Title	Joint data analysis in nutritional epidemiology: identification of observational studies and minimal requirements
Authors	Pinart, Mariona;Nimptsch, Katharina;Bouwman, Jildau;Dragsted, Lars O.;Yang, Chen;De Cock, Nathalie;Lachat, Carl;Perozzi, Giuditta;Canali, Raffaella;Lombardo, Rosario;D'Archivio, Massimo;Guillaume, MichSle;Donneau, Anne-Francoise;Jeran, Stephanie;Linseisen, Jakob;Kleiser, Christina;Nothlings, Ute;Barbaresko, Janett;Boeing, Heiner;Stelmach-Mardas, Marta;Heuer, Thorsten;Laird, Eamon;Walton, Janette;Gasparini, Paolo;Robino, Antonietta;Castano, Luis;Rojo-Martinez, Gemma;Merino, Jordi;Masana, Luis;Standl, Marie;Schulz, Holger;Biagi, Elena;Nurk, Eha;Matthys, Christophe;Gobbetti, Marco;De Angelis, Maria;Windler, Eberhard;Zyriax, Birgit-Christiane;Tafforeau, Jean;Pischon, Tobias
Publication date	2018
Original Citation	Pinart, M., Nimptsch, K., Bouwman, J., Dragsted, L. O., Yang, C., De Cock, N., Lachat, C., Perozzi, G., Canali, R., Lombardo, R., D'Archivio, M., Guillaume, M., Donneau, A.-F., Jeran, S., Linseisen, J., Kleiser, C., Nöthlings, U., Barbaresko, J., Boeing, H., Stelmach-Mardas, M., Heuer, T., Laird, E., Walton, J., Gasparini, P., Robino, A., Castaño, L., Rojo-Martínez, G., Merino, J., Masana, L., Standl, M., Schulz, H., Biagi, E., Nurk, E., Matthys, C., Gobbetti, M., de Angelis, M., Windler, E., Zyriax, B.-C., Tafforeau, J. and Pischon, T. (2018) 'Joint Data Analysis in Nutritional Epidemiology: Identification of Observational Studies and Minimal Requirements', Journal of Nutrition, 148(2), pp. 285-297. doi: 10.1093/jn/nxx037
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://academic.oup.com/jn/article/148/2/285/4913031 - 10.1093/jn/nxx037
Rights	

The Journal of Nutrition
Methodology and Mathematical Modeling

Joint Data Analysis in Nutritional Epidemiology: Identification of Observational Studies and Minimal Requirements

Mariona Pinart,[1] Katharina Nimptsch,[1] Jildau Bouwman,[2] Lars O Dragsted,[3] Chen Yang,[4] Nathalie De Cock,[4] Carl Lachat,[4] Giuditta Perozzi,[5] Raffaella Canali,[5] Rosario Lombardo,[6] Massimo D'Archivio,[7] Michèle Guillaume,[8] Anne-Françoise Donneau,[8] Stephanie Jeran,[1] Jakob Linseisen,[9,10] Christina Kleiser,[9] Ute Nöthlings,[11] Janett Barbaresko,[11] Heiner Boeing,[12] Marta Stelmach-Mardas,[12,13] Thorsten Heuer,[14] Eamon Laird,[15] Janette Walton,[16] Paolo Gasparini,[17,18] Antonietta Robino,[18] Luis Castaño,[19] Gemma Rojo-Martínez,[20,21] Jordi Merino,[22,23] Luis Masana,[22] Marie Standl,[24] Holger Schulz,[24] Elena Biagi,[25] Eha Nurk,[26] Christophe Matthys,[27] Marco Gobbetti,[28] Maria de Angelis,[29] Eberhard Windler,[30] Birgit-Christiane Zyriax,[31] Jean Tafforeau,[32] and Tobias Pischon[1,33,34,35]

[1]Molecular Epidemiology Research Group, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany; [2]TNO, Microbiology and Systems Biology Group, Zeist, Netherlands; [3]Department of Nutrition, Exercise, and Sports, University of Copenhagen, Copenhagen, Denmark; [4]Department of Food Safety and Food Quality, Ghent University, Ghent, Belgium; [5]CREA-NUT, Food and Nutrition Research Center, Rome, Italy; [6]The Microsoft Research–University of Trento Centre for Computational and Systems Biology (COSBI), Trentino, Italy; [7]Center for Gender-Specific Medicine, Istituto Superiore di Sanità, Rome, Italy; [8]Department of Public Health, University of Liège, Liège, Belgium; [9]Helmholtz Zentrum München, Institute of Epidemiology, Neuherberg, Germany; [10]Ludwig-Maximilians-Universität (LMU) München, Medical Faculty, Institute of Medical Information Processing, Biometry, and Epidemiology (IBE), Chair of Epidemiology at University Centre for Health Care Sciences at the Augsburg Clinic (UNIKA-T Augsburg), Ausburg, Germany; [11]Department of Nutrition and Food Sciences, University of Bonn, Bonn, Germany; [12]Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbrücke, Nuthetal, Germany; [13]Department of Pediatric Gastroenterology and Metabolic Diseases, Poznan University of Medical Sciences, Poznan, Poland; [14]Department of Nutritional Behavior, Max Rubner-Institut, Federal Research Institute of Nutrition and Food, Karlsruhe, Germany; [15]Centre for Medical Gerontology, School of Medicine, Trinity College Dublin, Dublin, Ireland; [16]School of Food and Nutritional Sciences, University College Cork, Cork, Ireland; [17]Department of Medical Sciences, University of Trieste, Trieste, Italy; [18]Institute for Maternal and Child Health–IRCCS "Burlo Garofolo", Trieste, Italy; [19]Spanish Biomedical Research Center in Diabetes and Associated Metabolic Disorders (CIBERDEM), Rare Diseases Networking Biomedical Research Centre (CIBERER), BioCruces-Hospital Universitario Cruces–The University of the Basque Country (Basque: Euskal Herriko Unibertsitatea/Spanish: Universidad del País Vasco (UPV/EHU)), Barakaldo, Spain; [20]Spanish Biomedical Research Center in Diabetes and Associated Metabolic Disorders (CIBERDEM), Madrid, Spain; [21]Unidad de Gestión Clínica (UGC) Endocrinology and Nutrition. Hospital Regional Universitario de Málaga, Institute of Biomedical Research in Malaga (IBIMA), Málaga, Spain; [22]Research Unit on Lipids and Atherosclerosis, Universitat Rovira i Virgili, Reus, Spain; [23]Diabetes Unit and Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA; [24]Helmholtz Zentrum München–German Research Center for Environmental Health, Institute of Epidemiology I, Neuherberg, Germany; [25]Department of Pharmacy and Biotechnology, Alma Mater Studiorum–University of Bologna, Bologna, Italy; [26]National Institute for Health Development, Tallinn, Estonia; [27]Department of Clinical and Experimental Medicine, The Katholieke Universiteit Leuven (KU Leuven) and Department of Endocrinology, University Hospitals Leuven, Leuven, Belgium; [28]Faculty of Science and Technology, Free University of Bozen, Bolzano, Italy; [29]Department of Soil, Plant and Food Sciences, University of Bari Aldo Moro, Bari, Italy; [30]Preventive Medicine, University Heart Center, University Hospital Hamburg-Eppendorf, Hamburg, Germany; [31]Institute for Health Services Research in Dermatology and Nursing (IVDP), University Medical Center Hamburg-Eppendorf, Hamburg, Germany; [32]Scientific Institute of Public Health, Brussels, Belgium; [33]Charité Universitätsmedizin Berlin, Berlin, Germany; [34]MDC/BIH Biobank, Max Delbrück Center for Molecular Medicine and Berlin Institute of Health, Berlin, Germany; and [35]German Centre for Cardiovascular Research (DZHK), Berlin, Germany

Abstract

Background: Joint data analysis from multiple nutrition studies may improve the ability to answer complex questions regarding the role of nutritional status and diet in health and disease.

Objective: The objective was to identify nutritional observational studies from partners participating in the European Nutritional Phenotype Assessment and Data Sharing Initiative (ENPADASI) Consortium, as well as minimal requirements for joint data analysis.

Methods: A predefined template containing information on study design, exposure measurements (dietary intake, alcohol and tobacco consumption, physical activity, sedentary behavior, anthropometric measures, and sociodemographic and health status), main health-related outcomes, and laboratory measurements (traditional and omics biomarkers) was developed and circulated to those European research groups participating in the ENPADASI under the strategic research area of "diet-related chronic diseases." Information about raw data disposition and metadata sharing was requested. A set of minimal requirements was abstracted from the gathered information.

Manuscript received August 24, 2017. Initial review completed September 21, 2017. Revision accepted November 6, 2017.
First published online February 27, 2018; doi: https://doi.org/10.1093/jn/nxx037.

