HHS Public Access

Author manuscript

Annu Rev Nutr. Author manuscript; available in PMC 2015 March 10.

Published in final edited form as:

Annu Rev Nutr. 2013 ; 33: 349–371. doi:10.1146/annurev-nutr-072610-145203.

Quantifying Diet for Nutrigenomic Studies

Katherine L. Tucker[1], **Caren E. Smith**[2], **Chao-Qiang Lai**[2], and **Jose M. Ordovas**[2]

[1]Department of Health Sciences, Northeastern University, Boston, Massachusetts 02115

[2]Nutrition & Genomics Laboratory, USDA Human Nutrition Research Center on Aging at Tufts University, Boston, Massachusens 02111

Abstract

The field of nutrigenomics shows tremendous promise for improved understanding of the effects of dietary intake on health. The knowledge that metabolic pathways may be altered in individuals with genetic variants in the presence of certain dietary exposures offers great potential for personalized nutrition advice. However, although considerable resources have gone into improving technology for measurement of the genome and biological systems, dietary intake assessment remains inadequate. Each of the methods currently used has limitations tliat may be exaggerated in the context of gene x nutrient interaction in large multiethnic studies. Because of the specificity of most gene x nutrient interactions, valid data are needed for nutrient intakes at the individual level. Most statistical adjustment efforts are designed to improve estimates of nutrient intake distributions in populations and are unlikely to solve this problem. An improved method of direct measurement of individual usual dietary intake that is unbiased across populations is urgently needed.

Keywords

dietary assessment; food frequency questionnaire; gene x diet interaction; nutrigenomics; personalized nutrition

INTRODUCTION

Genetic variation does not always affect individual disease risk directly, but rather the potential is expressed only in the presence of certain dietary conditions. Genetic up- or down-regulation of specific metabolic pathways may also explain variation in individual requirements for certain vitamins and minerals. The field of genomics often struggles to replicate consistently observations of relationships between genetic variants and health outcomes. It is likely, however, that at least some of these disparate results are due to unmeasured differences in population nutrient intake that are interacting with genes to allow or block their expression (65, 66, 93). As more of these important interactions are identified,

kl.tucker@neu.edu.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

it becomes clear that many efforts in nutritional epidemiology to identify direct dietary effects on health outcomes are weakened by this underlying “noise” that makes it difficult to get solid and replicable results. The result of any single diet and health outcome study represents the average effect within a range of dietary responses due to unmeasured genetic variation in regulation. Because of this variability—not only in individuals, but also in populations—studies attempting to replicate an association between a specific genetic polymorphism and health outcomes may or may not find it, depending on the overall level of intake of a moderating dietary factor. Future research combining an understanding of genetic variation and dietary intake, therefore, promises to clarify many previously controversial or conflicting results on diet and health. Furthermore, we expect that this research will lead to more effective personalized nutrition information that is based on improved understanding of individual requirements for specific nutrients or sensitivity to others. In many cases, however, these associations may not be identified because of substantial error in the dietary assessment.

Genomic methods are advancing rapidly, and studies of gene-by-diet interaction are proliferating, with great promise for improving our understanding of how nutrition influences metabolic pathways and how this regulation may affect the development of diet-related disease (25, 27, 28, 34, 39, 56, 70, 73, 81, 90, 113). However, although the measurement and analysis approaches used for describing the genome are rapidly improving, the measurement of dietary intake has remained cumbersome and error prone and is lagging behind this new technologic revolution. At the simplest level, most individual gene × diet interactions reflect effects of up- or down-regulation of specific metabolic pathways that lead to accumulated health effects over time. It is therefore important to recognize that dietary variables used in population-based studies should, in most cases, reflect long-term usual intake of the specific individual nutrient that may modify the effect. This requires careful attention to the selection of the dietary method and its validity. At the same time, future research will need to deal with increasing complexities, including multiple simultaneous gene × gene, nutrient × nutrient, and gene × nutrient interactions. This will require the continued development of new complex computational methods. Again, we are much further ahead in developing new methodologies for the analysis of massive amounts of genomic data than for the dietary data that may be linked with them.

The rapid growth of consortia seeking to combine genomewide association studies (GWAS) in differing populations in order to gain sufficient power to identify these interactions illustrates the limitations of existing phenotypic data and underscores the critical need for improved measures, particularly of diet. Available dietary data are measured with different methods and with differing precision and sources of error, leading to complications and uncertainties in these large efforts. In this article, we review the state of the art and current strengths and limitations of various measurement methods, and we discuss progress in improving dietary assessment along with research needs to advance this important area.

NUTRIGENOMICS

Variation in dietary response among individuals has been recognized for a long time, but our understanding of the specific contribution of genetic variation to those differences, other

than in clear cases of inborn errors of metabolism, is quite recent (81). One source of variability in response to diet is the difference in allelic frequency of particular variants between populations of different ethnicities (60). The interactions between genetic and dietary factors are complex and may appear at different stages of life and under differing circumstances (53, 67). Although many environmental factors interact with genes to affect disease risk, food intake is of particular importance, as it is a regular exposure that has been important to human development throughout evolution. With accelerating global migration and changes in the food supply, it is expected that some individuals as well as populations will have greater responses to specific dietary choices than others. In many cases, this genetic variation expressed in response to dietary intake is seen with intermediate markers of health, such as blood pressure, plasma lipids, glycemic response, and obesity (1, 27, 57).

As an example, one of the earliest genetic polymorphisms to gain widespread acknowledgment as a modulator of disease risk is apolipoprotein E (*APOE*), including risk of cardiovascular disease (CVD), cognitive decline and neurological disorders (19, 22, 46, 61, 64, 71), cancer (117), osteoporosis (2, 23), and earlier mortality (92). Epidemiologic studies have repeatedly shown that low-density lipoprotein (LDL) cholesterol (LDL-C) and apolipoprotein B concentrations are highest in individuals with the E4 polymorphism, intermediate in those with E3, and lowest in those with E2 alleles (82). However, it has been further noted that among those with the E4 polymorphism, LDL-C is particularly high in the presence of a diet that is high in fat, particularly saturated fat (21, 81, 83, 93). Thus, in many and perhaps most cases, the risk associated with genetic polymorphisms for the chronic conditions we see in epidemic proportions today may be expressed primarily in conditions of imbalance in dietary exposures relative to individual requirement.

Because our basic genetic structure has not changed in the past 30 years (epigenetics aside), the exponential growth in chronic conditions, including obesity, type 2 diabetes, CVD, and certain cancers, must be related to environmental exposures, and leading the list of important chronic exposures is dietary intake. Within the population, the distribution of risk is unequal, based on individual genetic response to the diets we are consuming and to the variation in the imbalances in dietary quality. This combination is likely to explain some of the major disparities in health risk seen across subpopulations. Despite progress in identifying important gene × diet interactions, studies often show conflicting and inconclusive associations with disease outcomes. Due to valid concern about false positive associations, the field is increasingly demanding replication of results across studies, and in many cases this has been more difficult than originally anticipated (48, 74, 78). The same is true for gene × diet interactions (21). Some heterogeneity is expected due to differences in prevalence of genetic polymorphisms and population differences in range of dietary intake. However, a major limitation and current barrier to progress in this area is the limited validity of the dietary phenotypes, as measured.

This becomes even more important when a pathway involves very specific and generally less well measured nutrients. As an example, proinflammatory leukotrienes, derived from arachidonic acid in the 5-lipoxygenase (5-LO) pathway, appear to be atherogenic. In a US population of adults (n = 470), a *5-LO* promoter polymorphism, seen in about 6% of the sample, was significantly associated with intima-media thickness of the carotid artery.

Importantly, however, the effect was significantly stronger among those with high intakes of arachidonic and linoleic acids and was weakened among those with high intakes of n-3 fatty acids (31). In contrast, a study in a Spanish population without dietary information did not detect a relationship between 5-LO variation and risk of myocardial infarction (MI) (42). However, the latter did not assess dietary intake, and it is possible that underlying dietary variation may have masked an association by having differential effects in subsets of the population. This gene was then tested in a Costa Rican population in relation to risk of MI (3). As in the Spanish study, and in contrast to the earlier US study, the allelic frequency did not differ significantly between cases and controls. Importantly, however, there was a gene × diet interaction similar to the one reported in the US study: Variant alleles were associated with 1.3 times higher MI risk in those with high (≥0.25 g/d) dietary arachidonic acid intake and 23% lower risk in those with lower (<0.25 g/d) arachidonic acid intake.

This illustrates both the importance of considering dietary factors when identifying genetic risk and of having detailed measures of usual intake. In the first study that found the interaction, fatty acid intake was measured with six detailed nonconsecutive 24-hour recalls (24HRs) over a period of 1.5 years, with detailed questions on foods consumed and types of preparation (31). Furthermore, these intakes were validated in a subset against plasma fatty acids. Six recalls over time provide a strong measure of usual dietary intake for most nutrients (excluding those with very high intraindividual variability, such as vitamin A), but this either is prohibitively expensive for very large studies, requiring six contacts with participants and extensive staff time to conduct the recalls, or is dependent on volunteer completion of a tool such as the National Cancer Institute's Automated Self-Administered 24-Hour Recall (ASA24) on multiple occasions, with likely missing days and noncompliance by large numbers of participants. In addition, the loss of data when using self-administered assessment tools is not likely to be random; those without access to computers and/or with low education levels are likely to be excluded.

