Web-Enabled and Improved Software Tools and Data Are Needed to Measure Nutrient Intakes and Physical Activity for Personalized Health Research

Phyllis J. Stumbo, Rick Weiss, John W. Newman, Jean A. Pennington, Katherine L. Tucker, Paddy L. Wiesenfeld, Anne-Kathrin Illner, David M. Klurfeld, and Jim Kaput

Introduction

Omic technologies open new avenues to produce discipline-specific research knowledge that is transforming biomedical research. Omics are generally defined as any high-information content analyses of a large number of analytical targets belonging to organizational subclasses such as genes, metabolites, proteins, and transcripts. Combining results from several omic technologies in 1 study to more fully characterize and quantify physiological processes and genetic makeup is fundamental to the promise of creating science-based, personalized nutrition, medicine, and healthcare.

The breadth of a population’s response to normal and stressful environmental exposures is defined by the extent of individual responses within that population, with genetic variation being a major contributor to this variance. Thus, the integration of datasets from deep phenotyping (1) and genotyping is necessary, but likely insufficient, to develop the knowledge base for personalized healthcare. Metabolic phenotypes, whether healthy or diseased, result from complex interactions between an individual’s genetic makeup and his or her environment. Because food is required for survival, nutrients and other bioactive compounds are among the most important environmental factors to alter the expression of genomic information [reviewed in (2,3)]. However, many genomic and omic studies fail to account for not only the energy intake but also the types and amounts of nutrients consumed (4). Moreover, the fate of consumed energy is determined by combining the energy expenditure and metabolic efficiency in fuel use of an individual, which also is affected by genetic makeup. Thus, the quality and quantity of the diet, in conjunction with energetic demand and fitness of the individual, are critical variables affecting health maintenance. The dramatic rise in obesity and nutrition-related chronic diseases in the past ~25 y confirms the importance of energy balance in maintaining health (5).

The “silo” nature of biomedical research (most advances develop vertically within the discipline rather than horizontally between and among multiple fields) is often blamed for the...
failure to integrate nutritional and physical activity (PA)\textsuperscript{13} data in human studies. Perhaps a more prominent impediment is the lack of centralized online tools and databases. Dietary assessment advances like the Automated Multiple Pass Method (AMPM) (6,13), a computerized, 24-h recall tool for assessing nutrient intake and adequacy, and other nutrient assessment instruments are currently seldom used as Web-based applications. Exceptions include the ASA24, which is a Web-enabled version of the AMPM instrument (7,8). PA assessments are similarly limited. Most nutritional and physical assessment tools for research are still paper based.

Assessment of both nutrition and activity have long histories of research and development, but the tools most widely used are self-reported, with the concomitant concern for a lack of objectivity required and the precision needed to capture the subtle differences predicted to affect genetic expression. A workshop was cosponsored by the FDA, NIH, and USDA and held at the USDA-ARS Beltsville facilities in spring 2009 to assess: 1) availability of nutrition, activity, and genetic tools for researchers; 2) what additional resources are needed; and 3) how to proceed to develop and deploy the necessary assessment tools for biomedical researchers. Ten scientists spoke on the promise and limitations of research techniques in their respective fields (Supplemental Table 1). This paper summarizes their presentations (posted at http://www.ars.usda.gov/research/programs/programs.htm?np_code = 107&docid = 18952&dropcache = true&cmode = preview).

Food composition databases

The current food databases. Food composition databases are a prerequisite for any type of dietary assessment methodology. Food composition databases are multi-functional and used to: 1) evaluate data from food consumption surveys/studies; 2) plan meals/diets; 3) perform dietetic counseling/teaching; 4) assess national food/food component adequacy; 5) guide nutritional regulatory and health policy decisions; 6) populate nutrition labels; 7) develop new food products; and 8) evaluate data from food consumption surveys and studies (9).

Food composition databases are works in progress, because the cost of analyzing chemical composition is about $12,000 per food (10) and the food supply is in a constant state of flux, with new foods and new formulations for old foods continually brought to market. In addition, the average supermarket contains ~45,000 products (11), with many new food products replacing older versions. Hence an accurate catalog of available products, the appropriate sampling and costs of analyses of foods, and the many new combinations of food components (e.g. ginseng iced tea) exceeds resources for compiling and maintaining a comprehensive food composition database.

More accurate and complete databases are needed not only for basic research but also for improved clinical trials, because the assessments of nutrient intakes and PA of participants are generally weak links in data acquisition and analysis. Improving the assessment of environmental factors for clinical and research studies will not be possible without further development of tools and techniques, especially those related to food composition databases and dietary assessment methodologies (12,13).

Several food composition databases exist, most notably the USDA Standard Reference (SR) food composition database (14). Selected reference and user U.S. databases and activities are summarized in Table 1. The Canadian Nutrient File is based on the US SR database, with revisions made to include Canadian-only foods and to adjust values to conform to Canadian levels of fortification and regulatory standards (15). The European Union has developed and continues to improve the comparable European Food Information Resource Network EuroFir database (16). The FAO of the United Nations (17) and various national governments have developed food composition tables for many countries worldwide, but data must often be extracted from unlinked flat files or from publications. Some of the country-specific food composition tables are based on extrapolations from databases in Western countries. Nevertheless, progress in developing national and international food composition databases is being made, as noted by the International Life Sciences Institute Crop Composition Database (18) and the aforementioned International Network of Food Data Systems and EuroFir databases. The International Food Data Conference meets every 2 y (19) to describe progress made around the globe to document the nutritional properties of the world food supply. The US National Nutrient Database Conference (NNDC) meets annually for a similar purpose within the US. A committee of the NNDC has compiled the International Nutrient Databank Directory, listing 32 databases currently available (20).

