Healthy diet metrics: a suitability assessment of indicators for global and national monitoring purposes

Eric O. Verger, Mathilde Savy, Yves Martin-Prevel, Jennifer Coates, Edward Frongillo, Lynnette Neufeld, Kuntal Saha, Chika Hayashi, Bridget Holmes, Chris Vogliano, et al.

HEALTHY DIET METRICS

A suitability assessment of indicators for global and national monitoring purposes

Eric O. Verger, Mathilde Savy, Yves Martin-Prével
Jennifer Coates, Edward Frongillo, Lynnette Neufeld
Kuntal Saha, Chika Hayashi, Bridget Holmes
Chris Vogliano, Elaine Borghi, Francesco Branca

World Health Organization
Food and Agriculture Organization of the United Nations
unicef
The Rockefeller Foundation

HEALTHY DIET METRICS

A suitability assessment of indicators for global and national monitoring purposes

Eric O. Verger, Mathilde Savy, Yves Martin-Prével
Jennifer Coates, Edward Frongillo, Lynnette Neufeld
Kuntal Saha, Chika Hayashi, Bridget Holmes
Chris Vogliano, Elaine Borghi, Francesco Branca

World Health Organization
Food and Agriculture Organization of the United Nations
unicef
The Rockefeller Foundation

Healthy diet metrics: a suitability assessment of indicators for global and national monitoring purposes/ Eric O Verger, Mathilde Savy, Yves Martin-Prével, Jennifer Coates, Edward Frongillo, Lynnette Neufeld et al.

ISBN 978-92-4-007213-8 (electronic version)
ISBN 978-92-4-007214-5 (print version)

Suggested citation. Verger EO, Savy M, Martin-Prével Y, Coates J, Frongillo E, Neufeld L et al. Healthy diet metrics: a suitability assessment of indicators for global and national monitoring purposes. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO.

Cataloguing-in-Publication (CIP) data. CIP data are available at http://apps.who.int/iris.

Contents

Acknowledgements

This technical report is an outcome of the Healthy Diets Monitoring Initiative led by the Food and Agriculture Organization of the United Nations (FAO), the United Nations Children's Fund (UNICEF) and the World Health Organization (WHO). The authors gratefully acknowledge all those who contributed to its development.

The report was conceptualized and led by the Core Group of the Healthy Diets Monitoring Initiative: Jennifer Coates, Associate Professor, Food and Nutrition Policy and Programs, Tufts University Friedman School of Nutrition Science and Policy, Boston, MA, United States of America (USA) and Co-chair, Diet Quality Working Group of the Technical Expert Advisory Group on Nutrition Monitoring (TEAM DQ WG); Edward Frongillo, Director, Global Health Initiatives, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA and Co-chair, TEAM DQ WG; Chika Hayashi, Senior Advisor, Monitoring and Statistics and Unit Chief, Nutrition Data, UNICEF, New York, USA; Lynnette Neufeld, Director, Food and Nutrition Division, FAO, Rome, Italy; Kuntal Saha, Technical Officer, Monitoring Nutritional Status and Food Safety Events, Department of Nutrition and Food Safety, WHO, Geneva, Switzerland; and Chris Vogliano, Technical Advisor, Food Systems, United States Agency for International Development (USAID) Advancing Nutrition, Arlington, VA, USA. The contributions of TEAM DQ WG to the conceptualization of the document are gratefully acknowledged.

This document was prepared by Eric O. Verger, Mathilde Savy and Yves Martin-Prével, consultants at the Institut de Recherche pour le Développement (IRD), Montpellier, France, with the assistance of Antoine de Reviers, research assistant, IRD.

The document was reviewed by the Strategic Planning Group of the Healthy Diets Monitoring Initiative: Victor Aguayo, Director, Child Nutrition and Development, UNICEF, NY, USA; Francesco Branca, Director, Department of Nutrition and Food Safety, WHO, Geneva, Switzerland; Jennifer Coates, Associate Professor, Food and Nutrition Policy and Programs, Tufts University Friedman School of Nutrition Science and Policy, Boston, MA, USA, and Co-chair, TEAM DQ WG; Sara Farley, Vice President, Food Initiative, The Rockefeller Foundation, New York, USA; Edward Frongillo, Director, Global Health Initiatives, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA and Co-chair, TEAM DQ WG; Chika Hayashi, Senior Advisor, Monitoring and Statistics and Unit Chief, Nutrition Data, UNICEF, New York, USA; Lynnette Neufeld, Director, Food and Nutrition Division, FAO, Rome, Italy; Kuntal Saha, Technical Officer, Monitoring Nutritional Status and Food Safety Events, Department of Nutrition and Food Safety, WHO, Geneva, Switzerland; and Shelly Sundberg, Senior Program Officer, Nutrition, Bill & Melinda Gates Foundation.

Anna Herforth (Research Associate, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA), Megan Deitchler (Director, Intake, Center for Dietary Assessment, FHI 360, Washington, DC, USA) and Carlos Monteiro (Professor, Department of Nutrition, Universidade Estadual de São Paulo, São Paulo, Brazil) shared their experiences and insights related to the development and operationalization of specific indicators, and Elaine Borghi (Unit Head, Monitoring Nutritional Status and Food Safety Events, Department of Nutrition and Food Safety, WHO, Geneva, Switzerland), Bridget Holmes (Nutrition and Food Systems Officer and Nutrition Assessment Team Leader, Food and Nutrition Division, FAO, Rome, Italy) and Mia Blakstad (Senior Associate, Food Initiative, The Rockefeller Foundation, New York, USA) reviewed earlier drafts of the report and provided comments.

The contributions of the participants at the Technical Expert Meeting on Harmonizing and Mainstreaming Measurement of Healthy Diets Globally, Bellagio, Italy, from 28 November to 2 December 2022 (see Annex 1) are also gratefully acknowledged.