In the Costa Rican study, dietary intake was measured with a carefully designed food frequency questionnaire (FFQ), with deliberate consideration in the design for the measurement of fatty acid intake (54). This is an excellent option for studies conducted within cultural groups, but the same questionnaire will not necessarily be valid for other populations without considerable revision. Most existing questionnaires have not focused specifically on the validity of individual fatty acids, and the detail collected is highly variable. Therefore, it is possible that another study, using a less sensitive FFQ, may not replicate the clear finding in Costa Ricans. Many questionnaires lack sufficient detail on the types of oils and fats used in cooking and added to salads or other foods, leading to substantial misclassification of intake of specific fatty acids. To the extent that the types of oil most frequently used differ by ethnic/culinary subgroup and are not captured, the data may not be simply diluted but the results also potentially biased. This example points to the ability to successfully detect true gene × diet interactions when sufficient attention to the validity of the dietary data is given. Unfortunately, most studies do not have such high-quality dietary measures of usual intake.

DIETARY ASSESSMENT METHODS

The validity of dietary data assessment varies by method and, within method, by nutrient, depending on the detail required and the time period of interest. At the simplest, dietary patterns may be assessed with a checklist of frequency of food group intakes. Although this is useful for ranking dietary quality, it is clearly inadequate for use with gene × diet interaction studies where metabolic pathways are affected by individual nutrients. Because genetic polymorphisms affect only a subset of any population, it is critical that nutrient intake be precise at the individual level. Otherwise, the extent of misclassification for use in interactions may attenuate the ability to detect true associations. Often, all nutrients that will be of potential interest in a large study are not identified at baseline, and in most cases, relative estimates improve with adjustment for total energy intake. Therefore, the ideal method will capture the full diet, including total energy intake. The latter is important to standardize nutrient exposure relative to total energy requirement and considers variation in needs that correlate with energy requirement, including, to some extent, body size and activity level (118).

The investigation of gene × diet interaction almost always requires large sample sizes to be able to achieve stability in relatively small subsets that may have both the polymorphism of concern and the dietary condition that reacts with it. The most commonly used approaches for collecting data at the population level include the FFQ and the 24HR, as in the examples noted above. A third method, weighed diet records, is often used in smaller human studies with educated and compliant volunteers but is less often used in large epidemiologic studies because of recognized limitations (37).

Diet Records

For many years, the weighed diet record was considered the gold standard for dietary intake assessment. This method continues to be used successfully in controlled metabolic studies with motivated trained volunteers. It has been used less frequently in large epidemiologic studies because of the high participant burden and the cost of data entry. One study that did use this method for many years is the Baltimore Longitudinal Study of Aging, a longitudinal study that began with a highly educated sample of volunteers (95). Selected European studies continue to favor the seven-day diet record. The Norfolk site of the European Prospective Investigation into Cancer and Nutrition (EPIC), for example, reported that they were able to obtain 93% completion of records from 25,000 participants attending their health-check day and that these records were generally of high quality due to instruction from trained nurses (for details, see http://www.srl.cam.ac.uk/epic/about/baselineII.shtml). Because of the cost of data entry and processing for this huge number of diaries, the study has been analyzing them only for cancer cases and matched controls, as needed. This success with diet records in a large study is unusual, however, and most of the other 23 EPIC centers chose to use only a FFQ for estimation of intake (41).

In addition to nonrandom loss of data due to respondent burden, the diet record has been shown to alter actual consumption because individuals focus on what they are eating and may consume smaller portion sizes or avoid complex foods that are difficult to weigh or describe (43). Thus, even when highly accurate for the day recorded, the weighed record

tends to underestimate and misrepresent usual intake. Data quality depends on individual effort at recording the details required, and this varies by education level, making the diet record a poor choice for populations low in literacy.

Twenty-Four-Hour Dietary Recall

A frequently used method of dietary assessment in large studies is the 24HR. Recent advances in data entry, such as the United States Department of Agriculture (USDA) automated multiple pass system, have greatly improved the validity of this measure for estimating actual previous day's intake (72). The USDA companion food model booklet (13) or other measuring aids help to improve portion size estimation (24). The open-ended nature accommodates diverse dietary patterns, making it the most valid method currently available for assessing recent intake of multiethnic groups or newly investigated populations. 24HRs are optimally administered in person by an interviewer, but direct entry by an interviewer over the phone has also been shown to have excellent validity at a reduced cost (14, 17). New computer technologies, such as the ASA24, allow direct individual input at minimal cost (100). To the extent that participants are willing and able to complete this assessment on their own, it provides a cost-effective option for large-scale data collection.

A major limitation of the 24HR is the day-to-day variability inherent in dietary intake (7, 8). Although this source of random error is not a threat to the validity of group mean intakes, the misclassification of individuals can be serious, attenuating the ability to detect true associations with outcome variables. Furthermore, intra/interindividual intake varies considerably by nutrient (8, 118). For nutrients dispersed in limited or infrequently consumed specific foods, intraindividual variation exceeds between-person variation to such an extent that many nonconsecutive days are needed to stabilize the within-person estimate of usual intake. Using data from a Finnish population, for example, one group concluded that 7–14 days are generally adequate to classify most nutrients, but that nutrients with high variability, such as vitamin A, will require 21 days or more to obtain a stable estimate (45).

To the extent that the error is truly random, statistical corrections may be used when at least two recalls, on nonconsecutive days, are available to estimate intra/interindividual variation. Correction for intraversus interindividual variation allows projection back (or deattenuation) to the true likely association in linear models with continuous outcome variables (7, 62). With dichotomous outcome variables, validation studies on a subset may be performed to provide measures of sensitivity and specificity for use in adjustment (36). However, straightforward corrections are not available for more complex analyses with categorical variables, such as those often used when studying gene × nutrient interaction.

One approach to improve the estimates of recalls with respect to high day-to-day variability is to use mixed methods, with two 24HRs supplemented by a propensity (frequency) questionnaire to capture exposure to episodically consumed foods (such as liver) that can be used in the regression equation separately as a control variable when relating individual nutrient intake to an outcome [e.g., vitamin A assessed from the average of the two recalls, holding frequency of liver intake constant (never, seldom, frequently.. .)]. The National Cancer Institute (NCI) has developed a two-step model to first estimate the probability of intakes using data from two recalls and incorporating portion size information, and then

fitting a model with the transformed recall data, adjusted for intake of episodically consumed foods from the FFQ (107). This helps to improve estimates of intake distributions, as it was designed to do, but this approach is not sufficient for assigning total intakes to individuals for precise categorization. A limitation of correcting for intra/interindividual variation is that day-to-day variability is not constant across populations. Thus, the use of the deattenuation formulas in multiethnic samples will calculate only the average variability and will not consider the differences in the extent of random variability across subsets. This weakens validity, particularly if the risk of the outcome or the prevalence of a genotype of interest also varies by ethnicity. An example may be rice, for which intra/interindividual variation may be low in some groups, such as Puerto Ricans, with regular daily intake, but high in populations where consumption is much less frequent (109).

Another concern, even within fairly homogeneous populations, is that underreporting on the 24HR is not, in fact, random (68), with evidence for greater underreporting by women than men (79), by obese versus nonobese individuals (52, 79, 108), by smokers versus nonsmokers (52), and among those concerned with providing socially desirable responses (79) or with restrained eating behavior (5, 6). Although the NCI is actively working on statistical algorithms to adjust for nonrandom error focused on national distribution estimates (30), these will not compensate for the different sources of bias among populations with different eating patterns, cultural attitudes toward dietary reporting, and underlying behaviors that may be associated with genotypic variation.

Food Frequency Questionnaires

FFQs have long been the method of choice for assessing dietary intake in large epidemiologic studies because of their efficiency and ability to assess usual intake in a single administration. However, FFQs may contain substantial systematic measurement error due to the finite food list, set portion sizes, weighting of contributions of individual foods to line items using overall assumptions, and lack of recipe detail, limiting their precision. However, FFQs have been shown to rank intakes successfully (118) provided they are carefully developed and validated within the population for which they are being used (49, 112). Most FFQs have been developed and validated either for a defined population, e.g., health professionals, or to be nationally representative, based on data from the National Health and Nutrition Examination Survey (11, 101). FFQs have shown relatively poor results in minority groups, however, including non-Hispanic blacks and Hispanics (4, 10, 99). For example, validity coefficients for energy intake in the multiethnic questionnaire used by Block in studies of mothers participating in the Women, Infants, and Children program were 0.44 for non-Hispanic white women but only 0.14 for Hispanic women (10); in a multiethnic cohort in Hawaii and Los Angeles, energy coefficients for non-Hispanic white, Hispanic, and non-Hispanic black men were 0.48, 0.33, and 0.16, respectively (99); and the Insulin Resistance Atherosclerosis Study showed energy validity coefficients for urban non-Hispanic whites of 0.61, relative to 0.37 for urban non-Hispanic blacks, and 0.27 for rural Hispanics (69). One study (4) reported weaker correlations between reported intakes of carotenoids from the NCI Diet History Questionnaire and plasma carotenoids for non-Hispanic blacks compared with non-Hispanic whites ($r = 0.21$ versus 0.47 for lutein + zeaxanthin and 0.17 versus 0.31 for β-carotene, respectively). It is likely that the adapted

FFQ failed to include sufficient detail on recipes and portion sizes that differ across groups owing to the compromises required to ensure that the FFQ was a reasonable length. FFQs that have been carefully designed for a specific population can be valid for that population (16, 84, 109). However, the use of different FFQs for different populations introduces additional biases. In some questionnaires, such as the Harvard FFQ, portion sizes are assumed at standard levels (121). In others, portion sizes are presented as small, medium, or large categories (11) or are presented from photographs (44). However, portion sizes are often culturally determined, and the lack of this quantification in mixed ethnicities is another source of bias (109).

