes are relatively small so it may be challenging to apply some of these methods in that context. Perhaps animal models are better suited to the mechanistic types of questions that can be answered with relatively small samples.

Dr. Ideker agreed that the two critical axes are data versus knowledge and added that transfer learning, a type of machine learning, can help translate across animal and human datasets. Animal data on drug response, which might be focused on a single causal mechanism, might not be translatable to humans, and high throughput screening often shows many misses, thus transfer learning is necessary. Dr. Bauer added that observational human research, as compared to animal studies that involve a homogeneous inbred strain, call for methods that account for the inability to control or hold a variable constant. Dr. Moffitt said that although animal research can explore methods for slowing aging, such studies are limited by the lack of community variables, which humans encounter, and a shorter animal lifespan.

Dr. Moffitt said that the Dunedin study shows that use of repeated measures can be used to track the pace of aging across multiple body systems and this could easily be done in animal models as well. She added that samples should be held for long periods of time to ensure there is no bias. There are countless quantitative approaches that can be used retrospectively to analyze data. Unfortunately, researchers in the United States have insufficient access to longitudinal work with repeated measures

over time that facilitate studying change in health over time and the layering in of more data as understanding grows. Dr. Rudin said that if you understand what the important variables are, you can sometimes determine when a model is not going to be generalizable. Further, interpretive models can be manipulated.

Dr. Simmons asked whether feeding more data into a model is more helpful with translation. Dr. Ideker said that overfitting should be avoided, especially when translating data from animals to humans, so it is best to begin with the simplest amount of data possible. He also said that researchers have stopped overfitting models to the White population. Dr. Bar-Joseph said that applying prior knowledge to identify generalizable variables can help determine the proper fit. Feature selection is critical, that is, finding the variables likely to be the most valuable for prediction and then optimizing them. Active learning in machine learning can be used for such optimization. Dr. Rudin said that matching requires minimizing the number of features, otherwise matches cannot be found. Dr. Bauer added that factor analysis can be a good way to identify the number of dimensions on which indicators load. Dr. David Amar from Stanford University warned about the potential for interpreting observational evidence from nonrepresentative samples, which could lead to collider bias, which can make associations between two or more variables that then affect the likelihood of an individual being sampled, and distorting associations between these variables.

In response to a question from the VideoCast audience about matching in other medical systems, Dr. Rudin said that regardless of how much data are needed, the variables being assessed have to be based on uniform and consistent data about the intervention being studied.

Dr. Simmons asked the panel to comment on which methodologies can be applied to longitudinal data. Dr. Rudin replied that matching can be used. Dr. Bar-Joseph said that clustering can be used with time series data, although it might not distinguish when an event occurred, just that it did. Dr. Bauer said that latent transition analysis can extend longitudinal data and look at longitudinal profile data, including trajectories across single or multiple domains. Dr. Moffitt commented that with longitudinal data, the ability to get repeated measures over time is important, even if in the early stages they do not appear to be informative. They might change over time, provide different results in a different population, or reveal a healthy volunteer bias.

In response to a question from the VideoCast audience about measuring the effect of human behavior and environment in supervised learning, Dr. Bar-Joseph replied that doing so requires collecting a lot of data, particularly on DNA variation, to determine whether a variable is informative.

Dr. Simmons asked if data from multiple longitudinal studies can be combined on overlapping measures. Dr. Amar responded that meta-analysis could use regression methods to analyze effect sizes across studies as long as the variance around the inputs is assessed as well. Dr. Bauer added that integrative data analysis aims to combine multiple studies to enable analyses that could not be done in any one study. The challenge is to harmonize measurements to enable direct comparisons.

Session Two—How To Study the Impact of Single Component Interventions or Manipulation on Interconnected Multiple Systems

Moderators: Bramaramba Kowtha, M.S., R.D.N., L.D.N., Office of Disease Prevention, NIH Office of the Director, and Elizabeth Barr, Ph.D., Office of Research on Women's Health, NIH Office of the Director

Dr. Kowtha provided a brief introduction for each of the speakers.

Impact of Sexual Trauma on the Interconnected Outcomes of Mental Health and Immune Response
Mimi Ghosh, Ph.D., George Washington University

Dr. Ghosh presented a list of physical and psychosocial stressors and noted that a person's perception of stress is a stressor. She said stress interacts with the brain, and the brain responds to stress through two pathways: the hypothalamic pituitary adrenal (HPA) axis, which results in cortisol production, and the sympathetic adrenomedullary (SAM) axis, which results in adrenaline production. Together, these pathways regulate the immune system. An aberrant immune response results in immune dysfunction at multiple levels, which has detrimental effects on mental and physical health.

Dr. Ghosh said she studies the relationship between stress and inflammation and how it affects the body. Acute, short-term stress is immunoprotective and beneficial. It enhances wound healing, provides a good vaccine response, and increases resistance to infection. Chronic and long-term stress are not good because they can result in an immunosuppressive phenotype or in excessive, uncontrolled immune responses. Those responses are harmful because they reduce wound healing, vaccine response, and resistance to infection and cancer.

Most of the studies on connections between stress, inflammation, and depression have been done in animals. Experimental studies have shown that stress is connected to increased inflammation, which has been connected to symptoms of depression. If the stressor is removed, inflammation and depression symptoms reduce. If stress occurs but inflammation is reduced, depression symptoms still reduce; and when depression is reduced, inflammation also reduces.

Dr. Ghosh said she has been investigating stress resulting from emotional, physical, or sexual abuse and how that stress affects mental health outcomes and immune dysfunction. She noted that cumulative abuse exposure has an additive effect. To demonstrate the interconnectedness of mental health and the immune system, she discussed data from three of her studies. All three are human observational studies that used minimal, descriptive statistics and had small sample sizes.

Dr. Ghosh showed data from a single case within the Trauma and HIV Risk: Investigating Stress and the Immune Disruption of the Vaginal Environment (THRIVE) study. The data compare timelines of a woman's known victimization events, her mental health indicators (e.g., for depression, resilience, and perceived stress), and biomarkers for inflammation taken from her cervicovaginal fluid. At Month 3, about the time of a known victimization event, the woman had poorer mental health and suppressed biomarkers. At Month 4, the data showed improved mental health and biomarker levels indicative of recovery. These data demonstrate a possible connection between mental health status and inflammation biomarkers.

In her next example, Dr. Ghosh showed data from a study of 19 women who had experienced acute sexual assault (i.e., rape) in the previous 4 days. This longitudinal study collected mental health data from surveys and inflammation biomarker data from cervicovaginal fluid. The results from this study were mixed. Between Visits 2 and 3, all the participants showed improved mental health, but their biomarker levels varied. However, Dr. Ghosh noted the biomarker data showed a downward trend toward recovery from Visit 1 to Visit 3.

In her third example, Dr. Ghosh showed data from the Women's Interagency HIV Study, which examined the effects of chronic sexual abuse and depression on immune biomarkers in women. In this study, biomarkers were collected from plasma and cervicovaginal fluid. Dr. Ghosh presented the data in a linear regression model (i.e., heat maps) to show associations between biomarkers, with depression and abuse as the predictor variables. For participants who experienced abuse and depression, data collected from plasma showed positive correlations between a small cluster of inflammatory biomarkers and negative correlations between several inflammatory and anti-inflammatory biomarkers. These patterns were not shown in the other participant groups (i.e., the control group and the groups that experienced only abuse or only depression) or in data collected from cervicovaginal fluid. These results demonstrate that the immune signature of biomarkers differs across conditions.

Dr. Ghosh stated that her methodologies allow researchers to study the effect of chronic exposure or a single event on human biomarkers and mental health, highlighting the interconnectedness of behavior and biology. However, these methodologies have many disadvantages. The results from observational data show association only, not causation. Sample sizes are small, and accessing, recruiting, and finding matching controls for participants is difficult. Confounding factors and biological variability cannot be adequately identified and controlled. Dr. Ghosh said she hopes future researchers will incorporate substudies into existing clinical trials, improve methodologies, and increase awareness of system interconnectedness.

Total-Body Positron Emission Tomography—A Transformative Tool for Quantitative Whole-Person Research
Ramsey D. Badawi, Ph.D., University of California, Davis

Dr. Badawi showed a total-body positron emission tomography (PET) video from 2018, which was the first occurrence of imaging an entire live, human body in three dimensions in real time. He briefly described PET, an imaging technique developed about 40 years ago. A radioactive tracer is injected and disperses through the body. As the radioactive material decays, it emits gamma rays, which are detected by the PET scanner. From those data, a three-dimensional image can be mathematically constructed. However, most PET scanners in use require administering a high dose of radiation to the body, and much of the signal is lost because the detector can only scan a portion of the body at one time.

Historically, PET scans have involved only single organs. In 2005, Dr. Badawi and colleagues proposed the development of a full-body PET scanner. If the entire body is scanned at one time, all of the radioactive signal can be captured simultaneously, less radioactive material is needed, and quantitative data can be collected from multiple organs at the same time.

Dr. Badawi said that a PET scanner detects a snapshot image of radioactive material concentration. However, data collected during a single snapshot do not account for factors that affect radioactive tracer concentration, such as tissue type or the metabolic stage of the radioactive material. Tracer kinetic modeling can help resolve that problem. A PET scan always includes a blood vessel, so arterial input function (i.e., the radioactivity concentration in the blood) can be measured. That measurement can be used to mathematically solve for several rate constants. Once the rate constants are known, the data collected from the PET scan can be differentiated, and images can be created.

Dr. Badawi shared several PET images: whole-body radioactive tracer concentration, blood volume, and glycolic input and delivery activity. These types of images have been used to determine cancer stages and to measure tumor response to chemotherapy. The images also can be used to examine systemic

effects of chemotherapy or any other intervention. For example, the images can show any effects of doxorubicin, a cancer treatment medication, on the heart. PET data can be used to collect information and create images from many angles to show cardiac motion and ejection fractions.

PET data can be used to examine the effects of chemotherapy on the brain, liver, kidneys, and entire body systems, all at the same time. Dr. Badawi's colleagues in Vienna, Austria, have been working toward developing a total-body connectome using a multiorgan segmentation technique that pulls PET data from individual organs. From that, they plan to generate a multiorgan functional connectivity map. Preliminary work in this area has been done among patients who have significant weight loss due to cancer. PET data also has been used in studies on the effects of meditation on metabolism and on the systemic effects of COVID-19, inflammatory arthritis, and acute myocardial infarction.

Preclinical Approaches for Whole Person Research: Lessons From the Molecular Transducers of Physical Activity Consortium (MoTrPAC)
Karyn Esser, Ph.D., University of Florida

Dr. Esser said the MoTrPAC is an NIH Common Fund initiative sponsored by the Office of the Director and several NIH Institutes, including the National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Arthritis and Musculoskeletal and Skin Diseases, and National Institute on Aging. The MoTrPAC conducts preclinical and clinical studies to discover and assemble a comprehensive map of molecular changes that occur in response to exercise. It plans to use that knowledge to help explain the benefits of exercise. The consortium also plans to establish a public, user-friendly database and tissue biorepository for future research. The consortium includes a bioinformatics center and several clinical centers, chemical analysis sites, and preclinical animal study sites.

Dr. Esser said one goal of the MoTrPAC preclinical studies was to develop a study design for a rat exercise intervention that would be standardized across three study sites and that harmonized with the consortium's plans for human clinical studies. Standardized environmental conditions included the light cycle, bedding, housing, and food for the rats. The rats all came from the same source, and common protocols for rat handling were established. Dr. Esser described the study's design and procedures and the types of data and tissues collected. She noted that the dates and times of tissue collection and freezing were recorded in case the researchers identified outlier data sets and needed to check the procedures on specific batches. The three different sites produced similar results, demonstrating that the procedures were well-matched, although there were some small, site-specific differences.