Results: Studies (12 cohort, 12 cross-sectional, and 2 case-control) were identified. Two studies recruited children only and the rest recruited adults. All studies included dietary intake data. Twenty studies collected blood samples. Data on traditional biomarkers were available for 20 studies, of which 17 measured lipoproteins, glucose, and insulin and 13 measured inflammatory biomarkers. Metabolomics, proteomics, and genomics or transcriptomics data were available in 5, 3, and 12 studies, respectively. Although the study authors were willing to share metadata, most refused, were hesitant, or had legal or ethical issues related to sharing raw data. Forty-one descriptors of minimal requirements for the study data were identified to facilitate data integration.

Conclusions: Combining study data sets will enable sufficiently powered, refined investigations to increase the knowledge and understanding of the relation between food, nutrition, and human health. Furthermore, the minimal requirements for study data may encourage more efficient secondary usage of existing data and provide sufficient information for researchers to draft future multicenter research proposals in nutrition. *J Nutr* 2018;148:285–297.

Keywords: nutritional phenotype, metadata, data integration, data sharing, observational studies

Introduction

The joint analysis of individual-level data from multiple nutrition studies may improve the ability to answer complex questions in studying the role of diet and metabolism in health and disease that individual studies are underpowered to examine (1). Moreover, a joint individual-level data analysis, unlike in study-level meta-analysis, offers the possibility to reuse data in new ways by combining individual data from different studies, thereby increasing the diversity of samples and the robustness of statistical subgroup analyses (i.e., increasing statistical efficiency and flexibility). This is particularly relevant for nutrition studies on biomarkers, because their laboratory analysis is usually expensive and joint data analysis may provide an efficient way of using existing biomarker data.

Although international research funders encourage sharing data to maximize discovery and innovation in public health, scientists are reluctant to engage in such initiatives. Reasons for their concern range from intellectual property rights to potential data misuse or misinterpretation, insufficient participant privacy, confidentiality safeguards to scientists, unfamiliarity with data management systems and metadata standards, and general lack of scientific culture for data sharing (2). In nutritional epidemiology, there are only a few examples of successful implementation of data integration platforms facilitating pooled analysis of individual-level data, including the Nutritional Phenotype database (www.dbnp.org) (3) and the EURopean micronutrient RECommendations Aligned (EURRECA) Network of Excellence (4). Yet, examples of joint data analyses with regard to nutrition and biomarkers are scarce (5).

A prerequisite to facilitating interpretation, comparison, reproducibility, and reuse of data is the identification of minimal information to add as metadata. Several initiatives have developed minimal information checklists containing a set of guidelines or recommendations for reporting data on specific high-throughput experimental technologies that have become a prerequisite for publication in journals (6–11). They follow a hierarchical structure developed by the Investigation, Study, and Assay (ISA) Commons (12), which is a growing community that uses the ISA metadata categories tracking framework to facilitate standards-compliant collection, curation, management, and reuse of data sets in an increasingly diverse set of life science domains (13). These checklists provide access to a range of background information that may help in interpreting results and having a better picture of the context of the study as well as the methods used, data collected, and conclusions drawn. So far, in the field of nutrition, only Minimum Information About a Nutrigenomics Experiment (MIAME/Nutr) was developed for array-based nutrigenomics experiments, as an extension of the MIAME standards (Minimum Information About a Microarray Experiment) (14). However, minimal information checklists for the integration of data sets from nutritional epidemiologic studies are lacking.

The aim of this study was to identify epidemiologic observational studies with a wealth of data and metadata, particularly on dietary assessment and traditional and omics biomarkers within the context of the European Nutritional Phenotype Assessment and Data Sharing Initiative (ENPADASI). These studies served as the basis for the identification of minimal information, hereafter referred to as minimal requirements to connect existing and future study (meta)databases and facilitate data exchange, data interpretation, and increasing the robustness of results from future joint data analysis in nutritional epidemiology—and for infrastructures that support such projects. In the present study, joint data analysis refers to either "pooled," if individual-level data from different studies are stored in a single central database, or "federated," if joint analyses of individual-level data from several studies are conducted without physically transferring their data into a single central database.

Supported by the European Nutritional Phenotype Assessment and Data Sharing Initiative (ENPADASI) and its infrastructure as part of the Joint Programming Initiative "A Healthy Diet for a Healthy Life" (JPI-HDHL), which is funded by national funding agencies in 9 European countries. This work was supported by the German Ministry of Food and Agriculture (BMEL) through the Federal Office for Agriculture and Food (BLE; grant 2814ERA01F). CL received a grant from the Research Foundation–Flanders (FWO; grant G0D4815N). CY is funded by a scholarship from the Chinese Scholarship Council.

Author disclosures: MP, KN, JB, LOD, CY, NDC, CL, GP, RC, RL, MD, MG, A-FD, SJ, JL, CK, UN, JB, HB, MS-M, TH, EL, JW, PG, AR, LC, GR-M, JM, LM, MS, HS, EB, EN, CM, MG, MdA, EW, B-CZ, JT, and TP, no conflicts of interest.

Supplemental Tables 1–7 are available from the "Supplementary data" link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/jn/.

Address correspondence to MP (e-mail: mariona.pinartgilberga@mdc-berlin.de).

Abbreviations used: CRP, C-reactive protein; DASH-IN, Data Sharing Initiative for Nutrition; ENPADASI, European Nutritional Phenotype Assessment and Data Sharing Initiative; HbA1c, glycated hemoglobin; ISA, Investigation, Study, and Assay.

Methods

Consortium assembly

This study was conducted within the framework of ENPADASI, a knowledge hub comprising 51 partners from 9 countries that aim to provide the open-access Data Sharing Initiative for Nutrition (DASH-IN) infrastructure with easy-to-follow instructions for data- and metadata-sharing processes and tools to address political, legal, and ethical barriers to enable joint data analyses (15, 16). It was created in response to a call by the Joint Programming Initiative "A Healthy Diet for a Healthy Life" (JPI-HDHL) within the strategic research area of "diet-related chronic diseases" (17). Interested research groups submitted an "Expression of Interest" letter to the Call Secretariat and networked to develop the ENPADASI program proposal, in which a list of studies (24 observational and 79 intervention or mechanistic) potentially available within the appointed ENPADASI members was provided.

Development of a tool for collecting study metadata

A template was developed to obtain meta-information from each identified observational study. The first version of the template was based on the work conducted in task 2.1.2 "Explorative secondary data analysis and further development of the dynamic and evolving framework of determinants of dietary behaviour" within the JPI-HDHL Determinants of Diet and Physical Activity (DEDIPAC) project (18). This template was then modified by a group of involved researchers to fit the purposes of ENPADASI. The final template was circulated to those ENPADASI partners who were concerned with identification of studies and data collection, and was further extended to JPI-project DEDIPAC partners. To include as many studies as possible, the inclusion criteria were very broad: cohort, case-control, or cross-sectional studies with data on nutrition (i.e., dietary assessment) and conducted in humans.

The template contained the following information: *1*) general information on the study (name of study, study web links, funding body, coordination center of the study), *2*) scope of the study, *3*) study design and recruitment, *4*) exposure measurements (dietary intake, alcohol and tobacco consumption, physical activity, sedentary behavior, anthropometric measurements, sociodemographic information, and health status), *5*) main health-related outcome, and *6*) laboratory measurements in biological samples (traditional biomarkers as well as omics biomarkers, such as proteomics, genomics, transcriptomics, and metabolomics). In addition, the template ascertained information on signed informed consents, ethics committee approval, and potential raw and metadata sharing (within and outside the ENPADASI consortium).

The template was circulated to the principal investigators from the 24 observational studies that expressed their interest when the call was launched. In addition, the template was circulated among the ENPADASI partners to identify more studies. The partners filled in the requested information for the respective study and returned the complete templates to the Max Delbrück Center for Molecular Medicine, where they were kept and aggregated into a final list of observational studies for integration in ENPADASI.

Identification of minimal requirements described in the form of a checklist

The concept of minimal requirements was defined in ENPADASI as the description of a set of metadata descriptors sufficient for clear interpretation and use of data, to improve study comparability, and finally, to aid in the development of reusable data-quality metrics. The foundation of the research infrastructure that is currently being created by the ENPADASI initiative is built on minimal requirements to which data should comply (including the necessary information on study design, endpoints, other measurements, data ownership, data availability, and ethical limitations).