Many compromises are made in designing the FFQ food list. Similar foods are grouped together, and the resulting nutrient content is based on a weighted average of these in the larger population. FFQs in epidemiologic studies have been shown to have excellent validity for ranking individual nutrient intakes— provided that the population evaluated is the same as that for whom the FFQ was designed (12, 89, 121). Unfortunately, many studies limit the time allotted to dietary data collection and select shorter instruments. However, short instruments should be used with caution, as validity is likely to be compromised. An example is the FFQ used by the Jackson Heart Study (15). By limiting the data collection time to 20 minutes, an existing questionnaire developed for the southern United States was shortened by collapsing food line items (112). Validation against four 24HRs (16) and plasma carotenoids (103) showed few differences in macronutrient assessment between the long and short forms, but better estimation of most micronutrients with the long form.

For nutrigenomic studies, these limitations may be exaggerated, as ethnicity may be related to genotype prevalence, risk of certain health outcomes, and considerably different dietary habits. If commonly used foods are not on the food list, then calories and associated nutrients are automatically underestimated. If portion sizes of commonly consumed foods are differentially misestimated by ethnicity, this will lead to further nonrandom misestimation of calories and associated nutrients. Further, recipe assumptions can have a dramatic effect. As an example, Puerto Ricans tend to eat large amounts of rice on most days and usually cook rice with oil (109). If not accounted for, this leads to a systematic underestimation of fatty acid intakes for this group. Without consideration of the ubiquitous use of tomato paste with beans, lycopene intake is seriously underestimated (110). Importantly, these errors are nonrandom and can lead to invalid conclusions, not merely attenuation in results. This bias could lead to more complex invalidity when considering gene × diet interactions.

Recalls or Records Versus Food Frequency Questionnaires

The 24HR has been promoted in recent years following observations from the Observing Protein and Energy Nutrition (OPEN) study that estimates of energy and protein intake assessed by multiple 24HRs were more strongly correlated with recovery biomarkers for energy and protein than were those of the NCI FFQ (94, 102). Further support for quantitative assessment was obtained when nutrient estimates from seven-day diet records versus FFQs were compared with doubly labeled water and urinary nitrogen in the UK component of the EPIC study (26). This was reinforced by findings of a significant

association between dietary fat intake and breast cancer when assessed by seven-day diet records but not by FFQs in EPIC-Norfolk (9) and in a subset of the Women's Health Initiative (WHI) nonintervention cohort (35).

Together, these observations led some to question the validity of the FFQ (58) and to endorse more quantitative methods. It is important to note some limitations of these findings (119, 120). First, concurrent measures of energy and protein intake would be expected to relate more to multiple 24HRs taken at the same time point than to the longer-term estimate made by the FFQ. Further, because of the semiquantitative nature of FFQ responses, most analyses are adjusted for total energy intake to improve the proportional ranking of nutrient intakes. This was evident in the protein measures, which, although presented both ways, were clearly improved for the FFQ following adjustment for total energy intake (94). Similar improvements would be expected with micronutrients, which could be more valid for usual intake with the FFQ than with a few 24HRs, which are likely to miss episodically consumed foods high in specific nutrient content. Comparisons of intake estimates from FFQs with biochemical indicators of nutrient status have shown good results for several micronutrients (51, 110, 111), suggesting that FFQs have the capability of accurately ranking them. A more recent study compared a four-day food record, three 24HRs, and a FFQ from 450 postmenopausal women in the WHI prospective cohort study with doubly labeled water and urinary nitrogen (87). As in the earlier studies, the food record showed a stronger association with energy and protein biomarkers than did the FFQ, with intermediate results for the 24HR. However, differences across methods were nonsignificant for protein density, and further, equations that included body mass index, age, and ethnicity substantially improved the estimates and reduced the differences in methods, suggesting that at least within the population studied, "any of the assessment procedures may yield suitable consumption estimates for epidemiologic study purposes" (87, p. 591).

Although expensive and thus unlikely to be used uniformly for large studies, two 24HRs plus a propensity questionnaire is likely the best option currently available for complex populations. This option is being used in the current National Heart, Lung, and Blood Institute Hispanic Community Health Study/Study of Latinos, as no validated FFQs have been demonstrated to cover all Hispanic groups without bias (97). FFQs remain the least expensive option for large studies and therefore are most widely used. However, as discussed below, FFQs differ considerably across studies, making it difficult to calibrate intakes across groups in large consortia analyses. Discussion on and investigation of the utility of the FFQ continues (58, 86), and it is clear that more work is needed to continue to better understand the properties of measurement error, particularly for diverse subsets of the population.

USING DIETARY DATA FROM DIVERSE STUDIES IN GENOMEWIDE ASSOCIATION STUDY CONSORTIA

The limitations in both quality and consistency of dietary data present great challenges to members of large consortia groups, who are increasingly attempting to examine diet × gene interactions in combined data sets across multiple studies or by conducting meta-analyses. The rationale for multistudy analyses of genetic and dietary factors in combination is

compelling. First, to reduce false positive results and increase scientific credibility, genetic relationships detected in a single population must be replicated in one or more independent populations; criteria for replication have been established (18). Additionally, in light of the very large sample sizes estimated to be necessary for detecting gene × diet interactions (47), meta-analyses represent one of the few effective ways to ensure adequate power. Genetic studies have derived tremendous benefits from consortia efforts that have detected novel genetic associations. However, for consortium-based interaction studies, now in their infancy, considerable challenges remain. The inherent heterogeneity of studies that span diverse geographic, age, and socioeconomic groups across the United States and multiple European countries is compounded by imprecision in dietary data. The extent to which reported dietary differences across cohorts reflect actual differences in intake versus imprecision in measurement of nutrients, foods, or food groups is unknown. The use of different FFQs or even completely differing methodologies of dietary data collection makes it difficult to be sure that data from different studies are on the same quantitative scale. The use of qualitative quintile cutoffs for assigning intakes to low, medium, and high levels is sometimes used, but these can have differing meanings in different populations, adding further distortion to the comparisons.

Consortia Example

Most large studies continue to use the FFQ as the only cost-effective, feasible method for large numbers of individuals. However, as discussed above, different FFQs are used, with different food lists, with or without portion sizes, and including or omitting specific ethnic foods. It remains unclear how combining food or nutrient intake data from these different questionnaires may affect analytic results. Collapsing data from a questionnaire designed to include Latino or Asian foods with a Harvard or Block questionnaire, for example, may lead to systematic biases that misclassify one group relative to the other. For this reason, consortia have approached these analyses carefully, and to date, have mostly limited analyses to single ethnic groups.

An example is the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, which was originally formed with five well-described longitudinal cohort studies in the United States and Europe to facilitate GWAS meta-analyses of genetic variation and health (88). Since then, further studies have been added, with a recent analysis using data from 15 cohorts with dietary measures. Examples of the range of dietary instruments and estimates of dietary fats data obtained through a variety of instruments in one CHARGE-based study are shown in **Table 1** (adapted from Reference 96). **Table 1** shows a range of estimated total fat intake (from 28.4% of total energy in the Framingham Heart Study to 36.3% in the Rotterdam Study) and polyunsaturated fat intake (3.4% of total energy in the Invecchiare in Chianti Study to 8.7 % in the Health, Aging, and Body Composition Study). In addition, the same FFQ used in the Framingham Study estimated higher total fat intake in the Nurses Health Study (33.4%). Although the various dietary instruments are likely to rank nutrient intakes relatively accurately within individual populations, the extent to which differences in fat intakes in these studies are real or are due to differences in assessment methods or the validity of assessment methods for the specific

population remains unknown. It remains unlikely, therefore, that such diverse data sets can achieve the precision required for valid combined analyses across global populations.

Three additional manuscripts from these early efforts examining gene × diet interaction in large numbers of individuals from diverse populations are illustrative of the need for detailed and accurate dietary data. One study examined whole-grain intake in interaction with 18 genetic loci associated with fasting glucose and/or insulin in 14 cohorts totaling more than 48,000 individuals (76). As noted in **Table 1**, these cohorts cover diverse populations, from the US-based Framingham Heart and Cardiovascular Health Studies to studies in Italy, Sweden, the Netherlands, Iceland, and Greece. Dietary assessment methods were diverse, including a 108-item Block FFQ, a 99-item version of the NCI FFQ, the 126-item Harvard FFQ, modified 66-item and 130-item Harvard FFQs, a 30-item dietary practice questionnaire, a diet history with a seven-day record and 168-item FFQ, a seven-day food record, a 236-item Italian FFQ, two nonconsecutive 24HRs, a two-step Dutch FFQ, and a 55-item Greek FFQ. As can be seen from this list of very different assessment methods, the detail and precision of estimates are likely to vary considerably. From these assessments, whole-grain intake estimates were obtained at each site. Available whole-grain information ranged from one line item (in the Italian FFQ) to eight (in the 126-item Harvard and the Dutch FFQs); these were expressed as number of servings per day. Findings showed that greater whole-grain intake was associated with lower fasting glucose, but no gene × diet interactions met the assigned significance criteria. One interaction, between whole-grain intake and the rs780094 variant in the *glucokinase regulatory protein* (*GCKR*) gene, approached significance, suggesting that whole-grain intake was associated with a smaller protective association with fasting insulin in the presence of the insulin-raising allele. It is possible that clearer results showing more interactions may have been seen had the quality of the dietary data been more consistent and detailed across cohorts.