Challenges and Opportunities From the Multiomic MoTrPAC Project
David Amar, Ph.D., Stanford University

Dr. Amar focused on the computation and statistical challenges of analyzing data collected in the MoTrPAC project. He provided a brief summary of the objectives of the MoTrPAC project and the types of data collected. The MoTrPAC project involves a single intervention (exercise) with many outcomes (millions of molecular features). The randomization method used provides some guarantee that the study successfully measured the cause and effect of exercise, but the study does not provide direct causal evidence for how the different molecules interact. Analysis tasks for the study included a differential analysis, which involved using a standard regression analysis to quantify the training effect on a measured molecular feature.

Dr. Amar described several statistical methods that were used on data collected from the MoTrPAC project. For example, an independent hypothesis weighting method was used to avoid false positives and increase statistical power. A clustering analysis was conducted, and 15 main patterns were identified. To help with biological interpretation, the clusters were further evaluated using separate pathway enrichment analyses on each ome. Maps were generated that show the main pathways in each cluster and whether the pathways are consistent across omes.

A network propagation analysis was used to integrate data sets with prior knowledge. This method leveraged power across the different omes to find localized structural modules that could represent biological responses. A network inference method was used to suggest novel pathways and cascades of signaling.

A sample-level analysis was conducted using data factorization. Some factors were cross-omic and some were omic-specific. Factorization can provide insight for interpretation or for exploratory analysis. Factorization also can be used for data quality control. For example, one factor was identified that had a unique pattern. Reverse engineering revealed contamination of the data, and those data were removed from the analysis

Dr. Amar noted that the methods described were used with data collected from the MoTrPAC's preclinical studies. Analysis of data collected from humans will require additional statistical tools and methods.

Panel Discussion

Dr. Barr began the discussion portion of Session Two and asked the panelists to respond to the following question:

> What methodologies can be used in preclinical models, human subjects, or both for studying interconnected systems, and what are the advantages/strengths/pros (e.g., ability to assess temporal dynamic range and responses) and limitations/cons (e.g., some approaches are considered fishing expeditions; need to adjustment for multiple comparisons, lack of causality analysis of machine learning) of these methodologies?

Dr. Ghosh said animal studies cannot be used to study HIV in women. Hormones have a huge effect in HIV infection. Mouse and monkey models for HIV are good, but they cannot be used when examining the effects of hormones on genital tract inflammation, because those models do not accurately reflect the hormone status of human women. Mice are not human, and monkeys do not menstruate the same way human women do. Animal models cannot be used when examining HIV sexual transmission, infection, or pathogenesis in humans.

Dr. Ghosh also commented that healthy volunteer bias is a problem in her field. Finding matching controls for sexual trauma survivors is not possible. She said that if studies collect samples for biomarkers in slightly different ways (e.g., from different locations), those differences can make huge differences in the data, which makes data synthesis difficult.

Dr. Badawi said the accuracy of the model is critical. He said the United States has several small animal PET scanners, and the analyses used for data derived from the PET scanner would be the same for humans or other animals. However, animal models such as mice can have very small organs compared

to the spatial resolution of the system, which could introduce bias. Also, imaging an animal in the same state as a human is challenging, because humans can lie flat and be still. Animals can be anesthetized, but that alters the state of their central nervous systems.

Dr. Esser said the MoTrPAC project was designed to harmonize preclinical and human clinical studies from its inception. With animal models, more variables can be controlled, which helps the informatics group of the project. However, humans offer more genetic diversity. The consortium chose rats for the animal studies because they provide more tissue per organ, and all the omes could work off the same organ.

Dr. Amar said mouse experiments do not have genetic heterogeneity. That heterogeneity is very useful for molecular analysis. Some analytic methods can only be used with large data sets, and the data sets for rodents are not large enough. For observational data, correlation does not mean causation; an observation may have many causal explanations. The MoTrPAC project plans to create an encyclopedic database with as little bias as possible so researchers can use that data to extract and test their own hypotheses. Dr. Amar noted that translating analyses from rodents to humans depends on the biological question to be answered. Also, rodents may have evolutionary limitations that make comparisons with humans impossible.

Dr. Badawi asked if the MoTrPAC project did any imaging on the rats used in the study. Dr. Esser said no but added that imaging the results of exercise would be valuable. Dr. Badawi suggested examples of physiological changes that could be observed using PET.

Dr. Kowtha asked the panelists if any of the methods discussed in Session One were applicable to their research and which methods might be the most useful for intervention trials.

Dr. Amar said the MoTrPAC project has used many of the methods discussed in Session One, especially the unsupervised methods. The methods discussed in Session One are useful for complex, interconnected outcomes with unknown causal connections.

Dr. Badawi said the whole person approach is new for the PET field. The field has a legacy of kinetic modeling. He said several of the methods mentioned in Session One would be highly applicable to his research.

Dr. Ghosh said the AME method would apply well to human studies. She added that meta-analysis would advance her field, although cooperation among researchers to share their data would be required. She commented that examining global data is sometimes criticized as being a fishing expedition and a study using that method would not receive funding, but fishing expeditions can uncover unanticipated results.

Dr. Barr asked Dr. Ghosh if engagement in trauma-informed research might mitigate some of the mental health impact of traumatic experiences for participants. Dr. Ghosh answered that she initially was concerned about asking participants to discuss their experiences, but the participants were very resilient and wanted to talk about the events, liked the support provided, and wanted to know the results of the study. She acknowledged that the resiliency of participants could contribute to bias—less resilient individuals would not choose to participate in a study.

Dr. Chen said the differences between pre- and postintervention could help probe connections in the systems. She asked the panelists about systems interconnections that could benefit from intervention analysis. Dr. Esser said that not all organs respond or adapt to exercise in the same manner, and a temporal pattern might help them understand the reasons for that. Dr. Amar said Session One and Session Two could inform each other. Evidence from theoretical research has shown that methods combining experimental and observational data can provide stronger results than separate analysis of each data set.

Dr. Badawi said methods that combine types of data could be a therapy paradigm. For example, researchers could make observations on healthy controls and a disease population, conduct an intervention, and then examine the change in the disease population. Dr. Amar commented that in causal inference theory, if two features are not correlated, there is no reason to design a randomized trial between them. Data sets can inform each other, and researchers could use that information to determine which experiments would not be helpful and should not be conducted. If enough data are removed, the resulting data set could be small enough to be informative.

Dr. Kowtha asked Dr. Amar how data and methods are coordinated among the many partners in the MoTrPAC project and when meaningful results might be revealed. Dr. Amar said the consortium uses a hierarchical organization of teams and divides tasks. They use tools such as Jira. Dr. Esser said having a coordinating center record and manage the data has been helpful. The coordinating center has a quality control process that enhances data integrity.

Dr. Barr presented a question from a VideoCast viewer: Can these methods be used to identify specific characteristics, such as biomarkers, of animal models that are translatable to modeling the human condition? Dr. Amar said MoTrPAC human experiments do not have results yet and making comparisons with another database to do an analysis would be difficult. No human multiomic data set exists for comparison.

Dr. Badawi said that a selection effect exists in some research, such as in drug development. Drugs tested in humans have already passed safety testing in other animals. Drugs that are poisonous to mice but not humans would never achieve the level of human research. He said expecting an animal model to replicate a human in every area is not reasonable. Any results from MoTrPAC animal research would need follow-up studies to determine if those results translated to a meaningful effect in humans. The phenotypes are different, and a benefit in rodents may not necessarily be a benefit in humans. For example, the basic muscular mechanics may be the same for a rodent and a human, but a rodent's motivation to exercise would not match a human's motivation, because the human mind has such a large influence on the body.

Dr. Ghosh said systems are too complex for researchers to discover one biomarker that identifies something specific or one model that answers all the research questions. Human biovariability makes it too challenging to translate many results from animal studies to human studies. Investigators need to focus on multiple components and interactions between biomarkers. They need to choose a specific aspect and use a model that will help the researcher examine that specific aspect.

Dr. Kowtha asked Dr. Badawi how total-body PET might help with translational research. Dr. Badawi said PET begins with molecular pathways in the tissues, but the data are acquired at the organ level. He said it is easy to conceive of a molecular marker that could be labeled with a radioactive material and then traced in a PET scanner, but finding molecules and markers that actually work well is difficult, and

developing new markers requires a significant amount of research. Dr. Badawi said PET is a tool, but research is needed to find new ways to use that tool and to develop new paradigms to further biological understanding of human health.

A VideoCast viewer asked Dr. Ghosh how her research identified the bidirectional effect of depression on inflammation. Dr. Ghosh clarified that the research was not hers, but rat studies have identified inflammatory biomarkers associated with depression, and when the depression was treated, the levels of those biomarkers reduced. Also, research has shown that when inflammatory cytokines are blocked, depressive symptoms reduce.

Roundtable Discussion I
Moderators: Wen Chen, Ph.D., M.M.Sc., NCCIH, and Judith Arroyo, Ph.D., National Institute on Minority Health and Health Disparities (NIMHD)

Ms. Law introduced Drs. Chen and Arroyo. Dr. Chen presented several topics for the roundtable discussion.

- What methodologies can we learn from other fields?
- What are the gaps and challenges in the current methodologies?
- What social determinants and health disparities need to be overcome?
- What are opportunities for innovation and further advancements in computational and analytic methods, data collection, and related technologies?

Dr. Chen briefly introduced Dr. Elaine Hsiao.

Toward Uncovering Molecular Mechanisms for Microbiome–Nervous System Interactions
Elaine Y. Hsiao, Ph.D., University of California, Los Angeles

Dr. Hsiao said the microbiome should be included as a factor in studies of human and animal health. She provided a brief background of research on microbiome interactions with the nervous system. Mechanistic studies have been conducted using animal models (mostly mice) that can be raised in germ-free, microbiome-controlled environments. Previous research has shown that manipulating microbiomes can generate reproducible changes in behavior. Also, microbiota abnormalities have been linked to several health conditions, such as anxiety, Alzheimer's disease, and multiple sclerosis. However, research has not indicated if these associations indicate a cause-and-effect relationship or if confounding factors are involved.

Dr. Hsiao shared a graphic of the signaling pathways of the microbiome-gut-brain axis. The microbiome can be used to study multiorgan interactions. Host genetics partially shape the microbiome, but it also responds to changes in environmental conditions, such as diet or stress. The microbiome interacts with the nervous system in the following ways:

- Absorbed metabolites enter the circulatory system and interact with peripheral neurons or cross the blood-brain barrier to enter the central nervous system and interact with glia.
- In the gut epithelium, the microbiome can interact directly with immune cells, generating immunomodulation. Research has shown that in conditions such as stroke and multiple sclerosis, immune cells can infiltrate the brain.

- Microbiome-based metabolites can directly regulate neuronal activity when subsets of vagal neurons extend into the gut epithelium and synapse directly with endocrine cells in the gut.
- Indirect interactions with the microbiome may occur through secondary organs. For example, microbes that affect liver metabolism may have downstream effects in the brain.

Dr. Hsiao discussed several mechanistic approaches to studying microbiomes. Model organisms can be raised in a germ-free environment and be used as tools for colonizing synthetically designed or human-transplanted microbial communities. Specific microbiomes can be isolated and studied at the organ, tissue, and cell levels. Genomic sequencing of stool samples can be used to isolate and profile microbiomes. This method could be used to determine which microbes respond to an intervention or correlate with a disease.

If the targeted function is known, investigators can examine the bacterial genomes individually and identify candidate species that may participate in that function. A microbiome of those candidate species could be generated for an *in vitro* study. *In vitro* models can be used to try to understand connections between the microbiome, brain, and behavior. Other methods include using "organ on a chip" models that integrate neurons, immune cells, and the microbiome with intestinal culture and using bioreactor systems to model interactions between medications, diet, and microbial communities.

Two approaches to translational microbiome research include integrating omics with metagenomics and transplanting human microbiomes into animals, although fidelity of the transplantation has been poor. In humans, methods include microbiome transplants, use of antibiotics to deplete most of the microbiome, and dietary interventions.

Dr. Hsiao summarized her research on the use of a ketogenic diet to treat refractory epilepsy. The diet is severe and has side effects, so patient retention is low. Her investigation used a mouse model. Mice with depleted microbiomes did not respond to the ketogenic diet. She identified specific bacterial species and molecules that predicted seizure protection and developed a microbiome-based intervention to protect against seizures.