To identify the minimal requirements for observational studies metadata entry, we followed the hierarchical structure developed by the ISA Commons (12). The "investigation" category describes the project context; the "study" category describes a unit of research, describing the subjects of study and how they are obtained; and the "assay" category describes any analytical measurement. We developed a checklist structure following the 3 ISA categories. The checklist was also developed in close collaboration with the researchers leading the development of study-quality descriptors for data from nutritional epidemiologic research described elsewhere (19) to avoid overlap. Briefly, the study-quality descriptors were identified after a literature review, a face-to-face meeting to discuss the descriptors found in the literature, and a consensus meeting to decide on the essential study-quality descriptors. The difference between the minimal requirements descriptors and the study-quality descriptors is that the former are mandatory and the latter are an optional set of questions. Both minimal requirements and study-quality descriptors (19) define the meta-database in ENPADASI.

Results

We identified 26 observational nutrition studies (20–47) conducted in Germany ($n = 8$), Italy ($n = 6$), Belgium ($n = 5$), Spain ($n = 3$), Ireland ($n = 2$), and Estonia ($n = 2$) (**Table 1**). The study designs were longitudinal ($n = 12$ cohort studies), cross-sectional ($n = 12$), and case-control ($n = 2$) studies. Twenty-two studies were population-based. Four were patient-based, of which 1 was a case-control study (41), 2 were cohort studies (31, 44), and 1 was a cross-sectional study (32). Twenty-four studies recruited adults, of which 7 also recruited subjects aged <18 y. Two longitudinal studies (23, 29, 30) recruited children only. Of the 26 observational studies, 4 are still ongoing (23, 29, 30, 33, 41) and 3 longitudinal studies recently started [Hamburg City Health Study (HCHS) and the German National Cohort (NAKO Gesundheitsstudie); Diet4MicroGut, Italy].

The information related to the following sections is based on the metadata provided in the templates and grouped according to the study design.

Assessment of exposures: dietary intake and covariates

Table 2 describes the studies with information on dietary intake and other covariates. All of the studies collected information on dietary intake, alcohol and tobacco consumption, physical activity, anthropometric measurements, socioeconomic status, and health status, with the exception of 1 case-control study (41) that had no data on tobacco consumption, 1 longitudinal study (44) that had no data on physical activity, and 1 cross-sectional study (40) that had no data on health status (**Supplemental Table 1**). Furthermore, all of the studies had data on health status, mainly on prevalent chronic diseases such as cardiovascular disease, cancer, diabetes, respiratory disease, chronic infectious disease, and neurodegenerative disease.

Case-control studies. Of the 2 studies from Italy, 1 collected information on dietary intake by using a semiquantitative FFQ, whereas the other study used food records obtained by self-completed questionnaires (Table 2).

Both studies had subjective data on physical activity, but only one also used accelerometers. One collected data on sedentary behavior objectively (35). Anthropometric measurements including weight and height were objectively measured in both studies, and one study also measured waist and hip circumference (Supplemental Table 1).

Cohort studies. Twelve studies collected information on dietary intake in the form of multiple ($n = 3$) 24-h dietary recalls, semiquantitative ($n = 7$) or qualitative ($n = 1$) FFQs, and food records ($n = 4$). Twenty-four-hour dietary recalls were conducted after face-to-face ($n = 1$) and telephone ($n = 1$) interviews or by self-completed questionnaires ($n = 1$). FFQs were

TABLE 1 List of the observational studies identified within the ENPADASI consortium[1]

Study (reference)	Country	Institution	Study design	Sample size, *n* (M/F)	Recruitment years	Target population
Food Consumption Survey 2014 (22)	Belgium	Scientific Institute of Public Health (WIV-ISP)	Cross-sectional	3200 (1600/1600)	2014–2015	General population; aged 3–64 y at recruitment
Health Interview Survey (25)	Belgium	Scientific Institute of Public Health (WIV-ISP)	Cross-sectional	10,600 (5300/5300)	2013	General population, aged 0–100 y at recruitment
NESCaV (21)	Belgium	University of Liège	Cross-sectional	1000 (500/500)	2010–2012	General population, aged 20–69 y at recruitment
Equilibre Alimentaire ou Equilibre Aliment-Terre? (40)	Belgium	University of Liège	Cross-sectional	188 (89/99)	2012	General population, aged 30–40 y at recruitment
INOGMA (44)	Belgium	University Hospitals Leuven	Cohort	54 (21/33)	2012–2014	Obese patients who had planned RYGB surgery at University Hospitals Leuven, Belgium; aged 18–60 y at recruitment
Baltic Nutrition and Health Survey (42)	Estonia	NIHD	Cross-sectional	2108 (902/1116)	1997	General population, aged 19–65 y at recruitment
National Dietary Survey 2014 (46, 47)	Estonia	NIHD	Cross-sectional	5031 (1843/3188)	2013–2015	General population, aged 3 mo–74 y at recruitment
NAKO (20)	Germany	Nationale Kohorte eV	Cohort	Expected: 200,000 (100,000/100,000)	2014 (ongoing)	General population, aged 20–69 y at recruitment
ActivE-Study[2]	Germany	MDC	Cross-sectional	50 (25/25)	2012–2014	General population, aged 20–69 y at recruitment
BVS II (36)	Germany	HMGU	Cross-sectional	1050 (442/608)	2002–2003	General population (German speaking), aged 13–80 y at recruitment
NVS II (45)	Germany	MRI	Cross-sectional	19,329 (8923/10,406)	2005–2007	General population, aged 14–80 y at recruitment
DONALD (23)	Germany	University of Bonn	Cohort	∼1500 (750/750)	Since 1985 (ongoing)	General population, aged 3 mo at recruitment
EPIC–Potsdam substudy (39)	Germany	DIfE	Cohort	815 (412/403)	2010	General population, aged 47–81 y at recruitment
GINIplus (29)/LISAplus (30)	Germany	HMGU	Cohort[3]	GINIplus: 5991 (2991/2839) LISAplus: 3094 (1584/1510)	GINIplus: 1995–1998 (ongoing) LISAplus: 1997–1999 (ongoing)	General population, recruited at birth
HCHS[2]	Germany	UKE	Cohort	45,000 (22,500/22,500)	Since May 2015	General population, aged 45–74 y at recruitment
TUDA (31)	Ireland	TCD; University of Ulster, Coleraine	Cohort	5186 (1699/3487)	2008–2012	Noninstitutionalized adults from either hospital clinics or the community, in subcohorts to focus on 3 common diseases of aging: cognitive dysfunction, bone disease, or hypertension; aged 60–102 y at recruitment
NANS (26)	Ireland	UCC; UCD	Cross-sectional	1500 (740/760)	2008–2010	General population (nonpregnant/lactating), aged >18 y at recruitment
NU-AGE (35)	Italy	University of Bologna	Case-control	1272 (557/715)	2012–2016	General population, free of major overt diseases; aged 65–79 y at recruitment

(*Continued*)

TABLE 1 *Continued*

Study (reference)	Country	Institution	Study design	Sample size, *n* (M/F)	Recruitment years	Target population
PRED-CC (41)	Italy	ISS	Case-control	48 (18/30)	2014–2016	Subjects undergoing abdominal surgery for colon cancer or benign conditions; aged 28–78 y at recruitment
INGI-FVG (33)	Italy	IRCCS Burlo Garofolo	Cohort	1444 (610/834)	Ongoing	General population, aged 18–99 y at recruitment
INGI-CARL (37)	Italy	IRCCS Burlo Garofolo	Cohort	538 (223/315)	2008	General population, aged 18–99 y at recruitment
SR (34)	Italy	IRCCS Burlo Garofolo	Cohort	692 (270/422)	2010–2013	General population, aged 18–99 y at recruitment
Diet4MicroGut (43)	Italy	UNIBA	Cohort	161 (67/94)	2013/2014	General population, aged 18–55 y at recruitment
Di@bet.es study (28)	Spain	CIBERDEM	Cross-sectional	~5000 (2000/3000)	2008–2010	General population, aged 18–100 y at recruitment
Pizarra study (38)	Spain	CIBERDEM	Cohort	~900 (360/540)	1996–2007	General population, aged 18–65 y at recruitment
METBANC (32)	Spain	CIBERDEM	Cross-sectional	~700 (250/450)	2007–2014	Patient-based study: patients at increased cardiovascular risk; aged 20–70 y at recruitment