A second analysis focused on the effect of zinc intake on the relationship between zinc transporter SLC30A8 and fasting glucose concentration (55). Dietary zinc was calculated from the same data sources, using the country- or study-specific nutrient databases. Only 5 of the 14 studies had data on zinc intake from supplements. Data were analyzed separately by study and presented as meta-analyses. Total zinc was associated with lower plasma glucose in four of the five studies with supplement information available, and in the meta-analysis of the five studies, but no association was seen when examining dietary zinc intake alone. Similarly, an interaction with the rs11558471 SLC30A8 variant was seen only in the studies with total (diet + supplement) zinc intake. This emphasizes the importance of complete information when estimating nutrient exposure and calls into question the validity of the dietary zinc estimates. When zinc supplements are included, the distribution is extended, more clearly identifying those with higher exposure.

Most recently, Nettleton et al. (75) published a meta-analysis examining a dietary quality score in interaction with the same 18 genetic loci previously identified by GWAS to influence fasting glucose and/or insulin in 15 international cohorts with 51,289 persons without diabetes (75). Because of the diversity of dietary methodology, they determined study-specific quartile rankings for intakes of whole grains, fish, fruit, vegetables, and nuts/seeds (favorable quality) and red/processed meats, sweets, sugared beverages, and fried

potatoes (unfavorable); these were combined to form a healthy diet score. Data were analyzed first with linear regression within studies and then by inverse-variance-weighted meta-analysis. As expected, the diet score was significantly inversely associated with fasting glucose and fasting insulin after adjustment for demographic factors, lifestyle, and body mass index, demonstrating that this semiquantitative ranking is sufficient for main effect associations, as has been demonstrated previously with dietary pattern analyses in many studies (29, 32). In contrast, however, no interactions were detected with the genetic loci. Nettleton et al. (75) conclude that focusing on genomic regions without, rather than those with, highly statistically significant associations from main effect GWAS may be more fruitful in identifying diet × gene interactions.

An alternative explanation for the lack of findings with dietary patterns is that gene × diet interactions may be difficult to see in general patterns because each pathway is associated with specific nutrients. The limited findings with whole grains and zinc also point to the likelihood that the size of the study may be less of a limiting factor than is precision of dietary intake measures. Together, these large efforts have been disappointing but clearly illustrate that approximate dietary measures will be unlikely to provide the information we need in order to discover gene × diet interactions. To understand and document these important interactions, it is necessary to invest in the detail required to obtain estimates with good precision at the individual level.

PhenX Recommendations

Several groups have thought about these issues, but further consideration is clearly needed to move the field forward. One of the groups that has been working to provide recommendations is the National Institutes of Health PhenX (Phenotypes and eXposures) working group, which has developed a tool kit to encourage standardized questionnaires or methods for obtaining phenotype data in large projects (98). Developed with support for the National Human Genome Research Institute, the group began its work to assist consortia in harmonizing data across studies. They developed a tool kit in 2006 and continue to refine it (**http://www.phenxtoolkit.org**). The stated goal of the program is “to identify and catalogue 15 high-quality, well-established, and broadly applicable measures for each of 21 research domains for use in GWA studies and other large-scale genomic research” (**http://www.genome.gov/27541903**).

One of those domains is nutrition and dietary supplements. The difficulty in arriving at ideal measures is illustrated by the group’s selection of 12 measures for this domain. These include specific questions to define breastfeeding (3 questions), caffeine intake (13 frequency questions), calcium intake (18 frequencies), dairy food servings (2 frequencies, only milk and cheese), use of dietary supplements (18 frequencies), fiber intake (17 frequencies), fruit and vegetable intake (9 frequencies), percent energy from fat (16 frequencies and formula), selenium (from serum), sugar intake (4 frequencies), vitamin D (serum), and total dietary intake (ASA24 and two 24HRs more than a week apart, one on a weekday and one on a weekend day). Measures specifically related to dietary intake assessment are listed in **Table 2**. The rationale for the recommendation of the ASA24 for this purpose is stated as:

> The 24-hour recall has long been regarded as the optimal methodology for collecting total dietary intake because it provides the highest-quality and least biased dietary data for a single day. This method allows collection of detailed intake and portion sizes. Because the data collection occurs after consumption, this method does not affect an individual's food choices on a given day. (PhenX Toolkit; https://www.phenxtoolkit.org/index.php?pageLink=browse.protocols&id=51200)

However, it is unclear whether the short-frequency screeners are valid for most populations. First, short-frequency screeners provide only a very rough approximation and population ranking for the targeted nutrients. As an example, the questionnaire that is supposed to provide percent energy from fat was selected because it has been validated and has low burden. A look at the questionnaire, however, shows a food list that is limited (https://www.phenxtoolkit.org/index.php?pageLink=browse.protocoldetails&id=50801). The only meats included are hot dogs, bacon, and sausage—no luncheon meats, hamburgers, or fried chicken; use of skim milk is included as protective, but use of whole milk or cream is not requested, etc. The questionnaire was developed from nationally representative data using a regression approach that maximally explained variance in percentage of energy from fat intake; thus, certain foods, such as fried chicken, were not included because at the time, and in the population tested, they did not explain large proportions of the variance (104). Estimates of percentage of energy from fat were computed from respondent-reported frequency responses in a nationally representative sample and assigned sex- and age-specific portion sizes and sex-specific regression coefficients using the USDA's 1994–1996 Continuing Survey of Food Intakes by Individuals. These types of equations, although useful in a similar population, are highly questionable when exported to different populations or even different time periods. Importantly, the equation assumes that the variation in these behaviors will correlate equally with the excluded foods for all members of the population. In fact, although this screener was validated against two nonconsecutive 24HRs in an older population (n = 401) that was 91% non-Hispanic white and where 72% had more than high school education (104), it did not hold up when used, as originally intended, to measure change in intake with intervention in a multisite diverse population, where differences between the screener and 24HRs were much greater at the site that included African American men and women. The screener significantly underestimated fat intake in African Americans in Georgia and approached significance for overestimating fat intake in Latinos in Boston (105). The positive predictive value for classifying women as consuming ≤30% energy from fat was 0.15 for African American women in Georgia versus 0.72 for non-Hispanic white women in Rhode Island. This finding is very important, as many gene × diet interaction analyses use this type of cutoff point. It is therefore questionable to recommend this fat screener for use in nutrigenomic studies.

The other short-frequency screeners all likely suffer from similar biases based on the populations from which they were developed and are unlikely to provide unbiased estimates across differing population groups. As such, their use is likely to obscure rather than inform a field in which it is clear that more—not less—precision is needed. As another example, the calcium questionnaire specifically states not to include small amounts of milk in coffee but does not allow for large amounts of whole milk in coffee, as used by Puerto Ricans; it also

does not include intake of small flsh, a major source of calcium in Asian populations. The dairy foods questionnaire asks only about milk and cheese and does not include yogurt or other dairy products. Finally the fruit and vegetable questionnaire, which should be more straight-forward, overestimated intake, particularly for African Americans in the multiethnic intervention study described above (85). The authors concluded, “Findings suggest multiple 24HR at multiple time points in adequate sample sizes remain the gold standard for FV reports. Biases in FVS estimates may reflect participants’ lifestyles and sociodemographic characteristics and require further examination in longitudinal samples representative of diverse populations” (85, p. 218S). A recent review further describes the limitations and challenges of using the NCI and other screeners for estimating fruit and vegetable intake (91).

Finally, although two ASA24 recalls is a reasonable recommendation, there is no discussion of the limitations of this approach for usual intake assessment or loss of participant data/poor quality data due to the requirement for self-administration. To date, only preliminary studies with educated beta testers have been completed. Although these are reported to show acceptable face validity (http://riskfactor.cancer.gov/tools/instruments/asa24/respondent/validation.html), no studies have been reported to show general population completion rates or quality of completion in differing ethnicities or literacy levels with use of the ASA24.

The PhenX group’s decisions reflect the efforts of an expert committee, illustrating the limitations in choices available to them. Although the group’s work presents a thoughtful approach to beginning to think about standardization, more work is needed to develop approaches that are uniform and valid across populations and reasonably cost-effective.

INCREASING COMPLEXITY IN ANALYSES

As a complement to dietary assessment, new approaches to complex systems are beginning to contribute to the understanding of gene × diet interactions on health (20). Bioinformatics, including proteomic and metabolomic data, has great potential for revealing the mechanisms involved in gene × nutrient interactions (25, 38, 116), including the importance of specific nutrients on cellular metabolism (59, 113). Systems biology (28, 57, 59, 90, 113) promises to help us move beyond simple gene × nutrient interactions toward global effects, where many genes and metabolites can be examined simultaneously (40). Although these exciting new technologies are receiving considerable attention, it is important to not forget that they will not be able to answer the questions we are asking without precise and accurate dietary information. Long-term exposures, particularly of diet, remain the key factors leading to outcomes under differing genetic environments. Advances in accurately capturing this behavior are critically needed if we are to utilize the new technologies effectively to inform policy and intervention.

Technological Advances in Measuring Usual Intake

Given the known limitations of current dietary assessment methods, several investigators are working to develop new approaches for measuring dietary intake, and technological advances are being used to reduce costs associated with collection and processing of dietary data (106). Web-based versions of FFQs have been developed (33, 80, 115, 122), but most

are similar to the paper version, with a defined food list. With the recognized need for information on diet, the Genes, Environment and Health Initiative of the National Institutes of Health has encouraged technology-driven methods to improve dietary intake, among other exposures, for use in large-scale studies (106). Examples of funded projects include a Web-based dietary recall for use in children, mobile phone photography to record foods consumed, a mobile phone application to record intake with voice recognition, and a device that can be worn to take pictures of food through image recognition (77). These innovations, along with improvements in biomarkers of dietary intake— including metabolomics and other techniques (63, 114)—hold promise for contributing to improved assessment through technology, but more innovations are needed to improve assessment of long-term usual intake for large numbers of individuals at a reasonable cost.

