CRITICAL REVIEWS IN FOOD SCIENCE AND NUTRITION
https://doi.org/10.1080/10408398.2025.2461237

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS

Check for updates

Personalized nutrition: perspectives on challenges, opportunities, and guiding principles for data use and fusion

Sharon M. Donovan[a], Mariette Abrahams[b], Joshua C. Anthony[c], Ying Bao[d], Maribel Barragan[e], Kate M. Bermingham[f], Gil Blander[g], Anna-Sigrid Keck[e], Bruce Y. Lee[h], Kristin M. Nieman[c], Jose M. Ordovas[i], Victor Penev[j], Machiel J. Reinders[k], Kris Sollid[l], Sumeet Thosar[e] and Barbara L. Winters[c]

[a]Personalized Nutrition Initiative, Carl R. Woese Institute for Genomic Biology, and the Department of Food Science and Human Nutrition, University of Illinois Urbana-Champaign, Urbana, IL, USA; [b]Qina, Ltd, Olhao, Portugal; [c]Nlumn, LLC, Princeton, NJ, USA; [d]Gies College of Business, University of Illinois Urbana-Champaign, Urbana, IL, USA; [e]Personalized Nutrition Initiative, Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA; [f]Zoe Ltd, London, UK, Department of Nutritional Sciences, King's College London, London, UK; [g]Segterra (InsideTracker), Cambridge, MA, USA; [h]Artificial Intelligence, Modeling, and Informatics for Nutrition Guidance and Systems (AIMINGS) Center, PHICOR, and the Center for Advanced Technology and Communication in Health (CATCH), at the City University of New York (CUNY), Graduate School of Public Health and Health Policy, New York City, NY, USA; [i]Nutrition and Genomics Laboratory, JM-USDA-Human Nutrition Research Center on Aging, Tufts University, Boston, MA, USA; [j]Edamam, LLC, New York, NY, USA; [k]Wageningen University & Research, Wageningen Social & Economic Research, the Hague, the Netherlands; [l]International Food Information Council, Washington, DC, USA

ABSTRACT

Personalized nutrition (PN) delivers tailored dietary guidance by integrating health, lifestyle, and behavioral data to improve individual health outcomes. Recent technological advances have enhanced access to diverse data sources, yet challenges remain in collecting, integrating, and analyzing complex datasets. To address these, the Personalized Nutrition Initiative at Illinois organized a workshop titled "Personalized Nutrition Data: Challenges & Opportunities," which gathered experts to explore three essential data domains in PN: 1) health and biological, 2) social, behavioral, and environmental, and 3) consumer purchasing data. Discussions underscored the importance of cross-disciplinary collaboration to standardize data collection, enable secure data sharing, and develop data fusion techniques that respect privacy and build trust. Participants emphasized the need for representative datasets that include underserved populations, ensuring that PN services are accessible and equitable. Key principles for responsible data integration were proposed, alongside strategies to overcome barriers to effective data use. By addressing these challenges, PN can enhance health outcomes through precise, personalized recommendations tailored to diverse population needs.

KEYWORDS

health data; behavioral data; consumer data; healthy equity; data privacy

Introduction

Personalized nutrition (PN) aims to tailor dietary guidance based on measures of health status in combination with their exposome, abilities, and preferences to help an individual change dietary and other lifestyle behaviors to achieve a health or functional benefit. PN often employs a multidisciplinary approach, integrating insights from nutrition science, clinical practice, genomics, epigenetics, systems biology, medicine, and behavioral science (Rozga, Latulippe, and Steiber 2020). Studies show that PN can help individuals with motivation, behavior change, and the attainment of health outcomes; however, more research is needed to demonstrate that PN is more effective than public health recommendations (Adams et al. 2020; Jinnette et al. 2021).

Despite the promise of PN, a significant challenge is data management. This includes effectively gathering, standardizing, integrating, securing, and analyzing the vast array of complex, interrelated data from both populations and individuals, which is essential to prioritize evidence-based health outcomes in PN. Harmonization of individual's data is required to support meaningful health recommendations that consider the context of that person's biology and lifestyle. The latter is critical as preferences and behaviors must be considered to provide recommendations and strategies that support behavior. Combining these strategies is needed for PN to support desired health outcomes. Technology and social media have greatly enhanced user data acquisition and provide a wealth of information relating to health status and lifestyle habits, social and behavioral information and purchasing habits. Each of these areas has important implications for designing PN programs that are engaging and effective. However, questions remain with respect to how to effectively and responsibly enable data fusion (integrating multiple data sources including information and knowledge

CONTACT Sharon M. Donovan sdonovan@illinois.edu

to enable reliable, precise, and valid results) across these areas, what expertise is needed, and what the implications are for safeguarding privacy (Shaik et al. 2024).

The rapid advancement of artificial intelligence (AI) and machine learning (ML) has greatly increased the speed, extent, and depth at which data can be analyzed, integrated, and leveraged. However, consistent data structuring is lacking across or even within sectors. There can be tremendous variability in the complexity, comprehensiveness, reliability, validity, and transparency of different AI and ML algorithms, models, and tools. The application of AI and ML to develop recommendations for PN requires cross-functional expertise to develop accurate algorithms, models, and tools, choose and organize the data sources, and generate appropriate analyses. If used appropriately, AI and ML can greatly facilitate PN. If used inappropriately with the wrong or misleading approaches, it could hinder PN. What will be important is to take a systems approach (consider multiple elements and data sources) to account for the complex systems that affect and are affected by nutrition (Kaplan et al. 2013; Bruce Y. Lee et al. 2021; van Ommen et al. 2017).

A major challenge is data access and sharing. Few organizations make their data publicly available. The 2023 State of Open Data global survey revealed that while many researchers support making research data openly available, they often lack the necessary resources and recognition for sharing their data (Science et al. 2023). Furthermore, private industry collects extensive data on their consumers' social, behavioral, lifestyle, and purchasing habits. However, this data is considered an important source of competitive advantage and is rarely made public.

To address opportunities and challenges related to the collection and use of data for PN programs, the Personalized Nutrition Initiative at the University of Illinois Urbana-Champaign held a virtual workshop on Aug 21, 2023, entitled, "Personalized Nutrition Data Challenges & Opportunities". Ten invited speakers recorded presentations focused on three key data sources that can inform PN recommendations: 1) health and biological data, 2) social, behavioral, and environmental data, and 3) consumer purchasing data (Appendix A). Each of these data types can support the PN system, which includes the collection and analysis of individual data, delivery of recommendations to the user, assessment of progress, and ongoing adjustments to recommendations to continue the cycle to improve health and functional outcomes (Donovan et al. under review; Figure 1). Presenters and discussants attended a live virtual session featuring a speaker question and answer session, followed by breakout sessions on each data type.

In this paper, we review how these data sources, which often overlap, can more effectively be collected and used to inform PN. We explore the challenges that prevent data in each sector from being more impactful and discuss the principles that support data fusion across these sectors while ensuring privacy protection. Workshop participants developed guiding principles for the future (Table 1). The workshop was an initial step toward realizing the long-term goals of collecting and using diverse

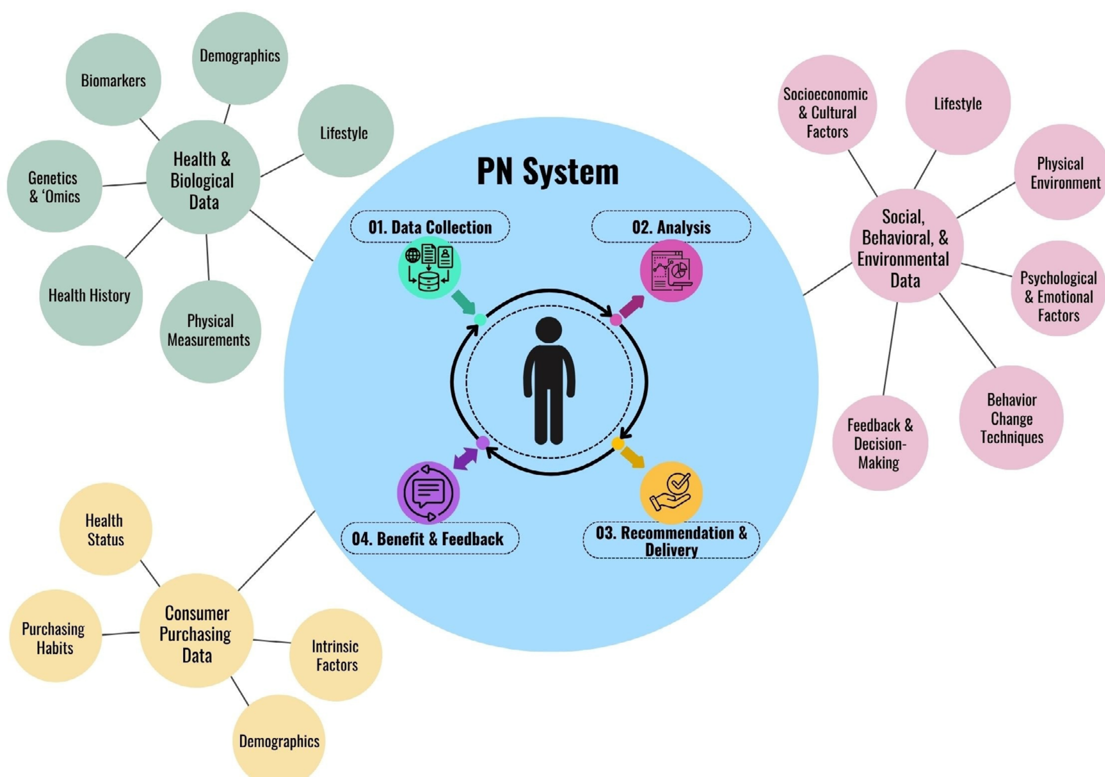

Figure 1. The main categories of data used in the personalized nutrition system. The primary data categories that feed into the PN system include health, biological, social, behavioral, environmental, and consumer purchasing data. Each category can be broken down into types which are exemplified here. Abbreviations: PN, personalized nutrition. Figure created using canva.com.

sources of data for developing and delivering effective PN to users and patients.

Health and biological data in PN

Precision medicine, which encompasses precision and/or personalized nutrition, seeks to tailor treatments to a subpopulation who have a common susceptibility to a particular disease or similar response to a particular drug (McGrath and Ghersi 2016). Precision medicine considers variability in genetic, socio-environmental, and lifestyle factors among subpopulations to propose precise therapies (Naithani et al. 2021). Health and biological data include information, such as demographics and lifestyle factors, as well as biomarkers, such as physical or anthropometric measurements, genetic variation, blood or urine biomarkers, and measures of the microbiome (Table 2). Emerging research on the gut-brain axis highlights the potential role of the microbiome in influencing dietary behaviors and overall mental and physical health. For instance, the bidirectional communication between the gut microbiota and the central nervous system can affect mood, stress response, and eating patterns, which are crucial considerations for personalized nutrition strategies. While a detailed discussion is beyond the scope of this manuscript, additional information can be reviewed elsewhere (Morais, Schreiber, and Mazmanian 2021). Biomarker data is most often associated with PN programs; however, other types of health data can be widely accessible, have defined health ranges, and provide insight into individual preferences and lifestyles. These factors are important to consider when designing accessible PN programs that can support positive behavior changes. Therefore, we consider various types of health and biological data below.

Demographic information provides a foundation for both public health recommendations and PN programs (Table 2). Key demographic variables include age, biological sex, ethnicity, and socioeconomic status (Alkerwi et al. 2015). These factors influence dietary requirements and recommendations, are associated with disease risk, and support the effectiveness of

Table 1. Guiding Principles for the integration and use of data in personalized nutrition.

1. Define the data architecture to standardize data collection, storage, sharing, and fusion	Use of multiple sources of data to support PN requires a consistent data management framework. A set of models, policies, rules, and standards governing data interoperability, management, and storage need to be established within an organization.
2. Employ a systems-based approach to identify cause-and-effect relationships	Effective integration of diverse data sources in PN requires a cross-disciplinary collaboration to ensure meaningful flow between systems for comprehensive analysis, including AI and ML, and adaptation to changing needs.
3. Incentivize data sharing	Establish specific rules or permissions with respect to data use or sharing. Identify terms for profit sharing from shared data discoveries.
4. Guarantee and ensure that control of individual data is maintained by the individual	PN users should have the right to access, correct, delete, and share personal data in a secure setting. The data should be governed by easy-to-understand agreements which outline ethical use that can be understood by an audience with no special or expert knowledge.
5. Ensure relevant data is used to create accessible and actionable PN solutions	PN solutions should be developed using data that is most relevant and impactful to help the end user achieve their health goals. This approach will make it possible to offer cost-effective solutions to allow access to all demographics of the population.

Abbreviations: AI, artificial intelligence; ML, machine learning; PN, personalized nutrition.