[1]The general population includes both random and convenience sampling designs. BVS II, Bavarian Food Consumption Survey II; CIBERDEM, Center in Diabetes and Associated Metabolic Disorders; DIfE, German Institute of Human Nutrition; DONALD, DOrtmund Nutritional and Anthropometric Longitudinally Designed Study; ENPADASI, European Nutritional Phenotype Assessment and Data Sharing Initiative; EPIC, European Prospective Investigation in Cancer and Nutrition; GINIplus, German Infant Study on the Influence of Nutrition Intervention; HCHS, Hamburg City Health Study; HMGU, Helmholtz Zentrum München–German Research Center for Environmental Health; INGI-CARL, Italian Network of Genetic Isolates–Carlantino; INGI-FVG, Italian Network of Genetic Isolates–Friuli Venezia Guilia; INOGMA, Influence of Obesity and Gastric Bypass on Medication Absorption; IRCCS, Istituto di Ricovero e Cura a Carattere Scientifico; ISS, Istituto Superiore di Sanità; LISAplus, Influences of Lifestyle-Related Factors on the Human Immune System and Development of Allergies in Children; MDC, Max Delbrueck Center for Molecular Medicine in the Helmholtz Association; METBANC, Clinical and Genetic Study of Patients with Major Cardiovascular Rick Factors; MRI, Max Rubner-Institut–Federal Research Institute of Nutrition and Food; NAKO, National Cohort; NANS, National Adult Nutrition Survey; NESCaV, Nutrition, Environment, and Cardiovascular Health; NIHD, National Institute for Health Development; NU-AGE, New Dietary Strategies Addressing the Specific Needs of an Elderly Population for Healthy Aging in Europe; NVS II, German National Nutrition Survey II; PRED-CC, Predicting Tumor Development Risk by an Integrated Approach Liking Diet-Related Inflammation to Colon Cancer; RYGB, Roux-en-Y gastric bypass; SR, Silk Road; TCD, Trinity College Dublin; TUDA, The Trinity, Ulster, Department of Agriculture aging cohort study; UCC, University College Cork; UCD, University College Dublin; UKE, University Medical Centre Hamburg-Eppendorf; UNIBA, University of Bari Aldo Moro; WIV-ISP, Wetenschappelijk Instituut Volksgezondheid - Institut Scientifique de Santé Publique.
[2]No reference available.
[3]GINIplus and LISAplus are 2 German birth cohorts whose data will be pooled to increase statistical power as methods are harmonized.

equally conducted after face-to-face interviews ($n = 4$) or self-completed questionnaires ($n = 4$), and food records were obtained by self-completed questionnaires ($n = 2$) or face-to-face interviews ($n = 2$). One Italian study (43) collected information on dietary intake by using all 3 instruments in a subset of participants (Table 2). Two studies used a self-completed questionnaire to collect data on food preferences (34, 37).

Eleven studies had subjective data on physical activity, of which 4 also used accelerometers. Three studies collected data on sedentary behavior, of which 1 collected subjective data only (24), and 2 collected both subjective and objective data (29, 30, 39). Anthropometric measurements, including height, weight, and waist and hip circumference measurements, were objectively measured in 11 studies. In addition, 2 studies measured these variables subjectively (self-reported) (Supplemental Table 1).

Cross-sectional studies. Twelve studies collected information on dietary intake, mainly in the form of multiple ($n = 4$) or single ($n = 2$) 24-h dietary recalls, semiquantitative ($n = 4$) or qualitative ($n = 2$) FFQs, and food records ($n = 6$). Twenty-four-hour dietary recalls were conducted after face-to-face ($n = 4$) and telephone ($n = 2$) interviews. FFQs were equally conducted after face-to-face interviews ($n = 4$) or self-completed questionnaires ($n = 4$), and food records were obtained by self-completed questionnaires ($n = 2$) or face-to-face interviews ($n = 1$). Three studies (32, 43, 46, 47) collected information on dietary intake by using all 3 instruments in at least a subset of participants. Three studies used additional questions to collect data on dietary intake: the German National Nutrition Survey II (NVS II) study from Germany (45) applied 24-h dietary recalls, diet history interviews, and weighing records; the Bavarian Food Consumption Survey II (BVS II; Germany) (36) asked a set of questions concerning nutritional knowledge; and the Health Interview Survey (Belgium) (25) asked questions on food consumption (Table 2).

All of the studies had subjective data on physical activity, of which 3 also used accelerometers. Eight studies collected data on sedentary behavior, of which 6 collected data subjectively only (21, 26, 36, 42, 45–47), and 2 also collected objective data [Food Consumption Survey 2004/2014 (23), Belgium; ActivE-Study, Germany]. Anthropometric measurements, including height, weight, and waist and hip circumferences, were subjectively measured in 7 studies (self-reported) and objectively measured in 11 studies (Supplemental Table 1).

Biological samples and laboratory measurements

Case-control studies. The studies had a variety of traditional and omics biomarkers available (**Table 3**). NU-AGE and PRED-CC studies (35, 41) collected blood (serum, plasma), of which

TABLE 2 Dietary intake assessments conducted in the observational studies participating in ENPADASI[1]

Study	Country	24-h recall	FFQ	Food records	Other
Food Consumption Survey 2014	Belgium	√ M	√ Qual.	—	—
Health Interview Survey	Belgium	—	—	—	√
INOGMA	Belgium	—	—	√	—
NESCaV	Belgium	—	√ SQ	—	—
Équilibre Alimentaire ou Équilibre Aliment-Terre?	Belgium	—	√ SQ	—	—
Baltic Nutrition and Health Survey 1997	Estonia	√ S	—	—	—
National Dietary Survey 2014	Estonia	√ M	√ SQ	√	—
NAKO	Germany	√ M	√ SQ	—	—
ActivE	Germany	—	—	√	—
BVS II	Germany	√ M	—	—	√
NVS II[2]	Germany	√ M	—	√	√
DONALD	Germany	—	—	√	—
HCHS	Germany	—	√ SQ	—	—
GINIplus and LISAplus	Germany	—	√ SQ	—	—
EPIC-Potsdam substudy	Germany	√ M	√ SQ	—	—
TUDA	Ireland	—	√ SQ	—	—
NANS	Ireland	—	—	√	—
NU-AGE	Italy	—	—	√	—
PRED-CC	Italy	—	√ SQ	—	—
INGI-FVG	Italy	—	√ SQ	—	—
INGI-CARL	Italy	—	—	—	√
SR	Italy	—	—	—	√
Diet4MicroGut	Italy	√ M	√ SQ	√	—
Di@bet.es	Spain	—	√ Qual	—	—
Pizarra	Spain	—	√ Qual	√	—
METBANC[2]	Spain	√ S	√ SQ	√	—

[1] BVS II, Bavarian Food Consumption Survey II; DONALD, DOrtmund Nutritional and Anthropometric Longitudinally Designed Study; ENPADASI, European Nutritional Phenotype Assessment and Data Sharing Initiative; EPIC, European Prospective Investigation into Cancer and Nutrition; GINIplus, German Infant Study on the Influence of Nutrition Intervention; HCHS, Hamburg City Health Study; INGI-CARL, Italian Network of Genetic Isolates–Carlantino; INGI-FVG, Italian Network of Genetic Isolates–Friuli Venezia Giulia; INOGMA, Influence of Obesity and Gastric Bypass on Medication Absorption; LISAplus, Influences of Lifestyle Related Factors on the Human Immune System and Development of Allergies in Children; M, multiple; METBANC, Clinical and Genetic Study of Patients with Major Cardiovascular Risk Factors; NAKO, German National Cohort; NANS, National Adult Nutrition Survey; NESCaV, Nutrition, Environment and Cardiovascular Health; NU-AGE, New Dietary Strategies Addressing the Specific Needs of an Elderly Population for Healthy Aging in Europe; NVS II, German National Nutrition Survey II; Qual, qualitative; PRED-CC, Predicting Tumor Development Risk by an Integrated Approach Linking Diet-Related Inflammation to Colon Cancer; S, single; SQ, semiquantitative; SR, Silk Road; TUDA, Trinity, Ulster, Department of Agriculture aging cohort study; √, dietary measurement available/collected.

[2] Not all the dietary measurements were available for all of the included participants.

one also collected urine and feces, and the other collected adipose tissue. Overall, traditional biomarkers were available in both studies. Lipids and lipoproteins (mainly HDL cholesterol, LDL cholesterol, and total cholesterol) and glucose and insulin (glucose and insulin in serum and glycated hemoglobin (HbA1c) in EDTA plasma) were measured in 1 study. Both studies had data on inflammatory markers such as IL-6 and TNF-α, but only one had data on C-reactive protein (CRP). Both studies had data on adipokines, such as adiponectin, although one measured this in serum and the other in adipocytes from adipose tissue biopsy samples. One study had data on leptin in either serum or plasma. Further details on biomarkers are described in **Supplemental Tables 2–6**.

Metabolomics, proteomics, and genomics and transcriptomics were available only in one of the studies. With regard to metabolomics, 1 study applied NMR and MS for concentration measurements in blood (serum, plasma) and urine (Supplemental Table 6). One study performed proteomic analyses. Although both studies had DNA available, only one collected genetic information by whole-genome sequencing (Table 3).