Dr. Arroyo shared NIMHD's research framework and introduced Dr. Marybel Robledo Gonzalez.

Health Equity and Interconnected Systems
Marybel Robledo Gonzalez, Ph.D., University of California, San Diego

Dr. Gonzalez said that a health equity perspective posits that upstream determinants, such as economic and social opportunities and home and community living conditions, influence health outcomes and should be the focus of interventions. Upstream conditions also influence the health-related choices that people make for themselves and their families.

Dr. Gonzalez summarized her research using data from the Adolescent Brain Cognitive Development (ABCD) Study. Her research examined 22 contextual environmental variables across several domains: economic insecurity, parental ecology, adverse childhood experiences, school and community, physiological health, and perinatal health. Previous research has shown that poverty is associated with negative cognitive outcomes. In her study of children aged 9 to 10, Dr. Gonzalez conducted a latent factor analysis using group factor analysis to reduce the dimensionality of the contextual variables. The investigation found three latent factors (i.e., general resources to adversity, youth perceived social

support, and perinatal well-being) that explained a distinct variance across the 22 measures. All of these measures were predictive of cognition and brain structure among the participants. The results showed that the gap between children in poverty and wealthier peers narrowed as the resource-to-adversity factor scores increased.

Dr. Gonzalez said that future studies need to ensure that data collection includes diverse populations in both race/ethnicity and socioeconomic demographics. The ABCD Study has successfully adapted retention and recruitment efforts to reach a specific sample population. She noted some assumptions that should be incorporated when developing research models for investigating racial/ethnic differences. She said race is a social construct, and socioeconomic differences among races reflect differences in social and economic opportunity. A broader social construct should be used when evaluating racial and ethnic differences, and social inequities such as racism, discrimination, and public policies are contributing factors. Also, researchers need to assume that social determinants of health, such as the sample population's living and working environments, can affect a study's outcomes.

Dr. Arroyo shared the roundtable discussion questions:

- How do the social determinants of health disparities (e.g., race/ethnicity, socioeconomic status, place) enter into the study of interconnected systems and the impact of single component interventions?
- Limitations in data collection could lead to bias in the application of machine learning/artificial intelligence-based models to underrepresented populations. What kind of data collection and data analysis techniques ensure equitable whole person research?

Dr. Arroyo asked Dr. Hsiao how social determinants of health might affect her models and asked for her thoughts about the data collection limitations in her research. Dr. Hsiao acknowledged that her investigations have faced those issues when evaluating the relevance of preclinical data to human conditions. For example, in Dr. Hsiao's work on epilepsy, the pediatric prospective study was conducted in a population that does not reflect the population with the largest prevalence of epilepsy. Socioeconomic factors are very important when considering the severity and negative consequences of epilepsy, but recruitment of reflective populations is a challenge. Dr. Arroyo reflected that Dr. Hsiao is aware that her sample population may be a problem when interpreting results and noted that a ketogenic diet as an intervention could also be a problem because it is expensive.

Dr. Arroyo asked Dr. Gonzalez how social determinants of health disparities could be integrated into studies to enrich the outcomes and the types of research questions that could be answered. Dr. Gonzalez said integrating those aspects into a research framework is important so the results will have translational impact. The ultimate goal is to improve the quality of life of people in the United States. The Hispanic population in the United States is very heterogeneous, both genetically and socioeconomically, so study results that include race as a factor will not apply to the entire U.S. Hispanic population. Interventions need to be introduced at the policy level and focus on improving conditions for entire populations rather than focus on changing individual habits within a smaller population. Dr. Arroyo commented that NIH has been working on ways to generate transformative initiatives, not incremental changes.

Dr. Arroyo asked Dr. Gonzales to comment on the second discussion question. Dr. Gonzales responded that the ABCD Study is a large, epidemiologically informed study that tried to recruit a demographic sample that represented the United States. Studies tend to recruit participants based on convenience,

but more effort should be made to recruit reflective populations, even if it adds complexity to the study design. She commented that awareness of the assumptions she mentioned earlier will help researchers interpret their data and draw conclusions from their results.

Dr. Chen asked panelists from all the sessions which methodologies could be used in a whole person research project to learn which systems are connected and how they are connected. Dr. Hsiao suggested applying multiomics to see which signatures correlate with specific outcomes. This process can be done first in animal models and then in human studies to validate if the findings from animals can also be observed in the human condition. Then, reductionist approaches can be used to target cells or molecules. Intersectional genetic approaches in animals, such as conditional knockouts to modulate the immune system, can be used to identify cell-type specificity. However, many interconnections are likely occurring at the same time, and researchers can intervene only on a short time scale, particularly if imaging is used.

Dr. Esser said research on a large scale, such as the MoTrPAC project, is needed. This type of research cannot be conducted in one laboratory or one study. Communication among researchers is essential to see connections and patterns. Standards need to be set so comparisons across labs is less challenging. She noted that funding would need to come from multiple sources and multiple review panels.

Dr. Badawi said examinations of the whole person look at an evolved, redundant web of interactions. Examining individual stimuli, mechanisms, and links may not be useful. Surrogates for good outcomes may need to be defined. He agreed that current funding mechanisms will not be effective because the scope of this research problem is too large, the expertise is too widespread, and the funding needs to go to collaborative studies.

Dr. Chen mentioned a recent Common Fund announcement of an intent to publish a funding opportunity announcement to solicit applications for a new initiative of the Stimulating Peripheral Activity To Relieve Conditions (SPARC) program. The initiative is intended to support a large, multisite study of the multiorgan effects of vagus nerve stimulation in humans.

Dr. Ghosh said that human studies and animal studies cannot use the same methodologies. For human studies, she advocated for developing noninvasive or less invasive methods that mimic the methods used in animal studies.

Dr. Chen asked the panelists for input on methods for measuring and assessing multisystem outcomes—specifically, methods that involve composite index measurements of multiple systems versus methods that involve network or clustering analyses. For example, a composite index score has been developed to measure the multisystem outcome of quality of life. Drs. Badawi and Esser commented that composite index measures were an interesting idea, especially for research on frailty and circadian rhythm.

Dr. Helene Langevin suggested that a composite evaluation could be used as a measure of health. For example, a fractal index is used to measure heart rate variability. Heart rate variability is a measure that reflects activity throughout the system. Dr. Langevin asked if a similar index could be developed that evaluates the health of the entire organism.

Dr. Hsiao said acquiring data is easier than understanding and interpreting those data. Researchers tend to focus on refining what is known rather than attempting to understand the unknown. Basic science research is needed to understand some of those data.

Dr. Chen agreed that whole person research is a combination of analysis and integration, and she introduced Dr. Emmeline Edwards for some closing remarks.

Closing Remarks
Emmeline Edwards, Ph.D., NCCIH

Dr. Edwards said she appreciated the rich discussions and was gratified to hear about the different methods that could be applied to whole person research. Session Two provided examples of potential studies that could be developed. She said researchers need to find ways to encourage interdisciplinary teams to work together to develop a program that facilitates translation of basic work to human studies. She agreed that a consortium approach or a research network approach is appealing. She said a new generation of scientists is needed to move whole person research forward. She thanked Dr. Langevin for her vision and the planning committee for its efforts.

Catherine Law ended Day 1 of the workshop.

Day 2: September 30

Session Three—How To Investigate the Impact of Multicomponent Interventions or Therapeutic Systems on a Single Outcome
Moderators: Ranjan Gupta, Ph.D., Fogarty International Center, NIH; Miya Whitaker, Psy.D., M.A., Office of Research on Women's Health, NIH

Dr. Gupta provided a brief introduction of the speakers for Session Three.

Methods for Designing Multicomponent Interventions Based on Naturopathy
Lynne Shinto, N.D., M.P.H., Oregon Health & Science University

Dr. Shinto began by explaining that naturopathic medicine shares key characteristics with other whole systems of medicine, such as traditional Chinese medicine, Ayurvedic medicine, and other traditional systems that include a holistic approach to diagnosis and treatment. The intent of a naturopathic doctor is to stimulate the self-healing capacities of the individual by using therapeutic modalities that include lifestyle management, herbs, nutritional supplements, homeopathy, physical medicine (physiotherapy, hydrotherapy, manipulation), and counseling, usually in conjunction with conventional medicine. The inherent complexity of whole system medicine makes it difficult to evaluate scientifically. Neuropathic medicine treatment combinations can change overtime with practitioner variability in administration, and some treatments are used frequently while others are rarely used. Further, assessments of clinical effectiveness might depend on a patient's desired outcome, for example, less pain, desired weight loss, or slower disease progression. Given this variability, Dr. Shinto said that rigorous scientific design elements assessing such interventions must reduce bias and increase reproducibility and model validity.

Dr. Shinto described the methods used to design a multicomponent intervention that models whole system naturopathic medicine using examples from two studies. The first, "Naturopathic Medicine in Multiple Sclerosis," assessed quality of life in multiple sclerosis (MS) patients who received interventions

designed to model the whole practice of naturopathy. Naturopathic medicine encompasses a broad range of modalities and may improve quality of life in patients with MS. The first goal was to define "best practice," that is, treatments deemed by naturopaths and naturopathic clinical experts to be the most valuable or beneficial for people with MS. Dr. Shinto and her colleagues conducted provider surveys and Delphi panels and solicited patient/community stakeholder input to inform a "best" practice intervention model. Through this process, naturopaths identified diet, essential fatty acids, and vitamins and minerals as valuable treatments for MS.

A Delphi panel then developed the treatment intervention for the clinical trial. It included diet, nutritional supplementations, methylcobalamine, and counseling, with the cohort randomized to three groups: usual care, naturopathic plus usual care, and educational visits with a nurse plus usual care. All participants had a definite diagnosis of relapsing-remitting MS. The intervention lasted 6 months, followed by blinded assessments and blinded data analysis. The primary outcome was quality of life (QOL SF-36), which was informed by a survey. Naturopathic medicine combined with usual care for MS showed a trend in improvement in the SF-36 subscale. This study's strengths included limited individualization, a randomized multicomponent treatment, model validity informed by experts, and reproducibility. Its limitations are the breadth of neuropathic care considered and whether they are the best treatments for MS, the fact that the study was not double-blind placebo-controlled, and whether the control groups were appropriate.

The "Meals, Mindfulness, and Moving Forward in First Episode Psychosis Study (M3 Study)," aimed to determine if a multicomponent active lifestyle intervention anchored in mindfulness can affect cardiometabolic risk in young people with first episode psychosis (FEP). The design of the intervention was informed by the Early Assessment Support Alliance (EASA), community special daycare clinics that screen and treat people with early signs of psychosis in the state of Oregon. Stakeholder meetings were convened to understand what activities could be included in an intervention and what data would be available from the clinics to assess outcomes. Special attention was paid to what this group of patients might find the most practical, feasible, and unintrusive. The intervention design included mindfulness practice, physical activity, diet and nutrition activity, and facilitated group discussions at the close of each session. Sessions were held for 6 weeks once a week for 4 hours. Outcomes were measured at baseline, 6 weeks, and 12 weeks based on quantitative and qualitative measures. A control group received usual care. Adherence was the primary outcome. This study found that the intervention was feasible with high attendance; 88 percent attended four out of six sessions. The study identified intervention components perceived as beneficial as well as barriers to sustaining healthy behaviors. All components of the intervention were perceived as beneficial and were used 6 weeks post study.

The qualitative findings support the program's ability to reduce stigma and foster resilience and self-efficacy short term. Barriers included cost of continuing diet and exercise training outside of M3. Participants expressed a desire for continued support of an active-lifestyle training program through an M3 booster session, reunion meetings, or peer support meetings focused on healthy lifestyle. Strengths of this study include a multicomponent and reproducible treatment intervention, qualitative assessment that informed the intervention components, and the fact that it was stakeholder informed. Its limits are that it was not double-blind placebo-controlled, treatment fidelity had not yet been established, it used a nonrandomized design, and the control group might not have been sufficient.