List of acronyms

AHEI-2010	Alternative Healthy Eating Index 2010
CHD	Coronary heart disease
CHNS	China Health and Nutrition Survey
CVD	Cardiovascular disease
DAPA	Diet, Anthropometry and Physical Activity
DASH	Dietary Approaches to Stop Hypertension
DQI-I	Diet Quality Index-International
DQQ	Diet Quality Questionnaire
FAO	Food and Agriculture Organization of the United Nations
FBDG	Food-based dietary guidelines
FFQ	Food frequency questionnaire
GDQS	Global Diet Quality Score
GDR	Global Dietary Recommendations
HDI	Healthy Diet Index
HICs	High-income countries
LMICs	Low- and middle-income countries
MDD-W	Minimum Dietary Diversity for Women
MPA	Mean probability adequacy
NHANES	National Health and Nutrition Examination Survey
NCDs	Noncommunicable diseases
MetS	Metabolic syndrome
NPNL	Non-pregnant and non-lactating
PANDiet	Probability of Adequate Nutrient Intake
ROC	Receiver operating characteristic
SDGs	Sustainable Development Goals
TEAM	Technical Expert Advisory Group on Nutrition Monitoring
UNICEF	United Nations Children's Fund
UPF	Ultra-processed foods
WHO	World Health Organization
WRA	Women of reproductive age

Executive summary

Diets are changing everywhere, and better measurement and monitoring of these changes at different levels is becoming increasingly important in view of the huge influence dietary patterns have on the health of both people and the planet. Yet there are no harmonized metrics for tracking how the healthfulness of diets around the world is evolving. The main objective of this report is to scientifically assess the validity, usefulness and fitness for purpose of existing healthy diet metrics as global and national monitoring indicators.

Healthy diet metrics are here defined as metrics derived from intakes of nutrients, foods or food groups, or of all three, with the aim of measuring one or several subconstructs of the construct of healthy diets ("nutrient adequacy", "nutrient density", "macronutrient balance", "diversity", "moderation", "favourable dietary pattern" and "food safety"). The focus is on apparently healthy populations from children (2–10 years), adolescents (11–17 years), adults (18–59 years) to the elderly (≥ 60 years) in low-, middle- and high-income countries.

Seven healthy diet metrics were identified based either on their wide use in the nutrition literature or potential for global and national monitoring: the Alternative Healthy Eating Index 2010, Dietary Approaches to Stop Hypertension score, Diet Quality Index-International, Global Diet Quality Score, Global Dietary Recommendations score, Minimum Dietary Diversity for Women and Nova ultra-processed foods (UPF) score. Detailed information related to the development and performance of these metrics was derived from a narrative review of the peer-reviewed and grey literature, as well as from exchanges with their developers.

Healthy diet metrics were critically and comparatively assessed through a set of criteria for their use as global and national monitoring indicators. Judgement criteria used to assess the metrics were content, cross-context equivalence, validity for initial purpose, sensitivity to change, test-retest reliability, internal consistency, population groups coverage, interpretation threshold, ease of computation and ease of interpretation. Judgement criteria used to assess the data collection tools were criterion validity, interviewer burden, respondent burden and collection time.

Four healthy diet metrics were identified as the most suitable for global and national monitoring: the Global Dietary Recommendations Score, Global Diet Quality score, Minimum Dietary Diversity for Women, and Nova UPF score. Some weaknesses were noticed for these metrics: the need to confirm the validity of the Global Dietary Recommendations score in various contexts, the relative length and complexity in collecting semiquantitative dietary information for computing the Global Diet Quality Score. The limitation of the Minimum Dietary Diversity for Women, and the Nova UPF score is that these indicators assess only "diversity" and "moderation", respectively, which are two of the subconstructs of the construct of healthy diets.

A major knowledge gap identified in this report concerns the population groups for which these metrics are applicable. Indicators are often developed for women of reproductive age

or an adult population, and it remains largely unknown whether they can be extended to other population groups such as pregnant or lactating women, adolescents and children, or the elderly. Further exploring how these metrics might be used in wider population groups and settings is a priority. Another knowledge gap is to determine empirically whether they can provide equivalent (i.e. comparable) assessments across populations. Filling these gaps remains a priority for developers of healthy diet metrics who are currently conducting or planning to conduct such studies.

1. Introduction

1.1. Background and rationale

Diets are changing everywhere, and measuring their characteristics at global, regional, national and subnational levels is becoming increasingly important in view of the huge influence dietary patterns have on the health of both people and the planet. Better measurement and monitoring of these changes are needed to support governments in establishing policies and programmes to promote healthy diets, to assess the effectiveness of their actions and to hold them accountable. Yet there are critical gaps in the global, regional, national and subnational monitoring of diet characteristics.

Currently there are no World Health Assembly global nutrition targets related to diets. Although healthy diets are integral to achieving the Sustainable Development Goals (SDGs), there are no harmonized metrics for tracking how diets around the world are evolving. While researchers have been developing concepts, metrics, methods and tools to characterize the healthfulness of diets, there remains a need to share experiences, improve collaboration and coordination, harmonize and complement efforts, develop a common agenda, and accelerate progress in assessing and monitoring dietary characteristics at global and national levels.

WHO and UNICEF set up the Technical Expert Advisory Group on Nutrition Monitoring (TEAM) in 2015 to advise on all levels of nutrition monitoring. In 2017, TEAM in turn established a Working Group on Diet Quality; the current TEAM workplan includes tasks related to facilitating progress in identifying metrics for global and national monitoring of diets. The Working Group conducted a landscaping exercise in 2019 to identify gaps and potential focus areas across four domains: 1) concepts related to diet quality and healthy diets; 2) global and national monitoring frameworks; 3) platforms; and 4) metrics *(1)*. The exercise resulted in a recommendation to convene actors working on methods and metrics for healthy diets to become aware of each other's work and collaborate towards identifying common elements and concordant themes in definitions and metrics. In further response to this need, TEAM and the FAO, with technical and logistical support from USAID Advancing Nutrition, hosted a Technical Consultation on Measuring Healthy Diets: Concepts, Methods and Metrics *(2)*. This consultation was held virtually from 18 to 20 May 2021, involving 85 experts from a wide range of institutions and geographic areas, all with different roles in the data value chain.

In 2022, in collective recognition of the challenges facing the coordination of efforts to generate diet quality metrics, WHO and UNICEF, via TEAM, joined forces with FAO to resolve these issues and make headway through the Healthy Diets Monitoring Initiative. The Initiative is guided by a Strategic Planning Group that includes the Nutrition Division Directors from the three UN agencies, as well as TEAM co-chairs and secretariat, and key donors. The Strategic Planning Group is responsible for providing overall strategic guidance, planning and coordination of the global effort to advance assessment of healthy diets for diverse purposes and oversee the activities of the Healthy Diets Monitoring Initiative.