Longitudinal studies. All of the studies collected biological samples for measurements. Eleven studies collected blood (serum, plasma), 7 collected urine, 5 collected saliva, 2 collected feces, and 1 collected nasal swabs. These studies had a variety of traditional and omics biomarkers available (Table 3). Overall, traditional biomarkers, metabolomics, proteomics, and genomics and transcriptomics were available in 10, 3, 1, and 6 of the studies, respectively. With regard to lipids and lipoproteins, 9 studies measured HDL cholesterol, 9 collected LDL cholesterol, 10 collected total cholesterol, and 4 collected TGs. Glucose and insulin in serum were measured in 8 and 4 studies, respectively. HbA1c in EDTA plasma was measured in 3 studies. Five studies had data

TABLE 3 Measurements assessed in samples from the observational studies participating in ENPADASI[1]

		Biomarker measurements							Genomics/transcriptomics	
Study name	Country	L	G/I	INFL	A	Other	Metabolomics	Proteomics	SNPs	GW data
Food Consumption Survey 2014	Belgium	—	—	—	—	—	—	—	—	—
Health Interview Survey	Belgium	—	—	—	—	—	—	—	—	—
INOGMA	Belgium	√	√	√	—	—	—	—	—	—
NESCaV	Belgium	√	√	√	—	√	—	—	—	—
Équilibre Alimentaire ou Équilibre Aliment-Terre?	Belgium	—	—	—	—	—	—	—	—	—
Baltic Nutrition and Health Survey 1997	Estonia	—	—	—	—	—	—	—	—	—
National Dietary Survey 2014	Estonia	—	—	—	—	—	—	—	—	—
NAKO	Germany	√	√	—	—	—	—	—	—	—
ActivE[2]	Germany	√	√	√	—	—	—	—	—	—
BVS II	Germany	√	√	√	√	√	—	—	√	—
NVS II	Germany	—	—	—	—	—	—	—	—	—
DONALD	Germany	√	√	√	√	—	—	—	—	—
HCHS[2]	Germany	√	√	—	—	—	—	—	—	—
GINIplus and LISAplus	Germany	√	√	√	√	√	—	—	—	—
EPIC-Potsdam substudy[2]	Germany	√	√	—	—	—	√	—	—	—
TUDA	Ireland	√	√	√	—	√	—	—	√	—
NANS	Ireland	√	√	√	√	√	√	√	√	—
NU-AGE	Italy	√	√	√	√	√	√	√	√	—
PRED-CC	Italy	—	—	√	√	√	—	—	—	√
INGI-FVG	Italy	√	√	—	—	—	√	—	—	√
INGI-CARL	Italy	√	√	—	—	—	—	—	—	√
SR	Italy	—	—	—	—	—	—	—	—	√
Diet4MicroGut[3]	Italy	—	—	—	—	—	√	√	—	√
Di@bet.es	Spain	√	√	√	—	√	—	—	√	—
Pizarra	Spain	√	√	√	√	√	—	—	√	—
METBANC	Spain	√	√	√	√	√	—	—	√	—

[1] A, adiposity; BVS II, Bavarian Food Consumption Survey II; DONALD, DOrtmund Nutritional and Anthropometric Longitudinally Designed Study; ENPADASI, European Nutritional Phenotype Assessment and Data Sharing Initiative; EPIC, European Prospective Investigation into Cancer and Nutrition; G/I, glucose/insulin; GINIplus, German Infant Study on the Influence of Nutrition Intervention; GW, genome-wide; HCHS, Hamburg City Health Study; INFL, inflammatory; INGI-CARL, Italian Network of Genetic Isolates–Carlantino; INGI-FVG, Italian Network of Genetic Isolates–Friuli Venezia Giulia; INOGMA, Influence of Obesity and Gastric Bypass on Medication Absorption; L, lipids; LISAplus, Influences of Lifestyle Related Factors on the Human Immune System and Development of Allergies in Children; METBANC, Clinical and Genetic Study of Patients with Major Cardiovascular Risk Factors; NAKO, German National Cohort; NANS, National Adult Nutrition Survey; NESCaV, Nutrition, Environment and Cardiovascular Health; NU-AGE, New Dietary Strategies Addressing the Specific Needs of an Elderly Population for Healthy Aging in Europe; NVS II, German National Nutrition Survey II; PRED-CC, Predicting Tumor Development Risk by an Integrated Approach Linking Diet-Related Inflammation to Colon Cancer; SNP, single nucleotide polymorphism; SR, Silk Road; TUDA, Trinity, Ulster, Department of Agriculture aging cohort study; √, Measurement assessed.

[2] DNA available but no measurements done.

[3] GWAS, Genome-wide associations; SNPs, Measurements other than single nucleotide polymorphisms.

on inflammatory markers, of which 4 measured CRP, IL-6 ($n = 3$), and TNF-α ($n = 1$), among others. Two studies had data on adipokines such as adiponectin or leptin in either serum or plasma. Further details on biomarkers are described in Supplemental Tables 2–6.

With regard to metabolomics ($n = 3$), 2 studies applied NMR for concentration measurements in serum ($n = 1$) and in urine, saliva, or feces ($n = 1$). The latter also applied MS in serum urine, saliva, or feces (Supplemental Table 6). One study measured metabolomics using AbsoluteIDQ p180 kit (Biocrates Life Sciences AG, Innsbruck, Austria) in plasma from a targeted group. Two of the 8 studies with DNA available have not yet performed any genotyping or genomic measurements [HCHS and European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam substudy, Germany]. However, 6 studies had available genetic information by either whole-genome sequencing ($n = 4$) or candidate single nucleotide polymorphisms ($n = 2$) (Table 3). One study (Diet4MicroGut, Italy) also performed metagenomic analyses with available DNA.

Cross-sectional studies. Six studies did not collect biological samples for measurements (22, 25, 40, 42, 45–47) and thus had no biomarker data available (Table 3). The other 6 collected blood (serum, plasma), of which 4 collected urine and 1 study collected hair. These studies had a variety of traditional and omics biomarkers available (Table 3). Overall, traditional biomarkers, metabolomics, proteomics, and genomics and transcriptomics were available in 5, 1, 1, and 4 of the studies, respectively. With regard to lipids and lipoproteins, 6 studies measured HDL cholesterol and 5 measured LDL cholesterol, of which 1 had information on LDL cholesterol through calculation and 5 measured total cholesterol. Moreover, 3 of these studies measured TGs ($n = 3$). Five studies measured glucose and insulin in serum, and 4 studies measured HbA1c in EDTA plasma. Five studies had data on inflammatory markers such as CRP ($n = 6$), IL-6 ($n = 4$), and TNF-α ($n = 4$), among others. Three studies had data on adipokines such as adiponectin or leptin in either serum or plasma. Further details on biomarkers are described in Supplemental Tables 2–6.

Metabolomics, proteomics and genomics assessments are listed in **Table 4** . One study (26) applied NMR for concentration measurements in urine. The same study performed proteomic analyses. One of the 5 studies with DNA available had not yet performed any genotyping or genomic measurements (ActivE-Study, Germany). However, 4 studies had available genetic information by candidate single nucleotide polymorphisms (Table 3).

Informed consent, ethics, and data sharing

All of the identified observational studies indicated that their study was approved by an ethics committee or by the National Data Protection Office (45) (**Supplemental Table 7**). All of the studies provided informed consent, with the exception of the Belgian Health Interview survey (25), for which informed signed consents were not required. Only 4 studies expressed a clear interest in storing and sharing raw data within the ENPADASI consortium, whereas 4 declared that they probably would. Three were likely to share data upon confirmation from the study board or an agreement form. Four studies agreed to partially share data. Eleven studies were not keen on storing or sharing raw data within the consortium, partially due to uncertainties related to ethics, data protection issues, and privacy, especially for the ongoing cohort studies. Some of these are still internally discussing the possibility of sharing. With regard to metadata, most of the studies agreed to share metadata for future joint data analysis within a federated database system, whereas 5 need confirmation.

Minimal requirements checklist

Table 4 describes a set of mandatory descriptors, totaling 41, following the ISA categories. The Investigation category contains 12 descriptors devoted to collecting metadata about the project context, informed consent and ethical issues (descriptors 1–7), data-sharing policy (descriptors 8–11), and data analysis permissions (descriptor 12). The Study category describes a set of 12 descriptors about the study design (descriptors 1–2), study subjects (descriptors 3–5 and 12), and recruitment (descriptors 6–11). The Assay category with 17 descriptors describes endpoints (descriptors 1–10), study samples (descriptors 11–14), and analytical measurements (descriptors 15–17).