Data sources for food composition. The U.S. food composition databases (21,22) are developed by searching the scientific literature for new analyzed food composition values, issuing contracts to universities and industries for laboratory analysis of specific components in food, securing analytical values and food label values from food industries for their products, searching for published values in other U.S. and international databases, through calculations (for foods with more than 1 ingredient), or using values from similar foods.

Although many nutrition journals publish food composition data, the Journal of Food Composition and Analysis (23) is the primary publication for nutrient analyses of food. The Journal of Food Composition and Analysis is an official publication of the International Network of Food Data Systems (24) of the United Nations University and is sponsored by the FAO of the United Nations. The journal publishes data on the chemical composition of human foods, analytical methods, and studies on the manipulation, storage, distribution, and use of food composition data. Numerous factors affect nutrient composition of food, making this peer-reviewed journal a valuable asset to the scientific community. Reviewers evaluate how samples were selected, stored, and prepared, whether plant or animal breed is identified, quality of the soil, if water and weather are adequately documented, appropriateness of analytical methods, quality control in the laboratory, number of samples analyzed, and statistical methods used in determining the values reported.

The U.S. National Food and Nutrient Analysis Program (25) is an interagency collaboration led by the Nutrient Data Laboratory of the USDA-ARS, which includes the NIH’s National Cancer Institute and the Office of Dietary Supplements, other NIH offices, institutes, and the FDA. These agencies support the USDA-ARS, who in turn contract food analyses from universities and private laboratories. Because the analyses of food are expensive, priorities are set for the foods and components to be analyzed. These decisions are based on

\textsuperscript{13} Abbreviations used: AMPM, Automated Multiple-Pass Method; ASA24, Automated Self-administered 24-h dietary recall; EAFUS, Everything Added to Food in the United States; EPIC, European Prospective Investigation into Cancer and Nutrition; FCD, Food Composition Database; FLAPS, Food Label and Packaging Survey; FNDDS, Food and Nutrient Database for Dietary Studies; GRAS, Generally Recognized as Safe; GWAS, genome-wide association studies; HRV, heart rate variability; NNDC, National Nutrient Databank Conference; PA, physical activity; SR, USDA Standard Reference.
<table>
<thead>
<tr>
<th>Database</th>
<th>Agency</th>
<th>Foods</th>
<th>Components</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dietary Supplements Ingredient</td>
<td>NIH Office of Dietary Supplements and USDA-ARS Nutrient Data Laboratory</td>
<td>Representative multivitamin/mineral supplements</td>
<td>Folic acid, vitamin C, retinol, β-carotene, vitamin E (α-tocopherol), calcium, iron, riboflavin, thiamin, niacin, vitamin B-6, vitamin B-12, vitamin D, vitamin K, phosphorus, potassium, copper, selenium, chromium, manganese, magnesium, zinc, and iodine</td>
<td><a href="http://dietarysupplementdatabase.usda.nih.gov/Searchable">http://dietarysupplementdatabase.usda.nih.gov/Searchable</a> and Downloadable in ACCESS, EXCEL</td>
</tr>
<tr>
<td>EARUS database</td>
<td>CFSAN of FDA</td>
<td>Ingredients added to manufactured foods</td>
<td>Administrative, chemical, and toxicological information on over 2000 substances listed as direct, indirect, color additive, GRAS additives. FDA has approved their use in foods or they have independent GRAS determinations.</td>
<td><a href="http://www.fda.gov/Food/FoodIngredientsPacking/ucm115326.htm">http://www.fda.gov/Food/FoodIngredientsPacking/ucm115326.htm</a></td>
</tr>
<tr>
<td>FLAPS</td>
<td>CFSAN of FDA; NIH Office of Dietary Supplements</td>
<td>Nationally representative sample of the processed packaged food products</td>
<td>Package and serving size, number of servings per package, nutrition values for 1 serving, ingredients, presence of health claims, nutrient content claim, structure function claims (e.g. calcium for strong bones), dietary guidance statements and other recommendations, food safety information, and other additional information (e.g. organic, genetically engineered, percent juice, vegetarian).</td>
<td><a href="http://www.fda.gov/Food/LabelingNutrition/ConsumerInformation/ucm122084.htm">http://www.fda.gov/Food/LabelingNutrition/ConsumerInformation/ucm122084.htm</a></td>
</tr>
<tr>
<td>FNDDS</td>
<td>USDA ARS Food Surveys Research Group</td>
<td>~7000</td>
<td>63 (missing values are imputed)</td>
<td><a href="http://www.ars.usda.gov/Services/docs.htm?docid=17031Searchable">http://www.ars.usda.gov/Services/docs.htm?docid=17031Searchable</a> and Downloadable in ACCESS, EXCEL</td>
</tr>
<tr>
<td>Nutrition Coordinating Center Food and Nutrient Database</td>
<td>University of Minnesota</td>
<td>18,000 foods (7,000 brand name products)</td>
<td>160 nutrient, nutrient ratios and other components</td>
<td><a href="http://www.ncc.umn.edu/products/ndsrfeatures.html">http://www.ncc.umn.edu/products/ndsrfeatures.html</a></td>
</tr>
<tr>
<td>SR22</td>
<td>USDA ARS Nutrient Data Laboratory</td>
<td>75,38</td>
<td>143</td>
<td><a href="http://www.ars.usda.gov/Services/docs.htm?docid=8964Searchable">http://www.ars.usda.gov/Services/docs.htm?docid=8964Searchable</a> and Downloadable in ACCESS, EXCEL</td>
</tr>
<tr>
<td>Special Interest Databases</td>
<td>USDA ARS Nutrient Data Laboratory</td>
<td>Varies</td>
<td>Added sugars, choline, flavonoids, isoflavones, proanthocyanidins, oxalic acid, fluoride, oxygen radical absorbance capacity</td>
<td><a href="http://www.ars.usda.gov/Services/docs.htm?docid=5121Searchable">http://www.ars.usda.gov/Services/docs.htm?docid=5121Searchable</a> and Downloadable in ACCESS, EXCEL</td>
</tr>
<tr>
<td>Total Diet Study</td>
<td>FDA</td>
<td>~280</td>
<td>11 nutritional elements, radionuclides, pesticide residues, industrial chemicals, 4 toxic elements</td>
<td><a href="http://www.fda.gov/Food/FoodSafety/Downloadable">http://www.fda.gov/Food/FoodSafety/Downloadable</a> in PDF</td>
</tr>
</tbody>
</table>

**TABLE 1**: Summary of U.S. food composition databases
the selection of foods that provide >75% of nutrients in the American diet and the need for new food component data to support novel research.