Table 2. Personalized nutrition data types and Examples.

Data Category/Type	Examples[a]
Health and Biological Data	
Demographic	Age, biological sex, ethnicity, and socioeconomic status
Physical measurements	Body weight, height, body mass index, hip and waist circumferences, fat and muscle mass, blood pressure, heart rate, oxygen saturation, respiration rate, and body temperature
Health history	Health background, family health history, preexisting conditions, medications, and previous medical treatments
Lifestyle	Smoking, sleep, activity (level and type), food intake, social determinants (e.g., education, work and economic stability, environment, healthcare access and quality, food access and quality, social support and justice)
Biomarkers	Lipids, triglycerides, cortisol, vitamins and minerals, hormones, glucose, insulin
Genetic and ’omics	Single nucleotide polymorphisms, posttranslational modifications, structural variants, microbiome
Social, Behavioral, and Environmental Data	
Socioeconomic and cultural factors	Social support, income, cost of living (food cost, housing, etc.), education, digital and nutrition/food literacy/skills employment, family members, social and cultural norms and expectations, spiritual/religious beliefs and preferences, time constraints, childcare access
Lifestyle	Sleep (quantity and quality), activity (level and type), food and supplement intake, food preferences, smoking, allergies, alcohol intake, meal timing
Physical environment	Living area, safety, transportation, access to grocery stores, available cooking equipment, restaurant options, health and services access
Psychological and emotional factors	Perceived motivators and barriers, stress, self-efficacy, health attitude, mental state (depression, anxiety, positivity, etc.), perceived health risk, resilience, intent, readiness to change
Behavior change techniques	Goal setting, shaping knowledge, social support, and taxonomies
Feedback & Decision-making	Actions based on recommendation/recommender systems
Consumer Purchasing Data	
Buying habits	Price, quantity, product preferences, brand, time and location of purchase
Intrinsic factors	Attitudes, behaviors, beliefs
Health status	Feelings, conditions, healthy history
Demographics	Age, gender, ethnicity, household occupants, and socioeconomic status

[a]This is intended to provide examples of the data types and is not inclusive. All data types would also potentially include email addresses, names, physical addresses, images, account numbers, internet protocol addresses, and other personal identifying information.

nutritional interventions and behavior change (Guo et al. 2022; Russell 2001). Health history information allows an understanding of an individual's health background, family history, preexisting conditions, medications, and previous medical treatments. This information helps identify potential dietary restrictions, contraindications, and areas needing special attention in the personalized plan which may require additional regulatory considerations (Donovan et al. under review). Lifestyle factors significantly influence nutritional needs and health outcomes (R. Ng et al. 2020; Serio, De Donno, and Valacchi 2023). Some key lifestyle factors, such as sleep, activity (level and type), or smoking, can overlap with social, behavioral, and consumer purchasing data.

Physical and anthropometric measurements are indicators of overall health and nutritional status. These include body weight, height, body mass index (BMI), hip and waist circumferences, body composition (fat and muscle mass), blood pressure, heart rate, oxygen saturation, respiration rate, and body temperature (Holmes and Racette 2021; World Health Organization 2008). These metrics help evaluate the impact of dietary interventions and monitor changes over time. Biomarkers are measurable indicators of biological processes that may inform health risk or status and can include a wide variety of inputs, for example, glucose, cortisol, vitamins, minerals, liver enzymes, and lipids (Picó et al. 2019; Strimbu and Tavel 2010). Genetic information, including single nucleotide polymorphisms (SNPs), post-translational modifications, structural variants (SVs), and other genetic markers can help identify individual variations that have the potential to affect nutrient metabolism, dietary responses, and disease susceptibility (Aruoma et al. 2019; Fulton et al. 2018; Karvela et al. 2024; Rodríguez-García and Gutiérrez-Santiago 2023; Wald, Law, and Morris 2002; Zhang et al. 2012). These data may support tailoring dietary recommendations to an individual's genetic profile. In PN, biomarkers and genetic material can be assessed from various biological samples including blood, urine, feces, buccal, hair, nail, and sweat. Advanced 'omics' technologies, including genomics, epigenomics, proteomics, metabolomics, and microbiomics, can be employed to generate a comprehensive profile of an individual's biological status (Chaudhary et al. 2021).

Application of health and biological data in PN: study examples

There has been a paradigm shift in how nutrition research is conducted and how data is captured. Opportunities now exist to collect data on a large scale, made possible by advances in technology and remote testing. Digital devices, such as phones and apps, clinical devices that provide real-time measurements of blood pressure, blood glucose, and heart rate, and the application of multi-omics technologies in research settings are now widely available. Data can now be collected with breadth, depth, and precision. In addition, more people are interested in being citizen scientists and providing data daily to further PN research efforts (Wiggins and Wilbanks 2019). However, one of the biggest challenges in PN is how these many factors interact and affect the ways in which individuals respond to diet. Furthermore, this information needs to be used to ensure the individual is ready, empowered, and willing to make behavior changes aligned with desired health outcomes.

By integrating these diverse types of data, PN programs can develop tailored plans that support individuals in reaching their goals and, in turn, positive health outcomes. This holistic approach recognizes the complexity of human biology and the multitude of factors influencing nutritional needs and responses. A recent example of this was a retrospective analysis of the "Season of Me" digital health program, which evaluated over 2000 participants (Zahedani et al. 2023). The program integrated data from wearables (e.g., activity and continuous glucose monitors), mobile apps, personalized recommendations, and virtual coaching. Participants showed significant improvements in weight loss, glucose control, eating habits, and increased physical activity. A randomized controlled trial assessed the impact of a DNA-based PN intervention (11 SNPs) on 148 individuals with non-diabetic hyperglycemia and showed significant reductions in fasting plasma glucose and HbA1c levels at 26 wks compared to standard UK population guidelines (Karvela et al. 2024; U.K. National Institute for Health and Care Excellence 2024). Westerman et al. (2018) analyzed biomarker data (e.g., lipids, liver enzymes, vitamins, minerals, blood cell counts, hormones) from 1,032 individuals who used a PN program that provided lifestyle recommendations based on serum biomarkers. Participation in the program demonstrated a trend toward normalization of biomarkers that were out-of-range at baseline.

In addition, there are existing and emerging large databases that are being developed, maintained, and utilized in the PN space. These databases may support an improved understanding of how individuals respond to diet and how this information can be used to inform PN programs and services. *All of Us* is a National Institutes of Health (NIH)-funded research program with the goal of accelerating research that will improve the health of Americans (Ginsburg, Denny, and Schully 2023; U.S. National Institutes of Health 2024a). *All of Us* research sites utilize the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), an open community data standard designed to standardize the structure and content of observational data to provide efficient analyses that produce reliable evidence (Observational Health Data Sciences and Informatics 2024). *All of Us* participants currently number more than 823,000, 80% of which are underrepresented in existing biomedical research. Participants answer detailed surveys, provide data on physical measurements, such as blood pressure, heart rate, weight, height, and BMI, blood or saliva and urine samples, and share data from wearable fitness devices, such as Fitbit®. The NIH Nutrition for Precision Health (NPH) program is the first large ancillary study of the *All of Us* Research Program (U.S. National Institutes of Health 2024b). The NPH program is designed to develop algorithms that predict individual responses to food and dietary patterns.

The PREDICT studies 1, 2, and 3 (Personalized Response Dietary Composition Trial) have been and are being conducted in collaboration with researchers at multiple research institutions (Zoe 2023). Over 100,000 participants have completed the PREDICT studies, which have collected an enormous amount of data to better understand the variables that affect food choices and ultimately affect health and assess the efficacy of personalized dietary advice over time. Even with the extensive collection in these databases, further advancements may include continued research collaborations and publications in peer-reviewed journals and public repositories, addition of other data sources (social and behavioral data) and implementation of industry-wide data standardization enabling greater comparison across cohorts.

Challenges in the use of health and biological data for PN

There are several challenges to the use of health and biological data in PN programs. There is a need for new biomarkers, better representation in data sets, and the ability to integrate data safely and securely from existing electronic medical records (EMR) and other health records. Several limitations in EMR data exist. First, standard EMR typically do not contain information on dietary intakes or other aspects of the exposome, which might be pertinent to PN. A second significant challenge is the absence of comprehensive and representative biomarkers that can accurately reflect an individual's nutritional and health status (Landberg et al. 2024). This gap impedes the ability to provide truly personalized recommendations. Many robust biomarkers are designed as markers for disease or disease risk. Other than glucose, most of these biomarkers have not been adapted to real-time measurement technologies to better understand how an individual responds over time or in response to dietary intake. This is further complicated by a lack of diverse representation (e.g., age, gender, ethnicity, race) in data sets and research. For example, genomic studies typically use populations of European ancestry and results may not be suitable for underrepresented populations.

Another challenge is that PN programs are desired by individuals to address performance and well-being but are often designed to target disease prevention or management. There is an opportunity to establish and evaluate new biomarkers for these outcomes (Wopereis et al. 2017). These also need to be representative of different populations and may require an integrated look at different biological systems and how they perform in response to different environmental stressors, e.g., exercise, and across ages and sexes.

Despite the willingness of people to serve as citizen scientists and share data, the integration of PN programs with EMR data raises concerns about privacy and security (Verma et al. 2018). Sensitive health information is at risk of unauthorized access and breaches, necessitating robust security measures. Ensuring compliance with regulations, such as the Health Insurance Portability and Accountability Act, is essential to protect patient data (U.S. Department of Health and Human Services 2024b). Additionally, ensuring compliance with regulations such as General Data Protection Regulation (GDPR) and the California Consumer Privacy Act can provide additional patient protections (European Commission 2018; U.S. State of California 2024). The new EU AI Act which came into force on August 1st, 2024, will further protect members of society, by holding companies who employ AI systems accountable (Future of Life Institute 2024). These regulations set stringent requirements for data protection and individuals' rights over their information, posing challenges in data management and sharing. The opportunity lies in balancing the need for comprehensive data to tailor nutritional advice with maintaining the confidentiality and integrity of personal information.

Social, behavioral, and environmental data in PN

Social, behavioral, and environmental data provide a real-life context of individuals lives, how and why decisions are made, how they use and receive feedback, and how their socioeconomic environment and other factors impact their present and future health (Short and Mollborn 2015). These factors include socioeconomic and cultural (e.g., spiritual/religious beliefs and preferences), lifestyle, physical environment, psychological, emotional, and motivational, behavior change techniques, and feedback and decision-making that impact why individuals make certain dietary and lifestyle choices (Reinders et al. 2023). This variability makes collecting and analyzing social, behavioral, and environmental data exciting and challenging (Figure 1). Numerous behavior change techniques (BCT) including goal setting, feedback, social support, and knowledge sharing are defined, but which ones work best in specific temporal and contextual settings is less clear (Michie, van Stralen, and West 2011; Villinger et al. 2019). Having organized and quality data on the multifactorial drivers of behavior change will facilitate delivering tailored advice targeting positive actions. Combined with health and biological data, social, behavioral, and environmental data can pinpoint when and how personalized advice should be communicated for lasting behavior change.

Collecting, using, and organizing, social, behavioral, and environmental data

Within the complexities of social, behavioral, and environmental data collection and application, there are two broad areas to consider: (1) an individual's inputs and the ethical processes to obtain the most accurate and complete data, and (2) identifying, organizing, and leveraging quality behavioral data to deliver beneficial behavioral change outcomes.

Collecting individual input

An individual's willingness to share their social, behavioral, and environmental data is necessary to understand their current behavior patterns and to inform recommendations and BCT. This information can include a range of survey questions such as:

- What are your barriers to change?
- What motivates you to change?
- Are you challenged in having sufficient money for your food?
- Do you eat alone or with family?
- Do you drink alcohol?
- Are you able to safely get to and from the grocery store?
- Do you have cultural food preferences?

To achieve comfort in providing such data, privacy risks must be addressed (Reinders et al. 2020). Systems for data protection, including anonymizing and securely storing data need to be in place and transparent to the individual (Stewart-Knox et al. 2015). Ethical practices on how data will be used or shared must be disclosed, and where appropriate, options for informed consent need to be in place.

Robust and scientific methodology should be used to ensure the accuracy and integrity of recommendations received by users of PN programs. Algorithms that use an individual's data need to be based on representative datasets. Additionally, the models should be developed in collaboration with behavior science experts, meaning the dataset used is applicable to the individual and population studied (Marques et al. 2023).

When collecting data, it is important to remember that the questions' sensitivity can impact consumers' willingness to share their data. The perceived risk to privacy is lessened when the questions asked of them are less intrusive, they know the data will be used to inform a personal message, and when individuals have more control over their personal information (Berezowska et al. 2015). Perceived risk may also be lower, and increased trustworthiness is reported, when health-related organizations and professionals (healthcare providers and dietitians/nutritionists) collect the data compared to commercial parties (Stewart-Knox et al. 2016).