Discussion

We identified 26 observational studies conducted in 6 European countries with data on dietary intake, biomarkers, and health outcomes. Their design was mainly cross-sectional or longitudinal. Most of the studies included adults, mainly from the general population. All of the studies had data on dietary intake, and 20 studies collected samples such as blood (serum, plasma), urine, or saliva, which are the most commonly used to measure omics biomarkers (48). Although most of the 20 studies had data on lipoproteins and glucose and insulin biomarkers available, metabolomics or proteomics profiles were determined less often and in different biological samples. Our results are in agreement with the fact that the methodology used for genomics is relatively mature, whereas the methodology used for the metabolome is still in its infancy (48, 49). From these studies, we identified a set of minimal requirements that each study should provide when uploading their study metadata into a metadatabase to allow for interpretation and comparability of the data.

The identified studies are heterogeneous in their design and used various types of dietary assessment methodologies. Nonetheless, we identified observational studies with comparable data on traditional biomarkers. To study nutritional phenotypes, diet-related subsets of metabolites, genes, and proteins can be used as biomarker profiles (49). Our studies collected limited and diverse omics data, which, at a first glance, may preclude the possibility to compare or join these types of data. The scarce omics data from the identified studies may be due to the current limitations in multi-omics integration (i.e., the cost of omics methodologies, computational integration of multidimensional omics data, and the diversity of samples used, which largely depends on the study aims and technical issues). Yet, 2 Italian studies (35, 43) measured metabolites by using both NMR and MS in urine samples, which may allow for comparability. In addition, a third study (26) from Ireland measured metabolites by using NMR in urine. The same 3 studies (26, 35, 43) also collected data on proteomics and genomics so that nutritional phenotypes could be characterized together with data related to anthropometrics and functional measures, such as physical activity, in the future. These results highlight the importance of including high-throughput technology approaches in nutritional epidemiology studies to better investigate the diet-health relation through joint data analysis.

The scientific community is increasingly pooling information from multiple studies to construct large databases (50). However, if subtle changes or differences are to be captured, standardized methodology and data formats are required (5). There are many ways to overcome such challenges, one of which is by means of data harmonization approaches. In the case of dietary intake, the use of standardized methodology may be helpful to collect more comparable data and to minimize residual confounding. However, the challenge remains when exploring regional differences in nutrition behavior. Thus, standardized regional nutrient databases are needed to reflect country-specific products. Because our results showed that data owners were reluctant to share raw data, we had to design a simple stepwise process for the generation of harmonized databases that can easily be implemented in future federated data analyses: *1*) consensus in the selection and definition of the list of variables requested in the research proposal to be circulated among the participating study partners in the form of a variable catalog, *2*) generation of databases by using the variables and the exact variable names and format measures described in the variable catalog for study integration, and *3*) creation of data dictionaries for each participating study. Data dictionaries derived from approved scientific research proposals will be stored in the metadatabase together with the successful research proposal as part of the uploaded files requested in the minimal requirements (Investigation category, descriptor 7). Cataloging such information will allow researchers to have an idea of the level of both homo- and heterogeneity across study designs as well as potential sample sizes available for analyses for future research proposals (51).

The identification of 41 minimal descriptors, together with the 32 study-quality descriptors, will facilitate the integration of data sets and enable querying the meta-information of studies stored in centralized repositories for future research proposals. In addition, the identified descriptors for minimal requirements were used to develop the Ontology for Nutritional Studies (52). The Ontology for Nutritional Studies aims to define a common language (controlled vocabulary) for study metadata as well as standardizing existing ontologies to facilitate data integration.

TABLE 4 Checklist for minimal requirements for study metadata[1]

Number	Descriptors	Options	Type
Investigation			
1	Full name of the study	Acronym if applicable	Fill in the blank
2	Country of the study	—	Fill in the blank
3	Description of the study aim within the investigation	—	Fill in the blank
4	Principal investigator (name) for the study described	—	Fill in the blank
5	Contact information of the contact person of the study/experiment	—	Fill in the blank
6	Funding body/bodies for the investigation	—	Fill in the blank
7	Upload if available, or provide the URL	Study web link for the investigation or study (URL)	Multiple choice
		Registration number of the study (i.e., clinicaltrials.gov)	—
		IRB/IEC approval number	—
		Informed consent	—
		Study protocol and any protocol deviation/amendments	—
		Questionnaires	—
		SOPs for samples collection	—
		Publications: type and DOI or file location	—
		Other: Please specify type of document (i.e., data dictionaries and research proposals)	—
8	Data-sharing policy: study terminated	Yes/no/ongoing	Multiple choice
		If ongoing, when are data going to be available? (DD/MM/YYYY)	—
9	Data-sharing policy: data	Publicly accessible	Multiple choice
		Not publicly accessible but available upon request	—
		Not publicly accessible	—
10	Aggregate data-sharing policy (i.e., descriptive statistics)	Publicly accessible	Multiple choice
		Not publicly accessible but available upon request	—
		Not publicly accessible	—
11	Metadata	Publicly accessible	Multiple choice
		Not publicly accessible but available upon request	—
		Not publicly accessible	—
12	Data analysis permission	With access to the raw data	Multiple choice
		Without access to the raw data (federated analysis)	—
Study			
1	Study design	Cohort	Multiple choice
		Cross-sectional	—
		Case-control	—
		Seroepidemiologic study	—
		Other: please specify study design	—
2	Provide a short description of the study	—	Fill in the blank
3	Study population	Recruited from the general population	Multiple choice
		No	—
4	Particular dietary, physiologic, or nutritional characteristics of target population	—	Fill in the blank
5	Population representativeness	National level (nationally representative)	Multiple choice
		Subnational level (covers multiple communities. i.e., >3 cities, >5 villages, or ≥1 provinces or states)	—
		Community level	—
6	Type of sampling	Probability sampling:	Multiple choice
		-Simple random sampling	—
		-Stratified random sampling	—
		-Multistage sampling	—
		Nonprobability sampling:	
		-Voluntary response sampling	—
		-Judgement sampling	—
		-Convenience sampling	—
		-Other: describe	—

(*Continued*)

TABLE 4 *Continued*

Number	Descriptors	Options	Type
7	Describe control group	—	Fill in the blank
8	Describe type of controls	—	Fill in the blank
9	Start/end of recruitment	DD/MM/YYYY–DD/MM/YYYY	Fill in the blank
10	Follow-ups	Describe time points and actions taken	Fill in the blank
11	Total number of participants recruited	Total; M; F	Fill in the blank
12	Age range of the study participants	—	Fill in the blank
Assay			
1	Method for dietary or nutritional assessment	Dietary records (innovative alternatives: PDA-technologies, mobile phone–based technologies, camera- and tape recorder–based technologies)	Multiple choice
		24-h recall (innovative alternatives: interactive computer-based technologies, Web-based)	—
		Screener/FFQ (innovative alternatives: interactive computer-based technologies, Web-based):	—
		-Qualitative (only frequency)	—
		-Semiquantitative	—
		-Quantitative	—
		Diet history	—
		Other: please specify	—
2	Reference of the main food-composition table used (or URL)	—	Fill in the blank
3	Type of food assessed	Food	Multiple choice
		Drinks	—
		Dietary supplements	—
4	Nutrient and food intake data	Unadjusted (preferred option)	Multiple choice
		Adjusted for total energy intake using:	—
		1. Density method	—
		2. Residual method	—
		Estimates of usual intake from short-term measurements	—
		Other: describe	—
5	Physical activity	Objective measurement	Multiple choice
		Name of the tool, provider, version, year of the version	questions & Fill in the blank
		Subjective measurement	—
		Name of the questionnaire, provider, version, year of the version	—
6	Tobacco use	Yes/no	Multiple choice
7	Alcohol consumption	Yes/no	Multiple choice
8	Anthropometry	Weight	Multiple choice
		Height	—
		Waist circumference	—
		BMI status (categories)	—
		Body fat percentage	—
9	Sociodemographic information	Yes/no	Multiple choice
10	Study outcomes and time points of assessment: health outcomes	—	Fill in the blank
11	Total number of sample donors (number of individuals with biological samples)	—	Fill in the blank
12	Type of biological samples and total number of sample donors per sample type	Whole blood	Multiple choice
		Serum	—
		Plasma	—
		Urine	—
		Saliva	—
		Feces	—
		Other: please specify	—
13	Fasting	Yes/no	Multiple choice
14	Relative time points of sampling event	—	Fill in the blank

(*Continued*)

TABLE 4 *Continued*

Number	Descriptors	Options	Type
15	Type of omics	Biomarkers	Multiple choice
		Metabolomics	—
		Proteomics	—
		Genomics	—
		Transcriptomics	—
16	Measurement (i.e., metabolite profiling)	—	Fill in the blank
17	Technology (i.e., MS, chromatography)	—	Fill in the blank

[1]DD/MM/YYYY, date/month/year; IRB/IEC, Institutional Review Board/Independent Ethics Committee; PDA, Personal Digital Assistant; SOP, Standard Operating Procedure.