Addition of a Mindfulness Component to a Conventional Lifestyle Intervention for Sustained Remission of the Metabolic Syndrome
Lynda Powell, Ph.D., M.Ed., Rush University

Dr. Powell said that progressive translational science has been developing to answer the question of why behavioral treatments do not get incorporated into clinical practice. There is a need to be more critical of behavioral science by seeking cross-disciplinary insights, establishing clinical significance, conducting progressive studies, obtaining multiple perspectives, and employing multidisciplinary methods. Dr. Powell described a progressive translational science model that was used to determine if adding a mindfulness component to diet and physical activity can achieve sustained remission of metabolic syndrome (MetS). MetS is increasing in the United States, with as many one-third of the population affected. Behavioral interventions have been shown to be effective but are not sustained over time.

The need for a stress component in the model is drawn from basic science observations that cognitive control deteriorates in the face of stress and from convergence of focus group insights from clinicians and applied behavioral scientists that stress undercuts intention to engage in healthy behaviors. Dr. Powell and colleagues convened a multidisciplinary group of researchers for a year, who collectively identified stress as a component in unhealthy behaviors that can lead to MetS. The Eat Love Move (EML) protocol was developed targeting diet, physical activity, and stress.

This proof-of-concept study of the three-component intervention produced MetS remission in 54 percent of treated participants after 2.5 years. Component analysis of the mechanisms of MetS remission status did not support alleviation of depressed symptoms, but a 7-year follow up revealed that participants perceived the most important skill in sustaining change was emotional nonreactivity (a mindfulness characteristic). Using principles of human-centered design, the next step in the research was to convene a small group of 10 volunteers with MetS to prepare and eat food together over 3 months. Observing this group revealed that fast, mindless eating was a needed target for intervention.

This progressive set of studies produced a hypothesized pathway where stress was replaced with a mindfulness component that targeted emotional nonreactivity and sensory awareness. Emotional nonreactivity focused on pausing before reacting and sensory awareness focused on nonjudging sensory experiences. The combination of the two components led to MetS remission in more than 40 percent of participants. Continued development will evaluate sustainability of mindfulness over 2 years using ecological momentary assessment and mechanisms of sustainability with change score analyses to determine if change in mindfulness over 6 months predicts MetS remission over 18 months, independently of, or in interaction with, change in diet and physical activity. Dr. Powell said that the analysis was not complex, a mixed effects logistic regression looked at main effects and interactions between mindfulness and conventional habits. This model is a reminder that randomized controlled trials are not always the gold standard, that is, multiple methods developed over time can provide useful insights into how a multicomponent intervention can affect a single outcome.

Achieving Intervention EASE (Effectiveness, Affordability, Scalability, and Efficiency) Using the Multiphase Optimization Strategy (MOST)
Linda Collins, Ph.D., New York University

Multicomponent behavioral and biobehavioral interventions are used widely for prevention and treatment of health problems, improvement of academic achievement, and promotion of health. Dr. Collins began by suggesting that to achieve greater public health impact, interventions should be developed in consideration of affordability, scalability, and efficiency, along with effectiveness from the outset. Further, interventions should be optimized to achieve ease of implementation.

Historically, interventions are typically developed and evaluated using a treatment package approach, in which the intervention is assembled *a priori* and evaluated by means of a two-group randomized controlled trial. This is a useful approach to assess the performance of the entire package but not necessarily the individual components. Knowing which combination of components will provide the most affordability, scalability, and efficiency is what is needed.

Dr. Collins defined affordability as the extent to which the intervention is effective without exceeding budgetary constraints, acknowledging there is little incentive for scientists to think about how much the intervention will cost and who will pay for it. Altering the intervention to make it less expensive is risky if you do not know which components would be the best ones to remove and still have an intervention that gives you the best predicted outcome. The consequence of this is that there are many interventions that have never been implemented because they are too expensive, or they end up being implemented but ad hoc modifications have been made that undermined their effectiveness.

Dr. Collins then defined scalability as the extent to which the intervention can be implemented widely with fidelity. The prevailing logic today in intervention development and evaluation is to first establish effectiveness and then assess scalability. An effective intervention might not be scalable if it is too complicated to implement in a community setting or it requires more staff attention than can be spared. Again, it is risky to make alterations when an intervention has been developed using the classical treatment package approach because you do not know which components to remove and which to keep.

Efficiency is the extent to which the intervention avoids wasting time, money, or other valuable resources. The way to achieve an efficient intervention is to develop one that is made up completely of components that have a detectable effect on the outcome or augment the effect of another component. The traditional approach of randomized controlled trials aims to find a significant effect, but there is little incentive to consider cost or scalability. That implicit convention is that inactive or counterproductive components are acceptable as long as there is a significant effect. Dr. Collins pointed out that consumer products such as cars or software are not developed in this way.

Dr. Collins introduced an alternative methodological framework for developing, optimizing, and evaluating behavioral and biobehavioral interventions. She defined optimization as the process of identifying a strategic balance of effectiveness against affordability, scalability, and efficiency. It is not about finding the absolute best, but the best you can actually use. In addition, there may not be one single optimal intervention. What is optimal may vary across settings and time.

The multiphase optimization strategy (MOST) framework is a principled approach that integrates ideas from behavioral science, multivariate statistics, engineering, health economics, and decision science. The three phases include preparation, optimization, and evaluation. MOST enables the investigator to balance intervention effectiveness, affordability, scalability, and efficiency to achieve intervention EASE (a strategic balance of **E**ffectiveness, **A**ffordability, **S**calability, and **E**fficiency). Using MOST, behavioral and biobehavioral interventions can be optimized to meet any objective chosen by the investigator. The objective may be to develop a cost-effective intervention, an intervention that achieves a specified level of effectiveness, the briefest intervention that achieves a minimum level of effectiveness, or any other reasonable goal. The MOST framework relies heavily on resource management by strategic choice of highly efficient experimental designs. MOST offers several benefits, including more rapid long-run improvement of interventions, without requiring a dramatic increase in research resources.

In closing, Dr. Collins described an example of its application, optimization of a smoking cessation intervention with the goal to arrive at an efficient smoking cessation intervention made up of all active components. This was considered a screening experiment because it was used in a trial that used a fractional or fractional factorial design.

Community Wise: Development of a Multilevel Intervention To Reduce Alcohol and Substance Misuse Among Formerly Incarcerated Men
Liliane Windsor, Ph.D., The University of Illinois at Urbana-Champaign

Dr. Windsor introduced her talk by noting that rather than trying to fit the world to science, we should be trying to fit science to meet the needs of the world. She has long been motivated by the desire to conduct research with real-world applications based on community-based participatory research (CBPR). She and her team searched for a social problem existing in a complex and messy social system. They aimed to find a population or more than one population that is interconnected and impacting one another. They then identified an issue that a funder would care about and created a team with the right expertise to conduct the intervention, including members with community knowledge. The next step was to look at the state of intervention science and models available to address CBPR.

Marginalized urban communities with higher rates of poverty and people of color face serious inequities related to alcohol and substance misuse (ASM) when compared to more affluent and White communities (e.g., higher incarceration and HIV/hepatitis C virus infection rates). These communities also have considerably less access to effective and affordable treatment of substance use disorders. With this focus in mind, the team settled on a multilevel intervention with multilevel outcomes. The intervention, Community Wise, is a multilevel behavioral group intervention created in partnership with service providers, residents of marginalized communities, and individuals with histories of substance use disorders and incarceration, to reduce health inequities related to ASM. Community Wise addresses social determinants of health (e.g., stigma, poverty, lack of treatment access, housing, and meaningful employment) and inequities related to ASM at the micro level (e.g., cognitive and behavioral processes), meso level (e.g., relationships with individuals and organizations), and macro level (e.g., political and cultural processes). The program builds on critical consciousness theory, which empowers individuals, organizations, and communities to address social determinants of health while changing individual behaviors (e.g., reducing alcohol and illicit drug use).

The intervention is manualized, delivered by a trained peer-facilitator, and includes 9 weekly sessions lasting 2 hours each. The MOST strategy and CBPR were employed to optimize Community Wise over the past 10 years. Importantly, the intervention was developed based on what is known to be effective, rather than conducting a pilot study and conducting an efficacy trial later. Challenges remain to implementing these frameworks in the real world and there are limitations to multilevel intervention science, for example, selecting a single outcome or selecting one outcome at each level of an intervention. Dr. Windsor ended with a plea for the scientific community to create interventions and approaches that are more user friendly and more efficient, and that recognize the inherent messiness of some research questions.

Identifying the Mechanisms Underlying Multicomponent Pain Interventions
Mark P. Jensen, Ph.D., University of Washington

Dr. Jensen began by noting that multidimensional problems, for example chronic pain, require multicomponent treatment. However, we often do not know which components are necessary for an effective outcome. Two analytic strategies for understanding the mechanisms underlying multicomponent pain interventions are mediation and cross-lagged panel design analyses.

Dr. Jensen emphasized that the current gold standard for chronic pain treatment is multicomponent treatment, because chronic pain both impacts and is impacted by the whole person. Interventions might include physical activity, cognitive therapy, education, coping skills training, family therapy, medication withdrawal, acceptance and commitment therapy, mindfulness-based cognitive therapy, mindfulness-based stress reduction, acupuncture, yoga, or tai chi. One strategy for increasing our understanding of these treatments is to examine the effects they have on whole-person mechanism variables (i.e., biological, psychological, and social factors that the treatments target for change) and the subsequent impact of these mechanism variables on outcome.

The challenge is determining which components should be included to create the most effective multicomponent pain treatment and which should be excluded; further, which components work best together, that is, synergistically, to enhance outcomes. Mechanism research can provide empirical findings to inform component selection and exclusion. It can elucidate why the treatment works and once you know why you can adapt treatment to include those components that engage the most important mechanisms. Mechanistic clinical trials are studies in which a complementary or integrative intervention with demonstrated efficacy for a population is studied to understand mechanisms of response, nonresponse, or risk of adverse effects of the efficacious intervention. The primary outcome variables are the mechanism variables.

NIH defines mechanistic studies as those designed to understand a biological or behavioral process, the pathophysiology of a disease, or the mechanism of action of an intervention. Both mediation and cross-lagged analyses allow investigators to determine: (1) if the multicomponent treatments influence specific mechanism variables, as hypothesized; and (2) if treatment-related changes in these mechanism variables are associated with treatment outcome.

Mediation analyses identify the mechanism variables affected by treatment. For example, is the level of mindfulness affected by a treatment that includes a mindfulness training component and is activity level affected by a treatment that includes a behavioral activation component. It explores whether any treatment-related changes in the mechanism variables are associated with treatment outcome. For example, how do the mediators of exercise, mindfulness practice, and food intake affect waist circumference, blood pressure, psychological function, and quality of life. A primary strength of this approach is that the coefficient is causal because the treatment condition is randomly assigned to the multicomponent treatment for control. Another strength is that it indicates the extent to which changes in the mediator are associated with changes in clinical outcome. A null finding can rule out variables as mechanism factors.

Dr. Jensen cited a study in which he and colleagues conducted a randomized control trial of 173 individuals with chronic pain. Participants were randomly assigned to a cognitive therapy self-hypnosis training arm, a multicomponent hypnotic cognitive therapy arm, or an education control arm. Seventeen mechanism variables were studied, and mediation analysis found significant effects for only two variables, which allowed the team to carve out factors that were less likely to influence outcome.

In contrast, cross-lagged panel analyses determine if changes in mechanism variables during treatment precede subsequent changes in outcome. It is a strategy that can describe reciprocal relationships, or directional influences, between variables over time and to understand how they influence each other. An example of this approach is a randomized control trial of 290 individuals experiencing chronic pain. Participants were randomly assigned to receive cognitive behavior therapy or education as a control. Two mechanism variables were examined, catastrophizing and pain self-efficacy beliefs. Both treatments decreased catastrophizing and increased pain self-efficacy beliefs. These changes that occurred during treatment were then associated with later improvements in pain. It was concluded that these treatments were mutually causal, consistent with the idea that in networks it is not one thing only affecting another but rather mutual causation. The weakness of this approach is that it requires measurement of many variables multiple times. If there are multiple potential mediators that have to be measured frequently, there might be cases of missing data, which can complicate analysis.