The food industry is required to provide nutritional labeling of foods. However, the nutrients reported are limited to macronutrients, sodium, vitamins A and C, iron, and calcium, whereas FCD may list 100 or more food components. When label databases are used as the only data source, most FCD components are incomplete. The ingredient label can be used as a recipe to generate complete food profiles using mathematical procedures to create recipes that match label-based macronutrient values using ingredients from a comprehensive (complete) database (26). An underappreciated limitation of food nutrient profiles for manufactured foods is that each company has an internal database of nutrient compositions; no national registry exists, making it difficult to compare across products and to track nutrient intakes for individuals. The FDA conducts Food Label and Package Surveys [FLAPS (27)] of representative food products that includes package size, serving size, number of servings per package, label nutritional values for 1 serving, product ingredients, allergen labeling, presence of health or nutrient content or structure function claims, and food safety information. FLAPS data are used to monitor manufacturers’ responses to food labeling regulations and guidance to support food safety and regulatory and policy decisions.

Quality of the databases. All food databases contain data merged from various sources, which are of uneven quality, reliability, representativeness, and accuracy. Although some variability can be minimized (choice of cultivar, region sampled), many environmental factors are inherently variable (temperature, rainfall, amount of sun), which could alter the chemical composition of the plants and therefore the analytical values. Different procedures for food processing, such as the removal of fat and the addition of vitamins and/or minerals may also produce the same product with variation in composition either by design or by the inherent variability in manufacturing methodology.

The values in FCD are influenced by differences in laboratory procedures such as the number of samples analyzed, preparation methods, instrument methods, and postanalyses statistical methodologies. FCD usually provide the mean but often without clear indications of the ranges, SD, and number of samples analyzed (the recommended minimum number of samples is 6) (10). Nationally representative data may differ from the individual food grown in different regions in a different environment with different times and conditions between harvest, market, and consumption. Because of the cost of analyses, databases are not comprehensive, and nutrient contents are often matched to the most similar food.

Improving food composition databases. The recognized limitations of the FCD provide guidance for their improvement (28). Metadata should be linked to each value documenting analytical procedures. For example, the number of samples, ranges, and standard deviations should be available. A composite database might be developed from existing databases by merging all data or by selecting the best-characterized datasets. This would require a consistent and common terminology for food names and descriptors (i.e. a standardized food ontology).

Because the demographics and food consumption patterns of many countries are changing, data for cultural or ethnic foods also need to be incorporated into most national FCD. Data for fast foods, restaurant, and manufactured foods should be incorporated and updated continually.

The omics revolution also has generated a need for linking bioactive food components in FCD to metabolomic, proteomic, and genomic data. Linking these data may seem premature given the state of gene-nutrient interaction datasets, but dynamic linkages may allow referencing the latest data or experimental needs. New methodologies, such as image recognition algorithms for photographic applications, also need to be documented (29).

Current methods for assessing nutrient intake

Comparable assessment methods for nutrient intake are fragmented by experiment, study, and by individual laboratory (30). Early nutrition research focused on small groups of individuals fed metabolic diets that were highly defined or characterized (31). These types of data could be stored in local databases, often on a laboratory computer. The growing interest and emphasis on personalization is likely to see a revival in these types of studies, even though the statistical analyses of results will be challenging. Many of the recent nutrition studies have focused on large populations or groups (32,33) and have used the FFQ as the dietary assessment method due to ease, time, cost, and ability to retrieve dietary consumption over time. The NHANES studies involving large numbers of people were done with 24-h recall assessments in person and a second 24-h recall via telephone. The weighed food record method and 7-d food record method have been used to assess usual food and nutrient intake per eating occasion, but the proposed number of days of food records required to establish confidence in usual intake of different nutrients varies depending on the nutrient but can be as few as 3 d (for energy intake assessment in groups) up to 474 d for vitamin A in individual males (34).

The Nutrition Data System for Research. The Nutrition Data System for Research (35) is a publicly available, Windows-based (but not Web-enabled) dietary intake assessment program designed for the collection and analyses of 24-h dietary recalls and the analysis of food records, menus, and recipes. This tool was developed with funding from the National Heart, Lung and Blood Institute to support their early studies in heart disease. Nutrients from food records and recalls are calculated in real time to provide data on ingredients, food, meal, and day. A dietary supplement assessment module is included so nutrient intake from both food and supplemental sources may be captured and quantified. The source of food composition information in the program is the NCC Food and Nutrient Database (36), which includes over 18,000 foods, among which are 7000 brand name products. More than 160,000 food variants are listed based on ingredient choices and preparation methods. Food group assignments are also provided. The database is updated annually to reflect marketplace changes and new analytic data.