SUMMARY

It is clear that advances in measuring dietary intake have lagged behind the rapid growth in sophisticated methods to characterize the diversity within the human genome. Dietary behavior is complex, and accurate assessment requires detailed questioning that considers the diversity in dietary patterns across populations. Combining data is challenging because methods used both within and across populations differ greatly in approach, detail, and timing of intake. The most open-ended approach remains the 24HR, which, provided that accurate dietary databases are available and data are carefully collected, is unbiased in terms of ethnicity or dietary pattern for short-term (prior-day) assessment. Because the effect of diet on gene expression usually influences health risk over long periods of time, however, accurate assessment of long-term usual intake is the goal. The major limitation of the 24HR is the large misclassification that occurs because of day-to-day variation; multiple recalls are needed to stabilize this error. Although techniques to estimate this attenuation due to random error are available for straightforward analyses at the population level, these do not correct for misclassification at the individual level; thus, categories used in gene × diet interactions remain poorly specified for diet, with resulting uncertainty of estimates.

As noted in the example above with zinc, supplement use must also be measured to capture fully nutrient intake for use in interaction analyses. The importance of specific nutrients for individual genetically affected metabolic pathways means that great precision is needed not only in food groups but also in nutrients. Most large studies do not currently contain good measures of dietary intake, and fewer still account for the diversity of diets within their populations to capture detailed differences in long-term intake at the individual level. Improving these shortcomings is essential if we are to fulfill the promise of truly understanding how our diets affect our metabolism and health in relation to our genome. The only way to enhance our understanding is to recognize that dietary data are important, to acknowledge that improving their precision will require resources and time, and to continue to study ways to improve the representativeness and validity of usual intake assessment for all groups without bias.

The use of two 24HRs, with the addition of a propensity questionnaire to adjust for sporadically consumed foods with large amounts of specified nutrients, is currently the best option for multipopulation studies. The openended nature of the recall allows for detail of

portion sizes and recipes in different groups. The adjustment for major occasional sources of nutrients can improve estimates, but individual rankings will still be affected. Considerably more attention to subgroups in the population is needed to ensure validity of results and inclusion of all appropriate high-nutrient-source foods in the propensity questionnaire. Further investigation of the meaning of statistical adjustment for specific foods that vary in consumption across ethnic subgroups is needed, rather than inclusion of these foods in the total usual intake estimate.

Unfortunately, even the inclusion of two 24HRs and a propensity questionnaire is considered too much of a respondent burden and too great a cost for many large studies. Therefore, single administration of a FFQ is likely to remain the most cost-effective tool of choice for large populations. It is important to note that not all FFQs are equal and comparable, particularly in measuring diverse diets. Many consortia suffer from decisions on how to compromise with differing measures of dietary assessment in different studies. To the extent that these have differing levels of validity for subgroups, errors may be multiplied and thus mask the ability to detect important associations.

CONCLUSION

Nutrition is probably the most important environmental factor that modulates the action of genes and the phenotypes being considered. This has been known for a long time but has been underestimated in scientific priorities (50). Accumulating data show that lack of replication of health outcome effects with genetic polymorphisms may be due to differences in levels of dietary exposure, allowing or suppressing the expression of the genetic tendency. Therefore, it is of paramount importance that genetic variation is considered in the context of dietary exposure. This paradigm constitutes the basis for nutritional genomics, with tremendous potential to yield results that could change the way dietary guidelines and personal recommendations are considered.

The interaction between genotype and dietary exposure—nutritional genomics—has the potential to change dietary disease prevention and to have a major impact on public health. It is clear that without consideration of diet, the discovery of genetic determinants of disease will be limited. Furthermore, although we cannot change our genes, diet is a modifiable risk factor. The promise of nutrigenomics is that personalized dietary advice, based on genotype, will allow changes to improve the health of populations. To do this will require increased understanding of the limitations of current dietary methodologies and the development of better approaches to obtain unbiased assessments of long-term usual intake in diverse groups.

LITERATURE CITED

1. Afman L, Muller M. Nutrigenomics: from molecular nutrition to prevention of disease. J. Am. Diet. Assoc. 2006; 106(4):569–76. [PubMed: 16567153]
2. Alagiakrishnan K, Juby A, Hanley D, Tymchak W, Sclater A. Role of vascular factors in osteoporosis. J. Gerontol. A Biol. Sci. Med. Sci. 2003; 58(4):362–66. [PubMed: 12663699]
3. Allayee H, Baylin A, Hartiala J, Wijesuriya H, Mehrabian M, et al. Nutrigenetic association of the 5-lipoxygenase gene with myocardial infarction. Am. J. Clin. Nutr. 2008; 88(4):934–40. [PubMed: 18842779]

4. Arab L, Cambou MC, Craft N, Wesseling-Perry K, Jardack P, Ang A. Racial differences in correlations between reported dietary intakes of carotenoids and their concentration biomarkers. Am. J. Clin. Nutr. 2011; 93(5):1102–8. [PubMed: 21389177]
5. Asbeck I, Mast M, Bierwag A, Westenhofer J, Acheson KJ, Muller MJ. Severe underreporting of energy intake in normal weight subjects: use of an appropriate standard and relation to restrained eating. Public Health Nutr. 2002; 5(5):683–90. [PubMed: 12372163]
6. Bathalon GP, Tucker KL, Hays NP, Vinken AG, Greenberg AS, et al. Psychological measures of eating behavior and the accuracy of 3 common dietary assessment methods in healthy postmenopausal women. Am. J. Clin. Nutr. 2000; 71(3):739–45. [PubMed: 10702167]
7. Beaton GH, Milner J, Corey P, McGuire V, Cousins M, et al. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Am. J. Clin. Nutr. 1979; 32(12):2546–59. [PubMed: 506977]
8. Beaton GH, Milner J, McGuire V, Feather TE, Little JA. Source of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Carbohydrate sources, vitamins, and minerals. Am. J. Clin. Nutr. 1983; 37(6):986–95. [PubMed: 6846242]
9. Bingham SA, Luben R, Welch A, Wareham N, Khaw KT, Day N. Are imprecise methods obscuring a relation between fat and breast cancer? Lancet. 2003; 362(9379):212–14. [PubMed: 12885485]
10. Block, G.; DiSogra, C. WIC Dietary Assessment Validation Study. Final Report. U.S. Dep. Agric., Food Nutr. Serv.; Alexandria, VA: 1994.
11. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am. J. Epidemiol. 1986; 124(3):453–69. [PubMed: 3740045]
12. Block G, Thompson FE, Hartman AM, Larkin FA, Guire KE. Comparison of two dietary questionnaires validated against multiple dietary records collected during a 1-year period. J. Am. Diet. Assoc. 1992; 92(6):686–93. [PubMed: 1607564]
13. Bodner JE, Haggerty ES, Ingwersen LA, Perloff BP, Moshfegh AJ. Methods used by Americans to estimate portion sizes of foods and beverages. J. Am. Diet. Assoc. 2003; 103(9 Suppl. 1) A-23 (Abstr.).
14. Bogle M, Stuff J, Davis L, Forrester I, Strickland E, et al. Validity of a telephone-administered 24-hour dietary recall in telephone and non-telephone households in the rural Lower Mississippi Delta region. J. Am. Diet. Assoc. 2001; 101(2):216–22. [PubMed: 11271695]
15. Carithers T, Dubbert PM, Crook E, Davy B, Wyatt SB, et al. Dietary assessment in African Americans: methods used in the Jackson Heart Study. Ethn. Dis. 2005; 15(4 Suppl. 6) S6–49–55.
16. Carithers TC, Talegawkar SA, Rowser ML, Henry OR, Dubbert PM, et al. Validity and calibration of food frequency questionnaires used with African-American adults in the Jackson Heart Study. J. Am. Diet. Assoc. 2009; 109(7):1184–93. [PubMed: 19559135]
17. Casey PH, Goolsby SL, Lensing SY, Perloff BP, Bogle ML. The use of telephone interview methodology to obtain 24-hour dietary recalls. J. Am. Diet. Assoc. 1999; 99(11):1406–11. [PubMed: 10570678]
18. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, et al. Replicating genotypephenotype associations. Nature. 2007; 447(7145):655–60. [PubMed: 17554299]
19. Cocco E, Sotgiu A, Costa G, Murru MR, Mancosu C, et al. *HLA-DR,DQ* and *APOE* genotypes and gender influence in Sardinian primary progressive MS. Neurology. 2005; 64(3):564–66. [PubMed: 15699400]
20. Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. Nature. 2003; 422(6934):835–47. [PubMed: 12695777]
21. Corella D, Ordovas JM. Single nucleotide polymorphisms that influence lipid metabolism: interaction with dietary factors. Annu. Rev. Nutr. 2005; 25:341–90. [PubMed: 16011471]
22. Crutcher KA. Apolipoprotein E is a prime suspect, not just an accomplice, in Alzheimer's disease. J. Mol. Neurosci. 2004; 23(3):181–88. [PubMed: 15181246]
23. Cusack S, Cashman KD. Impact of genetic variation on metabolic response of bone to diet. Proc. Nutr. Soc. 2003; 62(4):901–12. [PubMed: 15018490]
24. Cypel YS, Guenther PM, Petot GJ. Validity of portion-size measurement aids: a review. J. Am. Diet. Assoc. 1997; 97(3):289–92. [PubMed: 9060946]