Identifying, organizing, and leveraging behavioral change datasets

Behavioral data is highly fragmented and consists of "small data" (multiple limited datasets) from research studies, apps, or proprietary data collected by businesses (Reinders et al. 2023). Private behavior data exists in businesses that conduct exclusive, branded research, and with clinical research organizations and apps. Proprietary industry-generated data (e.g., Circana, Nielson, MacKinzie) provides a wealth of information on purchasing behavior and contributes significantly to understanding individual choices (Reinders et al. 2023). These data are often not captured and stored using standardized behavior systems and gaining access to them can be challenging (see "Consumer Purchasing Data in PN" section).

Data from research studies and apps are not always standardized by behavior theory models (e.g., Transtheoretical, Social Cognitive, Theory of Reasoned Action, Health Belief, Theory of Planned Behavior, etc.), and BCTs are not consistently or sufficiently detailed, making data fusion and analyses challenging (Reinders et al. 2023). Study designs grounded in theoretically-based standardized behavior models together with BCTs used in an app or a study, will provide data on what works in practice to deliver messaging that results in positive health outcomes (Buzcu et al. 2024). Implementation of this practice is still evolving.

Application of social, behavioral, and environmental data in PN: study examples

Initially, PN largely focused on generating insights based on biological and genetic markers with less emphasis on behavior data and actualizing the benefits obtained from applying BCT (Gibney and Walsh 2013; Reinders et al. 2020). However, insights on applications of behavior change in PN are advancing. There are several studies that have shown the efficacy of behavior change, with Food4Me being a foundational study (Celis-Morales et al. 2017). This six-month, seven-country randomized controlled study demonstrated that personalized advice was more beneficial than generic advice. Data showed the PN groups receiving individualized advice consumed significantly less red meat, salt, saturated fat, and increased their folate intake, and had a higher Healthy Eating Index score, compared to control. There was no evidence that adding phenotype or genotypic information enhanced the benefit. The Food4Me study illustrated the value of behavior change methods to improve dietary patterns. The study implemented BCTs *a priori*, with 17 embedded in the intervention design. The use of existing taxonomies facilitated standardization of techniques. It was a novel application of behavior change taxonomies and the first internet-based PN to use such a framework remotely (Macready et al. 2018).

A systematic review and meta-analysis by Villinger et al. (2019) supported the benefits of BCT (goals/planning, feedback/monitoring, shaping knowledge, and social support) with app-based mobile interventions reporting significantly improved nutrition health outcomes and behaviors (i.e., positive effects on blood pressure, lipids, and obesity indices). Another systematic review by Salas-Groves et al. (2023) investigated the effectiveness of behavior with nutrition apps in people with chronic diseases (cardiovascular, cancer, diabetes, and obesity). This study demonstrated that apps involving nutrition can significantly improve health outcomes in people with chronic diseases. However, data showed less than half of the studies were based on behavior change theory or constructs, less than 25% measured maintenance of health behavior change, highlighting a need to identify the effective and sustainable components of behavior change.

PN studies that include or focus on socioeconomic data are critical to building a representative dataset on behavior. Froome et al. (2024) conducted a systematic review of digital behavior change interventions (goal setting, problem-solving, behavior instruction, and prompts and cues) to improve fruit and vegetable intake in low-socioeconomic status (SES) primary school children. Their results identified a need to embed behavior and socioeconomic components to tailor SES-targeted interventions. Others have also highlighted the importance of appropriate study designs to address the potential for health inequalities (Lorenc et al. 2013). Ronteltap et al. (2022) contributed additional data on digital health interventions with lower SES populations in a systematic review of studies aimed at changing dietary behavior using BCTs based on Abraham and Michie (2008) taxonomy. Results showed the potential success of digital health interventions with goal setting and planning and information provision. They also identified the most often used BCTs and noted that social support was rarely applied. The review underlined that disadvantaged populations are understudied in the field and should be prioritized relative to the magnitude of health disparities.

Organizing and leveraging datasets for use in PN

Overall, studies point to the need for quality datasets that include details of the BCT used, including taxonomy, and outcome data. Advances have been made in extending and connecting BCT taxonomy, which specifies the active component of behavior change interventions, with formal ontology practices (Michie et al. 2013). BCT Ontology (BCTO) was developed as part of the Human Behavior Change Project, an open-source platform, which is a collaboration between behavior and computer scientists and data architects (Human Behaviour Change Project 2024; Marques et al. 2023; Michie et al. 2020). The BCTO provides standard terminology and a comprehensive classification system for behavior change data that can be used to synthesize evidence and generate novel insights into behavior change. It facilitates efficient evidence accumulation of "what works" in behavior change interventions across lifestyle disciplines and behavior domains. Another aim of the Human Behavior Change Project is to develop algorithms that predict behavior change outcomes from an intervention based on using automated information extraction reports (Mac Aonghusa and Michie 2020). This process includes training AI to recognize patterns and natural language to predict intervention outcomes that can be evaluated by behavioral scientists.

Challenges and opportunities for use of social, behavioral, and environmental data for PN

Research shows the need for details and clarity on BCT, and taxonomies used in studies to optimally move forward with AI and ML to create effective behavior change messages based on analyses of shared data. Key contextual elements need to be considered when compiling and merging data to use for AI include assessing accuracy, completeness, consistency, and ethical acquisition of each dataset. It is also critical to have quality standards, common means for codification, and documentation of where the data was obtained (M. Y. Ng et al. 2023).

AI can play a role in helping to generate effective nutrition and lifestyle messages both visually and with text. When building an AI repository for behavior, it will be paramount to generate messaging that is culturally, socioeconomically, and ethically applicable (M. Y. Ng et al. 2023). Feeding the tacit knowledge and blueprints currently used by dietitian-nutritionists in behavior change counseling, into algorithms will expedite generative AI models to create effective messaging and visuals to target long-term engagement (Reinders et al. 2023).

Consumer purchasing data in PN

Consumer purchasing data includes information on consumer's buying habits, such as price, quantity, product preferences, brand, time and location of purchase, preferred shopping channels, payment choices as well as demographics, health status and lifestyle practices (Figure 1) (Ramya and Mohamed Ali 2016). Consumer purchasing behavior is complex and influenced by many factors; hence data is generated from many fields including psychology, behavior science, marketing, and socioeconomics (Peighambari et al. 2016). It has the potential to significantly inform and enrich PN to facilitate beneficial health outcomes. Analyses of the data on its own, or through data fusion can characterize individual and subgroup nutrient intake and behavior and can be used to help optimize PN platforms and recommendations.

Consumer packaged goods (CPG) companies often conduct market research on consumers or purchase proprietary datasets from other companies for competitive advantages to drive sales, enhance the shopping experience, and for marketing and optimization of products (Bedrock Analytics 2022). In addition, other for-profit industries such as Mintel, Nielson, and Circana collect, analyze, and sell proprietary consumer data to CPG companies and other organizations. Significant costs are involved in accruing, managing, hosting, and organizing massive amounts of consumer data, which translate into fees for purchasing the data.

Consumer data industries also provide proprietary data for use by academia and government agencies through distinct access agreement contracts based on detailed and defined purposes of the research (U.S. Department of Agriculture 2024d). Nonprofit, 501(c)(3) organizations and other companies also generate consumer data through surveys that provide insights into factors influencing food choices providing key trends to enhance PN health benefits.

Consumer research data often receives less attention but has a tremendous ability to make PN more accessible by helping to fill the gap between intentions and actions. As it can be obtained from a variety of venues from retail sales and surveys to proprietary market research, challenges and opportunities are provided below in the context of specific examples.

Consumer purchasing: survey data on values, attitudes, and behavior

The International Food Information Council (IFIC), an example of a 501(c)(3), is an organization that collects data from American consumers to more effectively communicate information on nutrition, food safety and sustainable food systems. It supports a variety of food and nutrition stakeholders, including researchers, health professionals, government policymakers, and food industry representatives through its communications and education programs on consumer values, attitudes and behaviors. Their annual Food & Health Survey provides valuable data on how Americans connect food and food purchasing decisions to health and well-being (International Food Information Council 2024).

Survey results demonstrated the impact of stress, price, the economy, and social media on consumer attitudes, generational differences in health priorities, and the key purchase drivers for foods and beverages. Data shows taste (85%) continues to be the top driver for food and beverage purchases. Price (76%) and healthfulness (62%) ranked second and third in importance. Despite the consistently high impact of healthfulness on food and beverage purchasing decisions over the last decade, the percentage of consumers reporting themselves to be in excellent or very good health is declining in recent years with 57% reporting this in 2020, 52% in 2021, 55% in 2022, 49% in 2023, and 47% in 2024, underscoring a growing opportunity for a new approach such as tailored PN programs.

Results suggest income influences stress, happiness, and health status. Healthfulness surpasses price as a purchase driver for those with a household income >$100K per year. Only one-third of survey takers with household incomes <$20K describe their health as excellent or very good. Whereas households with income >$150K/year reported twice that level highlighting health disparities based on perceptions of health. IFIC data show a significant relationship between income and stress, as twice as many Americans with household income <$35K report being very stressed in the last six months, compared to those with household income >$75K (33% vs. 16%). This also highlights the need for accessible nutrition programs to support the mental and physical health needs of lower-income groups. Other factors like social media and emotional state influence food behaviors, highlighting a need for understanding the impact of these areas as part of PN solutions (International Food Information Council 2024).

More granular data would be beneficial in designing programs to target specific needs and income issues such as having consumer survey data relative to participation in the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) and the Supplemental Nutrition Assistance Program (SNAP) (U.S. Department of Agriculture 2024g, 2024h). It is recognized that answering questions relative to WIC and SNAP status would be voluntary and all means possible would need to be in place to combat disclosure risks. This could include informed consent and Institutional Review Board approval. Another future opportunity to fill the gap and address the accuracy and completeness of data would be to link consumer survey data with loyalty card shopping data to help determine if self-reported behaviors and goals matched purchases. This would also be a voluntary input from survey participants and accentuates having data security and privacy in place to gain and maintain consumer trust (Stewart-Knox et al. 2015).

Consumer purchasing data: application to a government Project

Another perspective and example of consumer data use is the United States Department of Agriculture (USDA) methodology to develop the Thrifty Food Plan (TFP) (U.S. Department of Agriculture 2023b, 2024i). The TFP is the lowest cost of four Food Plans developed by the USDA (Thrifty, Low-cost, Moderate-cost and Liberal Food Plans). The TFP outlines nutrient-dense foods and beverages, their amounts, and associated costs to support a healthy diet at-home on a limited budget and serves as the basis for setting maximum SNAP benefit allotments (U.S. Department of Agriculture 2024k). The USDA develops the TFP using an optimization model that selects quantities of food and beverage modeling categories, subject to a set of constraints, to yield a list of food and beverage quantities and costs that can be used to prepare healthy meals and snacks, called a market basket (Figure 2) (U.S. Department of Agriculture 2024i).

The methodology for developing the TFP is tailored specifically to meet requirements defined in Federal law, including that the TFP be based on current food prices, food composition data, consumption patterns, and dietary guidance (U.S. House of Representatives 2012). The TFP model uses government food composition data, dietary intake data, and dietary guidance together with proprietary retail scanner data (Circana OmniMarket Core Outlets) to achieve these requirements. The government food composition data is from the USDA Food and Nutrition Database Dietary Studies (FNDDS) and the Food Pattern Equivalents Database (U.S. Department of Agriculture 2023a, 2024a). The government dietary intake data is from What We Eat In America (WWEIA), the dietary component of the National Health and Nutrition Examination Study (NHANES)(U.S. Department of Agriculture 2022; U.S. Department of Health and Human Services 2024a). Dietary guidance for the most recent TFP reevaluation is from the Dietary Guidelines for Americans 2020–2025 (U.S. Department of Agriculture and and U.S. Department of Health and Human Services 2020). Current food price data is from proprietary retail scanner data provided to the USDA. Proprietary retail scanner data provides weekly transaction data collected through in-store scanners and includes all food items sold by a set of affiliated retailers. Store types include grocery stores/supermarkets, mass merchandizers, super stores, convenience stores, drug stores, dollar stores, and defense commissaries in urban, suburban, and rural communities (Muth et al. 2016) (Levin et al. 2018). These data are used in the TFP optimization model together with a set of constraints, which include energy requirements, nutrient recommendations, the Healthy U.S. Style Dietary Pattern Food Groups and subgroup amounts, and practicality based on current consumption patterns across 15 age-sex groups (Figure 2). The model is run iteratively to identify the lowest total cost at which a market basket can be identified that meets all constraints using the input data.