Although each approach described here has strengths and weaknesses, both allow investigators to identify the mechanisms that are more or less likely to explain multicomponent treatment benefits, allowing for greater understanding of the components that might be most important to include in order to maximize beneficial outcomes. However, we still know very little regarding the mechanisms responsible for the benefits of multicomponent pain management. A greater understanding of mechanisms could inform adaptations to maximize efficacy. Mechanism analyses can be added to any multicomponent randomized controlled trial by assessing outcomes and potential mechanism variables.

Mass Spectrometry Metabolomics To Identify Bioactives and Synergists in Botanical Medicines
Nadja Cech, Ph.D., University of North Carolina at Greensboro

Dr. Cech began by highlighting that botanical (plant-based) natural products, which are classified as dietary supplements in the United States, constitute a multibillion-dollar industry in North America. Regulation and quality control for this industry is an ongoing challenge. Myriad examples exist whereby commercially available botanical natural products are either intentionally or unintentionally adulterated or mislabeled, a situation that constitutes a major health concern for consumers. Even for correctly identified botanical natural products, composition differs depending on numerous factors, including variability in genetics, cultivation conditions, and processing methods. While there is general agreement that rigorous scientific studies are needed to evaluate the safety and efficacy of botanical natural products used by consumers, researchers conducting such studies face the challenge of dealing with inherently complex mixtures of variable composition. Unfortunately, many studies of botanical natural products are carried out with poorly characterized study material, such that the results are irreproducible and difficult to interpret.

Dr. Cech explained that the mission of the Center of Excellence for Natural Product-Drug Interaction Research is to “conduct clinical studies to evaluate potential interactions that occur when botanical natural products are co-consumed with conventional medications.” Dr. Cech cited the example of green tea, which many people consume. There have been reports that its consumption can alter the metabolism of pharmaceutical drugs that a person might be taking. Dr. Cech and her colleagues conducted a clinical study of 16 healthy volunteers who consumed green tea while also taking raloxifene for treating bone loss. Raloxifene was selected because it is a good example of a pharmaceutical that is metabolized in the liver, so it served as a good test case for studying how metabolism could be altered by green tea. They found a 30 percent decrease in the amount of raloxifene in the urine of participants consuming green tea, which highlights the importance of patients telling their health care providers about any dietary supplements they are taking.

Dr. Cech explained how the team selected which green tea product to include in further *in vitro* studies, as there are hundreds of products available. Relying on Consumer Reports and Amazon sales of products, which reflect wide use by the public, 34 samples were selected for comparison with mass spectrometry metabolomics approaches to capture and compare the chemical diversity of the complex botanical natural products. Three standards used by the National Institute of Standards and Technology were used to select similar green tea materials.

Untargeted metabolomics was used to identify the multiple constituents that may play a role in the biological activity of a given green tea leaf or plant and to uncover mechanisms by which biological effects occur. This approach measures differences in abundance and presence or absence of the small molecules (secondary metabolites) for a holistic profile to enable comparisons among complex samples. Principal component analysis was then used to transform the number of potentially correlated variables into a smaller set of uncorrelated variables. All of these steps allowed the researchers to select tea leaves for study that are most commonly consumed. Although this limits the applicability of the findings, Dr. Cech said there had to be a rationale for which supplement was selected and why.

The next step was to assess which chemical constituents are responsible for the activity of green tea by correlating the chemical composition of samples with their biological activity. Biochemometrics was used to integrate the biological and chemical datasets identified. This involves statistical analysis of chemical and biological data to identify markers of a particular biological activity. This generates a whole series of fractions that are analyzed in an *in vitro* enzyme inhibition assay. These combined analyses can be used to predict which compounds might be active. A selectivity ratio plot can be used to predict the inhibitory components of green tea, which is key to understanding how it inhibits metabolism. Molecules found to have this effect have to be validated with isolated catechins because the results are correlative.

In sum, these methods provide a mechanism for assessing the identity, quality, and variability of a botanical natural product before selecting it as a research material through mass spectrometry metabolomics. It shifts the focus from searching for marker compounds to markers of activity through biochemometrics. Dr. Cech and colleagues have published the strategies they use in comparison to other approaches, which could be useful for others aiming to identify which study material to include. In addition, her group has conducted research on the topic of synergy and the importance of having multiple components in a mixture to achieve a biological effect.

Panel Discussion

Drs. Gupta and Whitaker led the panel in a discussion of four questions:

- What methodologies can be used in preclinical models, human subjects, or both for studying multicomponent interventions or therapeutic systems, and what are the advantages/strengths/pros (e.g., ability to assess temporal dynamic range and responses) and limitations/cons (e.g., some approaches are considered fishing expeditions; need to adjustment for multiple comparisons, lack of causality analysis of machine learning) of these methodologies?
- What methodologies can be commonly used in both preclinical and clinical studies to study multicomponent interventions or therapeutic systems, and which types of data can be captured in both preclinical and human subject studies?

- Which computational and analytic methods are unique to preclinical studies vs. human subject research to study multicomponent interventions or therapeutic systems, and what types of data can be captured in preclinical models or human subjects uniquely?
- What rigorous methods exist for developing multicomponent interventions?

Dr. Shinto responded that the methods she described would not apply to animal or preclinical research; however, such studies could inform clinical analyses of multicomponent interventions.

Dr. Powell encouraged more use of quasi-experimental designs with one arm as a means of proof of concept. Dr. Collins agreed, saying there is a need to do more small formative studies early on. Failure is informative and provides data for future, more tightly controlled trials. She added that much can be learned from secondary data; thus, the field of whole person research must become more systematic in developing a coherent knowledge base.

Dr. Shinto was asked what happened to participants in the M3 study after it was completed. She replied that resources are available to help participants remain stabilized. She added that the effort was also a proof-of-concept study, and they should have published what did not work (e.g., lengthy, burdensome surveys, too many blood draws) because that informed the final design.

Dr. Whitaker asked Dr. Jensen to comment on efforts to expand the pain intervention to other populations. He acknowledged the need to develop chronic pain interventions that are culturally appropriate. Relatedly, Dr. Whitaker asked Dr. Windsor to provide some insights on cultural centering and gender sensitive approaches to multicomponent interventions and assessment approaches and approaches that can be used to make the intervention more potent. Dr. Windsor replied that her team screened four components for the substance misuse intervention and identified two that were not only more effective but also easier to deliver. This maximized the efficacy and efficiency of the combined intervention. Future analysis could use factorial designs to identify the impact of individual components and the interaction effect of those individual components, especially those that might be more promising for different populations, for example, more culturally targeted.

Dr. Cech was asked if the method she described for identifying active components of the botanical and then validating it could be used in clinical studies of multicomponent interventions to validate which components are active. She responded that determining which components of mixtures being used in clinical studies are active is challenged by the need to get approval to test botanical extracts. It might be necessary to do preclinical, *in vitro* animal studies first. Dr. Weber asked Dr. Cech if these multiple components could then be analyzed to determine their mechanistic effects, which could then be used for clinical studies. Dr. Cech responded that many of the methodological approaches discussed in this session could be combined in various ways to derive more meaningful interpretations and enhance validity.

Dr. Cech responded to a question from Dr. Chen about scaling up from preclinical to clinical studies. She said that the power of multivariate statistical analyses is that it can simplify the data analysis to identify groups that are relevant, which could help reduce sample size while retaining power. The same methods can be used to target the biological observations to be made.

Dr. Whitaker asked each of the panelists to comment on whether there is a need for proof-of-concept studies of interventions that are similar to effective interventions already studied, and how many

feasibility studies are needed to test similar interventions. Dr. Powell replied that the essence of a proof-of-concept study is the question of whether there is a plausible clinical signal. In terms of feasibility studies, the focus is on the feasibility of the trial protocol, for example, whether applying an existing untested protocol can be used to assess a different outcome. The key is figuring out what set of questions you want to ask and then linking the methods that already exist to finding an appropriate best answer. Dr. Collins agreed, saying if the question that you are trying to investigate empirically is clearly defined that can inform how much experimentation needs to be done. It is important to start with a clearly stated scientific question. Then the question drives the method rather than the method driving the question.

Dr. Jensen said that choosing which data to collect obviously depends on the question being asked but also relies on finding measures that are valid and reliable and have the fewest items to measure. He often adds additional questions or data points that might be important to identifying future potential areas of exploration. Dr. Windsor encouraged designing the study as a process and not a single activity. Defining the questions is as much art as it is science because you must consider the problem and its complexity. Dr. Collins added that a common theme here is that research should be iterative and discovery driven. Dr. Simmons noted that the NIH Stage Model for Behavioral Intervention Development focuses on the pipeline connecting translational basic research to clinical research. It could be a useful approach for much of the research being discussed in this workshop.

Dr. Gupta asked Dr. Cech if changing the mode of administration of green tea would change outcomes. Dr. Cech replied that to date, all of her studies have been *in vitro* using green tea extracts or fractions of extracts. Should the studies move to *in vivo* in either animals or humans, the mode of administration should be consistent with the extracts and the fractions that can be administered.

Dr. Collins responded to a question from Dr. Whitaker about patient or client burden in study design, saying it should be a consideration, with limits set on cost and time.

Dr. Jensen said that mediation analysis can be used for multicomponent interventions to determine which components are effective. However, it can be difficult to assess which clinically meaningful changes can be made in an intervention. He encouraged use of studies that estimate effect sizes rather than narrowly focus on p values. Dr. Weber asked how methods can be enhanced in hypothesis-generating studies. Dr. Jensen replied that it might be more palatable to peer reviewers to select a single primary outcome measure to be tested at a certain p value and then plan to do secondary analyses.

Session Four—How To Examine the Impact of Complex Multicomponent Interventions on Multisystem or Multiorgan Outcomes
Moderators: Yvonne Bryan, Ph.D., National Institute of Nursing Research, and Hye-Sook Kim, Ph.D., NCCIH

Dr. Bryan briefly introduced the speakers for Session Four.

The Microbiome and Metabolome as a Readout of Complex Interventions Throughout the Body
Rob Knight, Ph.D., University of California, San Diego

Dr. Knight provided some background information about the human microbiome. He said human bodies have about 30 trillion human cells and about 39 trillion microbial cells. The human gene catalog has about 20,000 human genes, and the microbial gene catalog has an estimated 2 to 20 million microbial

genes. Humans ignore 99 percent of their genes, and those are the genes that can be changed through interventions.

Dr. Knight said the microbiome is linked to diseases and treatment responses throughout the body, including in the brain. The microbiome in the gut reflects different exposures, ranging from diet and exercise to drugs. Microbes from the gut can communicate with the brain through the release of small molecules in the immune system and through direct signaling with the vagus nerve. The composition of the microbiome can indicate if specific medications, such as acetaminophen, digoxin, and anticancer checkpoint inhibitors, will be safe or effective for an individual. The microbiome has an effect on most ingested drug and food molecules. Personalized nutrition plans can be developed based on predictions for how an individual's microbiome will influence that person's glycemic response to specific foods.

Dr. Knight summarized previous research on the microbiome. A study that compared mice raised in germ-free versus conventional environments found that up to 50 percent of the molecules in distal organs, including the brain, can differ depending on the presence of a microbiome. Other research has shown that humans have been losing microbiome complexity through industrialization, including diet. Dr. Knight said his investigations have found that modern populations have less seasonal cycling in the gut microbiome than a Hadza hunter-gatherer population, and industrialized societies have fewer taxa and simpler microbiome communities.

Dr. Knight said that his colleague, Dr. Martin Blaser at Rutgers University, documented how antibiotics and other factors, such as a low-fiber diet, lead to attenuated microbiome diversity. In the 20th century, infectious diseases caused by single microbes have been brought under control, but the prevalence of chronic diseases has increased. Many chronic conditions have been linked to changes in the microbiome. Analogue models in mice have shown that manipulation of the microbiome can cause or cure these conditions.