25. Daniel H. Genomics and proteomics: importance for the future of nutrition research. Br. J. Nutr. 2002; 87(Suppl. 2):S305–11. [PubMed: 12088535]
26. Day N, McKeown N, Wong M, Welch A, Bingham S. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. Int. J. Epidemiol. 2001; 30(2):309–17. [PubMed: 11369735]
27. DeBusk RM, Fogarty CP, Ordovas JM, Kornman KS. Nutritional genomics in practice: Where do we begin? J. Am. Diet. Assoc. 2005; 105(4):589–98. [PubMed: 15800562]
28. Desiere F. Towards a systems biology understanding of human health: interplay between genotype, environment and nutrition. Biotechnol. Annu. Rev. 2004; 10:51–84. [PubMed: 15504703]
29. Djoussé L, Padilla H, Nelson TL, Gaziano JM, Mukamal KJ. Diet and metabolic syndrome. Endocr. Metab. Immune Disord. Drug Targets. 2010; 10(2):124–37. [PubMed: 20088815]
30. Dodd KW, Guenther PM, Freedman LS, Subar AF, Kipnis V, et al. Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. J. Am. Diet. Assoc. 2006; 106(10):1640–50. [PubMed: 17000197]
31. Dwyer JH, Allayee H, Dwyer KM, Fan J, Wu H, et al. Arachidonate 5-lipoxygenase promoter genotype, dietary arachidonic acid, and atherosclerosis. N. Engl. J. Med. 2004; 350(1):29–37. [PubMed: 14702425]
32. Esposito K, Maiorino MI, Ceriello A, Giugliano D. Prevention and control of type 2 diabetes by Mediterranean diet: a systematic review. Diabetes Res. Clin. Pract. 2010; 89(2):97–102. [PubMed: 20546959]
33. Eur. Food Propensity Quest. Dept. Epidemiol., Ger. Inst. Hum. Nutr. Potsdam-Rehbrücke. 2013 https://nugo.dife.de/efbo/portal/en.
34. Ferguson LR. Nutrigenomics: integrating genomic approaches into nutrition research. Mol. Diagn. Ther. 2006; 10(2):101–8. [PubMed: 16669608]
35. Freedman LS, Potischman N, Kipnis V, Midthune D, Schatzkin A, et al. A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. Int. J. Epidemiol. 2006; 35(4): 1011–21. [PubMed: 16672309]
36. Freedman LS, Schatzkin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional cohort studies. J. Natl. Cancer Inst. 2011; 103(14):1086–92. [PubMed: 21653922]
37. Gersovitz M, Madden JP, Smiciklas-Wright H. Validity of the 24-hr. dietary recall and seven-day record for group comparisons. J. Am. Diet. Assoc. 1978; 73(1):48–55. [PubMed: 659761]
38. Gibney MJ, Walsh M, Brennan L, Roche HM, German B, van Ommen B. Metabolomics in human nutrition: opportunities and challenges. Am. J. Clin. Nutr. 2005; 82(3):497–503. [PubMed: 16155259]
39. Gillies PJ, Krul ES. Using genetic variation to optimize nutritional preemption. J. Nutr. 2007; 137(1 Suppl.):270–74S.
40. Go VL, Butrum RR, Wong DA. Diet, nutrition, and cancer prevention: the postgenomic era. J. Nutr. 2003; 133(11 Suppl. 1):3830–36S.
41. Gonzalez CA. The European Prospective Investigation into Cancer and Nutrition (EPIC). Public Health Nutr. 2006; 9(1a):124–26. [PubMed: 16512959]
42. Gonzalez P, Reguero JR, Lozano I, Moris C, Coto E. A functional Sp1/Egr1-tandem repeat polymorphism in the 5-lipoxygenase gene is not associated with myocardial infarction. Int. J. Immunogenet. 2007; 34(2):127–30. [PubMed: 17373938]
43. Goris AH, Westerterp KR. Underreporting of habitual food intake is explained by undereating in highly motivated lean women. J. Nutr. 1999; 129(4):878–82. [PubMed: 10203564]
44. Hankin JH, Wilkens LR, Kolonel LN, Yoshizawa CN. Validation of a quantitative diet history method in Hawaii. Am. J. Epidemiol. 1991; 133(6):616–28. [PubMed: 2006649]
45. Hartman AM, Brown CC, Palmgren J, Pietinen P, Verkasalo M, et al. Variability in nutrient and food intakes among older middle-aged men. Implications for design of epidemiologic and validation studies using food recording. Am. J. Epidemiol. 1990; 132(5):999–1012. [PubMed: 2239915]
46. Hatters DM, Peters-Libeu CA, Weisgraber KH. Apolipoprotein E structure: insights into function. Trends Biochem. Sci. 2006; 31(8):445–54. [PubMed: 16820298]

47. Hein R, Beckmann L, Chang-Claude J. Sample size requirements for indirect association studies of gene-environment interactions (G × E). Genet. Epidemiol. 2008; 32(3):235–45. [PubMed: 18163529]
48. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat. Genet. 2001; 29(3):306–9. [PubMed: 11600885]
49. Jackson MD, Walker SP, Younger NM, Bennett FI. Use of a food frequency questionnaire to assess diets of Jamaican adults: validation and correlation with biomarkers. Nutr. J. 2011; 10:28. [PubMed: 21477338]
50. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 1961; 3:318–56. [PubMed: 13718526]
51. Jacques PF, Sulsky SI, Sadowski JA, Phillips JC, Rush D, Willett WC. Comparison of micronutrient intake measured by a dietary questionnaire and biochemical indicators of micronutrient status. Am. J. Clin. Nutr. 1993; 57(2):182–89. [PubMed: 8424386]
52. Johansson G, Wikman A, Ahren AM, Hallmans G, Johansson I. Underreporting of energy intake in repeated 24-hour recalls related to gender, age, weight status, day of interview, educational level, reported food intake, smoking habits and area of living. Public Health Nutr. 2001; 4(4):919–27. [PubMed: 11527517]
53. Junien C. Impact of diets and nutrients/drugs on early epigenetic programming. J. Inherit. Metab. Dis. 2006; 29(2–3):359–65. [PubMed: 16763902]
54. Kabagambe EK, Baylin A, Allan DA, Siles X, Spiegelman D, Campos H. Application of the method of triads to evaluate the performance of food frequency questionnaires and biomarkers as indicators of long-term dietary intake. Am. J. Epidemiol. 2001; 154(12):1126–35. [PubMed: 11744518]
55. Kanoni S, Nettleton JA, Hivert MF, Ye Z, van Rooij FJ, et al. Total zinc intake may modify the glucose-raising effect of a zinc transporter (SLC30A8) variant: a 14-cohort meta-analysis. Diabetes. 2011; 60(9):2407–16. [PubMed: 21810599]
56. Kaput J, Perlina A, Hatipoglu B, Bartholomew A, Nikolsky Y. Nutrigenomics: concepts and applications to pharmacogenomics and clinical medicine. Pharmacogenomics. 2007; 8(4):369–90. [PubMed: 17391074]
57. Keusch GT. What doomics mean for the science and policy of the nutritional sciences? Am. J. Clin. Nutr. 2006; 83(2):520–22S.
58. Kristal AR, Peters U, Potter JD. Is it time to abandon the food frequency questionnaire? Cancer Epidemiol. Biomarkers Prev. 2005; 14(12):2826–28. [PubMed: 16364996]
59. Kussmann M, Raymond F, Affolter M. OMICS-driven biomarker discovery in nutrition and health. J. Biotechnol. 2006; 124(4):758–87. [PubMed: 16600411]
60. Lai CQ. Adaptive genetic variation and population differences. Prog. Mol. Biol. Transl. Sci. 2012; 108:461–89. [PubMed: 22656388]
61. Laskowitz DT, Fillit H, Yeung N, Toku K, Vitek MP. Apolipoprotein E-derived peptides reduce CNS inflammation: implications for therapy of neurological disease. Acta Neurol. Scand. Suppl. 2006; 185:15–20. [PubMed: 16866906]
62. Liu K, Stamler J, Dyer A, McKeever J, McKeever P. Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. J. Chronic Dis. 1978; 31(6–7):399–418. [PubMed: 711832]
63. Lloyd AJ, Beckmann M, Haldar S, Seal C, Brandt K, Draper J. Data-driven strategy for the discovery of potential urinary biomarkers of habitual dietary exposure. Am. J. Clin. Nutr. 2013; 97(2):377–89. [PubMed: 23269817]
64. Mahley RW, Weisgraber KH, Huang Y. Apolipoprotein E4: a causative factor and therapeutic target in neuropathology, including Alzheimer's disease. Proc. Natl. Acad. Sci. USA. 2006; 103(15):5644–51. [PubMed: 16567625]
65. Masson LF, McNeill G. The effect of genetic variation on the lipid response to dietary change: recent findings. Curr. Opin. Lipidol. 2005; 16(1):61–67. [PubMed: 15650565]
66. Masson LF, McNeill G, Avenell A. Genetic variation and the lipid response to dietary intervention: a systematic review. Am. J. Clin. Nutr. 2003; 77(5):1098–111. [PubMed: 12716659]