Within the scope of the FAO-UNICEF-WHO Healthy Diets Monitoring Initiative and building on the outputs of the 2021 Technical Consultation, the aim of the TEAM Diet Quality Technical Working Group in collaboration with FAO is to advance at least two key workstreams in 2022. The first is to identify, and reach consensus on, the constructs and subconstructs of a healthy diet. The second is to scientifically assess the validity, usefulness and fitness for purpose of existing healthy diet metrics for population assessment in global and national monitoring. This work is expected to provide guidance for governments and others seeking to assess the healthfulness of diets and serve as the basis for spotlighting one or more metrics for global and national monitoring.

1.2. Healthy diets and healthy diet metrics

1.2.1. Healthy diets

A healthy diet can be defined as a diet that promotes growth and development, and prevents malnutrition in all its forms *(3)*. Malnutrition refers to undernutrition, such as wasting, stunting, underweight or deficiencies in vitamins or mineral, as well as to obesity and dietary factors that increase the risk of noncommunicable diseases (NCDs) such as heart disease, stroke, diabetes and certain cancers. What a healthy diet exactly entails will vary according to individual characteristics (e.g. age, gender, lifestyle and level of physical activity), cultural context, locally available foods and dietary habits. However, the main subconstructs of the construct of healthy diets remain the same.

For this report, we used the subconstructs of the construct of healthy diets defined from the work carried out by the working group led by Edward Frongillo *(4)*:

- **Nutrient adequacy**: sufficient quantity and quality of nutrient intake compared with nutrient requirements to meet dietary needs, without excess.
- **Nutrient density**: amount or relative proportion of nutrients per weight of food, per unit of energy (often 100 kcal), or per serving.
- **Macronutrient balance**: balance of energy-yielding macronutrients: carbohydrates, proteins and fats.
- **Diversity**: diets composed of a variety of foods derived from diverse food groups.
- **Moderation**: limited intake of foods related to chronic diseases, including refined grains, red and processed meats, and sugar-sweetened foods and beverages.
- **Food safety**: foods are free of microbial pathogens, foodborne macroparasites, toxins and chemicals.

In addition to these subconstructs, we used a subconstruct called "**favourable dietary pattern**" from the *2020 Dietary guidelines for Americans (5)*: "common characteristics of dietary patterns associated with positive health outcomes include relatively higher intake of vegetables, fruits, legumes, whole grains, dairy, lean meats and poultry, seafood, nuts, and unsaturated vegetable oils, and relatively lower consumption of red and processed meats, sugar-sweetened foods and beverages, and refined grains". This subconstruct serves as the basis for several healthy diet metrics and corresponds to diets that are themselves based on several subconstructs. In this report, we have considered the subconstruct "favourable dietary pattern" to be more holistic (and therefore equivalent to the construct of a healthy diet) compared to other, more specific subconstructs. In this report, healthfulness of diets is used as a synonym for healthy diets and relies on the same subconstructs.

1.2.2. Metric, measure and indicator

We use the term "metric" to refer broadly to measures and indicators. However, it is important to differentiate between a **measure** and an **indicator**. According to Frongillo et al., measures assign numbers to people or things to represent the relative amounts of a property while indicators reflect the presence or absence of a given property *(6)*.

This difference is illustrated by the following example: the height of a child is a measure, whereas whether or not a child is stunted (i.e. height below -2 standard deviations of the sex- and age-appropriate growth standard) is an indicator.

1.2.3. Healthy diet metrics

Like Miller et al. *(7)*, we defined a healthy diet metric as a metric derived from intakes of nutrients, foods or food groups, or all of them, with the aim of measuring one or several subconstructs of the construct of healthy diets: "nutrient adequacy", "nutrient density", "macronutrient balance", "diversity", "moderation", "favourable dietary pattern" and "food safety". This definition excludes dietary metrics that involve non-dietary factors as a core component (e.g. physical activity), and dietary metrics that summarize only a single nutrient or food (unless they are considered as part of a suite of metrics). For this report, we focused on metrics collecting individual-level data.

1.3. Population-level uses of healthy diet metrics

For this report, and based on the different typologies observed *(2, 8, 9)*, five main types of population-level uses for dietary measures and indicators are considered.

- **Identification**: healthy diet population-level assessment, cross-country comparisons, identification of at-risk population groups (based on geographical location, sociodemographic or economic characteristics, etc.) and possibly population-level targeting for an intervention.
- **Surveillance and monitoring**: healthy diet monitoring of selected population groups (e.g. children) at global, national or even subnational levels, which includes tracking of population-level changes across time as well as within- and cross-country comparison of changes.
- **Policy and programme design**: informing decisions related to policies, programmes, establishing population-based standards, regulations or targets, and preparing communication messages and instruments to improve the healthfulness of people's diets.
- **Evaluation**: measuring the impact (positive or negative) of any kind of policy or intervention or effects of altered external circumstances (e.g. shocks) on the healthfulness of diets in selected population groups at global, national or even subnational levels.
- **Research**: increasing scientific knowledge on the determinants of healthy diets and their relationships with health outcomes, and identifying options for interventions.

Identification, as well as surveillance and monitoring, usually require standardized approaches that are simple, rapid and feasible, and allow for comparability over time and geography, thereby providing a population-level snapshot of diet quality. However, policy and programme design, evaluation and research usually require more complex, resource-intensive and fit-for-purpose approaches, which may take priority over data comparability beyond the scope of any specific study. While standardized approaches developed for identification and surveillance and monitoring can be adapted for policy and programme design, evaluation and research, more specific approaches developed for policy and programme design, evaluation and research on the other hand are generally less often used for identification or surveillance and monitoring. Nevertheless, it should be noted that approaches developed for policy and programme design, evaluation and research can provide insights into how to strengthen identification as well as surveillance and monitoring.

In this report, we adapted the concepts of global and national monitoring by Hosseinpoor et al. *(10)* as follows. **Global healthy diet monitoring** can be described as a comparison of the healthfulness of diets across countries, and is a useful practice to track progress on international initiatives such as the SDGs. The results of global healthy diet monitoring may inform high-level decisions about resource allocation and identify areas in need of additional support. Global-level monitoring enables benchmarking among countries, prompting poorly performing countries to recognize areas for improvement, and to draw lessons from the success stories of better-performing countries. **National healthy diet monitoring** serves as the basis for global monitoring, but also goes beyond it in providing a more in-depth understanding of the healthfulness of diets: it can be disaggregated for diverse subgroups within a country and, depending on specific objectives, may require more detailed information. National healthy diet monitoring can be tailored to investigate healthy diet metrics and dimensions of dietary healthfulness that are relevant within a particular country, but not across multiple contexts. The process of national monitoring can make use of the best available data sources at the national level although these same sources may not be appropriate for global monitoring.