Dr. Knight said that diet is an important multifactorial intervention, and NIH-funded research has shown that long-term, but not short-term, diet can reshape the microbiome. This research sequenced microbe DNA and used the phylogenetic distance metric UniFrac to compare microbial communities. Results were summarized on a distance matrix, and a principal coordinates analysis (PCoA) and hierarchical clustering techniques were used. This investigation found that the diet effect exceeded the genotype effect.

Dr. Knight said his research has found that long-term diets high in proteins or carbohydrates correlate with specific microbes. However, a short-term diet intervention did not result in much individual-level change. Other researchers have shown rapid effects of a short-term diet on the human gut microbiome, but these investigations only show the magnitude, not the duration, of a specific diet's effect on the microbiome. When the same data are processed using different techniques, the results show no consistency in the direction of change across participants, and the effect size is small in relation to the standard variation among individuals.

Dr. Knight summarized a study of fecal transplants in people with *Clostridioides difficile (C. diff),* which showed that the stool of people with *C. diff* differs from healthy stool. Four participants received fecal transplants, and within a few days their symptoms vanished. This dramatic change has not been seen with drug or diet interventions.

Dr. Knight said that a using a PCoA method assumes linearity and does not work well with sparse or compositional data. A compositional tensor factorization allows for more subtle signals and time series data. Other extensively used methods in this field include random forest classifiers and regression models.

Dr. Knight commented that an analysis of the skin microbiome can identify a person's age within about 3 years, and an analysis of the gut microbiome can identify age within about 10 years. His research projects have been investigating if diet-based interventions can address the difference between a person's microbiome and chronical ages. Other research conducted by Dr. Knight includes the Alzheimer's Gut Microbiome Project, which examines links between the microbiome diet and Alzheimer's disease. This research, supported by NIH, leverages existing infrastructure from organizations such as the National Centralized Repository for Alzheimer's Disease and Related Dementias, National Alzheimer's Coordinating Center, Alzheimer's Disease Genetics Consortium, and Alzheimer's Disease Sequencing Project. The investigation has collaborated with the American Gut Project to use tools such as at-home collection kits for microbiome and blood samples.

Dr. Knight said that all these research results need to be integrated for longitudinal analysis, because longitudinal studies are critical for determining causality.

N-of-1 and Aggregated N-of-1 Studies for Exploring Multicomponent Intervention Effects on Multiple Health Outcomes
Nicholas J. Schork, Ph.D., The Translational Genomics Research Institute

Dr. Schork defined an N-of-1 clinical trial as a trial that examines the utility of an intervention for an individual, rather than a trial that focuses on population effects. For example, in an N-of-1 trial of blood pressure, the participant might receive different drugs, one at a time, with washout periods between drugs. Data on many variables would be collected throughout the entire process. Results from this hypothetical trial could show which drug was the most effective at reducing blood pressure in that one person. The individual acts as a control and active participant.

Dr. Schork said that this type of study can use many statistical strategies, such as randomization, blinding, multiple crossovers, and multivariate analyses, to achieve greater scientific rigor. Results from these trials can be aggregated for meta-analyses to determine common factors among participants who responded to an intervention.

Dr. Schork discussed methods for analyzing data from N-of-1 trials. He said for the purposes of simplifying his explanation, he made the assumption of continuous time, although that may not be the situation in a real analysis of response data. An N-of-1 study would typically include a treatment response variable and an intervention. Data analysis would be conducted comparing responses to the intervention versus response to the placebo. A simple t-test could be used on the measures collected, but that would not account for underlying factors, such as nutrients, microbiome species, or counseling activities, associated with the intervention that might contribute to the response. For example, a study examining the response to a nutritional intervention of four nutrients would not account for the participant's natural exposure to those nutrients. The analysis would need to compare the active intervention periods versus the placebo periods and examine the difference in the average levels (dose) for the factors that contribute to the response. Also, variation can occur within each placebo and active period.

Dr. Schork said one way of analyzing the data is to condition the factor effects on the treatment effect and use a dummy variable for the placebo period. The relationships between the levels of those factors in the outcome could be examined. Such an analysis does not have the power to detect a relationship between two factors that are not causally related but would have good power to detect factors that are causally related. If enough data implicate that a factor is associated with a response, a sequential probability rate could be used. Dr. Schork summarized some of the advantages and disadvantages to using sequential tests of time on response effects.

Dr. Schork said clinical trials can be used to identify causal relationships between variables, not just associations. Causality likely has a temporal component. A Granger regression analysis and other analysis techniques such as stochastic differential equations could be used to identify causal relationships between individual factors implicated in responses to a multifactor intervention. In an evaluation of the power of detecting effects, the power will be greater for a system of variables rather than one variable.

Dr. Schork summarized his research on the systematic influence of a polypharmacy intervention on a woman's sleep. The participant had a serious sleep disorder and was being treated with multiple medications. She was weaned off all her medications, and three drugs were reintroduced in different dosages and in different combinations. The investigators measured sleep quality and other factors, such as anxiety and depression. They conducted a regression analysis with sleep quality phenotypes as dependent variables and dose and drug as independent variables. The results showed that the participant had better quality sleep when all interventions were removed, but she still had depressive episodes and other conditions that deserved attention.

Dr. Schork described a series of N-of-1 studies that used a stress-relief app (Stop, Breathe, and Think) as an intervention. App users self-reported their current mood, selected from a menu of meditation exercises, and self-reported mood after completing the meditation exercise. Dr. Schork's investigation examined the long-term effects of the app on base mood. He wanted to identify which combinations of base mood and meditation exercises resulted in the strongest positive or negative change in mood. The meditations that people choose had very individual effects. Some meditations had only a marginal effect on mood, and some had a large effect. He analyzed the data to determine which components of the intervention appeared to work best for individuals who reported specific base moods.

Dr. Schork said that personalized medicine will require novel and more appropriate ways of testing the effect of complex, multifactor interventions on an individual's health as a whole. N-of-1 clinical trials can help researchers understand the different factors implicated in a complex intervention.

Multicomponent Interventions: An Organizing Framework for Selecting an Experimental Design
Inbal Nahum-Shani, Ph.D., University of Michigan

Dr. Nahum-Shani said she would be presenting a framework to help investigators decide which experimental design to use for multicomponent interventions. These designs all focus on a single outcome, but they could be extended to accommodate multiple outcomes. The framework includes five key questions and four experimental approaches.

Dr. Nahum-Shani said the first question is: Will the intervention include multiple components? A component is any aspect of intervention that can be separated for investigation. If the answer to this question is no, the research design would typically be a randomized controlled trial that compares the

effectiveness of that intervention with a suitable alternative. If the answer is yes, the next question to answer is: Is the investigator unsure about which components to include in the intervention? If the answer to this question is no, a randomized controlled trial is a good study design. If the answer is yes, the next question to answer is: Is the investigator unsure about one component or multiple components in the proposed intervention? If the uncertainty involves only one component, a trial could be conducted that compares the effect of the intervention with and without that specific component.

Dr. Nahum-Shani said that if an investigator is unsure about more than one component of the intervention, the next question to answer is: Is the investigator unsure about the timing of the intervention? Issues regarding timing could include when to deliver or not deliver a specific component of an intervention, the best time for delivering a component, and which components should be offered at which times. If the investigator has confidence about the timing of the intervention components, a factorial design for the study is a good option. Factorial designs are randomized trials that include more than one factor, and the levels of each factor cross with the levels of the other factors to form a design with multiple experimental conditions.

If an investigator is unsure about intervention timing, Dr. Nahum-Shani said the next question to answer is: Are the intervention's components designed to address conditions that change slowly (e.g., over a few weeks or months) or rapidly? Slowly changing conditions are addressed through adaptation, which involves collecting ongoing and changing information from an individual in an attempt to decide if components should be modified and how they should be modified. Interventions that use this technique are called adaptive interventions. Adaptive interventions are designed to achieve both long-term (distal) and short-term (proximal) outcomes. The proximal outcomes are pathways (mediators) through which the distal outcome can be achieved.

Dr. Nahum-Shani said that if the investigator wants to address conditions that change slowly, a sequential, multiple assignment, randomized trial (SMART) is an appropriate design. In a SMART, stages of randomization correspond to different points in time for introducing or adapting the intervention components. Previous research has shown that SMART designs can be highly efficient and help answer questions about the selection and adaptation of components designed to address conditions that change relatively slowly over time.

If an investigator wants to address conditions that change rapidly, Dr. Nahum-Shani suggested using a micro-randomized trial (MRT) design. Sense2Stop, a smoking cessation intervention that continuously monitors a person's physiology to detect stress, is an example of a study with an MRT design. When the monitoring device detects stress, the participant receives an intervention. MRTs used for rapidly changing conditions also involve adaptation, although the adaptation occurs rapidly and frequently. This type of adaptive intervention is called a "just-in-time" intervention. MRTs involve sequential randomizations in which each participant is randomized between the intervention options at each decision point. Each person may be randomized thousands of times in the trial.

Dr. Nahum-Shani commented that scientific questions should motivate an experiment's design; the design should not dictate the study's objectives. Her framework helps investigators match the design of a study to the scientific questions they want answered.

Using Systems Science for a Multifaceted Multioutcome Whole-of-Community Intervention To Prevent Childhood Obesity

Ross A. Hammond, Ph.D., The Brookings Institution

Dr. Hammond said complex systems science is a set of tools developed from a variety of quantitative and qualitative perspectives. It can be used to design intervention studies that are multifaceted and have multiple outcomes. This method is primarily used in the social and behavioral science fields, rather than for the biological sciences.

Dr. Hammond summarized his research, which examined an intervention to prevent childhood obesity, a complex problem. The system that drives obesity involves multiple pathways and determinants, many measurable factors, much heterogeneity, interdependent variables, and many causal relationships among variables. Obesity also involves adaptation over long periods of time, so the sustainability of any intervention needs to be considered.

Dr. Hammond said that the challenges associated with obesity create problems if a researcher attempts to use conventional statistical or analysis tools such as randomized control trials. The obesity field has accelerated use of tools from complex systems science. Three key themes for effective obesity prevention interventions have been identified. The first theme is that obesity is a systems problem, and interventions need to have multiple components, levels, and strategies to address the many different pieces of the problem.

For the second theme, Dr. Hammond said that previous research has shown that a community-level approach is an effective method for addressing the problem. Community-level interventions have enough power to change physical environments, but the scale is small enough that policies and strategies can be coordinated across sectors. The third theme is that interventions need to be tailored to the context—what works in one place may not work in another place. No single solution or approach fits all individual circumstances or contexts. A precision prevention approach uses tools from precision medicine and techniques from the underlying science to effectively prevent chronic disease by tailoring to context, including the context of individual biology.

Dr. Hammond summarized the NIH-funded Childhood Obesity Modeling for Prevention and Community Transformation (COMPACT) study. The study investigated upstream implementation and how to create lasting change in a community's policies, practices, and environment. The COMPACT study examined previous research on community-level interventions to learn which aspects were successful and which were not. Most of the successful interventions included a group of community stakeholders that designed and promoted the intervention. The COMPACT study developed a Stakeholder-Driven Community Diffusion theory, which proposes bringing together the right people, having the right conversations, and reaching the right groups to develop a workable, effective, community-level obesity prevention intervention.

Dr. Hammond said that the COMPACT study developed measurement tools to operationalize this theory. It developed computational predictive models to help with a network analysis to select the right people for engagement and the right kinds of activities. The COMPACT study included multiple intervention elements, methodologies, and outcomes. It examined how individual characteristics are distributed across a mathematical network structure within a community. The study used a variety of tools, such as group model building, systems mapping, and social network analysis. The study tested its models using existing data sets and new data collected with its new, customized tools.

Dr. Hammond noted that the approach developed in the COMPACT study incorporated the importance of local context and tailoring intervention elements to those contexts. Change agents who could catalyze

community action were identified. Those agents focused on the network properties of individual settings and on distribution across those networks. Dr. Hammond said this approach can be used in social science or biological settings and can be used to examine treatments or prevention. He added that complex systems tools can be used to help investigators design studies, analyze the power of an approach, and make testable predictions about causality.