One challenge with the model is that the proprietary retail scanner data is not directly compatible with the FNDDS/WWEIA data. To address this, the proprietary retail scanner data is linked with FNDDS/WWEIA data using the Purchase to Plate Suite, which includes the Purchase to Plate Crosswalk (PPC), and Price Tool (PPPT) (U.S. Department of Agriculture 2024b). The PPC links FNDDS food codes to Universal Product Codes (also called barcodes) in the retail scanner data and assigns evidence-based adjustment factors (i.e., moisture loss or gain in cooking) to convert food as-sold, in the retail scanner data, to as-consumed forms found in the WWEIA. The PPPT then provides average unit prices for FNDDS food codes in their as-consumed forms based on the retail scanner data. Note the average unit prices provided by the PPPT and used in the TFP optimization model do not include other food-preparation costs such as time, energy to transport, store, and prepare food or the skill level or time to prepare it (Carlson et al. 2020).

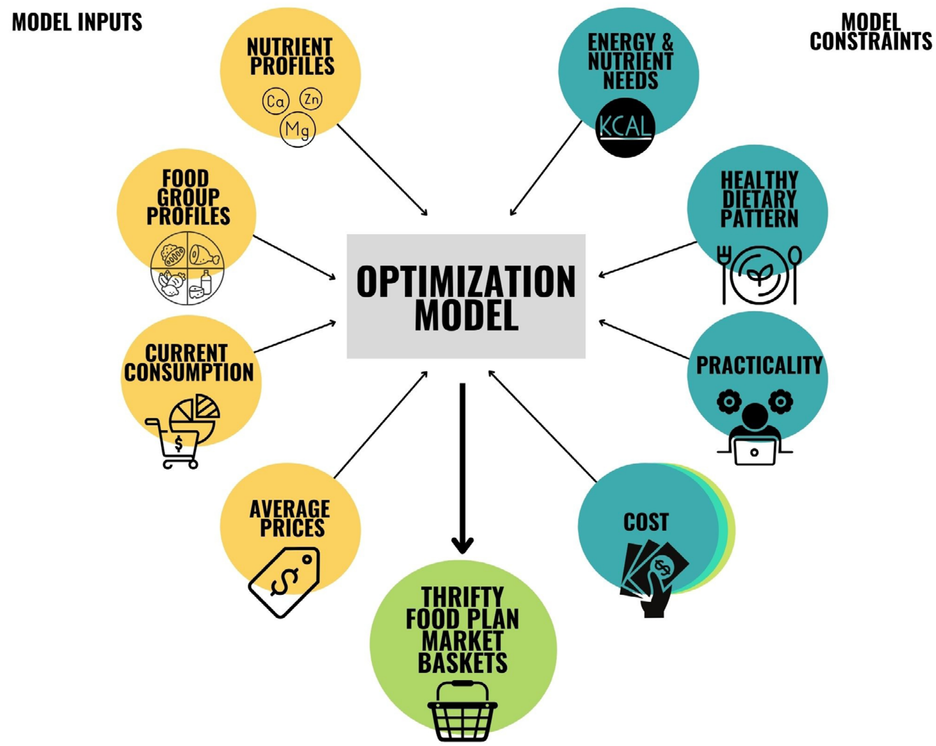

Figure 2. Multiple datasets convergence example: the thrifty food plan model. This figure is modified from the Thrifty Food Plan, 2021, published by the USDA. The TFP is the product of an optimization model that includes multiple inputs and constraints, and ultimately outlines types and amounts of nutrient-dense foods and beverages and associated costs that can be purchased on a limited budget (U.S. Department of Agriculture 2024i). The cost constraint is set iteratively (represented by the layering) to identify the lowest cost at which the optimization model can solve and yield market baskets. Figure created using canva.com.

Further consideration in developing the TFP includes the data's timeliness. As stipulated by the Agriculture Improvement Act of 2018, the USDA must reevaluate the TFP every 5 years, with the last update in 2021 and the next anticipated in 2026 (U.S. Congress 2018). Federal law also stipulates that the food prices, food composition data, consumption patterns, and dietary guidance on which the TFP is based must be *current*. In 2021, when the TFP was developed, the most current data on food composition and consumption was from 2017-18. However, the most current food price data was from 2015–2016 because the PPS for 2017-18 was still being developed. The proprietary retail scanner data takes significant time to collect, clean, store, and maintain, and development of the PPS relies on availability of both the food composition data and the retail scanner data for a given period. This results in a lag in availability. To ensure the TFP reflects current food prices, the food price data provided by the 2015-16 PPPT were adjusted for inflation to reflect June 2021 price levels before running the optimization model (U.S. Department of Agriculture 2024i). The USDA also updates the cost of the TFP each month to reflect inflation using the Consumer Price Index for All Urban Consumers for specific food items that align with the foods and beverages in the TFP market basket (U.S. Department of Agriculture 2024j).

Expenditures on food eaten away from home climbed from $1.3 trillion in 2022 to $1.5 trillion in 2023. Food away from home expenditures accounted for 58.5% of total food expenditures in 2023 (U.S. Department of Agriculture 2018, 2024c). Eating away from home data is not currently included in the TFP market basket model based on the purpose and the scope of the project per Federal Law. This is a significant food purchase behavior that, if used, would significantly impact the TFP model.

Another gap in this model is the available food composition databases have somewhat limited ethnic, geographical, and regional food composition analyses. Increased diversity has markedly impacted the types and how food eaten in the U.S., including more ethnic and culturally relevant foods, could better inform the model on U.S. eating behaviors (Dong and Stewart 2022). Differences in food availability by region and composition of these food products are not all available in FNDDS. Data is not collected from small culturally tailored stores, such as smaller Asian or Hispanic markets. The proprietary retail scanner database only collects data from retailers having annual sales ≥$2 million (Levin et al. 2018). If food composition and pricing data were available from these food outlets, it could potentially provide TFP costs that are more reflective of the complete distribution of foods and beverages purchased and consumed in the U.S.

The USDA continues to develop, integrate, and modernize databases to resolve issues such as the lag in available food price data and the need for more cultural, regional and geographical foods. Since the 2021 TFP reevaluation, the proprietary retail scanner database has a streamlined process for providing users access to the most currently available retail scanner data. The USDA has also published an initial study plan for the 2026 reevaluation of the TFP, indicating a commitment to identifying data and information needs to inform the next reevaluation of the TFP and ensure that the inputs into the optimization model reflect the latest science and evidence. To accomplish this, the USDA initiated a separate, multifaceted research plan to address data, and information needs for future reevaluations, including a study on the optimization model, which will inform the TFP 2026 final study plan (U.S. Department of Agriculture 2024f).

Consumer purchased proprietary data: academic example

Proprietary household and scanner data can be a rich source to understand food purchasing habits and their relationship to health outcomes. Working through their different complexities is a potential value to developing PN programs. As an example, three proprietary consumer purchasing data sources (NielsenIQ Homescan Consumer Panel, the NielsenIQ Retail Scanner data, and the Circana MedProfiler survey) were used to investigate consumer purchasing behavior (Bao et al. 2018, 2024) as it relates to NIH body mass index (BMI) categories (Muth et al. 2016; U.S. Department of Agriculture 2024d; Weir and Jan 2024). Circana MedProfiler Survey is an annual survey conducted with Nielsen HomescanIQ panelists; a nationally representative sample of 100,000 households who have agreed to answer survey questions on health concerns, medical conditions, diet, and lifestyle (U.S. Department of Agriculture 2024e). The Circana MedProfiler collects self-reported data on height, weight, and questions about having, treating, or concerns with health conditions and behaviors (e.g., rate overall health/dietary behaviors on a Likert-type scale). The NielsenIQ Homescan survey tracks detailed grocery purchases on a household level. The NielsenIQ Retail Scanner data provides data on the store level of weekly prices, features, and displays (U.S. Department of Agriculture 2024d). Working with these three datasets, results in a laborious process when attempting to link them. Collecting and integrating these datasets highlights gaps in inconsistent naming conventions, missing values, and sometimes differing assessment time points.

Other considerations for different types of data in PN

Self-reported data

A significant body of self-reported data using multiple methods. For example, self-reported dietary intake can come from food recalls, histories, food frequencies, food images, etc. The diversity of these self-reported methods impacts data harmonization across studies. Lifestyle practices are also self-reported and subject to positive and negative reporting bias (Eyles and Mhurchu 2009; Taylor et al. 2019). Self-reported input is by no means perfect but has been the state of practice although new technologies are rapidly evolving. An overreliance on self-reported data that cannot be verified may lead to errors due to individual's memory bias or social influences. These errors may be further compounded when applying AI tools, which could draw conclusions from incorrect data. Advances in smartwatches and phones, sensing technologies (health trackers), and commercial information (e.g., loyalty cards and retail apps) can add a greater degree of accuracy to lifestyle behavior and food intake (Reinders et al. 2023).

What consumers report can differ from their actual behavior. Consumers' food purchasing practices can be collected at the supermarket. However, it may not completely depict what is happening at home. As data shows, 30% to 40% of food in the home is wasted (U.S. Department of Agriculture 2024l). While there is some data suggesting which foods are consumed and which foods go to waste there is less known on an individual level (Conrad 2020; Springer Nature Ltd 2024). This leads to difficulty in identifying a clear association between diet and health based on food purchases alone. Determining the best and most accurate ways to accommodate what individuals consume away from home is one of the gaps in collected data sets. Overall, looking toward healthcare data guidelines and standards, is an opportunity to adapt already established and utilized practices instead of reinventing for PN. In the future, new technologies for more objective inputs can help overcome these challenges.

Wearable data

Wearable devices are transforming PN by enabling real-time, noninvasive monitoring of lifestyle and key biomarkers, such as activity, glucose, electrolytes, and vitamins (Brothers et al. 2019; Sempionatto et al. 2021; Shi et al. 2022). These sensors have the potential to bridge the gap between traditional food intake assessment and real-time molecular-level feedback, offering insights into nutrient metabolism, individual health status, and nutritional needs based on biological lifestyle, and environmental factors. For instance, continuous glucose monitoring and sweat sensoring have demonstrated potential in guiding dietary interventions and managing metabolic conditions (Sempionatto et al. 2021; Shi et al. 2022). Despite these advancements, challenges such as ensuring sensor accuracy, integrating multimodal data, and addressing privacy concerns remains.

Timely access to data, research applications, and representative databases

Time to access data can negatively impact its relevance and application. For example, research and other institutions have specific rules as to how Material Transfer Agreements work. The time required to execute these agreements can vary greatly depending on the companies' and institutions' policies and delay data analysis. Proprietary data transfers from industry to academia are often complex due to diverse objectives in the organizations. A company may want rights to review academic work and retain ownership of new inventions stemming from use of their data with academic institutes having similar goals on intellectual property. Also, freedom to publish can be an issue with academia not able to compromise its mission to disseminate knowledge and businesses not willing to compromise their proprietary dataset use.

PN research would also greatly benefit from more longitudinal data that tracks individuals' health and dietary patterns to understand how personalized programs affect health outcomes over time. However, such data is lacking due to the high cost and logistical challenges of long-term studies. The available datasets may also lack representation of different demographics limiting the generalizability of findings. Most studies tend to involve participants from similar socioeconomic backgrounds, which may not reflect the broader population. Ensuring that datasets are diverse and inclusive is essential for developing accessible PN strategies that are effective and applicable across different socioeconomic groups.

Guiding principles for collecting and integrating data into PN

In the sections above, we presented the challenges and opportunities related to the application of health, biological, social, behavioral, environmental, and consumer purchasing data to the PN field. To optimize PN programs and services, these different areas must come together to support the most robust possible recommendations that can inform positive behavior change and drive desired outcomes. However, there are system-wide challenges related to how this can best be accomplished. For this reason, we highlight a series of five guiding principles (Table 1) that can help support responsible data collection and integration that will help drive innovation while maintaining individual and organizational rights and support greater access to personalized solutions.

Define the data architecture to standardize data collection, storage, sharing, and fusion

A significant barrier to data use in PN is the inability to effectively link and integrate diverse data sources (Figure 3). Challenges are encountered when using data sets with inconsistent naming conventions, missing values, a lack of behavioral data, and sometimes time lags in obtaining datasets. Data from genomics, metabolomics, dietary intake, and lifestyle factors often exist in silos, making it challenging to derive comprehensive insights. An appropriate data architecture, a multi-scale system to manage data, must be created. Data architecture is defined as a set of models, policies, rules, and standards governing data interoperability and management within organizations for alignment in the PN field. This structure of standards, policies, and rules regulates what data is collected and how it is stored, organized, integrated, and used. Accordingly, data architecture requires cross-disciplinary expertise and collaboration. Advanced data integration techniques and interoperable systems across public and private sectors are needed to understand individual nutritional needs and responses. There is a need for consistent naming conventions, data quality, more detailed questions that would allow for more accurate measures and a faster turnover rate of data collection. The overall use of data that supports recommendations requires robust data management frameworks and interoperability standards to ensure flow between systems to allow comprehensive analysis including AI and ML, for personalized recommendations and adaptation to changing needs.