1.4. Focus, objectives and report structure

The ***main objective*** of this report is to scientifically assess the validity, usefulness and fitness for purpose of existing healthy diet metrics for use as global and national monitoring indicators. The **specific objectives** of this report are to: 1) compile empirical evidence related to the development and performance of existing healthy diet metrics; 2) describe and compare healthy diet metrics according to a set of criteria related to validity, usefulness and fitness for purpose; and 3) identify knowledge gaps, priority needs and opportunities to further investigate metric performance.

This report focuses on metrics collecting individual-level data measuring healthy diets at population level and designed for an apparently healthy population (i.e. free from acute or chronic illness) from the elderly (≥ 60 years) to adults (18–59 years), adolescents (11–17 years) and children (2–10 years) in any context, both low-and middle-income (LMIC) and high-income countries (HIC). Furthermore, this report primarily focuses on a suitability assessment of healthy diet metrics for their use as global monitoring indicators on two counts: global monitoring indicators can be more easily adopted as national monitoring indicators than the reverse, and achieving consensus around global monitoring indicators is ranked as a high priority by the Healthy Diets Monitoring Initiative.

This report is structured as follows: **Section 2** sets out the methodology employed to select the healthy diet metrics that are evaluated in this report and to define the set of criteria for assessing healthy diet metrics. **Section 3** describes the selected healthy diet metrics by compiling empirical evidence related to their development and performance, and presents their critical evaluation. **Section 4** presents a critical and comparative assessment of the selected healthy diet measures for use as global and national monitoring indicators. **Section 5** includes potential limitations of this assessment and **Section 6** identified knowledge gaps, priority needs and opportunities to further investigate metric performance. **Section 7** presents the conclusions from this assessment.

1.5. Approach for content development

The Core Group of the Healthy Diets Monitoring Initiative conceptualized and led the production of this document. The Strategic Planning Group of the Healthy Diets Monitoring Initiative and the TEAM Diet Quality Working Group contributed to the conceptualization and revision of the document. This document was prepared by a group of consultants.

Detailed information related to the development and performance of the healthy diet metrics was derived from a narrative review of the peer-reviewed and grey literature. Additional information was collected through email exchanges and/or videoconference meetings with the developers of the healthy diet metrics (see **Annex 2** for the list of developers who were contacted).

The participants (see Annex 1) of the Technical Expert Meeting on "Harmonizing and Mainstreaming Measurement of Healthy Diets Globally" including the metric developers and subject experts provided technical comments that were incorporated in the document. All external experts submitted to WHO a declaration of interest (see Annex 3).

2. Methodology

2.1. Selection of healthy diet metrics

We selected metrics collecting individual-level data measuring healthy diets at population level based on their inclusion in the May 2021 technical consultation on measuring healthy diets *(2)*, the importance of their use as measures or indicators of healthy diets in the nutrition literature *(7, 11)*, and/or their potential for global dietary identification, surveillance or monitoring of healthy diets. The following metrics were selected:

- The updated **Alternative Healthy Eating Index** (AHEI-2010), a frequently used measure of diet quality based on the level of consumption of 11 foods and nutrients predictive of chronic disease risk that incorporates current scientific evidence on the relationships between diet and health.
- The **Dietary Approaches to Stop Hypertension** (DASH) score, a frequently used measure of diet quality measuring adherence to a dietary pattern that substantially reduces both systolic and diastolic blood pressure among hyper- and normotensive individuals.
- The **Diet Quality Index-International** (DQI-I), a frequently used measure of diet quality based on four subconstructs of healthy diets ("diversity", "nutrient adequacy", "moderation" and "macronutrient balance") and designed for international comparisons.
- The **Global Diet Quality Score** (GDQS), a recent food group-based metric of diet quality that has been validated against outcomes representative of two forms of malnutrition in diverse regions. This metric was included in the May 2021 technical consultation on measuring healthy diets *(2)*.
- The **Global Dietary Recommendations (GDR)** score, a recent indicator of healthy diet patterns that adhere to global dietary recommendations. This metric was included in the May 2021 technical consultation on measuring healthy diets *(2)*. The GDR score was developed in tandem with a measurement tool, the **Diet Quality Questionnaire (DQQ)**. Other food group-based indicators can be derived from the DQQ, and some of them are presented jointly in this report but were not compared to other metrics.
- The **Minimum Dietary Diversity for Women** (MDD-W), a food group-based dietary diversity indicator for non-pregnant women of reproductive age (WRA) that has been shown to reflect micronutrient adequacy. This metric was included in the May 2021 technical consultation on measuring healthy diets *(2)*.
- The **Nova UPF score** for the consumption of ultraprocessed foods (UPF), a quick and practical metric with a good potential for reflecting the dietary share of UPF. This metric was included in the May 2021 technical consultation on measuring healthy diets *(2)*.

2.1.1. Exclusion of other healthy diet metrics

Because the main objective of this report is to assess the suitability of healthy diet metrics for use as global and national monitoring indicators, we excluded many healthy diet metrics that are valid for purposes other than global monitoring, such as policy and programme design, and evaluation and research in specific contexts. Without listing all these excluded indicators, we present below a list of the main types of indicators that we excluded and the reasons for exclusion.

Nutrient-based healthy diet metrics. While metrics like the Mean Adequacy Ratio *(12)*, Mean Probability Adequacy *(13)* and the Probability of Adequate Nutrient Intake (PANDiet) *(14)* are valid measures of the nutritional adequacy of diets that have been used in diverse contexts, they were excluded because they rely entirely on food composition tables or databases that many developing and some developed countries do not have *(15)*, a lack which is obviously a major hindrance to their use for global dietary monitoring.

Healthy diet metrics based on national dietary guidelines. While metrics like the Healthy Eating Index in the USA *(16–19)* or Programme National Nutrition Santé-Guidelines Score in France *(20, 21)* are valid measures of adherence to national dietary guidelines, they were excluded because they explore dimensions of the healthfulness of diets that may be relevant within a particular context but not universally.