Use of FFQ and other diet surveys compatible with omics studies. The FFQ is a technique commonly used in epidemiology to assess usual intake over a period of time. Available instruments have several formats and methods of administration, with paper-based products (37) gradually being supplemented by online versions (38). Each have strengths and limitations, but for now, the FFQ remains the most cost-effective tool for usual food intake in large populations (39). However, assessment methods and databases needed to study geographically or ethnically diverse groups within feasible time frames, costs, and respondent burden are lacking. For the foreseeable future, individualized nutrition interventions need to be food based to translate into counseling for clients, because ethnic
foods are poorly represented in current tools. Comparisons of the FFQ and multiple recalls with recovery biomarkers like doubly labeled water and urinary nitrogen have demonstrated limitations, particularly of the FFQ for macronutrient intake (40). Primary limitations are assumptions made for the sake of simplicity: the defined food list and, when included, defined limitations in portion size. Additional assumptions about recipes and preparation may bias results toward or against a particular regional population or ethnic group, because decisions are generally made to represent the majority of the population. Assessments in multi-ethnic populations have shown poor results in African American and Hispanic groups (41–43).

Even when major foods consumed by these populations are added, recipes and portion sizes usually are not adequately adjusted. The current tools based on validations in the non-Hispanic white majority U.S. population is of considerable concern, because these differences are not random and may lead to biases in diet-disease relationships if the populations under study do not match this demographic.

The FFQ in nutrigenomics. The acceleration of major research efforts toward gene-diet interactions highlights the limitations of current assessment methods and potential biases in assessment across groups that may be associated with other phenotypes, adding to the risk of false findings. Although multiple 24-h recalls allow for a full description of dietary differences, they remain expensive, are often difficult to administer, and do not capture usual long-term intake. Web-based technology addresses many of these limitations.

As with most technologies for research, the goal for Web-enabled tools is to meet the needs of both study participant and researcher. Comprehensive, paper-based FFQ using exhaustive food lists result in response fatigue and subsequent error-prone manual entry into databases. Online interviewing could permit users to skip foods in whole groups that they do not eat, such as meats when preliminary questions indicate a vegetarian style of eating. By using these skip patterns, the food list, portion size options, and food preparation methods for the appropriate foods can be greatly expanded to meet the requirements of diverse groups without substantially increasing the response time. This should allow precision similar to that achieved with a multi-pass 24-h recall but for usual intake rather than for previous day only. Using large groups of 24-h recalls, validated FFQ have been designed specifically for underserved ethnic groups, including Caribbean-origin Hispanics, Mexican Americans, and African Americans in the southern United States. Two of us (R.W. and K.L.T.) are using these data to add ethnic foods consumed in the US to an existing FFQ in such a way that all ethnic foods will be available to all users, but skip patterns will be available to eliminate excess length.

AMPM for food recalls for large groups. The USDA’s AMPM was developed and is continually improved by the USDA Food Surveys Research Group (6) and has been continuously used since 2002 as the dietary interview method for What We Eat in America, the NHANES national survey (44). The collection consists of 2 d of dietary intake data for over 9000 individuals every 2 y. Day 1 of the dietary recall is administered in person and day 2 telephonically. Bilingual interviews are available. The AMPM involves 5 steps: 1) Quick List, a listing of all foods and beverages; 2) Forgotten Foods, probes for forgotten foods in 9 categories (50% of respondents remembered foods when asked); 3) Time and Occasion, collected for each food and beverage; 4) Detail Cycle, collects description of each food, additions, amount eaten, source, and whether eaten at home, reviews each occasion and intervals between occasions; and 5) a Final Probe that provides a final opportunity to recall foods. Other programs and research initiatives also rely on the USDA, AMPM (Supplemental Table 2).

AMPM assessed mean energy intakes within 11% of energy expenditure as measured by doubly labeled water experiments in a large sample of adults and within <3% in normal-weight participants (6). Further research is needed to determine what may be contributing to the underreporting observed in overweight and obese participants.

Web-based, automated, self-administered 24-h dietary recall. The interviewer-administered 24-h recall such as the AMPM is regarded as an optimal tool for dietary data collection. However, the statistical modeling of usual intake requires at least 2 recalls per study participant and then generates data appropriate only for group analysis, which is hardly adequate for making genetic comparisons unless on a broad scale. Recalls traditionally require administration by trained professionals, making them too costly and impractical for large-scale studies. The US National Cancer Institute began developing a Web-based, automated, self-administered 24-h dietary recall (ASA24) in 2005. The goals were to create an engaging, easy-to-use application to mimic the quality of an in-person interview and to provide the software application via a public Web site (http://riskfactor.cancer.gov/tools/instruments/asa24/info.html) for use by investigators and clinicians at minimal or no cost. Many days of intake are still required for valid genetic studies, but the ASA24 will make this practice possible.

ASA24 is modeled after the AMPM for collecting 24-h dietary recalls but is a self-administered tool with a tutorial, multiple help screens, an animated agent who guides participants, and computer-generated audio cues. The list from which respondents identify yesterday’s foods and beverages is based on the AMPM interview system. After listing their foods, respondents are asked detailed questions about food descriptions and amounts consumed. The software includes pictures of foods in multiple portion sizes to aid in portion size estimation. Access to ASA24 is managed via the ASA24 Researcher Web site (45), where researchers and clinicians can configure options for their studies. These include specifying recall dates, number of recalls allowed or desired, and makeup dates for each participant. The Web site provides researchers with a tool to select among ASA24 optional modules such as dietary supplement use, where meals were eaten and with whom, and whether the TV was viewed at each meal. The investigator may request either immediate or batched data analyses to obtain nutrient and food group intake estimates for study participants based on USDA’s Food and Nutrient Database for Dietary Studies (FNDDS) and My Pyramid Equivalents Database.

The development of a 24-h recall that could be unannounced, automated, and self-administered makes feasible the administration of multiple recalls in large-scale epidemiologic studies, behavioral trials, or clinical research, thus enhancing investigators’ and practitioners’ ability to accurately assess dietary intakes.