One of the foremost goals in advancing PN programs is effective data fusion or the process of integrating multiple data sources to produce more useful information than any individual data source. This involves integrating and making accessible data from diverse technological sources such as wearables, apps, and EMRs. Achieving harmonization in this integration is crucial to ensure that data from these varied sources can be utilized effectively, efficiently, and ethically (M. Y. Ng et al. 2023). Standardization of cohorts across different countries can be difficult with individual variations in dietary habits, genetic backgrounds, and reporting bias. External factors such as public health recommendations, regulations, and healthcare systems can also make standardization difficult. These factors necessitate the development of industry-wide standardized protocols to ensure that data from different populations can be compared and analyzed consistently. However, this raises the question as to which group or industry would set the standard. As noted, multiple types of data need to be integrated. This could prove difficult as organizations will have standard nomenclature and practices for data management, requiring flexibility for all parties involved.

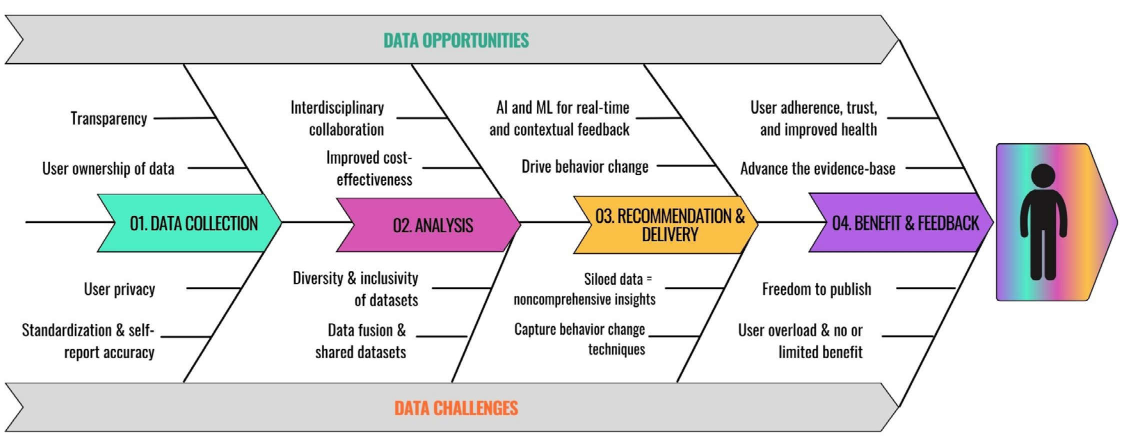

Figure 3. Data opportunities and challenges applied to the personalized nutrition system. The figure illustrates the four phases of the PN business model and the opportunities and challenges around data. The four phases include 1. Data collection, 2. Analysis, 3. Recommendations and Delivery, and 4. Benefit and Feedback. Figure created using canva.com.

Employ a systems-based approach to identify cause-and-effect recommendations

Several data challenges and opportunities support the need for a systems biology approach (van Ommen et al. 2017). Nutrition and health depend on multiscale systems such as biological processes, individual behaviors, social interactions, the local environment, social and economic forces, and the worldwide environment that may affect (for example) the price or availability of crops, all of which can change over time. Fully characterizing these systems and their interactions is difficult with traditional research and data collection methods alone. This goes beyond data architecture and requires an understanding of how these systems work together and influence each other. In this way, "big data", offers new opportunities for PN. Today there are more advanced computational systems compared to two or three decades ago. The key to unlocking the application of big data, so that it provides better insights into improving PN, is combining the data using new systems methods (B. Y. Lee et al. 2021). While AI can help us better understand complex systems, simply identifying an association using AI for a "top-down approach," i.e., starting with data and identifying patterns, trends, and associations, does not consider all the factors that may have come into play. Associations play a role in the discovery, but a systems approach, where the entirety of an individual's biological and environmental systems are taken into account, is necessary to identify cause and effect for diet and health. This requires cross-functional expertise when collecting and organizing the data structure for these tools. Leveraging expert knowledge in algorithm development will expedite generative AI models to create useful information in effective messaging and visuals for long-term engagement (Reinders et al. 2023).

Incentivize data sharing

The regulatory bodies that would be responsible for overseeing and setting guidelines for data sharing have not been identified. As a result, challenges exist for standardizing protocols for making data from disparate sources more widely available. Part of the challenge may be lack of consistency with the data. If standardized rules were applied, then the approach to regulation may be clearer. The lack of standard data architecture further hinders opportunities for data integration and fusion as well as the ability to use diverse data sets for the application of AI and PN solutions.

The proprietary nature of data collected by various entities also presents a significant barrier to advancing the evidence base in PN. Companies safeguard their data, which hinders the sharing of information that could enhance research and development. Further, the protected nature of a company's and research institution's data hinder standardization of the data collection and management. Encouraging open data sharing while protecting intellectual property (IP) rights remains a complex issue that needs addressing. Publishing data in peer-reviewed journals and public repositories can promote transparency, collaborative efforts, and IP, but it requires balancing commercial interests and scientific openness. One way to help incentivize companies for data sharing to support PN is agreeing to terms on profit sharing for discoveries. Specific rules or permissions can be established with respect to data use or sharing. By demonstrating the value of data sharing between organizations, other collaborations will likely follow.

Guarantee and ensure that control of the individual's data is maintained by the individual

While it is important to help incentivize organizations to share data, control of an individual's data ultimately rests with that individual. People often lack trust in the ability of service providers and technology to ensure the secure handling of personal health data (Stewart-Knox et al. 2013). While consumers deem health-related organizations, such as physicians and dietitians, as trustworthy, they place less trust in other commercial parties (Reinders et al. 2020). The European Union has the strongest laws regarding the privacy of personal data. In 2018, the GDPR policy was implemented, which applies to all companies that process the personal data of Europeans regardless of their location (European Commission 2018). As it stands now, USA data protection varies by state. The American Data Privacy and Protection Act, in committee, would provide more protections for personal data (U.S. Congress 2022). The bill sets limits on a company's collection, processing, and transfer of personal data to what is reasonably necessary to provide products and services. Furthermore, it establishes consumer protections that include the right to access, correct, and delete personal data. Each person should maintain the right to their data with informed consent that favors the individual's rights rather than corporate interests. Groups of individuals can also come together in Health Data Cooperatives where data are owned and controlled collectively by citizens. The data can be made available for public research or sold for commercial interests, but it is governed in a manner to ensure individual privacy (Tanwar et al. 2021).

Ensure relevant data is used to create accessible and actionable PN solutions

Integrating health and biological, social, behavioral, environmental, and consumer purchasing data into PN programs on a large scale is hindered by high costs, logistical challenges, and socioeconomic disparities. The expense associated with comprehensive data collection and storage, including genomic, metabolomic, and dietary information, makes it difficult to provide cost-effective PN services. Moreover, the infrastructure required to analyze and interpret this data is substantial, further escalating costs. Factors such as income, education, and physical and social environment can influence an individual's ability to benefit from PN services. Addressing these disparities is essential to ensure that all individuals, regardless of their background, can access and benefit from the biological and health data that can be utilized in PN programs.

One of the biggest challenges and opportunities in PN offerings is simplifying the information provided to users and making the data they receive actionable. More data does not always equate to better in this case. However, observing trends in their resting heart rate over a 2-week period, as well as disrupted sleep patterns and long work hours provides insights that allow for timely interventions that can improve their physical and emotional wellbeing. By identifying the most impactful data to help the end user achieve their health goals, it is possible to streamline and offer more cost-effective solutions. This is particularly important for delivering solutions for lower SES individuals and underrepresented minorities. These groups often experience higher rates of chronic diseases of lifestyle and could benefit greatly from personalized products and services that provide learning resources and recommendations.

Summary and recommendations

Applications and future directions

The field of PN is evolving rapidly as result of advancements in data science, machine learning, and wearable technology. These developments are enabling more precise and personalized dietary recommendations, supporting improved health outcomes across diverse populations. The applications of PN span clinical, consumer, and public health domains, with growing potential to enhance preventive healthcare, chronic disease management, and behavioral modification strategies.

Companies, academic researchers, and the government must be adaptable to incorporating emerging findings and technologies. Establishing standards across sectors will be necessary to realize the full value of data and help assure that PN products and services can support evidenced-based health outcomes. The PN field would benefit from a committee formed by the National Institutes of Health or National Academy of Sciences to recommend guidelines for best practices in PN data collection. Having these recommendations in place will ensure that data is AI-ready to support data fusion.

Expanding PN applications through advanced data fusion

Future PN applications will increasingly rely on integrating diverse data sources such as genetic information, real-time biometrics from wearable devices, social, behavioral, and environmental factors and consumer purchasing data to providing a more holistic view of individual health. For instance, the integration of wearable and behavioral data with traditional health metrics could enable dynamic dietary recommendations that adapt to a user's physiological state, activity levels, and environment in real-time.

Establishing shared data stores will enable wider access to data, allowing researchers and industry professionals to support data analyses that may not be currently in use. This would enable the formation of centralized repositories where data is accessible to authorized parties and facilitate broader analysis and innovative solutions. For this to work, it is critically important to identify incentives to share data. Incentivization (i.e., funding resources, direct benefits to data owners) needs to align with company's and researchers' interests and needs to compel changes and facilitate advances (M. Y. Ng et al. 2023).

Leveraging AI for precision and scalability

AI and ML have the potential to transform PN by analyzing large datasets to identify complex patterns that inform personalized dietary interventions. Predictive algorithms can refine recommendations by anticipating an individual's response to specific dietary changes, while natural language processing could facilitate personalized communication strategies. AI is emerging as a transformative tool capable of analyzing extensive datasets, performing tasks like portion size and calorie estimation via image recognition, and reducing reliance on self-reported data. However, its applications in the nutrition field remains underexplored, with potential benefits, challenges, and privacy concerns warranting further investigation (Sosa-Holwerda et al. 2024). To maximize AI's impact on PN, future research should prioritize transparency, fairness, reliability, validity, privacy, and accountability in algorithm design, ensuring recommendations are evidence-based and applicable to diverse populations. In January 2021, the NIH held the "Precision Nutrition: Research Gaps and Opportunities" workshop, which discussed ways that AI/ML methods have facilitated and can advance precision nutrition (Lee et al. 2022). This workshop led to the launch of NIH's NPH program a year later, which included the establishment of the Artificial Intelligence, Modeling, and Informatics for Nutrition Guidance and Systems (AIMINGS) Center. The AIMINGS Center serves as an AI center for the NPH consortium and includes three projects developing various AI/ML models and algorithms that account for the complex systems affecting diet, nutrition, and health to help better tailor diets and nutritional recommendations to different people and circumstances (Mabry et al. 2022; U.S. National Institutes of Health 2024b).

Building collaborative frameworks for privacy and security

Addressing the challenges of data interoperability, privacy, and integration in PN will require collaboration across sectors, including healthcare, academia, technology, and regulatory bodies. Establishing interdisciplinary partnerships can accelerate the development of standardized frameworks for secure data collection, sharing, and analysis, supporting broader PN applications (Figure 4). Collaborations with government agencies and industry stakeholders can facilitate the development of PN policies that protect consumer data while encouraging innovation. Data scientists and engineers need to work closely with nutrition researchers, dietitians, and behavior scientists to identify and leverage existing datasets and develop shared ontologies and classifications to address issues successfully.