67. Mathers JC. Nutritional modulation of ageing: genomic and epigenetic approaches. Mech. Ageing Dev. 2006; 127(6):584–89. [PubMed: 16513160]
68. Maurer J, Taren DL, Teixeira PJ, Thomson CA, Lohman TG, et al. The psychosocial and behavioral characteristics related to energy misreporting. Nutr. Rev. 2006; 64(2):53–66. Part 1. [PubMed: 16536182]
69. Mayer-Davis EJ, Vitolins MZ, Carmichael SL, Hemphill S, Tsaroucha G, et al. Validity and reproducibility of a food frequency interview in a multi-cultural epidemiology study. Ann. Epidemiol. 1999; 9(5):314–24. [PubMed: 10976858]
70. Milner JA. Nutrition in the "omics" era. Forum. Nutr. 2007; 60:1–24. [PubMed: 17684397]
71. Mortensen EL, Hogh P. A gender difference in the association between *APOE* genotype and age-related cognitive decline. Neurology. 2001; 57(1):89–95. [PubMed: 11445633]
72. Moshfegh AJ, Rhodes DG, Baer DJ, Murayi T, Clemens JC, et al. The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. Am. J. Clin. Nutr. 2008; 88(2):324–32. [PubMed: 18689367]
73. Muller M, Kersten S. Nutrigenomics: goals and strategies. Nat. Rev. Genet. 2003; 4(4):315–22. [PubMed: 12671662]
74. Munafo MR. Credible genetic associations? Int. J. Mol. Epidemiol. Genet. 2010; 1(1):31–34. [PubMed: 21537450]
75. Nettleton JA, Hivert MF, Lemaitre RN, McKeown NM, Mozaffarian D, et al. Meta-analysis investigating associations between healthy diet and fasting glucose and insulin levels and modification by loci associated with glucose homeostasis in data from 15 cohorts. Am. J. Epidemiol. 2013; 177(2):103–15. [PubMed: 23255780]
76. Nettleton JA, McKeown NM, Kanoni S, Lemaitre RN, Hivert MF, et al. Interactions of dietary whole-grain intake with fasting glucose- and insulin-related genetic loci in individuals of European descent: a meta-analysis of 14 cohort studies. Diabetes Care. 2010; 33(12):2684–91. [PubMed: 20693352]
77. NIH Genes Environ. Health Initiative (GEI). NIH; Bethesda, MD: 2013. http://www.gei.nih.gov./
78. Novelli V, Viviani Anselmi C, Roncarati R, Guffanti G, Malovini A, et al. Lack of replication of genetic associations with human longevity. Biogerontology. 2008; 9(2):85–92. [PubMed: 18034366]
79. Novotny JA, Rumpler WV, Riddick H, Hebert JR, Rhodes D, et al. Personality characteristics as predictors of underreporting of energy intake on 24-hour dietary recall interviews. J. Am. Diet. Assoc. 2003; 103(9):1146–51. [PubMed: 12963942]
80. NutritionQuest. Block assessment tools. NutritionQuest; Berkeley, CA: 2013. http://www.nutritionquest.com/assessment/
81. Ordovas JM, Corella D. Nutritional genomics. Annu. Rev. Genomics Hum. Genet. 2004; 5:71–118. [PubMed: 15485344]
82. Ordovas JM, Litwack-Klein L, Wilson PW, Schaefer MM, Schaefer EJ. Apolipoprotein E isoform phenotyping methodology and population frequency with identification of apoE1 and apoE5 isoforms. J. Lipid Res. 1987; 28(4):371–80. [PubMed: 3585172]
83. Ordovas JM, Mooser V. The APOE locus and the pharmacogenetics of lipid response. Curr. Opin. Lipidol. 2002; 13(2):113–17. [PubMed: 11891412]
84. Pakseresht M, Sharma S. Validation of a quantitative food frequency questionnaire for Inuit population in Nunavut, Canada. J. Hum. Nutr. Diet. 2010; 23(Suppl. 1):67–74. [PubMed: 21158964]
85. Peterson KE, Hebert JR, Hurley TG, Resnicow K, Thompson FE, et al. Accuracy and precision of two short screeners to assess change in fruit and vegetable consumption among diverse populations participating in health promotion intervention trials. J. Nutr. 2008; 138(1):218–25S.
86. Prentice RL. Dietary assessment and the reliability of nutritional epidemiology reports. Lancet. 2003; 362(9379):182–83. [PubMed: 12885475]
87. Prentice RL, Mossavar-Rahmani Y, Huang Y, Van Horn L, Beresford SA, et al. Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. Am. J. Epidemiol. 2011; 174(5):591–603. [PubMed: 21765003]

88. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ. Cardiovasc. Genet. 2009; 2(1): 73–80. [PubMed: 20031568]

89. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am. J. Epidemiol. 1992; 135(10):1114–26. discussion 1127–36. [PubMed: 1632423]

90. Rist MJ, Wenzel U, Daniel H. Nutrition and food science go genomic. Trends Biotechnol. 2006; 24(4):172–78. [PubMed: 16488035]

91. Roark RA, Niederhauser VP. Fruit and vegetable intake: issues with definition and measurement. Public Health Nutr. 2013; 16(1):2–7. [PubMed: 22475520]

92. Rontu R, Ojala P, Hervonen A, Goebeler S, Karhunen PJ, et al. Apolipoprotein E genotype is related to plasma levels of C-reactive protein and lipids and to longevity in nonagenarians. Clin. Endocrinol. Oxf. 2006; 64(3):265–70.

93. Rubin J, Berglund L. Apolipoprotein E and diets: a case of gene-nutrient interaction? Curr. Opin. Lipidol. 2002; 13(1):25–32. [PubMed: 11790960]

94. Schatzkin A, Kipnis V, Carroll RJ, Midthune D, Subar AF, et al. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. Int. J. Epidemiol. 2003; 32(6):1054–62. [PubMed: 14681273]

95. Shock, NW.; Greulich, RC.; Andres, R.; Arenberg, D.; Costa, PT., Jr., et al. NIH Publ. No. 84-2450. U.S. Dep. Health Hum. Serv.; Bethesda, MD: 1984. Normal Human Aging: The Baltimore Longitudinal Study of Aging.

96. Smith C, Ngwa J, Tanaka T, Qi Q, Wojczynski M, et al. Lipoprotein receptor-related protein 1 variants and dietary fatty acids: meta-analysis of European origin and African American studies. Intl. J. Obes. 2013 In press.

97. Sorlie PD, Aviles-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglus ML, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. Ann. Epidemiol. 2010; 20(8):629–41. [PubMed: 20609343]

98. Stover PJ, Harlan WR, Hammond JA, Hendershot T, Hamilton CM. PhenX: a toolkit for inter-disciplinary genetics research. Curr. Opin. Lipidol. 2010; 21(2):136–40. [PubMed: 20154612]

99. Stram DO, Hankin JH, Wilkens LR, Pike MC, Monroe KR, et al. Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles. Am. J. Epidemiol. 2000; 151(4):358–70. [PubMed: 10695594]

100. Subar AF, Kirkpatrick SI, Mittl B, Zimmerman TP, Thompson FE, et al. The Automated Self-Administered 24-hour dietary recall (ASA24): a resource for researchers, clinicians, and educators from the National Cancer Institute. J. Acad. Nutr. Diet. 2012; 112(8):1134–37. [PubMed: 22704899]

101. Subar AF, Krebs-Smith SM, Cook A, Kahle LL. Dietary sources of nutrients among US adults, 1989 to 1991. J. Am. Diet. Assoc. 1998; 98(5):537–47. [PubMed: 9597026]

102. Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. Am. J. Epidemiol. 2001; 154(12):1089–99. [PubMed: 11744511]

103. Talegawkar SA, Johnson EJ, Carithers TC, Taylor HA, Bogle ML, Tucker KL. Carotenoid intakes, assessed by food-frequency questionnaires (FFQs), are associated with serum carotenoid concentrations in the Jackson Heart Study: validation of the Jackson Heart Study Delta NIRI Adult FFQs. Public Health Nutr. 2008; 11(10):989–97. [PubMed: 18053294]

104. Thompson FE, Midthune D, Subar AF, Kipnis V, Kahle LL, Schatzkin A. Development and evaluation of a short instrument to estimate usual dietary intake of percentage energy from fat. J. Am. Diet. Assoc. 2007; 107(5):760–67. [PubMed: 17467371]

105. Thompson FE, Midthune D, Williams GC, Yaroch AL, Hurley TG, et al. Evaluation of a short dietary assessment instrument for percentage energy from fat in an intervention study. J. Nutr. 2008; 138(1):193–99S.