The Web-based European Food Propensity Questionnaire. A similar series of developments in dietary software is underway in Europe. The objectives of the European Food Propensity Questionnaire (EFPQ [46]) are to develop a Web-based, multilingual questionnaire and to investigate the feasibility of using the EFPQ in combination with 3 monthly telephone-administered
24-h dietary recalls in a multi-centric pilot study including 5 European countries.

The multilingual EFPQ was initially based on a validated German Food Frequency Questionnaire developed for use in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study (47) consisting of 102 food items and standardized portion sizes. Non-German experts were invited to review an English-translated, Web-based version of the initial FFQ to determine whether the FFQ reflects the dietary habits in their countries. The experts could decide online with predefined choices whether each item should be kept, modified to a country-specific food, or deleted. Additional comments could be entered into free text areas on each screen. Based on these evaluation results and further communications, the final EFPQ was constructed and translated into German, Estonian, Norwegian, Italian, Spanish, and Basque languages.

The common EFPQ queries the frequency of consumption of 116 foods, food groups, and beverages during the preceding year and asks about selected food preparation preferences, food choices, and the use of nutritional supplements and medications. The food item queries are supported by pictograms of approximate portion sizes. The EFPQ was built as part of a tool set for epidemiological studies and the Web-based EFPQ was designed to enable flexible and easy navigation with freedoms similar to the paper format. This allows generating, from the Web-based EFPQ, a paper-based version to also satisfy study participants not using the Web. An encrypted connection is used to ensure a secure electronic transmission of the data. Four hundred participants from 5 ongoing cohort studies in Estonia (Estonia Genome Centre, Tartu), Italy (EPIC-Florence), Spain (EPIC-San Sebastian), Norway (EPIC-Norway, Tromsoe), and Germany (EPIC-Potsdam) were randomly selected and invited to complete the EFPQ in their own language by using either the Web-based or paper format.

Response rates for the Web-based EFPQ vary among the study centers from 34.9% in Germany to 70.0% in Norway and 92.9% in Estonia, indicating the high potential of Web-based instruments for dietary assessment. In addition, the use of a combination of traditional and new technical dietary assessment instruments seems to be a feasible and promising approach for future dietary assessment methodology.

The Web technology-assisted collaborative effort made it possible to develop a common European FFQ in a relatively short time. The Web-based EFPQ was shown to be a feasible and suitable dietary assessment instrument in multi-centric cohort studies. Further development to refine its methodology will make it suitable for use in other countries.

Consumer activities and tools. The consumer has access to a large number of internet- or phone-based applications to gain knowledge or track nutrient intake and physical activities. For example, the iPhone has applications for tracking food intake and accessing a food information database, several for healthy recipes and fast food energy counting (48). The USDA’s online Pyramid Tracker (49) enables consumers to enter food records and track their intake and weight, if desired, on a continuous basis. This tool uses the same FNDDS database used by ASA24 and the AMPM. A commercial application for consumers developed by Safeway supermarket called Food Flex uses the food market’s club card to track a household’s food, and therefore nutrient, purchases (50). This Web-based program provides charts and graphs to illustrate how the nutritional value of foods purchased at Safeway and its affiliated stores compares to the federal government’s recommended nutritional intake for people of similar age, sex, and activity levels. Nutrient values of the food purchases use USDA data and information on the food labels. The point-of-purchase data can be downloaded to a PDF. Whereas a single food store visit may not be an accurate assessment of nutrient intakes for a household, food purchases for a longer period of time (≥3–4 mo) may provide valuable data for assessing nutrient intakes. To mitigate the obvious limitation that purchases are for a household rather than an individual, algorithms are available that associate food purchases with age, sex, income status, and other demographic factors across tens of thousands of individuals (51). For example, one store may have over 2000 customers and 150,000 transactions within a year (3). Safeway has ~1700 stores, resulting in ~3.4 million customers. Customers purchase not only food but also daily necessities and lifestyle products that help assign them to demographic groups (52). This information is as yet unavailable to research or health care providers but constitutes a potential new source of epidemiological data.

The NIH Genes, Environment, and Health Initiative

Several additional projects are underway in the US to add versatility to the ASA24 concept. As part of exposure biology research, the NIH initiated the GEI (53) and awarded grants to improve measures of diet, PA, biological exposures, and psychosocial measures for use in future large-scale population studies. Four funded GEI projects are using technology to further enhance the accuracy and speed and minimize the intrusiveness of existing diet assessment systems: the Food Intake Recording System version 4, a Web-based, self-administered recall, like ASA24, adapted for children; the Mobile Phone Food Record, software integrating mobile phone digital photography, image processing, and a nutrient database allowing a participant to record foods consumed and pertinent food item details; the Food Intake Visual and Voice Recognizer, a mobile phone digital image recording system with voice recognition to collect food details; and Wearable Device, a software system with a digital camera embedded in a wearable necklace or button that takes pictures of food when swallowing is detected and identifies foods through image recognition processing.

These electronic systems are at different stages of development but hold promise for improving and economizing measurement of food intake in the future.

The state of measuring PA

Assessing nutrient and energy intake is necessary but not sufficient to determine optimal nutrition for health, because energy expenditure dictates metabolic needs. Although many epidemiological studies use self-reported PA assessments, these, like self-reported food intake, are subject to error. Questionnaires are commonly used to obtain information about a respondent’s PA in epidemiological studies, population surveillance, intervention studies, and counseling sessions. A typical questionnaire approach may ask about frequency and duration of a broad and varied class of PA, such as “moderate activities done for at least 10 min.” Many differences are likely in interpretation between what the researcher needs to know and what the respondent reports. Examples of potential discrepancies in PA reporting are actual vs. typical vs. ideal frequency, and absolute (based on energy expenditure) vs. relative (subjective effort) intensity. Cognitive challenges for respondents...
include estimating duration of accumulated short variable activity bouts or providing a single estimate of frequency and duration for varying activities in a pattern such as combining walks at noon 3 times/wk for 20 min with a 2-h hike on the weekend.