Healthy diet metrics based on the traditional Mediterranean diet pattern. The widespread use of measures of diets based on the Mediterranean diet pattern in both Mediterranean and non-Mediterranean countries would argue in favour of considering such metrics in this report. Nevertheless, decades of research on and different approaches to Mediterranean diet pattern assessment have created large variability in the components and scoring systems of these metrics *(22)*, making it difficult to select a healthy diet metric based on the traditional Mediterranean diet pattern for this report. In addition, even metrics that intend to adapt the principles of the traditional Mediterranean diet pattern to non-Mediterranean countries such as the Alternative Mediterranean Diet Scale *(21)* cannot explore dimensions of the healthfulness of diets that may be relevant for all contexts.

Healthy diet metrics based on the EAT-*Lancet* Planetary Health Diet. In 2019, the EAT-Lancet Commission on healthy diets from sustainable food systems published a universal healthy reference diet based on extensive literature with the aim of both being environmentally sustainable and preventing diet-related chronic diseases and mortality *(24)*. Although such a universal healthy reference diet could provide a relevant basis for a suitable healthy diet metric for use as a global and national monitoring indicator, there is as yet no consensus on how best to operationalize a measure of adherence to this regime. Indeed, no less than four metrics have been recently developed with different scoring systems, applied in different contexts and associated with different outcomes: the EAT-Lancet score was applied in the United Kingdom of Great Britain and Northern Ireland (United Kingdom) and associated with ischaemic heart disease and diabetes, but not with either stroke or mortality *(25)*; the World Index for Sustainability and Health score was applied in Viet Nam and associated with some food and nutrient intakes *(26)*; the Planetary Health Diet Index was applied in Brazil and associated with overall diet quality *(27)*; and the EAT-Lancet index was applied in Sweden and associated with lower risk of mortality *(28)*. As a result, we excluded these metrics from our assessment.

2.2. Definition of the set of criteria to assess healthy diet metrics

The candidate metrics of healthy diets were assessed according to a set of criteria based on a modified version of a framework used for the critical appraisal of health assessment tools *(29)*, health literacy measures *(30)*, and nutrition and food literacy measures *(31)*. This set of criteria is described below.

2.2.1. Criteria for examination of the metric

Content. The assessment of metric content was based on the number of subconstructs of the construct of healthy diet assessed by the metric. We defined three levels, with two (specific vs multiple) differing in terms of the number of subconstructs assessed, and one (holistic) in which the subconstruct "favourable dietary pattern" was assessed.

- **Specific**: only one or two subconstructs of the construct of healthy diets are assessed.
- **Multiple**: more than two subconstructs of the construct of healthy diets are assessed.
- **Holistic**: the metric assesses adherence to dietary patterns that have been proven to be favourable for health (e.g. the DASH-style diet).

Cross-context equivalence

We used the same definition as Frongillo et al. *(6)* to assess the cross-context equivalence of the metric.

- **Construct equivalence** means that the same construct is measured across contexts, even if the measures used are not identical. In this case, the constructs measured are comparable. For example, the construct of food management strategies in response to household food insecurity is equivalent across contexts, but the items needed to measure this construct will differ markedly across contexts.
- **Item equivalence** means that the same construct is measured across contexts and the content of each item used is perceived and interpreted in the same way across contexts. In this case, the same items used across contexts mean, and are interpreted, in the same way. For example, the nine items used to assess household food insecurity in the Household Food Insecurity Access Scale were developed to be item equivalent across contexts.
- **Measurement equivalence** means that the constructs, items and units are the same across contexts (i.e. the difference in scores between two individuals means the same across contexts). In this case, the order of households or individuals is comparable across contexts. For example, a study has recently demonstrated the Healthy Eating Index to be measurement equivalent for children aged 1–2 years old in Brazil *(32)*.
- **Scalar equivalence** means the same as measurement equivalence but, in addition, the definition of zero is the same across contexts. In this case, average scores and prevalence values are comparable. For example, only three of the nine items in the Household Food Insecurity Access Scale that assess hunger are scalar equivalent.

Validity for initial purpose. This criterion can be defined as the extent to which a metric captures the construct it is trying to assess for a particular purpose, given that a metric is only valid with respect to how it was designed to be used and how particular groups of people are interpreted *(33)*. For this report, assessment of the validity of the metric was based on information collected about the rationale for its development, its intended uses and its content. We defined three levels of adequacy between what the metric intends to measure (rationale, data use and metric content) and how it was validated.

- **High**: there is strong evidence that the metric indicates what it intended to assess, with an adequate a priori study design applied to the appropriate context.
- **Medium**: according to published articles, the metric only partially indicates what it intended to assess (e.g. only some of the targeted subconstructs, or valid only in one context when cross-context validity is pursued).
- **Low**: there is limited evidence that the metric indicates what it intended to assess (e.g. a low confidence level in published results or in study design; or limited scope of the study as compared to its intended use).

Sensitivity to change. Ability of an instrument to measure a change in state that would be biologically relevant.

Test-retest reliability. Measure of the consistency of results in a test or other assessment instrument over time, given as the correlation of scores between the first and second administrations (or other statistics assessing the agreement between two measures). It provides an estimate of the repeatability stability of the construct being assessed.

Internal consistency. Degree of interrelationship or homogeneity among items that constitute a composite score, such that they are consistent with one another when measuring the same construct or dimension (usually tested by Cronbach's alpha coefficient, a measure of the average strength of association between all possible pairs of items contained within a set of items used to build the score).

Population groups covered. Has the metric been developed for specific population groups and/or for use among both men and women in apparently healthy populations (i.e. free from acute or chronic illness), ranging from the elderly (≥ 60 years) to adults (18–59 years), adolescents (11–17 years) and children (2–10 years)?

Interpretation threshold. Does the measure have a defined threshold to interpret its value as good (as opposed to poor) so that an indicator can be built?

Ease of computation. Assessment of the ease of computation (i.e. level of complexity to compute the metric) is based on the number and difficulty of steps needed to compute the metric. We considered that each step described below incorporates its own cleaning and verification parts. Four levels of ease were defined which differ in terms of the need to use a food composition table, the need to quantify food consumption and the need to manually assign food, ingredient and/or nutrient to the appropriate metric item.

- **High**: assign points per respondent; calculate the score; tabulate the indicator.
- **Medium**: **assign food and ingredients to the appropriate metric item**; assign points per respondent; calculate the score; tabulate the indicator.
- **Low**: **process the dietary data to estimate gram consumption of food and ingredients per respondent**; assign food and ingredient to the appropriate metric item; sum the quantity of consumption per metric food group per respondent; assign points per respondent; calculate the score; tabulate the indicator.
- **Very low**: **merge food composition database; process the dietary data to estimate gram consumption of food, ingredients, and nutrients per respondent**; assign food, ingredient and nutrient to the appropriate metric item; sum the quantity of consumption per metric item per respondent; assign points per respondent; calculate the score; tabulate the indicator.