Study metadata should be made available for data analysis for a positive and long-lasting impact on the value of collective scientific outputs (5). However, minimal requirements should be succinct enough to facilitate practical reporting without losing detail.

The ENPADASI consortium is currently developing a metadatabase with the use of Mica (53), a software tool designed by the Maelstrom Research Group (www.maelstrom-research.org) using the OBiBa (Open Source Software for BioBanks) software suite (http://www.obiba.org/) that allows cataloging the collected observational studies through their metadata. The mandatory minimal requirements together with the optional quality descriptors (19) have been uploaded in Mica (54). The DASH-IN system currently has a central platform for metadata (Mica) server in Bari (54), enabling the retrieval of study metadata from a stable location. Users can interrogate or interpret metadata from the identified studies by means of an initial set of ontologies in agreement with the FAIR principles to ensure that the data is Findable, Accessible, Interoperable, and Reusable (15). A strength of the ENPADASI consortium is that we have direct contact with data managers, allowing for more transparency. Data managers are in charge of uploading the required metadata about their study data, which, in turn, will facilitate data integration and comparability among studies by decreasing the likelihood of making errors in the selection of appropriate data for a study proposal.

The identified observational studies may serve as the basis for defining case studies on specific nutritional epidemiologic research questions that will test the integration of studies within the DASH-IN. Case studies will contribute to the optimization of processes with regard to data access, information technology infrastructure, solutions for intellectual property issues, methods of harmonization of existing data, as well as standardization of terms and ontologies within a protected environment.

The DASH-IN will *1*) show the numerous benefits of connecting those nutritional epidemiologic studies of similar design by creating opportunities for collaborative and multicenter research, *2*) increase the efficiency of secondary usage of existing data, *3*) improve the generalizability of results, and *4*) improve the validity of comparative studies. We therefore envision that the DASH-IN may enhance the capacity of nutritionists, biologists, epidemiologists, and clinicians to carry out high-impact research with particular emphasis on omics data (transcriptomics, proteomics, metabolomics) to study the complex relation and interaction between nutrition and human health through genes and metabolome profiles (55, 56) once these data become available.

In conclusion, we identified 26 observational studies in nutritional epidemiology with data on dietary intake, biomarkers, and health outcomes for federated data analysis. In addition, we compiled a list of minimal requirements for the development of a database to store metadata. The minimal requirements and the identified studies may encourage more efficient secondary usage of existing data and promote collaboration initiatives to conduct joint data analysis that may help to better understand the role of diet and metabolism in the development and prevention of chronic diseases.

Acknowledgments

The authors' responsibilities were as follows—MP, KN, and TP: designed the research; MP, KN, LOD, and TP: conducted the research; GP, KN, CY, NDC, MS-M, HB, RL, CL, RC, MdA, MG, A-FD, SJ, JL, CK, UN, JB, HB, MS-M, TH, EL, JW, PG, AR, LC, GR-M, JM, LM, MS, HS, EB, EN, CM, MG, MdA, EW, B-CZ, and JT: provided essential materials; MP: wrote the manuscript; TP, LOD, and JB: had primary responsibility for final content; and all authors: read and approved the final manuscript.

References

1. Perrino T, Howe G, Sperling A, Beardslee W, Sandler I, Shern D, Pantin H, Kaupert S, Cano N, Cruden G, et al. Advancing science through collaborative data sharing and synthesis. Perspect Psychol Sci 2013;8:433–44.
2. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, Manoff M, Frame M. Data sharing by scientists: practices and perceptions. PloS One 2011;6:e21101.
3. Baccini M, Bachmaier EM, Biggeri A, Boekschoten MV, Bouwman FG, Brennan L, Caesar R, Cinti S, Coort SL, Crosley K, et al. The NuGO proof of principle study package: a collaborative research effort of the European Nutrigenomics Organisation. Genes Nutr 2008;3:147–51.
4. Van't Veer P, Grammatikaki E, Matthys C, Raats MM, Contor L. EURRECA-Framework for Aligning Micronutrient Recommendations. Crit Rev Food Sci Nutr 2013;53:988–98.
5. van Ommen B, Bouwman J, Dragsted LO, Drevon CA, Elliott R, de Groot P, Kaput J, Mathers JC, Muller M, Pepping F, et al. Challenges of molecular nutrition research 6: the nutritional phenotype database to store, share and evaluate nutritional systems biology studies. Genes Nutr 2010;5:189–203.
6. Cwiek-Kupczynska H, Altmann T, Arend D, Arnaud E, Chen D, Cornut G, Fiorani F, Frohmberg W, Junker A, Klukas C, et al. Measures for interoperability of phenotypic data: minimum information requirements and formatting. Plant Methods 2016;12:44.
7. Merino-Martinez R, Norlin L, van Enckevort D, Anton G, Schuffenhauer S, Silander K, Mook L, Holub P, Bild R, Swertz M, et al. Toward global biobank integration by implementation of the Minimum Information About BIobank Data Sharing (MIABIS 2.0 core). Biopreservation Biobanking 2016;14:298–306.
8. Scudamore CL, Soilleux EJ, Karp NA, Smith K, Poulsom R, Herrington CS, Day MJ, Brayton CF, Bolon B, Whitelaw B, et al. Recommendations for minimum information for publication of experimental pathology data: MINPEPA guidelines. J Pathol 2016;238:359–67.