Future research and development needs

To realize the full potential of PN, future research should focus on creating accessible and inclusive data repositories that capture the variability in diet, genetics, and health across different demographics. Large-scale, longitudinal studies that include

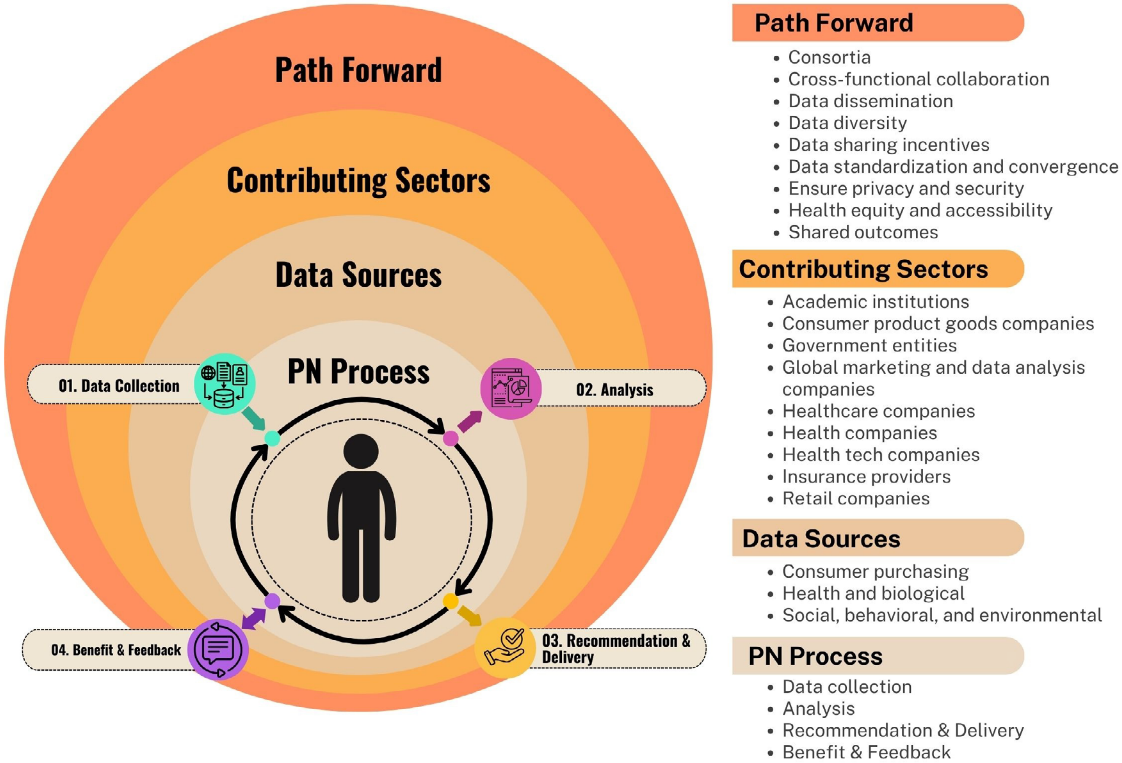

Figure 4. Layers of personalized nutrition data: a path forward. Abbreviation: PN, personalized nutrition. Figure created using canva.com.

underrepresented groups are essential to ensure that PN recommendations are effective and equitable. Additionally, developing standardized protocols for data sharing and privacy will be critical to fostering trust and promoting broader adoption of PN programs.

Toward accessible and actionable PN solutions

As PN continues to grow, a key priority will be ensuring that personalized dietary guidance is accessible and actionable for all individuals, regardless of socioeconomic status. Future PN models should prioritize user-friendly interfaces and personalized feedback that empower individuals to make informed dietary choices. Simplifying complex health insights into practical recommendations can enhance user engagement, making PN a valuable tool for health improvement at both individual and community levels.

A future example of PN data integration

In the future, a virtual personalized nutritionist could access integrated data sets including an individual's real-time blood biomarkers, daily food intake, food likes and dislikes, and budget. With sensor technology, the virtual personalized nutritionist could track the individual's exhaled compounds through a breath sensor and changes in microbiome status through a "smart toilet", to make dietary recommendations to improve the gut microbiota. The virtual personalized nutritionist could suggest a nearby restaurant for the next meal based on GPS data and menu items that fit their budget, personal dietary needs, and preferences. The virtual personalized nutritionist could also assess what is in the individual's refrigerator. While trying to decide whether to eat out or cook at home during the individual's commute, the virtual personalized nutritionist could provide an at-home option with a recipe based on the ingredients in the refrigerator. If two additional ingredients are needed, they could be ordered and delivered via a drone. Ultimately, summarized information could also be available to the individual's healthcare providers to provide information on adherence to dietary, lifestyle, and medical recommendations between office visits. This approach would support a continuum of care with the potential to reduce healthcare costs.

Conclusion

When transforming data sources into PN solutions, it is vital to prioritize desired outcomes from the beginning. By starting with the end in mind, PN programs can be designed to address specific goals, ensuring that all efforts are aligned toward achieving these outcomes. This approach helps in maintaining focus, reducing cost and participant burden, and directing resources toward the most impactful areas. By addressing these challenges and barriers, PN programs can become more effective, driving more targeted recommendations, and leading to more impactful health outcomes.

Acknowledgements

We thank the Planning Committee members for their input on this event's agenda, speakers, and topics. The Planning Committee members included: Valerie Benoit, Mars Wrigley; Tristin Brisbois, PepsiCo (Co-Chair); José Ordovas, Jean Mayer USDA Human Nutrition

Research Center on Aging at Tufts University; Barbara Schneeman, University of California, Davis; Alison Steiber, Academy of Nutrition and Dietetics (Co-Chair); Michelle Stout, Amway; Elizabeth Tilak, Givaudan, and Michelle Tucker, General Mills. We also thank workshop speakers Kevin Meyers Mathieu, the Center for Nutrition Policy and Promotion, U.S. Department of Agriculture, Food and Nutrition Service, and Sheri Schully, the *All of Us* Research Program, National Institutes of Health, for their valuable contributions and insights, as well as for reviewing, editing, and confirming the accuracy of the information presented in relation to their respective discussions. Finally, we thank the Personalized Nutrition Initiative External Advisory Committee for their guidance throughout the process and Densie Webb, PhD, RD for supporting the initial efforts in bringing this manuscript together.

Author contributions

All authors contributed to manuscript development. SMD and ASK directed the workshop. ASK, BLW, JCA, KMN, and SMD developed the concept for the manuscript. BLW, DW, JCA, and KMN led manuscript development. MA, YB, MB, KMB, GB, BYL, KMM, VP, MJR, SS, KS, and ST were speakers during the workshop on which the manuscript content was based upon. All authors read and approved the final manuscript. The views expressed in this manuscript are those of the authors and should not be construed to represent any official determination or policy of the entities of the U.S. Government or affiliate companies.

Disclosure statement

MA, KMB, GB, BYL, VP, MJR, and KS received speaker honorariums from the University of Illinois Urbana-Champaign. MA, JCA, GB, and MJR: External Advisory Committee member for the Personalized Nutrition Initiative at Illinois. JCA, KMN, and BLW work at Nlumn; the company compensated by the University of Illinois at Urbana-Champaign to draft the manuscript based on the workshop presentations and discussions. SMD reports a relationship with General Mills Inc. and PepsiCo, Inc. that includes consulting or advisory, speaking, and lecture fees. JCA reports a relationship with Dairy Management Inc. that includes consulting or advisory and travel reimbursement.

Funding

This session was funded by the Personalized Nutrition Initiative at the University of Illinois Urbana-Champaign through internal funds and External Partner Program membership fees. The 2023–2024 External Partner Program members were Amway, Archer Daniels Midland Co, General Mills, Givaudan, Mars Wrigley, National Dairy Council, PepsiCo, and Pharmavite. Representatives from the External Partners contributed to planning the agenda and identifying speakers, participating in the live session, and reviewing the manuscript.

Abbreviations

AI	Artificial Intelligence
AIMINGS	Artificial Intelligence, Modeling, and Informatics for Nutrition Guidance and Systems
BCT	Behavior Change Techniques
BMI	Body Mass Index
CPG	Consumer Product Goods
EMR	Electronic Medical Record
FNDDS	Food and Nutrition Database Dietary Studies
GDPR	General Data Protection Regulation
IFIC	International Food Information Council
IP	Intellectual Property
IRI	Information Resource, Inc.
ML	Machine Learning
NIH	United States National Institutes of Health
NHANES	United States National Health and Nutrition Examination Study
NPH	Nutrition for Precision Health
PN	Personalized Nutrition
PPC	Purchase to Plate Crosswalk
PPPT	Purchased to Plate Price Tool
PTM	Posttranslational Modifications
SNAP	Supplemental Nutrition Assistance Program
SNPs	Single Nucleotide Polymorphisms
SVs	Structural Variants
TFP	Thrifty Food Plan
USDA	United States Department of Agriculture
WIC	Women, Infants & Children Program
WWEIA	What We Eat in America

References

Abraham, C., and S. Michie. 2008. A taxonomy of behavior change techniques used in interventions. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association* 27 (3):379–87. doi: 10.1037/0278-6133.27.3.379.

Adams, S. H., J. C. Anthony, R. Carvajal, L. Chae, C. S. H. Khoo, M. E. Latulippe, N. V. Matusheski, H. L. McClung, M. Rozga, C. H. Schmid, et al. 2020. Perspective: Guiding principles for the implementation of personalized nutrition approaches that benefit health and function. *Advances in Nutrition (Bethesda, Md.)* 11 (1):25–34. doi: 10.1093/advances/nmz086.

Alkerwi, A., C. Vernier, N. Sauvageot, G. E. Crichton, and M. F. Elias. 2015. Demographic and socioeconomic disparity in nutrition: Application of a novel correlated component regression approach. *BMJ Open* 5 (5): E 006814. doi: 10.1136/bmjopen-2014-006814.

Aruoma, O. I., S. Hausman-Cohen, J. Pizano, M. A. Schmidt, D. M. Minich, Y. Joffe, S. Brandhorst, S. J. Evans, and D. M. Brady. 2019. Personalized nutrition: translating the science of nutrigenomics into practice: Proceedings From the 2018 American College of Nutrition Meeting. *Journal of the American College of Nutrition* 38 (4):287–301. doi: 10.1080/07315724.2019.1582980.

Bao, Y., M. Osborne, E. Wang, and E. C. Jaenicke. 2018. *BMI, food purchase, and promotional sensitivity*. SSRN. doi: 10.2139/ssrn.3260896.

Bao, Y., M. Osborne, E. Wang, and E. C. Jaenicke. 2024. BMI/Obesity and consumers' price sensitivity: Implications for food tax policies. *PNAS Nexus* 3 (6): Pgae190. doi: 10.1093/pnasnexus/pgae190.

Bedrock Analytics. 2022. Nielsen, IRI and SPINS: Navigating the CPG data syndicators. https://www.bedrockanalytics.com/blog/nielsen-iri-spins-navigating-cpg-data-syndicators/.

Berezowska, A., A. R. H. Fischer, A. Ronteltap, I. A. van der Lans, and H. C. M. van Trijp. 2015. Consumer adoption of personalised nutrition services from the perspective of a risk–benefit trade-off. *Genes & Nutrition* 10 (6):42. doi: 10.1007/s12263-015-0478-y.

Brothers, M. C., M. DeBrosse, C. C. Grigsby, R. R. Naik, S. M. Hussain, J. Heikenfeld, and S. S. Kim. 2019. Achievements and challenges for real-time sensing of analytes in sweat within wearable platforms. *Accounts of Chemical Research* 52 (2):297–306. doi: 10.1021/acs.accounts.8b00555.

Buzcu, B., Y. Pannatier, R. Aydoğan, M. Schumacher, J.-P. Calbimonte, and D. Calvaresi. 2024. A framework for explainable multi-purpose virtual assistants: A nutrition-focused case study. In *Post-Proceedings of the 6th International Workshop on EXplainable and TRAnsparent AI and Multi-Agent Systems*.

Carlson, A., K. Kuczynski, T. Pannucci, K. Koegel, E. T. Page, C. E. Tornow, and T. P. Zimmerman. 2020. U.S. Department of Agriculture. Econonmic Research Service. Estimating Prices for Foods in the National Health and Nutrition Examination.

Celis-Morales, C., K. M. Livingstone, C. F. Marsaux, A. L. Macready, R. Fallaize, C. B. O'Donovan, C. Woolhead, H. Forster, M. C. Walsh, S.

Navas-Carretero, et al. 2017. Effect of personalized nutrition on health-related behaviour change: Evidence from the Food4Me European randomized controlled trial. *International Journal of Epidemiology* 46 (2):578–88. doi: 10.1093/ije/dyw186.

Chaudhary, N., V. Kumar, P. Sangwan, N. C. Pant, A. Saxena, S. Joshi, and A. N. Yadav. 2021. Personalized nutrition and -omics. *Comprehensive Foodomics* :495–507. doi: 10.1016/b978-0-08-100596-5.22880-1.

Conrad, Z. 2020. Daily cost of consumer food wasted, inedible, and consumed in the United States, 2001–2016. *Nutrition Journal* 19 (1):35. doi: 10.1186/s12937-020-00552-w.

Dong, D., and H. Stewart. 2022. U.S. Department of Agriculture. Economic Research Service. Racial and Ethnic Diversification Will Likely Shape U.S. Food Demand and Diet Quality. https://www.ers.usda.gov/amber-waves/2022/april/racial-and-ethnic-diversification-will-likely-shape-u-s-food-demand-and-diet-quality/.

European Commission. 2018. Data protection – rules for business and organisations. https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations_en.

Eyles, H. C., and C. N. Mhurchu. 2009. Does tailoring make a difference? A systematic review of the long-term effectiveness of tailored nutrition education for adults. *Nutrition Reviews* 67 (8):464–80. doi: 10.1111/j.1753-4887.2009.00219.x.