Precisely Practicing Medicine From 700 Trillion Points of Data
Atul Butte, M.D., Ph.D., University of California, San Francisco

Dr. Butte summarized the characteristics and reach of the University of California and the University of California Health (UC Health) system. He said the system acts as one health enterprise and shares data, including electronic health records. Data from the six academic health centers in the system are collected daily and transferred to a central data warehouse. The system has collected almost 10 years of longitudinal data on many patients. More recent data have been collected from more than 7 million patients. In addition to collecting health-oriented data, the system collects claims data and location data and manages the death index for the state of California.

Dr. Butte shared an area deprivation index map of California, which depicted socioeconomic status at the neighborhood level based on income, education, employment, and housing quality. He noted that location data are valuable when examining social determinants of health. The index is significantly associated with adverse health outcomes in patients. Area deprivation index data can be combined with physiological data collected from patients. Research has shown that even when factors such as age, sex, and race and ethnicity are controlled, patients who live in the most disadvantaged areas have the worst health outcomes.

Dr. Butte said this type of data is called "real-world data" and has been receiving a lot of attention. He shared a list of 21 uses for real-word data, including evaluating the safety of drugs after they receive U.S. Food and Drug Administration (FDA) approval, informing the design of clinical trials, comparing cost effectiveness of interventions, and evaluating the quality of medical care.

The data sets in the UC Health system range from fully identifiable data to legally deidentified data. To use these data, researchers need to be affiliated with the University of California or partner organizations, and a data use agreement must be signed. The data are maintained and executed on the Cloud and cannot be downloaded, although scripts can be uploaded and run.

Panel Discussion

Dr. Kim began the discussion portion of Session Four and asked Dr. Knight for his responses to the discussion questions.

- What methodologies can be used in preclinical models, human subjects, or both for studying the impact of multicomponent interventions or therapeutic systems on multisystem or multiorgan outcomes, and what are the advantages/strengths/pros (e.g., ability to assess temporal dynamic range and responses) and limitations/cons (e.g., some approaches are considered fishing expeditions; need to adjustment for multiple comparisons, lack of causality analysis of machine learning) of these methodologies?
- What methodologies can be commonly used in both preclinical and clinical studies to study the impact of multicomponent interventions or therapeutic systems on multisystem or multiorgan

outcomes, and which types of data can be captured in both preclinical and human subject studies?

- Which computational and analytic methods are unique to preclinical studies vs. human subject research to study the impact of multicomponent interventions or therapeutic systems on multisystem or multiorgan outcomes, and what types of data can be captured in preclinical models or human subjects uniquely?

Dr. Knight said that all of the methodologies he mentioned in his presentation can be applied to preclinical models or human subjects, with the exception of the whole-body metabolomic and microbiome studies, which require dissection of the animal. He said he has had a lot of success generalizing microbiome findings between humans and animals for conditions such as inflammatory bowel disease. He said his investigators are still working on translating temporal and spatial scales, especially between mouse models and humans. For example, in a mouse circadian study, does the time scale of the intervention need to be adjusted? Understanding differences in time scale between species has been a major challenge for increasing generalizability.

Dr. Shinto said the UC Health system has large integrative medicine centers. She asked Dr. Butte if data on integrative medicine use could be separated for analysis. Dr. Butte said yes, but integrative therapies do not have standardized nomenclature or National Drug Code (NDC) numbers, making it a challenge to harmonize similar data across institutions within the system. Also, those records would only include interventions that were ordered by health professionals, not self-reported therapies.

Providing an example, Dr. Butte said for therapeutic use of marijuana, the drug may not have an NDC, and the vendor and manufacturer might be unknown. Harmonizing those types of data is a challenge, but it is also an opportunity to advocate for national standards for therapeutic interventions so large data sets can be built. Some self-reported patient information may be captured in a practitioner's clinical notes, but translating the notes into usable data is another challenge that the University of California, San Francisco has been working to resolve. A methodology for that project has been approved, and the university is ready to scale the effort across the entire UC Health system.

Dr. Weber commented that eventually behavioral or other types of interventions will be prescribed and be in electronic health records. Dr. Butte said about 20 of those interventions have been approved by the FDA and can be prescribed, but they still have not been standardized in electronic health records. If practitioners prescribe an app, they have no way to track it and generate real-world evidence for it.

Dr. Kim asked Dr. Nahum-Shani if SMART and MRT designs could be used to examine the effect on multiple outcomes. Dr. Nahum-Shani said yes, typically a SMART or MRT is used with a single outcome, but various methods could be used in the planning and analysis of the design to incorporate multiple outcomes. A composite outcome could be used, but then assumptions would need to be made. Also, significant effects can be washed out if a composite outcome is used. Weighting could be used, but that requires prioritizing outcomes and establishing the weights, although those potentially could be extracted from existing empirical evidence. Each solution has advantages and disadvantages. Dr. Nahum-Shani recommended choosing a desired outcome based on the ultimate goal and planning the trial for that outcome. Other, less important outcomes could be included in secondary or exploratory analyses.

Dr. Kim asked Dr. Butte about the need for data curation in the UC Health database and the levels of effort needed to clean data so they can be used effectively. Dr. Butte responded that medicine is a messy world and electronic health records accurately capture that mess. The medical records will have contradictions because that is how medicine is practiced. Trivial mistakes, such as data entered into the wrong field, can be fixed. Also, computational tools can be used to extract data from clinical notes. Curation tasks can be prioritized based on need. For example, before COVID, data on extracorporeal membrane oxygenation was not a priority and had not been harmonized across UC Health institutions. COVID changed that, and the institutions worked together to harmonize those data elements quickly.

Dr. Yu asked Drs. Schork and Nahum-Shani for their comments on study design choice for hypothesis generation versus hypothesis testing. She also asked if they expected to see applications of N-of-1 and MRT methods for behavioral or mHealth interventions. Dr. Schork responded that N-of-1 and aggregated N-of-1 trials could be used for almost any intervention, including preclinical studies. Multiple N-of-1 trials of an intervention could be conducted in locations with different socioeconomic environments, and the differences in response might be attributable to a social setting. Or, the responses could be factored into a meta-analysis. As long as enough data on any one individual are generated to state with confidence that the individual is a responder or nonresponder, that N-of-1 trial can be included in an aggregate analysis. However, in most traditional clinical trials, not enough data are collected on any one individual, so the investigator cannot unequivocally establish if a person is a responder or nonresponder.

Dr. Nahum-Shani agreed and added that N-of-1 and MRT designs traditionally answer different types of scientific questions. N-of-1 trials are similar to standard randomized trials, and they are useful because the person acts as a control and fewer participants need to be recruited. MRTs are used to answer questions about sequences and treatments. Studies that use MRTs often involve scientific questions about carryover effects, an intervention's changing effect over time, or the constancy of an effect. Those questions could not be answered with an N-of-1 trial, because its placebo periods require a washout of the intervention.

Dr. Chen asked for clarification about the washout periods in an N-of-1 trial. For many interventions, researchers investigate both short-term and long-term effects. For example, a study of a meditation intervention may investigate what happens in the body during the meditation practice and how the body's response to meditation may change after 4 to 8 weeks of meditation training. She asked if an N-of-1 trial could be used for fast, online, mechanistic analyses.

Dr. Schork said yes, washout periods are not always appropriate in N-of-1 trials. For example, in an investigation of medications for a person with a severe, acute condition, there is no time to test one medication at a time with washout periods. He added that sequential regression procedures can be used to do real-time tests to make decisions about which subset of factors in the therapeutic construct influence a person's response. However, with that method, a way to quantify the contribution of the different factors is needed. For example, therapeutic drug monitoring studies use this method for drugs that have a small therapeutic window for dose. In those studies, the level of the drug in an individual's plasma is measured, not the amount of drug administered, and those levels are correlated with the individual's response. Those kinds of studies can be used for real-time assessments.

Dr. Kim asked Dr. Hammond for his recommendations for building multicomponent interventions at the community level using the methods he described. Dr. Hammond said that for obesity prevention interventions, the COMPACT study developed a package of tools that outlines the method for identifying

the most important aspects of the social context, measuring them, and using that information to select nodes in the network and members of the community. Specific sets of processes catalyze change in that community by leveraging resources already present. He said that for other kinds of multicomponent interventions, the agent-based modeling technique and the computation simulation models he described earlier are very different ways of conducting scientific research, but they are useful for extrapolating beyond where data can be collected for practical or ethical reasons.

Dr. Weber asked Dr. Hammond for his suggestions for how researchers could find people who can do that kind of modeling or who might have those kinds of data. Dr. Hammond responded that finding people with the right technical skills to do the modeling is more important than finding the right kinds of data. He said that training in complex systems methods, including agent-based modeling, is now part of the accreditation process in schools of public health. Several clearinghouses have been created that connect people with topical domain questions expertise with people who have modeling skills. NIH and the Centers for Disease Control and Prevention in the past offered a summer training institute to give researchers exposure to these tools to learn enough to work with an expert, although that has been discontinued. Dr. Hammond said he teaches a similar type of class at Washington University in St. Louis, and there is another similar class in Michigan.

Dr. Kim asked Dr. Hammond about tailoring and context in the COMPACT study during development of the intervention and its implementation in the various sites. Dr. Hammond said much research has shown that context is important in chronic disease prevention. For example, methods that may work in a rural setting will not work in an urban setting. To understand how to translate a successful intervention from one context to another, researchers need the ability to measure aspects of context. He said he has designed some custom network science tools and survey tools to help researchers understand how individual characteristics are distributed across a community network. Researchers can use simulation models that will explain how an intervention needs to adapt to work equally well in a different environment.

Dr. Kim asked all the panelists if the proposed tools and frameworks will help researchers understand systems. Dr. Knight said that including microbes in studies will provide a lot of information about a system. Studying microbes and the co-metabolism performed by the host and the microbe together has been useful for the formulation of mechanistic hypotheses that can be tested in animal models. Dr. Schork said the right measurements and the right kind of modeling can help researchers understand how systems works, but the investigations come with large logistical and financial challenges. Dr. Hammond said the answer to that question depends on how the system boundary is defined. Attempting to understand an entire system all at once is too challenging. A better approach would be to build pieces of knowledge in a modular way and then connect them to create a patchwork.

Dr. Butte commented that society is a moving target. Modeling a system as it exists today may become irrelevant as the system's environment changes. Dr. Nahum-Shani said understanding how multiple components affect a whole system is a fantastic goal, but a more important goal is engaging people cognitively and physically with interventions that will improve the system.

Dr. Kim asked Dr. Schork if he has used any methods that examine multiple or global outcomes. Dr. Schork said yes, analyses can have multiple dependent variables and multiple outcomes, but the statistical techniques are more complicated. He provided an example of an N-of-1 study that originally had multiple outcomes and benefited the participant. The results analysis, however, was complicated by

another variable—the patient's active engagement in the intervention. Using a bottom-up approach in a study may give researchers the ability to gain insights about individual patients that could benefit them.

Dr. Badawi commented that Dr. Schork's approach is similar to the practice of precision medicine rather than a study. Dr. Schork agreed that N-of-1 studies are designed to benefit an individual person through objective methods. However, logistics and cost do not allow for conducting N-of-1 studies on everyone with a health issue.

Roundtable Discussion II
Moderators: Craig Hopp, Ph.D., NCCIH, and Wendy Weber, N.D., Ph.D., M.P.H., NCCIH

Ms. Law introduced Drs. Hopp and Weber. Dr. Hopp presented the key issues for the roundtable discussion:

- Social determinants and health disparities
- Application of methodologies from other fields
- Gaps and challenges in the current methodologies
- Opportunities for innovation and further advancements in computational and analytic methods, data collection, and related technologies

Dr. Hopp briefly introduced Dr. Scott Mist.