9. York WS, Agravat S, Aoki-Kinoshita KF, McBride R, Campbell MP, Costello CE, Dell A, Feizi T, Haslam SM, Karlsson N, et al. MIRAGE: the minimum information required for a glycomics experiment. Glycobiology 2014;24:402–6.
10. Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, Ball CA, Binz PA, Bogue M, Booth T, et al. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. Nat Biotechnol 2008;26:889–96.
11. Lachat C, Hawwash D, Ocke MC, Berg C, Forsum E, Hornell A, Larsson C, Sonestedt E, Wirfalt E, Akesson A, et al. Strengthening the Reporting of Observational Studies in Epidemiology-Nutritional Epidemiology (STROBE-nut): an extension of the STROBE statement. PLoS Med 2016;13:e1002036.
12. Sansone SA, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, Fang H, Neumann S, Tong W, Amaral-Zettler L, et al. Toward interoperable bioscience data. Nat Genet 2012;44:121–6.
13. Gonzalez-Beltran A, Maguire E, Sansone SA, Rocca-Serra P. linkedISA: semantic representation of ISA-Tab experimental metadata. BMC Bioinformatics 2014;15:S4.
14. Sansone SA, Rocca-Serra P, Tong W, Fostel J, Morrison N, Jones AR. A strategy capitalizing on synergies: the Reporting Structure for Biological Investigation (RSBI) working group. Omics 2006;10:164–71.
15. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, et al. The FAIR guiding principles for scientific data management and stewardship. Scientific Data 2016;3:160018.
16. Murtagh MJ, Turner A, Minion JT, Fay M, Burton PR. International data sharing in practice: new technologies meet old governance. Biopreserv Biobank 2016;14:231–40.
17. Joint Programming Initiative. A Healthy Diet for a Healthy Life, JPI HDHL [Internet]. [cited 2017 Jul 18]. Available from: http://healthydietforhealthylife.eu/.
18. Lakerveld J, van der Ploeg HP, Kroeze W, Ahrens W, Allais O, Andersen LF, Cardon G, Capranica L, Chastin S, Donnelly A, et al. Towards the integration and development of a cross-European research network and infrastructure: the DEterminants of DIet and Physical ACtivity (DEDIPAC) Knowledge Hub. Int J Behav Nutr Phys Act 2014;11:143.
19. Yang C, Pinart M, Kolsteren P, Van Camp J, De Cock N, Nimptsch K, Pischon T, Laird E, Perozzi G, Canali R, et al. Perspective: essential study quality descriptors for data from nutritional epidemiologic research. Adv Nutr 2017;8:639–51.
20. The German National Cohort: aims, study design and organization. Eur J Epidemiol 2014;29:371–82.
21. Alkerwi A, Guillaume M, Zannad F, Laufs U, Lair ML. Nutrition, environment and cardiovascular health (NESCAV): protocol of an inter-regional cross-sectional study. BMC Public Health 2010;10:698.
22. Bel S, Van den Abeele S, Lebacq T, Ost C, Brocatus L, Stievenart C, Teppers E, Tafforeau J, Cuypers K. Protocol of the Belgian food consumption survey 2014: objectives, design and methods. Arch Public Health 2016;74:20.
23. Buyken AE, Alexy U, Kersting M, Remer T. [The DONALD cohort: an updated overview on 25 years of research based on the Dortmund Nutritional and Anthropometric Longitudinally Designed study.] Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2012;55:875–84 (in German).
24. De Filippis F, Pellegrini N, Vannini L, Jeffery IB, La Storia A, Laghi L, Serrazanetti DI, Di Cagno R, Ferrocino I, Lazzi C, et al. High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. Gut 2016;65:1812–21.
25. Demarest S, Van der Heyden J, Charafeddine R, Drieskens S, Gisle L, Tafforeau J. Methodological basics and evolution of the Belgian health interview survey 1997–2008. Arch Public Health 2013;71:24.
26. Feeney EL, Nugent AP, Mc Nulty B, Walton J, Flynn A, Gibney ER. An overview of the contribution of dairy and cheese intakes to nutrient intakes in the Irish diet: results from the National Adult Nutrition Survey. Br J Nutr 2016;115:709–17.
27. Gesquiere I, Augustijns P, Lannoo M, Matthys C, Van der Schueren B, Foulon V. Barriers in the approach of obese patients undergoing bariatric surgery in Flemish hospitals. Obes Surg 2015;25:2153–8.
28. Gutierrez-Repiso C, Soriguer F, Rojo-Martinez G, Garcia-Fuentes E, Valdes S, Goday A, Calle-Pascual A, Lopez-Alba A, Castell C, Menendez E, et al. Variable patterns of obesity and cardiometabolic phenotypes and their association with lifestyle factors in the Di@bet.es study. Nutr Metab Cardiovasc Dis 2014;24:947–55.
29. Harris C, Flexeder C, Thiering E, Buyken A, Berdel D, Koletzko S, Bauer CP, Bruske I, Koletzko B, Standl M. Changes in dietary intake during puberty and their determinants: results from the GINIplus birth cohort study. BMC Public Health 2015;15:841.
30. Kohlboeck G, Glaser C, Tiesler C, Demmelmair H, Standl M, Romanos M, Koletzko B, Lehmann I, Heinrich J. Effect of fatty acid status in cord blood serum on children's behavioral difficulties at 10 y of age: results from the LISAplus study. Am J Clin Nutr 2011;94:1592–9.
31. Laird E, McNulty H, Ward M, Hoey L, McSorley E, Wallace JM, Carson E, Molloy AM, Healy M, Casey MC, et al. Vitamin D deficiency is associated with inflammation in older Irish adults. J Clin Endocrinol Metab 2014;99:1807–15.
32. Merino J, Kones R, Ferre R, Plana N, Girona J, Aragones G, Ibarretxe D, Heras M, Masana L. Negative effect of a low-carbohydrate, high-protein, high-fat diet on small peripheral artery reactivity in patients with increased cardiovascular risk. Br J Nutr 2013;109:1241–7.
33. Mezzavilla M, Iorio A, Bobbo M, D'Eustacchio A, Merlo M, Gasparini P, Ulivi S, Sinagra G. Insight into genetic determinants of resting heart rate. Gene 2014;545:170–4.
34. Mezzavilla M, Vozzi D, Pirastu N, Girotto G, d'Adamo P, Gasparini P, Colonna V. Genetic landscape of populations along the Silk Road: admixture and migration patterns. BMC Genet 2014;15:131.
35. Santoro A, Pini E, Scurti M, Palmas G, Berendsen A, Brzozowska A, Pietruszka B, Szczecinska A, Cano N, Meunier N, et al. Combating inflammaging through a Mediterranean whole diet approach: the NU-AGE project's conceptual framework and design. Mech Ageing Dev 2014;136–137:3–13.
36. Schaller N, Seiler H, Himmerich S, Karg G, Gedrich K, Wolfram G, Linseisen J. Estimated physical activity in Bavaria, Germany, and its implications for obesity risk: results from the BVS-II study. Int J Behav Nutr Phys Act 2005;2:6.
37. Sorice R, Bione S, Sansanelli S, Ulivi S, Athanasakis E, Lanzara C, Nutile T, Sala C, Camaschella C, D'Adamo P, et al. Association of a variant in the CHRNA5-A3-B4 gene cluster region to heavy smoking in the Italian population. Eur J Hum Genet 2011;19:593–6.
38. Soriguer F, Almaraz MC, Garcia-Almeida JM, Cardona I, Linares F, Morcillo S, Garcia-Escobar E, Dobarganes MC, Olveira G, Hernando V, et al. Intake and home use of olive oil or mixed oils in relation to healthy lifestyles in a Mediterranean population: findings from the prospective Pizarra study. Br J Nutr 2010;103:114–22.
39. von Ruesten A, Feller S, Bergmann MM, Boeing H. Diet and risk of chronic diseases: results from the first 8 years of follow-up in the EPIC-Potsdam study. Eur J Clin Nutr 2013;67:412–9.
40. Winandy M. Equilibre alimentaire ou équilibre aliment-terre? School of Medicine, Univeristy of Liège, Liège, Belgium; 2012 (in French).
41. Del Corno M, D'Archivio M, Conti L, Scazzocchio B, Vari R, Donninelli G, Varano B, Giammarioli S, De Meo S, Silecchia G, et al. Visceral fat adipocytes from obese and colorectal cancer subjects exhibit distinct secretory and omega6 polyunsaturated fatty acid profiles and deliver immunosuppressive signals to innate immunity cells. Oncotarget 2016;7:63093–105.
42. Pomerlau J MM, Robertson A, Vaask S, Pudule I, Grinberga D, Kadziauskiene K, Abaravicius A, Bartkeviciute R. Nutrition and lifestyle in the Baltic Republics: summary report. Copenhagen (Denmark); 1999.
43. Ferrocino I, Di Cagno R, De Angelis M, Turroni S, Vannini L, Bancalari E, Rantsiou K, Cardinali G, Neviani E, Cocolin L. Fecal microbiota in healthy subjects following omnivore, vegetarian and vegan diets: culturable populations and rRNA DGGE profiling. PloS One 2015;10:e0128669.
44. Gesquiere I, Foulon V, Augustijns P, Gils A, Lannoo M, Van der Schueren B, Matthys C. Micronutrient intake, from diet and supplements, and association with status markers in pre- and post-RYGB patients. Clin Nutr 2017;36:1175–81.
45. Heuer T, Krems C, Moon K, Brombach C, Hoffmann I. Food consumption of adults in Germany: results of the German National Nutrition Survey II based on diet history interviews. Br J Nutr 2015;113:1603–14.
46. National Institute for Health Development (Estonia); Nurk E, Nelis K, Saamel M, Martverk M, Nelis L. National Dietary Survey among 11–74 years old individuals in Estonia. EFSA Supporting Publications 2017;14:1–22.

47. National Institute for Health Development (Estonia);Nurk E, Nelis K, Saamel M, Martverk M, Jõeleht A, Nelis L. National Dietary Survey among children up to ten years old and breastfeeding mothers in Estonia. EFSA Supporting Publications 2017;14: 1–25.
48. Fito M, Melander O, Martinez JA, Toledo E, Carpene C, Corella D. Advances in integrating traditional and omic biomarkers when analyzing the effects of the Mediterranean diet intervention in cardiovascular prevention. Int J Mol Sci 2016;17:1–20.
49. Zeisel SH, Freake HC, Bauman DE, Bier DM, Burrin DG, German JB, Klein S, Marquis GS, Milner JA, Pelto GH, et al. The nutritional phenotype in the age of metabolomics. J Nutr 2005;135: 1613–6.
50. Fortier I, Burton PR, Robson PJ, Ferretti V, Little J, L'Heureux F, Deschenes M, Knoppers BM, Doiron D, Keers JC, et al. Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. Int J Epidemiol 2010;39: 1383–93.
51. Doiron D, Burton P, Marcon Y, Gaye A, Wolffenbuttel BH, Perola M, Stolk RP, Foco L, Minelli C, Waldenberger M, et al. Data harmonization and federated analysis of population-based studies: the BioSHaRE project. Emerg Themes Epidemiol 2013;10:12.
52. European Nutritional Phenotype Assessment and Data Sharing Initiative (ENPADASI). Ontology for Nutritional Studies [Internet]. [cited 2017 Jul 18]. Available from: https://github.com/enpadasi/Ontology-for-Nutritional-Studies.
53. OBiBa. Open Source Software for BioBanks [Internet]. [cited 2017 Jul 18]. Available from: www.obiba.org.
54. Mica Server Web Application [Internet]. [cited 2017 Jul 18]. Available from: http://mica.cloud.ba.infn.it/.
55. Harttig U, Travis AJ, Rocca-Serra P, Renkema M, van Ommen B, Boeing H. Owner controlled data exchange in nutrigenomic collaborations: the NuGO information network. Genes Nutr 2009;4:113–22.
56. Astarita G, Langridge J. An emerging role for metabolomics in nutrition science. J Nutrigenet Nutrigenom 2013;6:181–200.