106. Thompson FE, Subar AF, Loria CM, Reedy JL, Baranowski T. Need for technological innovation in dietary assessment. J. Am. Diet. Assoc. 2010; 110(1):48–51. [PubMed: 20102826]

107. Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, et al. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. J. Am. Diet. Assoc. 2006; 106(10):1575–87. [PubMed: 17000190]

108. Tooze JA, Subar AF, Thompson FE, Troiano R, Schatzkin A, Kipnis V. Psychosocial predictors of energy underreporting in a large doubly labeled water study. Am. J. Clin. Nutr. 2004; 79(5): 795–804. [PubMed: 15113717]

109. Tucker KL, Bianchi LA, Maras J, Bermudez OI. Adaptation of a food frequency questionnaire to assess diets of Puerto Rican and non-Hispanic adults. Am. J. Epidemiol. 1998; 148(5):507–18. [PubMed: 9737563]

110. Tucker KL, Chen H, Vogel S, Wilson PW, Schaefer EJ, Lammi-Keefe CJ. Carotenoid intakes, assessed by dietary questionnaire, are associated with plasma carotenoid concentrations in an elderly population. J. Nutr. 1999; 129(2):438–45. [PubMed: 10024624]

111. Tucker KL, Mahnken B, Wilson PW, Jacques P, Selhub J. Folic acid fortification of the food supply. Potential benefits and risks for the elderly population. JAMA. 1996; 276(23):1879–85. [PubMed: 8968013]

112. Tucker KL, Maras J, Champagne C, Connell C, Goolsby S, et al. A regional food-frequency questionnaire for the US Mississippi Delta. Public Health Nutr. 2005; 8(1):87–96. [PubMed: 15705249]

113. van Ommen B, Stierum R. Nutrigenomics: exploiting systems biology in the nutrition and health arena. Curr. Opin. Biotechnol. 2002; 13(5):517–21. [PubMed: 12459347]

114. Vernocchi P, Vannini L, Gottardi D, Del Chierico F, Serrazanetti DI, et al. Integration of datasets from different analytical techniques to assess the impact of nutrition on human metabolome. Front. Cell Infect. Microbiol. 2012; 2:156. [PubMed: 23248777]

115. Viocare Vioscreen. VioFFQ. Viocare; Princeton, NJ: 2013. http://www.viocare.com/vioffq.aspx

116. Wang J, Li D, Dangott LJ, Wu G. Proteomics and its role in nutrition research. J. Nutr. 2006; 136(7):1759–62. [PubMed: 16772433]

117. Watson MA, Gay L, Stebbings WS, Speakman CT, Bingham SA, Loktionov A. Apolipoprotein E gene polymorphism and colorectal cancer: gender-specific modulation of risk and prognosis. Clin. Sci. Lond. 2003; 104(5):537–45. [PubMed: 12529167]

118. Willett, W. Nutritional Epidemiology. 2nd. Oxford Univ. Press; New York: 1998.

119. Willett W. Invited commentary: OPEN questions. Am. J. Epidemiol. 2003; 158(1):22–24.

120. Willett WC, Hu FB. Not the time to abandon the food frequency questionnaire: point. Cancer Epidemiol. Biomarkers Prev. 2006; 15(10):1757–58. [PubMed: 17021351]

121. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am. J. Epidemiol. 1985; 122(1):51–65. [PubMed: 4014201]

122. Wong SS, Boushey CJ, Novotny R, Gustafson DR. Evaluation of a computerized food frequency questionnaire to estimate calcium intake of Asian, Hispanic, and non-Hispanic white youth. J. Am. Diet. Assoc. 2008; 108(3):539–43. [PubMed: 18313437]

Table 1

Dietary assessment and dietary fat intakes in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium*

Cohort	Country	*n*	Dietary assessment method	Total fat % energy	Saturated fat % energy	Polyunsaturated fatty acid % energy
Atherosclerosis Risk in Communities Family Heart Study	USA USA	2,980 9,189	Interviewer-administered 66-item food frequency questionnaire (FFQ) modified from the Willett 61-item FFQ. Participants were asked to indicate average consumption over the past year, with 9 frequency categories, from never or <1 time per month to ≥6 times per day. Standard portion sizes were given as a reference.	33.2 ± 6.8 30.5 ± 7.5	12.2 ± 3.1 11.2 ± 3.2	5.1 ± 1.5 4.5 ± 1.4
Cardiovascular Health Study	USA	3,222	Self-administered 99-item picture-sort version of the National Cancer Institute FFQ in 1989–90 and the 131-item Willett FFQ in 1992–93. Participants were asked to indicate average consumption over the past year, with 9 frequency categories, from never to ≥5 times per week (picture sort) or <1 time per month to ≥6 times per day (Willett FFQ).	32.3 ± 6.0	10.3 ± 2.2	7.4 ± 2.2
European Prospective Investigation of Cancer Norfolk Fenland	UK	2,353	Self-administered 130-item FFQ with lists of foods from the average British diet. Participants were asked to indicate average consumption over the past year, with 9 frequency categories, from never or <1 per month to ≥6 times per day. Portion sizes were specified. Intakes were estimated in grams per day.	32.3 ± 5.7 33.3 ± 5.9	12.3 ± 3.2 12.3 ± 3.2	6.2 ± 2.0 6.5 ± 1.8
Framingham Heart Study Nurses Health Study	USA USA	6,374 3,065	Self-administered 126-item Willett FFQ. Participants were asked to indicate average consumption over the past year, with 9 frequency categories, from never or <1 per month to ≥6 times per day. Portion sizes were specified.	28.4 ± 6.1 33.2 ± 5.6	11.0 ± 3.0 11.8 ± 2.5	5.8 ± 1.6 6.3 ± 1.6
Genetics of Lipid-Lowering Drugs and Diet Network	USA	1,120	Interviewer-administered NCI 124-item Diet History Questionnaire with questions for portion size.	35.5 ± 6.7	11.9 ± 2.7	7.6 ± 2.1
Health, Aging, and Body Composition	USA	1,499	Interviewer-administered 108-item Block FFQ. Participants were asked to indicate average consumption over the past year, with 9 frequency categories, from never to every day. Portion size was estimated using wood blocks, food models, standard kitchen measures, and flashcards.	32.9 ± 7.6	9.4 ± 2.5	8.7 ± 2.8
Health Professionals Follow-up Study	USA	2,326	Self-administered 131-item Willett FFQ. Participants were asked to indicate average consumption over the past year, with 9 frequency categories, from never or <1 time per month to ≥6 times per day. Standard portion sizes were given as a reference.	32.8 ± 6.4	11.3 ± 2.8	6.1 ± 1.6

Cohort	Country	*n*	Dietary assessment method	Total fat % energy	Saturated fat % energy	Polyunsaturated fatty acid % energy
Invecchiare in Chianti (Aging in the Chianti Area)	Italy	1,100	Interviewer-administered 236-item FFQ. Participants were asked to specify the size of the portion usually consumed in comparison to a range of portions shown in colored photographs.	31.0 ± 5.1	10.4 ± 2.2	3.4 ± 0.7
Multi-Ethnic Study of Atherosclerosis	USA	2,289	Self-administered 120-item Block FFQ, modified to include ethnic-specific foods.	33.4 ± 7.2	11.0 ± 3.3	7.0 ± 2.0
Rotterdam Study	Netherlands	4,576	**a.** Self-administered simple food list (no frequency or portion size). **b.** Interviewer-administered FFQ. Participants were asked to indicate average consumption over the past year, with 9 frequency categories, from never or <1 time per month to ≥6 times per day. Portion sizes were asked.	36.3 ± 6.1	14.4 ± 3.2	6.9 ± 1.1
Young Finns Study	Finland	1,762	Modified 131-item FFQ developed by the Finnish National Institute for Health and Welfare.	32.8 ± 4.8	11.8 ± 2.4	5.3 ± 1.1

* Adapted from (96).

Table 2

PhenX[a] Toolkit recommendations[b]

Measure	Features	Validity information	Limitations
5-Factor Screener[c]	Frequency of intake for specific food items for past month Age- and sex-specific median portion sizes are assigned	Indirect measurement error model versus multiple 24 HRs for 260 men and 222 women in the OPENs[d] study	■Estimation equations based on US National Health Interview Survey ■"…estimates may be less useful for populations with more ethnic diets, including Asian and possibly Latino populations"[c]
Calcium intake (mg/d)	Frequencies of 20 items used in weighted regression equations	Deattenuated r = 0.59 for men and 0.44 for women	■ Regression weights and selected foods likely to differ in different populations
Dairy food (servings/d)	Frequency of milk and cheese intake summed and used in regression equations	Deattenuated r = 0.64 for men and women	■ Does not include yogurt ■ Does not capture milk in coffee or tea
Fiber intake (g/d)	Frequencies of 20 items used in weighted regression equations	Deattenuated r = 0.52 for men and 0.54 for women	
Fruits and vegetables (servings/d)	Pyramid portions of 9 fruit and vegetable items calculated and then estimated in regression equations	Deattenuated r = 0.58 for men and 0.73 for women	
Added sugars (tsp/d)	Median tsp of sugar summed from 4 items and then estimated with regression equations	Deattenuated r = 0.68 for men and 0.66 for women	
%Energy from fat	Questionnaire with frequencies for 15 food items + 2 use questions. Amount from added solid fats is calculated and then included with other frequencies in weighted regression equations	r = 0.52 to 0.77 among men and 0.36 to 0.59 among women in the BCC[e] study. Positive predictive value for classification <30% energy from fat ranged from 0.15 for women in Georgia to 0.71 for women in Rhode Island	■ Does not include oils, snack chips, most meats, fried chicken, pizza, etc. ■ Regression equations likely to be highly population specific ■ "If accurate assessment of diet… is essential, a more detailed instrument… may be warranted"[e]
Caffeine intake	Frequencies of 13 items, including coffee, tea, and soft drinks; with small, medium, and large portion options		
Total dietary intake	24HR using the ASA24[f]		■ Misclassification from day-to-day variability likely limits use in complex interactions

[a] PhenX, Phenotypes and eXposures; collaborative funded by the National Human Genome Research Institute.

[b] Recommendations for food intake only.

[c] PhenX Toolkit; https://www.phenxtoolkit.org/index.php?pageLink%20=%20browse.protocoldetails&id%20=%2050202.

[d] OPEN, Observing Protein and Energy Nutrition study.

[e] BCC, Behavioral Change Consortium (105).

[f] ASA24, Automated Self-Administered 24-Hour Recall; http://riskfactor.cancer.gov/tools/instruments/asa24/.