Ease of interpretation. Assessment of the ease of interpretation of the metric (i.e. the level of complexity when interpreting the metric) was based on the ability to categorize whether or not the monitored population has a healthy diet. We defined two levels depending on whether interpretation of the metric is a continuous score or a multi-class/dichotomous indicator.

- **High**: the metric allows for a dichotomous interpretation, e.g. according to the metric λ, 55% of the population has a healthy diet while 45% does not or for a multi-class interpretation, e.g. according to the metric λ, 45% of the population has a high risk for unhealthy diet outcomes, 35% has a moderate risk for unhealthy diet outcomes and 20% has a low risk for unhealthy diet outcomes.
- **Low**: the metric provides a continuous score without a predefined reference to interpret its values (e.g. good as opposed to poor) and can only be interpreted as the higher the better.

2.2.2. Criteria for examination of the data collection tool

Criterion validity. Extent of agreement between the collection tool to derive data for the candidate metric and a reference or gold standard method (e.g. quantitative 24-hour dietary recall or weighed food record).

Interviewer burden. Total amount of perceived effort, both physical and cognitive, and level of expertise an interviewer must provide to complete an interview to specifications.

Respondent burden. The degree to which a survey respondent perceives his or her participation in the survey as difficult, time-consuming or emotionally stressful is known as respondent burden. Interview length, cognitive complexity of the task, required respondent effort, frequency of being interviewed, and stress due to psychologically invasive questions can all contribute to respondent burden in any survey.

Collection time. Time required for data collection per respondent.

2.3. Collection of information

Detailed information related to the development and performance of the seven healthy diet metrics was derived from a narrative review of the peer-reviewed and grey literature. Additional information was collected through email exchanges and/or videoconference meetings with the developers of the healthy diet metrics (see Annex 2 for the list of metric developers who were contacted).

3. Description of candidate healthy diet metrics

This section describes the selected healthy diet metrics by compiling the empirical evidence related to their development and performance (information taken directly from the metric developers' publications is printed in quotation marks), and presents their critical evaluation (Table 1).

3.1. Summary of the development and performances of healthy diet metrics

Table 1. Summary of the development and performances of healthy diet metrics

	AHEI-2010	DASH	DQI-I	GDQS	GDR score	MDD-W	Nova UPF score
Uses	Policy and programme design, evaluation and research (and potentially identification, monitoring and surveillance)	Research only	Identification, monitoring and surveillance, policy and programme design, evaluation and research	Identification, monitoring and surveillance, policy and programme design, evaluation and research	Identification, monitoring and surveillance, policy and programme design, evaluation and research	Identification, monitoring and surveillance, policy and programme design, evaluation and research	Identification, monitoring and surveillance, policy and programme design, evaluation and research
Content	Holistic	Holistic	Multiple	Holistic	Holistic	Specific (diversity)	Specific (moderation)
Cross-context equivalence	Measurement equivalence	Measurement equivalence	Scalar equivalence	Scalar equivalence	Scalar equivalence	Scalar equivalence	Measurement equivalence
Validity for initial purpose	High	High	Medium	High	High	High	High
Population groups	Men and women aged ≥ 18 years	Men and women aged ≥ 18 years	Men and women aged ≥ 20 years	NPNL WRA (and men to some extent)	Men and women aged ≥ 15 years	NPNL and lactating WRA	Men and women aged ≥ 18 years
Context	1 context only: USA	1 context only: USA	2 contexts only: USA and China[a]	14 contexts in Africa, the Americas, and Asia	2 contexts only: USA and Brazil[b]	9 datasets from 6 countries	1 context only: Brazil[b]
Collection tool	Standard tool[c]	Standard tool[c]	Standard tool[c]	Dedicated tool to be validated + Standard tool[c,d]	Dedicated tool validated for 3 LMICs + Standard tool[c]	Dedicated tool validated for 3 LMICs + Standard tool[c]	Dedicated tool validated for Brazil[b]
Ease of computation	Very low	Very low	Very low	High/Low[e]	High	High/Medium[f]	High
Ease of interpretation	Low	Low	Low	Medium	High	High	Low (High)[g]

[a] Other contexts were explored later;
[b] Other contexts are under investigation;
[c] Standard tools are quantitative 24-hour dietary recall
[d] Food frequency questionnaires;
[e] High when the GDQS app is used to collect data, low when using quantitative 24-hour dietary recall or food frequency questionnaires;
[f] High when using the list-based method or the DQQ to collect data, medium when using open recall method;
[g] To be confirmed in ongoing research.

3.2. Alternative Healthy Eating Index 2010 (AHEI-2010)

Detailed information related to the development and performance of the AHEI-2010 was found in the peer-reviewed article by Chiuve et al. (2012) *(34)*. We found additional information related to the validity and use of the AHEI-2010 in other populations and contexts in a recent, peer-reviewed meta-analysis by Schwingshackl et al. (2018) *(35)*. Requests for clarification and confirmation of certain information were sent to the developers by email and responses received.

3.2.1. Rationale and purpose

The AHEI-2010 is a measure of diet quality that was developed for the USA context, mainly by a research team from Harvard School of Public Health and Harvard Medical School. This measure is an update of the Alternative Healthy Eating Index that was created in 2002 by the research team. According to the developers, "the AHEI-2010 was designed as a new measure of diet quality that incorporates current scientific evidence on diet and health. It is an update of the Alternative Healthy Eating Index that was created in 2002 as an alternative to the Healthy Eating Index (which quantified adherence to the 1995 Dietary Guidelines for Americans) and was based on foods and nutrients predictive of chronic disease risk. Higher scores of the original AHEI have been found to be strongly associated with lower risk of major chronic disease, CVD, diabetes, heart failure, colorectal and estrogen-receptor-negative breast cancer, and total and cardiovascular mortality" *(34)*. We concluded that the purpose of the AHEI-2010 was to measure at population level adherence to a dietary pattern that has been proven favourable for health in the USA context.

3.2.2. Uses

No specific use of the AHEI-2010 has been reported by its developers. Nevertheless, the developers mentioned that because "the AHEI-2010 was more strongly associated with chronic disease risk than closer adherence to the 2005 Dietary Guidelines", this suggests "future revisions of Dietary Guidelines for Americans" *(34)*. We concluded that the AHEI-2010 could be used for research and that there is a lack of information to evaluate whether it is suitable for other uses.