Systems Sciences and Whole Person Research Methodologies
Scott D. Mist, Ph.D., M.Ac.O.M., Oregon Health & Science University

Dr. Mist presented a graphic of the different tools within systems science that can be used to model complex systems. He said systems science is methodological design analysis of complex systems. Complex systems have feedback loops in which the system's output is used as an input for future operations. Complex behaviors emerge from positive and negative feedback loops. In a system with multiple feedback loops, making a change in one of those loops may or may not make a change in the overall system.

Dr. Mist described the reconstructability analysis methodology called discrete multivariate modeling (DMM). DMM is an extension of log-linear methods. This method can be used for fuzzy systems, set theoretic models, information models, neutral systems, directed systems (i.e., systems with independent and dependent variables), and time-dependent systems. DMM overlaps with log-linear methods, but it is useful for large sets of models. DMM also can be used for nonlinear systems.

Dr. Mist described using a dual diagnosis within whole systems research. A dual diagnosis may be needed in a study to apply the results to both Western medicine and an allopathic field, such as traditional Chinese medicine. Using a dual diagnosis places greater importance on the reliability of the diagnoses.

Dr. Weber presented the research framework from the National Institute on Minority Health and Health Disparities and introduced Dr. Irene Headen.

Methodological Considerations: Incorporating Social and Structural Determinants of Health Into Research

Irene Headen, Ph.D., M.S., Drexel University

Dr. Headen said the drive to generate actionable evidence on the social determinants of health has steadily evolved. She provided the World Health Organization's definition of social determinants of health: the conditions in which people are born, grow, work, live, and age, and the wider set of forces and systems shaping the conditions of daily life. She said that identifying structural determinants of health is key to advancing health equity research. Structural factors of health include the policies, practices, cultural norms, and institutions that define the distribution of the social factors of health.

Dr. Headen said that operationalizing social and structural determinants requires weighting factors such as level of influence, time, multidimensionality, and interconnectedness. In her research, Dr. Headen said she uses longitudinal cohort studies to understand how neighborhood deprivation is embedded and accumulates within individuals and populations over time to affect health outcomes. She said she uses measurements and databases that can be integrated and layered to create a comprehensive typology of neighborhood context. She also has used latent class analysis as a clustering technique to identify neighborhood typology.

Dr. Headen said she has used multilevel models to understand the interconnectedness and dynamic nature of how neighborhoods reproduce structural inequities for populations over time. She added that methods that examine the research process are integral to rigorous science in the area of social and structural determinants of health. Researchers need to have awareness of how societal and historical contexts shape science, who conducts the research, and the implications for public health practice.

Dr. Weber presented the roundtable discussion questions and asked all the workshop speakers for their thoughts.

- How do the social determinants of health disparities (e.g., race/ethnicity, socioeconomic status, place) enter into the study of multicomponent interventions and their impact on interconnected multisystem outcomes?
- What are limitations in data collection/measurement that could lead to bias in the application of the methods to study multicomponent intervention impact on interconnected multisystem outcomes in underrepresented populations. What kind of data collection and data analysis techniques minimize bias to ensure equitable whole person research?

Dr. Headen said that to fulfill the promise of a whole person research agenda, research needs to be translatable to the real-world context of populations that have a disproportionate impact of health burdens. Structural barriers are going to be present, but researchers need to ask how they can optimize their studies and how they can address the contexts of structural determinants and structural racism or sexism. She said that researchers have to work with communities in a nonpaternalistic way, because people in those communities have established responses to the social and structural determinants.

Dr. Windsor said the political dialogue component of her intervention was designed to address racism, classism, and sexism as direct influencers of the target outcomes, including substance misuse and relationships with others and the community. A component of the intervention is a capacity-building project to increase health equity, in which participants engage in community organizing and activities they identify as problems. However, optimizing the intervention has methodological limitations because it is designed for substance misuse, and the capacity-building aspect may not translate at the

community level for other outcomes. An auto-optimization study could be used for that particular process.

Dr. Mist said that medicine can only affect 10 to 20 percent of the changeable outcomes for a person. Researchers need to find ways to bring communities into their studies because the community is an area in which most lives can be changed.

Dr. Hopp asked the speakers to comment on methodologies from other fields and the gaps and challenges of the current methodologies. He asked where NCCIH could invest to bring about the most change.

Dr. Schork said that trial design is an area that needs innovation. He suggested developing platform interventions. If a researcher has profiles for groups of patients, the groups could be categorized by likelihood of responding to specific treatments. Dr. Mist agreed that study design is an important area of innovation. He said most designs make assumptions of linearity, and researchers need to focus on how to design studies that address nonlinearity, nonlinear grouping, grouping that can have partial membership, simple behaviors that have complex outcomes, and agent-based modeling. He said those aspects of design are areas of growth.

Dr. Headen commented that the field needs to create infrastructures that permit bridging across different disciplines. Cross-disciplinary methodological innovation needs to be encouraged early in researchers' careers.

Dr. Windsor said that to foster innovation, funding for methodologists is needed to develop solutions to some of the challenges posed at this workshop. A paradigm shift away from traditional scientific perspectives of individual-level outcomes and randomized controlled trials is needed. Dr. Hopp asked the other speakers to share their thoughts about the traditional NIH paradigm for research. He suggested that perhaps investigators believe that paradigm is required for rigorous research. He asked how whole person research could be conducted in a nontraditional way but still be rigorous. Dr. Windsor responded that NIH can take a leadership role through its peer reviewers and its scientific community and emphasize that other perspectives are important and need to be equally valued.

Dr. Hammond commented that all the tools discussed at the workshop have been used in other fields, and lessons can be learned about how those tools permeated those fields. NIH has fostered the development and propagation of tools to get them to pass review panels. He noted that researchers tend to separate the biological from the social and behavioral, but the problems people care about almost always transcend those areas and involve mechanisms that cross or bring together levels of scale.

Dr. Shinto said that NIH could help by funding feasibility studies. Dr. Hopp said that NCCIH wants to learn from its studies to inform future studies.

Dr. Nahum-Shani responded to the comment that the methods discussed at the workshop are not new. She agreed but said the methods have not been well-disseminated. Methodologists could help by releasing information about their methods in a way that others can use them without assistance. For example, code, software, and documentation should be readily available to the broader community.

Dr. Jensen commented that review panels are inherently conservative. To encourage investigators to try nontraditional methodologies, the requests for applications (RFAs) might need to be structured very specifically so applications could get through the study section. A highly specific RFA might force reviewers to accept nontraditional methodologies.

Dr. Schork said he would like to ask the speakers and panelists which measure could represent a whole person phenotype. How complex would that measure need to be to capture whole person health?

Dr. Gupta commented that the variability in genetics and the microbiome are very relevant in a global context. Researchers in other countries and cultures may employ different methodologies. He asked the speakers and panelists to comment on the role of global cooperation in whole person research. Dr. Windsor said that global research cooperation is critically relevant. The COVID-19 pandemic has demonstrated the importance of global interconnectedness and health issues. Dr. Headen said cross-national studies can increase understanding of differences within populations. Researchers need to find ways to leverage global research to understand outcomes and the evolutions of those outcomes. In an increasingly global society, everything is codependent and interrelated.

Dr. Weber responded to the comment about developing a metric that could be developed to measure whole person health. She said that Dr. Moffit's research created outcome measures for the pace of aging. Dr. Weber suggested that pace of aging might be a measure of whole person health. She asked the speakers and panelists if a summary measure that crosses different outcomes would be helpful. Dr. Headen said yes, although researchers need to be careful not to re-create biases that impede effectiveness across populations. Researchers need to have flexibility when developing such a measure and ensure it answers the questions being asked. Efforts would need to be made to work with communities to help them understand how that measure translates to health. Dr. Windsor commented that some outcomes, such as poverty or excess food, are known to affect people at the individual and macro levels.

Dr. Nahum-Shani said she agrees with the concept of flexibility and suggested that the proposed funding mechanism be flexible as well. Novel scientific questions and novel interventions can be developed using standard experimental approaches and methodologies. Randomized controlled trials can be used to answer novel questions. The novelty of the scientific questions should be emphasized first, and the methodology should follow to answer those questions.

Dr. Weber briefly introduced Dr. Bruce Lee.

Workshop Synthesis: Whole Person Research Methods
Bruce Y. Lee, M.D., M.B.A., City University of New York

Dr. Lee summarized the highlights of the workshop, beginning with a definition for whole person health and a brief discussion of Dr. Langevin's opening remarks and NCCIH's strategic plan. He said a theme presented throughout the workshop was that focusing on a single disease, condition, or body part will not solve the major health problems in the United States. Investigators need to understand that everything is connected, and their research needs to focus on the health of the whole person.

Dr. Lee said another theme mentioned throughout the workshop was that correlation does not equal causation. Correlations reflect only a superficial layer of a deeper, complex system that might be affecting many types of outcomes. People are complex systems, and they live in complex systems.

People have individual behaviors, but they are affected by many social, environmental, and physical factors.

Dr. Lee reviewed some of the problems associated with developing interventions without using a systems approach: The intervention may be a temporary solution, be unsustainable, ignore follow-up effects, have unintended consequences, and waste time, effort, and resources.

Dr. Lee said that researchers need to collaborate and share information with other fields, but having a lot of data can be challenging because those data need to be curated, organized, and analyzed in a useful way. Failing to understand the complexity of data can lead to bias and inaccurate conclusions.

Dr. Lee said that another common theme from the workshop was rethinking approaches to clinical studies. Standard research methods may not be useful in all situations. Studies that have multiple components and outcomes may require more complex methodologies. Some studies require a top-down approach that examines patterns and associations, but others may need a bottom-up approach that examines the mechanisms that make up the system. Computer modeling can be used to help understand and address complex systems. Dr. Lee provided an example from his research on the impact of warning labels on sugar-sweetened beverages. Computer modeling of different cities helped his team understand why the impact of the warning labels varied across cities. In another example, a computer model of a virtual infant revealed that formula-feeding recommendations were not healthy for all infants because the infant population is too heterogeneous.

Dr. Lee noted that research on complex systems requires specialized methods and approaches. No single method or approach will apply to multiple systems or resolve all the issues within one system. However, different methods and approaches can be used together to help uncover different parts of the system. Maps and models help investigators see how the parts of a system fit together. Models can be used to help plan clinical studies. The results from those clinical studies can be used to update the map and model and inform future clinical studies. Eventually the knowledge of the different parts of the system can be connected to improve understanding of the entire system.

To better leverage different types of methods, Dr. Lee suggested using a hybrid connectors, glue, and mortar approach. This approach involves breaking down the separation of research fields and finding people, organizations, and even funding mechanisms that can serve as connectors across disciplines. These connectors work together as the glue to holistically address research questions.

Dr. Lee concluded by saying that researchers investigating whole person health can coexist because there is a role for everyone. Investigators need to receive training in some of the nontraditional methods and approaches. To advance the whole person research field, people and organizations need to understand that systems are complex, and they need to work collaboratively to learn about those systems and develop helpful and equitable interventions.

Closing Remarks
Helene Langevin, M.D., Director, NCCIH

Dr. Langevin thanked Dr. Lee for his overview of the workshop and asked the panelists and speakers if they had any questions. She said this workshop met and exceeded NCCIH's expectations. She thanked the workshop planning committee, the NCCIH communications and information technology (IT) teams, and the speakers.

Dr. Langevin said that the panelists and speakers rose to the challenge with their engaging presentations and their engagement with each other. The workshop was designed to bring people from different fields together. Dr. Langevin said she hopes cross-fertilization will result from the discussions and that a scientific community will grow around the theme of whole person research. She said every presentation included elements of integration and synthesis, and every speaker showed a determination to address complex topics. Dr. Langevin referred to a comment from Dr. Hammond that researchers need to place boundaries around their research questions even when moving the research toward integration. Entire systems cannot be integrated in one study; a step-by-step approach is needed. Eventually, however, researchers should aim to complete the picture as much as possible.

Dr. Langevin said that NCCIH will be producing a summary of the workshop, which will be posted on the NCCIH website. The NIH VideoCast also will be available on the website. She said she is looking forward to what happens next.

Ms. Law thanked everyone and reminded the viewers that the workshop was recorded and would be available on the NIH VideoCast website.