PA is a complex behavior that occurs in multiple contexts (occupation, transportation, household tasks, and recreation or sport) and can be expressed in various ways, such as energy expenditure, activity time, or number of steps. Self-report or objective measurement instruments and outcomes should be selected based upon what is relevant for the researcher’s purpose. Different measures may be appropriate for assessing activity level for a population compared with those useful for individuals. Questionnaires that are valid measures of the activity level of a population are not necessarily valid to measure change in activity level of individuals.

Heart rate variability. Quantitative assessments of PA can be done with commercially available heart rate monitors. However, more informative analyses would include heart rate variability (HRV), the beat-to-beat alterations in heart rate mediated by the autonomic nervous system. HRV is affected by physical conditioning but also by psychological and physiological disorders (54–57). The past 5 y have seen an increase and broadening of research incorporating HRV assessment well beyond medical diagnostics and its initial accepted consumer area of athletic training (including use by professional sports teams to maximize training efficiency and safety) (58–61). HRV work has been extended to nutrition, disease management (54,62), mental health, anxiety therapy, meditation, exercise in the elderly, and the effect on diet in the obese (63). Virtually all research utilizing HRV has relied on a small number of “snapshots” varying from 5 min to 48 h over the study interval of weeks to months. Snapshots fail to capture autonomic dynamics much as a 2-h film is poorly represented by a handful of still photographs. Continuous long-term HRV measurement and analysis in health management studies promises to improve the understanding of environmental influences on cancer causation and prevention. Thus, incorporating HRV into biomedical research studies and real life assessments should greatly enhance the quality and utility of data collected and the ability to develop personalized healthcare.

Activity measuring tools. A within-person comparison of minutes of moderate or higher intensity activity from 3087 adults (age ≥20 y) in the 2003–2004 NHANES demonstrated differences between activity reported on a questionnaire and measured by accelerometer worn for 4–7 d. The accelerometer was worn for at least 10 h/d and the questionnaire asked about PA in transportation, household tasks, and sports or activities in recreation or leisure over the previous 30 d. Although accelerometers may miss some movements that may be verbally reported, the accelerometer would capture occupational movement that is not part of the questionnaire. Overall, between 38 and 73% of participants had zero accelerometer-recorded bouts of moderate or higher intensity activity, but their mean self-reported moderate or higher intensity activity ranged from 43 to 65 min/d. Among participants with some measured moderate or higher intensity activity in bouts of 8–10 min, mean self-reported activity time exceeded accelerometer time by 49–67 min/d (64). Cross-classification ranking across quantiles of accelerometer and reported minutes indicated poor agreement, especially among those with low accelerometer minutes. Self-reports of PA behavior differ conceptually from movement measured by an accelerometer. However, these findings suggest that self-reports provide poor quantitative estimates of PA movement as captured by an accelerometer and also lead to substantial misclassification relative to an objective measure of activity time.

Commercial products. Consumers now have a large array of products and Web services to measure PA. Heart rate monitors, accelerometers, pedometers, and other devices can record large datasets to monitor activity. Many of the products allow for uploading data to databases with online graphing and interpretation. As with dietary assessment tools, these resources are generally not used by researchers. However, decreasing costs and increasing attention by developers to data quality are making objective measures of PA more feasible and useful for large-scale research deployment.

Metabolomics in diet and nutrition assessment

Omics technologies are providing novel tools to empirically assess dietary intake, metabolic changes in response to diet, and associations between variances in such phenotypic responses and underlying changes in the genetic code (65–67). In particular, metabolomics is a field devoted to the systematic study of small molecules present in a biological system originating from the cellular processes of that system. Therefore, metabolomic protocols geared toward the broad assessment of small molecules within a sample, coupled with multivariate statistics, are uniquely suited to support the characterization of phenotypic response to diet and nutrition. Such tools can provide quantitative assessments of dietary intake based on measurements of dietary components and/or their metabolites in body fluids, may yield a means to quantitatively assess fitness, and have the potential to reduce the costs associated with measuring nutrients and other small molecules in foodstuffs (68–70).

Metabolomics assessments of nutrient intake in humans. The ability to quantitatively survey all metabolites within biofluids and to assign them to specific dietary and/or biochemical sources will require considerable effort, but advances are being made especially in the use of plant secondary metabolites (65,66). Whereas nontargeted analyses generate complex small molecule profiles containing a mixture of identifiable and unidentifiable compounds, this information can be interpreted with respect to short-term dietary intakes, as would be revealed by 24-h recall assessments. Therefore, biomarkers of consumption could provide complementary tools to 24-h recalls that, although perhaps less comprehensive, could provide cost-effective longitudinal assessments of the variance in cohort diets. To date, urine has been reported as a suitable matrix for the detection of many water-soluble markers of recent dietary intake, whereas plasma and saliva may be less useful for these targets (71–74). Plasma, tissue, and erythrocyte fatty acid profiles are classically applied metabolite-based multivariate biomarkers of longer term dietary lipid intake (75–78).