Because the AHEI seems to be a complex and resource-intensive metric (see following sections), we concluded that it could be used for policy and programme design, evaluation and research. Because the AHEI was developed for the USA context, which has the resources and expertise to integrate this complex and resource-intensive metric into its monitoring systems (e.g. National Health and Nutrition Examination Survey (NHANES)), we concluded that it could be used for identification, monitoring and surveillance – but only in the USA context.

3.2.3. Scoring system

The AHEI-2010 is composed of 11 items related to food groups or nutrients (Table 2). All AHEI-2010 components are scored from 0 (worst) to 10 (best), and the total AHEI-2010 score ranges from 0 (non-adherence) to 110 (perfect adherence) *(34)*.

3.2.4. Criteria for examination of the metric

3.2.4.1. Content

The developers mentioned that "the AHEI-2010 is an update of the Alternative Healthy Eating Index that was based on foods and nutrients predictive of chronic disease risk" and "although some components differ, the AHEI and AHEI-2010 captured a similar dietary pattern" *(34)*. They also mentioned that "the AHEI-2010, which explicitly emphasizes high intakes of whole grains, polyunsaturated fatty acids, nuts, and fish and reductions in red and processed meats, refined grains, and sugar-sweetened beverages, was associated with lower risk of chronic diseases" *(34)*. We concluded that the AHEI-2010 assesses adherence to a dietary

Table 2. Scoring system of the Alternative Healthy Eating Index 2010[a]

Component	Criteria for minimum score (0)	Criteria for maximum score (10)
Vegetables[b], servings/day	0	≥5
Fruit[c], servings/day	0	≥4
Whole grains[d], g/day	0	
Women		75
Men		90
Sugar-sweetened beverages and fruit juice[e], servings/day	≥1	0
Nuts and legumes[f], servings/day	0	≥1
Red/processed meat[g], servings/day	≥1.5	0
trans Fat[h], % of energy	≥4	≤0.5
Long-chain (n-3) fats (EPA+DHA)[i], mg/day	0	250
PUFA[j], % of energy	≤2	≥10
Sodium[k], mg/day	Highest decile	Lowest decile
Alcohol[l], drinks/day		
Women	≥2.5	0.5–1.5
Men	≥3.5	0.5–2.0
Total	**0**	**110**

AHEI, Alternate Healthy Eating Index; DHA, docosahexaenoic acid; EPA, eicosapentaenoic acid; PUFA, polyunsaturated fatty acids.

[a] Values are means ± 6 SD unless otherwise noted. Researchers are invited to re-create and use the AHEI-2010 score in their own data.

[b] Vegetable consumption has been associated with lower risk of cardiovascular disease (CVD) and some cancers. Green leafy vegetables in particular may lower risk of diabetes. All vegetables on the food frequency questionnaire were included, except for potatoes (including French fries) because they are not associated with lower risk of chronic disease risk in epidemiologic studies and are associated with increased risk of diabetes. We considered 5 servings/day as ideal, which reflects the upper range of current dietary guidelines and is consistent with intervention studies of intermediate CVD risk factors. One serving is 0.5 cup of vegetables or 1 cup of green leafy vegetables (1 cup = 236.59 g).

[c] Fruit consumption has been associated with lower risk of CVD and some cancers. We included only whole fruit in our definition, because fruit juice is not associated with lower risk of CVD or cancer and may increase risk of diabetes. We considered 4 servings/day to be ideal, which is consistent with the upper range of current dietary guidelines. One serving is 1 medium piece of fruit or 0.5 cup of berries (1 cup = 236.59 g).

[d] Greater consumption of whole grains is associated with lower risk of CVD, diabetes and colorectal cancer. Conversely, refined grains are not associated with lower risk and may increase risk of diabetes, coronary heart disease (CHD) and other chronic diseases. We used grams of whole grains, which accounts for the variability of the percentages of whole grain in various "whole grain" products. One serving of a 100% whole-grain product (i.e. 0.5 cup of oatmeal or brown rice) contains 15–20 g of whole grains (per dry weight). We considered 75 g/day to be optimal (5 servings/day) for women and 90 g/day (6 servings/day) to be optimal for men on the basis of current guidelines for total grains.

[e] Intake of sugar-sweetened beverages, including soda and fruit drinks, is associated with increased risk of weight gain and obesity, CVD and diabetes. We included intake of fruit juice in this category, given the positive association with risk of diabetes and lack of beneficial effects on CVD or cancer. The association with pancreatic cancer risk is not well established. We considered 1 serving/day to be the least optimal on the basis of the associations in the literature. One serving is 8 oz (1 oz = 28.35 g). Nuts, legumes and vegetable protein (e.g. tofu) are important sources of protein and contain important constituents such as unsaturated fat, fibre, copper, magnesium, plant sterols and other nutrients. Nuts and other vegetable proteins have been associated with lower risk of CVD, especially when used as a substitute for other protein sources, such as red meat.

[f] Nuts are also associated with lower risk of diabetes and weight gain, whereas their relation to cancer is inconclusive. We considered 1 serving/day to be ideal on the basis of the AHEI recommendations and the current literature. One serving is 1 oz (1 oz = 28.35 g) of nuts or 1 tablespoon (15 mL) of peanut butter.

[g] Consumption of red meat and processed meats is associated with greater risk of CHD, especially when substituted for nuts, poultry or fish. Red meat and/or processed meats are also associated with higher risk of stroke, diabetes and colorectal and other cancers. Less than 1 serving/month was considered to be ideal, with an upper limit of 1.5 servings/day. One serving is 4 oz of unprocessed meat or 1.5 oz of processed meat (1 oz = 28.35 g).

[h] *trans*-Isomers of fatty acids, formed by partial hydrogenation of vegetable oils to produce margarines and vegetable shortening, are associated with higher risk of CHD and diabetes. Cutoffs are consistent with original AHEI cutoffs for *trans* fat.

[i] One serving of fish per week, specifically of species high in long-chain (n-3) fatty acids EPA + DHA, is strongly protective against fatal cardiac arrhythmias and sudden cardiac death and may lower the incidence of other CVD. EPA + DHA were associated with lower risk of diabetes in some, but not all, studies, and the relation with cancer risk is unclear. Because of the strength and consistency of fish and EPA + DHA on cardiac arrhythmias and CVD, we included this nutrient in the AHEI-2010 score. The cutoff for optimal intake (250 mg/day) is two 4-oz servings of fish/week, which is consistent with current guidelines (1 oz = 28.35 g).