Froome, H., K. L. Cheung, W. Martin, and E. Norris. 2024. Identifying and characterising digital behaviour change interventions to improve fruit and vegetable intake in low-socioeconomic status primary school children: A systematic review. *BMC Digital Health* 2 (1):29. doi: 10.1186/s44247-024-00085-w.

Fulton, J. L., P. C. Dinas, A. E. Carrillo, J. R. Edsall, E. J. Ryan, and E. J. Ryan. 2018. Impact of genetic variability on physiological responses to caffeine in humans: A systematic review. *Nutrients* 10 (10):1373. doi: 10.3390/nu10101373.

Future of Life Institute. 2024. The EU artificial intelligence act: Up-to-date developments and analyses of the EU AI Act. https://artificialintelligenceact.eu/.

Gibney, M. J., and M. C. Walsh. 2013. The future direction of personalised nutrition: My diet, my phenotype, my genes. *The Proceedings of the Nutrition Society* 72 (2):219–25. doi: 10.1017/S0029665112003436.

Ginsburg, G. S., J. C. Denny, and S. D. Schully. 2023. Data-driven science and diversity in the All of Us Research Program. *Science Translational Medicine* 15 (726):eade9214. doi: 10.1126/scitranslmed.ade9214.

Guo, J., X. Huang, L. Dou, M. Yan, T. Shen, W. Tang, and J. Li. 2022. Aging and aging-related diseases: From molecular mechanisms to interventions and treatments. *Signal Transduction and Targeted Therapy* 7 (1):391. doi: 10.1038/s41392-022-01251-0.

Holmes, C. J., and S. B. Racette. 2021. The utility of body composition assessment in nutrition and clinical practice: An overview of current methodology. *Nutrients* 13 (8):2493. doi: 10.3390/nu13082493.

Human Behaviour Change Project. 2024. Behavior change theory database. https://www.humanbehaviourchange.org/.

International Food Information Council. 2024. 2024 IFIC Food & Health Survey. https://foodinsight.org/wp-content/uploads/2024/06/2024-IFIC-Food-Health-Survey.pdf.

Jinnette, R., A. Narita, B. Manning, S. A. McNaughton, J. C. Mathers, and K. M. Livingstone. 2021. Does personalized nutrition advice improve dietary intake in healthy adults? A systematic review of randomized controlled trials. *Advances in Nutrition (Bethesda, Md.)* 12 (3):657–69. doi: 10.1093/advances/nmaa144.

Kaplan, G., G. Bo-Linn, G. Carayon, P. Pronovost, P. Rouse, W. Reid, P, and Saunders, R. 2013. Bringing a systems approach to health. *NAM Perspectives.* 3 (7). https://nam.edu/perspectives-2013-bringing-a-systems-approach-to-health/. doi: 10.31478/201307a.

Karvela, M., C. T. Golden, N. Bell, S. Martin-Li, J. Bedzo-Nutakor, N. Bosnic, P. DeBeaudrap, S. de Mateo-Lopez, A. Alajrami, Y. Qin, et al. 2024. Assessment of the impact of a personalised nutrition intervention in impaired glucose regulation over 26 weeks: A randomised controlled trial. *Scientific Reports* 14 (1):5428. doi: 10.1038/s41598-024-55105-6.

Landberg, R., P. Karra, R. Hoobler, E. Loftfield, I. Huybrechts, J. I. Rattner, S. Noerman, L. Claeys, V. Neveu, N. H. Vidkjaer, et al. 2024. Dietary biomarkers—An update on their validity and applicability in epidemiological studies. *Nutrition Reviews* 82 (9):1260–80. doi: 10.1093/nutrit/nuad119.

Lee, B. Y., J. M. Ordovás, E. J. Parks, C. A. M. Anderson, A. L. Barabási, S. K. Clinton, K. de la Haye, V. B. Duffy, P. W. Franks, E. M. Ginexi, et al. 2022. Research gaps and opportunities in precision nutrition: An NIH workshop report. *The American Journal of Clinical Nutrition* 116 (6):1877–900. doi: 10.1093/ajcn/nqac237.

Lee, B. Y., M. C. Ferguson, S. N. Cox, and P. H. Phan. 2021. Big data and systems methods: The next frontier to tackling the global obesity epidemic. *Obesity (Silver Spring, Md.)* 29 (2):263–4. doi: 10.1002/oby.23062.

Levin, D., D. Noriega, C. Dicken, A. M. Okrent, M. Harding, and M. Lovenheim. 2018. U.S. Department of Agriculture Economic Research Service: Examining food store scanner data: A comparison of the IRI InfoScan data with other data sets, 2008–2012. https://www.ers.usda.gov/webdocs/publications/90355/tb1949_summary.pdf?v=3080.7.

Lorenc, T., M. Petticrew, V. Welch, and P. Tugwell. 2013. What types of interventions generate inequalities? Evidence from systematic reviews. *Journal of Epidemiology and Community Health* 67 (2):190–3. doi: 10.1136/jech-2012-201257.

Mabry, P. L., N. P. Pronk, C. I. Amos, J. S. Witte, P. T. Wedlock, S. M. Bartsch, and B. Y. Lee. 2022. Cancer systems epidemiology: Overcoming misconceptions and integrating systems approaches into cancer research. *PLoS Medicine* 19 (6):e1004027. doi: 10.1371/journal.pmed.1004027.

Mac Aonghusa, P., and S. Michie. 2020. Artificial intelligence and behavioral science through the looking glass: Challenges for real-world application. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine* 54 (12):942–7. doi: 10.1093/abm/kaaa095.

Macready, A. L., R. Fallaize, L. T. Butler, J. A. Ellis, S. Kuznesof, L. J. Frewer, C. Celis-Morales, K. M. Livingstone, V. Araújo-Soares, A. R. H. Fischer, et al. 2018. Application of behavior change techniques in a personalized nutrition electronic health intervention study: Protocol for the web-based food4me randomized controlled trial. *JMIR Research Protocols* 7 (4):e87. doi: 10.2196/resprot.8703.

Marques, M. M., A. J. Wright, E. Corker, M. Johnston, R. West, J. Hastings, L. Zhang, and S. Michie. 2023. The behaviour change technique ontology: Transforming the behaviour change technique taxonomy v1. *Wellcome Open Research* 8:308. doi: 10.12688/wellcomeopenres.19363.1.

McGrath, S., and D. Ghersi. 2016. Building towards precision medicine: Empowering medical professionals for the next revolution. *BMC Medical Genomics* 9 (1):23. doi: 10.1186/s12920-016-0183-8.

Michie, S., J. Thomas, P. Mac Aonghusa, R. West, M. Johnston, M. P. Kelly, J. Shawe-Taylor, J. Hastings, F. Bonin, and A. O'Mara-Eves. 2020. The human behaviour-change project: An artificial intelligence system to answer questions about changing behaviour. *Wellcome Open Research* 5(:122. doi: 10.12688/wellcomeopenres.15900.1.

Michie, S., M. M. van Stralen, and R. West. 2011. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science: IS* 6 (1):42. doi: 10.1186/1748-5908-6-42.

Michie, S., M. Richardson, M. Johnston, C. Abraham, J. Francis, W. Hardeman, M. P. Eccles, J. Cane, and C. E. Wood. 2013. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine* 46 (1):81–95. doi: 10.1007/s12160-013-9486-6.

Morais, L. H., H. L. Schreiber, and S. K. Mazmanian. 2021. The gut microbiota–brain axis in behaviour and brain disorders. *Nature Reviews. Microbiology* 19 (4):241–55. doi: 10.1038/s41579-020-00460-0.

Muth, M. K., M. Sweitzer, D. Brown, K. Capogrossi, S. Karns, D. Levin, A. M. Okrent, P. Siegel, and C. Zhen. 2016. U.S. Department of Agriculture. Economic Research Service. Understanding IRI household-based and store-based scanner data. Technical Bulletin 1942. https://www.ers.usda.gov/webdocs/publications/47633/57105_tb-1942.pdf?v=42473.

Naithani, N., S. Sinha, P. Misra, B. Vasudevan, and R. Sahu. 2021. Precision medicine: Concept and tools. *Medical Journal, Armed Forces India* 77 (3):249–57. doi: 10.1016/j.mjafi.2021.06.021.

Ng, M. Y., A. Youssef, A. S. Miner, D. Sarellano, J. Long, D. B. Larson, T. Hernandez-Boussard, and C. P. Langlotz. 2023. Perceptions of data set experts on important characteristics of health data sets ready for machine learning: A qualitative study. *JAMA Network Open* 6 (12):e2345892-e2345892. doi: 10.1001/jamanetworkopen.2023.45892.

Ng, R., Sutradhar, R. Yao, Z. Wodchis, W. P. Rosella, and L. a C. 2020. Smoking, drinking, diet and physical activity-modifiable lifestyle risk factors and their associations with age to first chronic disease. *International Journal of Epidemiology* 49 (1):113–30. doi: 10.1093/ije/dyz078.

Observational Health Data Sciences and Informatics. 2024. Standardized data: The OMOP common data model. https://ohdsi.org/data-standardization/.

Peighambari, K., S. Sattari, A. Kordestani, and P. Oghazi. 2016. Consumer behavior research: A synthesis of the recent literature. *Sage Open* 6 (2):2158244016645638. doi: 10.1177/2158244016645638.

Picó, C., F. Serra, A. M. Rodríguez, J. Keijer, and A. Palou. 2019. Biomarkers of nutrition and health: New tools for new approaches. *Nutrients* 11 (5):1092. doi: 10.3390/nu11051092.

Ramya, N., and S. A. Mohamed Ali. 2016. Factors affecting consumer buying behavior. *International Journal of Applied Research* 2 (10):76–80. https://www.allresearchjournal.com/archives/?year=2016&vol=2&issue=10&part=B&ArticleId=2683.

Reinders, M. J., A. D. Starke, A. R. H. Fischer, M. C. D. Verain, E. L. Doets, and E. J. Van Loo. 2023. Determinants of consumer acceptance and use of personalized dietary advice: A systematic review. *Trends in Food Science & Technology* 131:277–94. doi: 10.1016/j.tifs.2022.12.008.

Reinders, M. J., E. P. Bouwman, J. van den Puttelaar, and M. C. D. Verain. 2020. Consumer acceptance of personalised nutrition: The role of ambivalent feelings and eating context. *PloS One* 15 (4):e0231342. doi: 10.1371/journal.pone.0231342.

Rodríguez-García, C., and F. Gutiérrez-Santiago. 2023. Emerging role of plant-based dietary components in post-translational modifications associated with colorectal cancer. *Life (Basel, Switzerland)* 13 (2):264. doi: 10.3390/life13020264.

Ronteltap, A., A. J. Bukman, G. E. Nagelhout, R. C. J. Hermans, K. Hosper, A. Haveman-Nies, R. Lupker, and C. A. W. Bolman. 2022. Digital health interventions to improve eating behaviour of people with a lower socioeconomic position: A scoping review of behaviour change techniques. *BMC Nutrition* 8 (1):145. doi: 10.1186/s40795-022-00635-3.

Rozga, M., M. E. Latulippe, and A. Steiber. 2020. Advancements in personalized nutrition technologies: Guiding principles for registered dietitian nutritionists. *Journal of the Academy of Nutrition and Dietetics* 120 (6):1074–85. doi: 10.1016/j.jand.2020.01.020.

Russell, R. M. 2001. Factors in aging that effect the bioavailability of nutrients. *The Journal of Nutrition* 131 (4 Suppl):1359s–61s. doi: 10.1093/jn/131.4.1359S.

Salas-Groves, E., S. Galyean, M. Alcorn, and A. Childress. 2023. Behavior change effectiveness using nutrition apps in people with chronic diseases: Scoping review. *JMIR mHealth and uHealth* 11:e41235. doi: 10.2196/41235.

Science, D., Hahnel, M. Smith, G. Schoenenberger, h, Scaplehorn, N, and Day, L. 2023. The state of open data 2023. https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_2023/24428194.

Sempionatto, J. R., V. R. Montiel, E. Vargas, H. Teymourian, and J. Wang. 2021. Wearable and mobile sensors for personalized nutrition. *ACS Sensors* 6 (5):1745–60. doi: 10.1021/acssensors.1c00553.

Serio, F., A. De Donno, and G. Valacchi. 2023. Lifestyle, nutrition, and environmental factors influencing health benefits. *International Journal of Environmental Research and Public Health* 20 (7):5323. doi: 10.3390/ijerph20075323.

Shaik, T., X. Tao, L. Li, H. Xie, and J. D. Velásquez. 2024. A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom. *Information Fusion* 102:102040. doi: 10.1016/j.inffus.2023.102040.

Shi, Z., X. Li, Y. Shuai, Y. Lu, and Q. Liu. 2022. The development of wearable technologies and their potential for measuring nutrient intake: Towards precision nutrition. *Nutrition Bulletin* 47 (4):388–406. doi: 10.1111/nbu.12581.