Current efforts are underway to evaluate the utility of metabolomics as a direct tool for assessing dietary status and intake. One such project is the Metabolomics to Characterize Dietary Exposure, or MEDE, Study being conducted in the United Kingdom as a joint effort by Aberystwyth University and Newcastle University (65). These researchers have found that a limited snapshot of an individuals’ plasma metabolome can be evaluated in a high throughput fashion and is quite stable over
time, but is influenced by diet. Thus, a person sampled on 2 different days in an equivalent nutritional state after feeding a defined diet was more similar to themselves than to other individuals (personal communication, J.C. Mathers, University of Newcastle upon Tyne, UK). Moreover, as a whole, the “directional” shifts in metabolic profiles produced by a standardized meal were very similar in different participants, suggesting that these nutritional “metabolic trajectories” hold information pertaining to diet and nutritional state. Similarly, metabolic fingerprints of plasma in the postprandial state after feeding low- and high-glycemic diets show distinct shifts in phospholipid and small molecule clearance rates (J.W. Newman, unpublished data), suggesting that postprandial time of sampling must also be carefully considered. As a corollary, these findings also argue that to assess variances in nutritionally responsive phenotypes, the use of a defined, and possibly standardized, diet will be necessary.

Open access metabolite/food component databases. To facilitate the application of metabolomic analyses as a phenotyping tool, broadly accessible and searchable databases of food components and metabolites will be needed. Such tools when coupled with broad-spectrum, high throughput quantitative analytical platforms may also allow the efficient and routine assessment of cultivar, if not crop, nutrient composition (79,80). The Pheno-Explorer database of polyphenols consisting of over 60,000 food composition data from the literature (81) is one such example. The current total of mean content values for 502 compounds (glycosides, esters, or aglycones) in 452 foods is available for download. Foods in this database containing polyphenols include grains (generic categories for flour and cereal), fruit, vegetables, seeds, and over 100 spices and herbs. Future applications of such a database could support the measurement of differences in food composition due to differences in cultivation, irrigation, harvesting, storage, and manufacturing practices, as well as cultivar genetics. Currently, the food items do not correspond to foods in common food composition databases and thus do not expand the capability of dietary assessment tools in common use. To address these needs, The Human Metabolome Project (82–84) is developing a searchable database containing all chemical compounds found in foods. A beta version of The FooDB (85) has been released. The FooDB is a Web-accessible database launched with an initial content derived from the FDA’s Everything Added to Food in the United States (EAFUS) list (86). The EAFUS database has information related to food additives and contains administrative, chemical, and toxicological information in over 2000 substances listed as direct, indirect, color additives, and Generally Recognized as Safe (GRAS) additives. This database unfortunately does not indicate which foods contain these compounds, nor the amount present. Regulations give acceptable levels, but not amount actually used, nor does it suggest which foods are likely to contain acceptable amounts. Such data are available only from the manufacturer and have not been added to databases commonly used to assess intake.

Ultimately, an important component of the nutrient databases will be datasets generated by metabolomic analyses of serum and urine linked with reliable assessments of nutrient intakes. A limitation of metabolomic analyses is that many of the biomarkers may respond to short-term nutrient intakes, but health outcomes are typically affected by long-term dietary exposures. Genomic data are likely to be included as well, because individuals differ in their metabolism based on inheritance of genetic information, and these differences may influence the quantitative measures of metabolites. Although these datasets may appear to complicate both the construction of the databases and the analyses of the data, dimensionality reduction and classification algorithms may permit a more complete understanding of the information contained within measured metabolite concentrations.

Metabolomics in fitness and energy assessment. In addition to the ability to influence dietary assessment, nutritional metabolomics may also provide opportunities to assess both fitness and energetic efficiency. For instance, the impact of dietary macronutrient content on the level and clearance of specific metabolic markers in exercised people has been shown to both reflect fitness and health status (69). Moreover, the status of energy metabolism influences metabolite profiles in biofluids that can be perturbed by disease states, a fact used diagnostically for decades to, for instance, identify inborn errors in metabolism (87). With the development of sensitive metabolomic platforms and diverse metabolomic databases, the spectrum of variance in the population’s metabolic profiles will become more refined, as will the development of novel indicators of metabolic fitness (88).

The need for dietary assessment in genome-wide association studies. Although genome-wide association studies (GWAS) have identified hundreds of genetic variants implicated in any of several human diseases and related phenotypes (89), the sum total of loci identified as risk variants explains only a small proportion of the measured phenotype. Because such studies have rarely considered environmental influences on the expression or function of the tested polymorphisms, 2 parallel avenues can be argued for future investigations: adding environmental (i.e., dietary and lifestyle) information to future GWAS and gathering known gene-by-environment interactions into a common database. Although both of these activities have begun, some 150 published or soon-to-be published GWAS, along with many proposed studies, fail to describe gene-by-environment interactions. Most studies of human diseases of metabolic origin (e.g., dyslipidemia and type 2 diabetes) describe these interactions with nutrients on a macro level. These include total energy intake and intakes of carbohydrate, protein, and fat. The latter may be parsed to PUFA or monounsaturated fatty acid or even levels of (n-3) or (n-6) PUFA. Linking metabolomic analyses, dietary assessment databases, and gene-by-environment interaction databases will provide novel tools and approaches to unravel the complexities governing population-wide linkages between nutritional factors and disease risk.

What is needed by the research community

The research community needs online and freely (or inexpensively) accessible software tools with databases that store, manage, and allow retrieval of nutrient intakes and PA measurements for individuals. These tools need to be tested under a variety of circumstances, so they accurately estimate intake for a diverse population. Accessibility to these new instruments should follow the lead of the newly available ASA24 24-h recall tool, which is freely available, extensively studied, and openly documented. Commercial and academic resources are also available that need to be tested and made available to researchers for minimal cost so the best tool can be selected without limits based on budget.
Although there have been many improvements in techniques to better assess diet and PA behaviors, there are still gaps in generating accurate measures with low burden on the individual. As previously discussed, we now have ongoing efforts to assess diet using new computer techniques that improve 24-h recalls, FFQ, and written food record collection. Additional research is under way that combines multiple techniques to reduce the error of one technique by incorporating another technique (i.e. combining 24-h recalls with FFQ).