Short, S. E., and S. Mollborn. 2015. Social determinants and health behaviors: Conceptual frames and empirical advances. *Current Opinion in Psychology* 5:78–84. doi: 10.1016/j.copsyc.2015.05.002.

Sosa-Holwerda, A., O.-H. Park, K. Albracht-Schulte, S. Niraula, L. Thompson, and W. Oldewage-Theron. 2024. The role of artificial intelligence in nutrition research: A scoping review. *Nutrients* 16 (13):2066. https://www.mdpi.com/2072-6643/16/13/2066. doi: 10.3390/nu16132066.

Springer Nature Ltd. 2024. Food loss and waste. *Nature Food* 5 (8):639–doi: 10.1038/s43016-024-01041-7.

Stewart-Knox, B., A. Rankin, S. Kuznesof, R. Poínhos, M. D. Vaz de Almeida, A. Fischer, and L. J. Frewer. 2015. Promoting healthy dietary behaviour through personalised nutrition: Technology push or technology pull? *The Proceedings of the Nutrition Society* 74 (2):171–6. doi: 10.1017/S0029665114001529.

Stewart-Knox, B., J. Markovina, A. Rankin, B. P. Bunting, S. Kuznesof, A. R. H. Fischer, I. A. van der Lans, R. Poínhos, M. D. V. de Almeida, L. Panzone, et al. 2016. Making personalised nutrition the easy choice: Creating policies to break down the barriers and reap the benefits. *Food Policy*. 63:134–44. doi: 10.1016/j.foodpol.2016.08.001.

Stewart-Knox, B., S. Kuznesof, J. Robinson, A. Rankin, K. Orr, M. Duffy, R. Poínhos, M. D. de Almeida, A. Macready, C. Gallagher, et al. 2013. Factors influencing European consumer uptake of personalised nutrition. Results of a qualitative analysis. *Appetite* 66:67–74. doi: 10.1016/j.appet.2013.03.001.

Strimbu, K., and J. A. Tavel. 2010. What are biomarkers? *Current Opinion in HIV and AIDS* 5 (6):463–6. doi: 10.1097/COH.0b013e32833ed177.

Survey: The Purchase to Plate Price Tool. https://www.ers.usda.gov/webdocs/publications/99295/tb-1955_summary.pdf?v=5122.8.

Tanwar, A. S., N. Evangelatos, J. Venne, L. A. Ogilvie, K. Satyamoorthy, and A. Brand. 2021. Global open health data cooperatives cloud in an era of COVID-19 and planetary health. *Omics: A Journal of Integrative Biology* 25 (3):169–75. doi: 10.1089/omi.2020.0134.

Taylor, R. M., R. L. Haslam, T. L. Burrows, K. R. Duncanson, L. M. Ashton, M. E. Rollo, V. A. Shrewsbury, T. L. Schumacher, and C. E. Collins. 2019. Issues in measuring and interpreting diet and its contribution to obesity. *Current Obesity Reports* 8 (2):53–65. doi: 10.1007/s13679-019-00336-2.

U.K. National Institute for Health and Care Excellence. 2024. NICE guidelines. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines.

U.S. Congress. 2018. Agriculture improvement act of 2018. https://www.congress.gov/bill/115th-congress/house-bill/2.

U.S. Congress. 2022. H.R.8152 - American data privacy and protection act. https://www.congress.gov/bill/117th-congress/house-bill/8152/text.

U.S. Department of Agriculture, and U.S. Department of Health and Human Services. 2020. Dietary guidelines for Americans, 2020-2025. 9th ed. https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf.

U.S. Department of Agriculture. 2018. Economic research service. America's eating habits: Food away from home. https://www.ers.usda.gov/webdocs/publications/90228/eib-196.pdf.

U.S. Department of Agriculture. 2022. Argricultural research service. What we eat in America. https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/wweianhanes-overview/.

U.S. Department of Agriculture. 2023a. Agricultural research service. Food patterns equivalents database. https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/fped-databases/.

U.S. Department of Agriculture. 2023b. Food and nutrition service. SNAP and the thrifty food plan. https://www.fns.usda.gov/snap/thriftyfoodplan.

U.S. Department of Agriculture. 2024a. Agricultural research service. food and nutrient database for dietary studies. https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/fndds/.

U.S. Department of Agriculture. 2024b. Economic research service. Data products: purchase to plate. https://www.ers.usda.gov/data-products/purchase-to-plate/.

U.S. Department of Agriculture. 2024c. Economic research service. Data products: total food spending reached $2.6 trillion in 2023.

https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=58364.

U.S. Department of Agriculture. 2024d. Economic research service. Food prices, expenditures, and establishements: Using proprietary data. https://www.ers.usda.gov/topics/food-markets-prices/food-prices-expenditures-and-establishments/using-proprietary-data/.

U.S. Department of Agriculture. 2024e. Economic research service. Food-related data sources. https://www.ers.usda.gov/about-ers/partnerships/strengthening-statistics-through-the-icars/food-related-data-sources/.

U.S. Department of Agriculture. 2024f. Food and nutrition service. Research policy documents. https://www.fns.usda.gov/research/policy.

U.S. Department of Agriculture. 2024g. Food and nutrition service. Special supplemental nutrition program for women, infants, and children (WIC). https://www.fns.usda.gov/wic.

U.S. Department of Agriculture. 2024h. Food and nutrition service. Supplemental nutrition assistance program (SNAP). https://www.fns.usda.gov/snap/supplemental-nutrition-assistance-program.

U.S. Department of Agriculture. 2024i. Food and nutrition service. Thrify food plan, 2021. https://www.fns.usda.gov/cnpp/thrifty-food-plan-2021.

U.S. Department of Agriculture. 2024j. Food and nutrition service. USDA food plans: Monthly cost of food reports. https://www.fns.usda.gov/cnpp/usda-food-plans-cost-food-monthly-reports.

U.S. Department of Agriculture. 2024k. USDA food plans. https://www.fns.usda.gov/research/cnpp/usda-food-plans.

U.S. Department of Agriculture. 2024l. Why should we care about food waste? https://www.usda.gov/foodlossandwaste/why.

U.S. Department of Health and Human Services. 2024a. Centers for disease control. National Center for Health Statistics. National Health and Nutrition Examination Survey. https://www.cdc.gov/nchs/nhanes/index.htm.

U.S. Department of Health and Human Services. 2024b. Health information privacy. https://www.hhs.gov/hipaa/index.html.

U.S. House of Representatives. 2012. Supplemental nutrition assistance program. 7 U.S.C. § 2012(u). https://uscode.house.gov/view.xhtml?req=(title:7%20section:2012%20edition:prelim)%20OR%20(granuleid:USC-prelim-title7-section2012)&f=treesort&edition=prelim&num=0&jumpTo=true.

U.S. National Institutes of Health. 2024a. All of Us Research Program. https://allofus.nih.gov/.

U.S. National Institutes of Health. 2024b. Nutrition for precision health, powered by the All of Us Research Program. https://commonfund.nih.gov/nutritionforprecisionhealth.

U.S. State of California. 2024. Department of Justice. California Consumer Privacy Act (CCPA). https://oag.ca.gov/privacy/ccpa.

van Ommen, B., T. van den Broek, I. de Hoogh, M. van Erk, E. van Someren, T. Rouhani-Rankouhi, J. C. Anthony, K. Hogenelst, W. Pasman, A. Boorsma, et al. 2017. Systems biology of personalized nutrition. *Nutrition Reviews* 75 (8):579–99. doi: 10.1093/nutrit/nux029.

Verma, M., R. Hontecillas, N. Tubau-Juni, V. Abedi, and J. Bassaganya-Riera. 2018. Challenges in personalized nutrition and health. *Frontiers in Nutrition* 5:117. doi: 10.3389/fnut.2018.00117.

Villinger, K., D. R. Wahl, H. Boeing, H. T. Schupp, and B. Renner. 2019. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: A systematic review and meta-analysis. *Obesity Reviews: An Official Journal of the International Association for the Study of Obesity* 20 (10):1465–84. doi: 10.1111/obr.12903.

Wald, D. S., M. Law, and J. K. Morris. 2002. Homocysteine and cardiovascular disease: Evidence on causality from a meta-analysis. *BMJ (Clinical Research ed.)* 325 (7374):1202–6. doi: 10.1136/bmj.325.7374.1202.

Weir, C. B., and A. Jan. 2024. BMI classification percentile and cut off points. In *StatPearls*. Treasure Island, FL: StatPearls Publishing LLC.

Westerman, K., A. Reaver, C. Roy, M. Ploch, E. Sharoni, B. Nogal, D. A. Sinclair, D. L. Katz, J. B. Blumberg, and G. Blander. 2018. Longitudinal analysis of biomarker data from a personalized nutrition platform in healthy subjects. *Scientific Reports* 8 (1):14685. doi: 10.1038/s41598-018-33008-7.

Wiggins, A., and J. Wilbanks. 2019. The rise of citizen science in health and biomedical research. *The American Journal of Bioethics: AJOB* 19 (8):3–14. doi: 10.1080/15265161.2019.1619859.

Wopereis, S., J. H. M. Stroeve, A. Stafleu, G. C. M. Bakker, J. Burggraaf, M. J. van Erk, L. Pellis, R. Boessen, A. A. F. Kardinaal, and B. van Ommen. 2017. Multi-parameter comparison of a standardized mixed meal tolerance test in healthy and type 2 diabetic subjects: The PhenFlex challenge. *Genes & Nutrition* 12 (1):21. doi: 10.1186/s12263-017-0570-6.

World Health Organization. 2008. Waist circumference and waist-hip ratio: Report of a WHO expert consultation. https://www.who.int/publications/i/item/9789241501491.

Zahedani, A. D., T. McLaughlin, A. Veluvali, N. Aghaeepour, A. Hosseinian, S. Agarwal, J. Ruan, S. Tripathi, M. Woodward, N. Hashemi, et al. 2023. Digital health application integrating wearable data and behavioral patterns improves metabolic health. *NPJ Digital Medicine* 6 (1):216. doi: 10.1038/s41746-023-00956-y.

Zhang, X., Q. Qi, C. Zhang, S. R. Smith, F. B. Hu, F. M. Sacks, G. A. Bray, and L. Qi. 2012. FTO genotype and 2-year change in body composition and fat distribution in response to weight-loss diets: The POUNDS LOST Trial. *Diabetes* 61 (11):3005–11. doi: 10.2337/db11-1799.

Zoe. 2023. Whitepaper: The PREDICT program. https://zoe.com/whitepapers/the-predict-program.

Appendix A

Data Type and Speakers	Title and Affiliations at the time of the Workshop	Presentation Title
Consumer Purchasing Data		
Kris Sollid, RD	Senior Director, Nutrition Communications, International Food Information Council	Nourishing choices: Influential dynamics of personal food decisions
Kevin Meyers Mathieu, MPH	Economist, Center for Nutrition Policy and Promotion, Food and Nutrition Service, United States Department of Agriculture	Consumer purchasing, data linkage, analysis and opportunities: The Thrifty Food Plan, 2021
Ying Bao, PhD	Assistant Professor, Business Administration, University of Illinois Urbana-Champaign	Body mass index, food purchase and promotional sensitivity
Health and Biological Data		
Kate Bermingham, PhD	Postdoctoral Research Assistant, King's College London and Senior Scientist, ZOE	Personalized nutrition data challenges and opportunities: Health and biological data
Gil Blander, PhD	Founder and Chief Scientific Officer, InsideTracker	Lessons learned from analysis of InsideTracker user data
Sheri Schully, PhD	Deputy Chief Medical and Scientific Officer, *All of Us* Research Program, National Institutes of Health	The *All of Us* Research Program: Data collection and harmonization
Bruce Y. Lee, MD, MBA	Professor of Health Policy and Management, City University of New York School of Public Health	Personalized nutrition data challenges and opportunities: Incorporating a systems approach
Social & Behavioral Data		
Mariette Abrahams, PhD, MBA	Founder and CEO, Qina Ltd., Olhao, Portugal	In a sea of data -We're drowning of inaction
Victor Penev, BA, MBA	Founder and CEO, Edamam LLC	Understanding food composition and nutrient profile to support personalized dietary recommendations
Machiel Reinders, PhD	Senior Researcher, Consumer Behavior and Marketing, Wageningen University and Research, Wageningen Economic Research, The Netherlands	Data in personalized nutrition: Challenges and opportunities for behavioral data
Background on Current Personalized Nutrition Guidelines		
Sumeet Thosar, MS	Bioengineering, University of Illinois Urbana-Champaign and currently second-year medical student, Kansas City University	Personalized nutrition data sharing guidelines and regulatory guidelines
Maribel Barragan, PhD, RD	Personalized Nutrition Initiative, University of Illinois Urbana-Champaign	