Similarly, PA assessment has been improving primarily by the development of small wearable devices that measure multiple biometrics relating to energy expenditure. These biometrics include motion in 1 dimension (pedometer), 2 or 3 dimensions (accelerometer), skin conductance, heat flux from skin, and near skin temperature (as examples) that accurately reflect metabolic rate, which is directly related to energy expenditure. Current efforts also include translating these PA measures into specific activities that may have health importance, e.g. activities using upper compared with lower body muscles or aerobic compared with strength training and not just a total energy expenditure measure. Some of the challenges that remain include improving compliance with using these devices and reducing the size, cost, and number of devices an individual must use.

A concerted effort is necessary to support the continued enhancements of these assessment methods. Equally important, a new effort is needed to create an environment that fosters a global collaborative culture. This environment will enable better sharing of data and new techniques.

Standards need to be established for diet, PA, and omics data interfaces that support free data exchange. A technology infrastructure could enable access to a “best of class” set of tools through an application programming interface and provide a standard protocol workflow management and sharing of data for cross-study comparisons. Several examples of discipline-specific workbenches provide models for the nutrition community plans. The NIH National Library of Medicine’s National Center for Biotechnology Information has 29 separate tools and databases for genome, proteomes, transcriptomics, epidemiological data, and other specialized resources (90). Other life sciences workbenches include The Cancer Bioinformatics Grid (91), the Biomedical Informatics Research Network (92), which includes a neuroimaging resource and tool link (93), and the Phenx project (94), which is establishing consensus measures for phenotypes and exposures link to similar resources. These workbenches are academic or governmental sponsored and freely available for collaboration by researchers around the world. Each of these efforts provides specific examples of tools or concepts that can be adapted by and for the nutrition community (Table 2).

The way forward

With the exception of some food composition databases, the nutrition community lags behind many disciplines in providing validated, easily accessible tools and interoperable databases. Tools and databases are needed for researchers, clinicians, and the general public. For the research community, these tools will foster research and improve the reliability of science. Because many databases require controlled vocabularies and standardized data elements, a natural consequence of the construction of tools and databases will be the development of best practices. The Minimal Information for Biological and Biomedical Investigations Initiative (95,96) first started by the gene expression community (Minimal Information in Microarray Experiments) (97–99) lists 30 disciplines that are developing minimal reporting standards. The nutrigenomics community is represented in this effort but is not yet fully functional (100).

We propose the collaboration of the nutrition and nutrigenomic research communities to assist in the implementation of a centralized infrastructure that will provide a framework for access to “best in class” dietary and physical assessment tools and information that facilitates proper tool usage and procedures, storage, analysis, and sharing of data. An initial effort would be to create a Web site that informs the omics community of the current assessment tools and possible data storage. The site must be marketed to those who develop the newest assessment techniques as a mechanism for distribution of their efforts and collaboration among developers.

The initial list of tools could include Web-based tools such as the National Cancer Institute’s ASA24 24-h recall, Viocare’s

<table>
<thead>
<tr>
<th>Resource</th>
<th>Agency</th>
<th>Relevant features</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Center for Biotechnology Information</td>
<td>NIH National Library of Medicine</td>
<td>At least 29 databases or software tools, government funded, freely accessible, common rules and language, interoperable data and database can be downloaded</td>
<td><a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a></td>
</tr>
<tr>
<td>Cancer Bioinformatics Grid</td>
<td>National Cancer Institute</td>
<td>Discipline (cancer)-specific information and tools, common vocabularies, tissue banks, clinical trials, management tools, integrative (cancer) research workplace nutrition terminology</td>
<td><a href="https://cabig.nci.nih.gov/">https://cabig.nci.nih.gov/</a></td>
</tr>
<tr>
<td>Biomedical Informatics Research Network</td>
<td>NIH National Center for Research Resources (NCRR)</td>
<td>Provides data-sharing infrastructure, software tools, strategies and advisory services, biomedical imaging and genetics, hardware, and data under the control of user</td>
<td><a href="http://www.bimcommunity.org/">http://www.bimcommunity.org/</a></td>
</tr>
<tr>
<td>PhenX-Consensus Measures for Phenotypes and Exposures</td>
<td>NIH National Human Genome Research Institute</td>
<td>Integrates genetic and epidemiological research, recommended minimal set of biological measures, domains, and tools currently developed within PhenX are demographics, anthropometrics, cardiovascular, alcohol, tobacco and other substances, nutrition and dietary supplements, PA and physical fitness, and environmental exposures.</td>
<td><a href="http://www.phenx.org">http://www.phenx.org</a></td>
</tr>
<tr>
<td>PharmGkb – Pharmacogenomics Knowledge Base</td>
<td>NIH funded, managed by Stanford University Medical Center</td>
<td>Collects and disseminates knowledge of human genetic variation on drug response data organized in genes, variants, pathways, drugs and small molecules, and diseases.</td>
<td><a href="http://www.pharmgkb.org/">http://www.pharmgkb.org/</a></td>
</tr>
</tbody>
</table>
VioScreen FFQ, accelerometer interfaces, and HRV collection tools. The dietary assessment tools will need expansion to deal with different ethnic and world populations. Expanded efforts will focus on the mechanism to set standards, integrate datasets of multiple areas, and controls for sharing data and results within the community. These efforts will be aided by and coordinated with the international community’s effort to develop a nutritional phenotype database (101) and nutrition toolkit (102).

Acknowledgments

We thank John Mathers, Institute for Ageing and Health, Newcastle University, Newcastle upon Tyne, UK, for commentary. J.K. and D.M.K. organized the workshop; R.W., J.W.N., J.A.P., K.L.T., P.L.W., and A.K.I wrote the paper; and P.J.S. and R.W. had primary responsibility for final content. All authors read and approved the final manuscript.

Literature Cited