[j] Replacing saturated fats with polyunsaturated fats leads to positive changes in lipid profiles, is associated with a lower risk of CHD, and may lower risk of type 2 diabetes. Furthermore, a low-fat diet had no beneficial effects on CVD risk factors, lipid profile or blood pressure and did not reduce the risk of CVD, breast cancer, colon cancer or total mortality. We gave the highest score to individuals with 10% of total energy intake from PUFA on the basis of current guidelines from the USDA and the AHA. PUFA does not include EPA or DHA intake.

[k] High sodium intake has been associated with higher blood pressure, and salt-preserved foods are associated with greater risk of stomach cancer, CVD and total mortality. Furthermore, sodium-reduced diets significantly lowered blood pressure and CVD risk in clinical trials. Large reductions in sodium intake, to levels recommended by the United States Department of Agriculture, may prevent a substantial number of new cases of CHD. The cutoffs for sodium were based on deciles of distribution in the population, due to lack of brand specificity in the FFQ to accurately estimate absolute intake. Values in the lowest decile were 1112 mg/day in women and 1612 mg/day in men and in the highest decile were 3337 mg/day in women and 5271 mg/day in men at baseline.

[l] In moderation, alcohol may be consumed as a part of an overall healthy diet. Moderate alcohol consumption has been associated with lower risk of CHD, dementia, diabetes and all-cause and CVD mortality. However, in heavier quantities, alcohol increases the risk of certain cancers and has other health and social implications such

Table 2. *Continued*

as alcohol dependence and alcohol-related injuries. Furthermore, many adults choose not to drink for various reasons. Thus, we assigned the highest score to moderate, and the worst score to heavy alcohol consumers. Non-drinkers received a score of 2.5. We used gender-specific cutoffs, because the health effects of alcohol are seen at lower quantities in women than in men. One drink is 4 oz of wine, 12 oz of beer or 1.5 oz of liquor (1 oz = 28.35 g).

Source: Chiuve SE et al., 2021 *(34)*.

pattern that has been proven to be favourable for health and thus assesses the subconstruct "favourable dietary pattern".

We assessed the content of the AHEI-2010 as holistic.

3.2.4.2. Cross-context equivalence

The AHEI-2010 was originally developed for the USA context and has been evaluated in this specific context *(34)*.Subsequently, this metric has been used unchanged in other contexts, for example Europe and China *(7)*, although it was not designed, developed and validated to be equivalent across contexts. Nevertheless, because one item of the AHEI-2010 depends on deciles of distribution in the studied population, we can conclude that the AHEI-2010 is at best measurement equivalent in contexts in which it has already been used (mainly in China and the United States, and in Europe).

We assessed the AHEI-2010 as measurement equivalent.

3.2.4.3. Validity for initial purpose

In an analysis of longitudinal quantitative dietary data from men aged 40–75 years (Health Professionals Follow-Up Study) and women aged 30–55 years (Nurses' Health Study) in the USA, the developers found "that the AHEI-2010 was associated inversely with risk of major chronic disease in both women and men in age-adjusted and multivariate models". They also found that "higher AHEI-2010 scores were inversely associated with risk of cardiovascular disease and diabetes, that the AHEI-2010 was more strongly associated with risk of coronary heart disease (CHD) than stroke, and that the AHEI-2010 was inversely associated with risk of total cancer in women but not in men" *(34)*.

The initial purpose of the AHEI-2010 was to assess adherence to a dietary pattern favourable for health in the context of improving the Dietary Guidelines for Americans. We concluded that the developers used an adequate a priori study design (longitudinal design with large sample sizes and multi-adjusted regressions) to demonstrate that this metric indicates adherence to a dietary pattern inversely associated with risk of major NCDs in two cohorts in the USA.

The validity for the initial purpose of the AHEI-2010 was assessed as high.

<u>Other elements of validation of the AHEI-2010</u>

Sensitivity to change of the AHEI-2010 has been evaluated using cross-sectional, nationally representative quantitative dietary data from individuals aged ≥ 20 years old in the USA (trend analysis with NHANES data from 1999 to 2010) *(36)*. Nevertheless, this type of study design cannot demonstrate sensitivity to change. Test-retest reliability and internal consistency have not been reported as being assessed *(34)*. In recent meta-analyses, it was found that diets that score highly on AHEI-2010 were associated with a significant reduction in the risk of all-cause mortality, cardiovascular disease, cancer, type 2 diabetes, and neurodegenerative disease in different cohorts (mainly from the United States, but also China and Europe) *(35)*, which confirms that the AHEI-2010 assesses adherence to a dietary pattern favourable for health.

3.2.4.4. Population groups covered

All analyses presented for validation of the AHEI-2010 are for men and women aged ≥ 18 years (including elders) *(34, 35)*. Additional research is needed to explore how the AHEI-2010 could be adapted for other population groups such as adolescents and children (Table 3).

3.2.4.5. Interpretation threshold

While no interpretation threshold was proposed by the developers of the AHEI-2010, other authors have proposed to "categorize overall diet quality as high (AHEI ≥65/110) or low (AHEI< 65/110) using a cut-off that was chosen based on the observation that individuals with a score of 65/110 and above are at a lower risk of major chronic disease compared with those with a lower score" *(34, 37)*. During our email exchange, the developers confirmed that there is no interpretation threshold for the AHEI-2010. They also mentioned that might be possible to establish such a threshold, which would clearly be helpful to some people. We concluded that additional research is needed to define an interpretation threshold for the AHEI-2010.

3.2.4.6. Ease of computation

The process of computing the AHEI-2010 appears to be complex and requires multiple steps as follows: cleaning the data; merging the food composition database; processing the dietary data to estimate gram consumption of food, ingredients and nutrients per respondent; assigning food, ingredients and nutrient to the appropriate metric item; summing the quantity of consumption per metric item per respondent; assigning points per respondent; calculating the score; and tabulating the indicator.

The ease of computation of the AHEI-2010 was assessed as very low.

3.2.4.7. Ease of interpretation

The AHEI-2010 can be interpreted by applying a possible range of 0 to 110 (the higher the score the better adherence to the dietary pattern), and by assigning a potential cut-off of 65 to determine the percentage of the population at a lower or higher risk for NCDs (to be confirmed).

Table 3. Population groups and applicability of the AHEI-2010

Population group	Applicability
Children (2–10 years)	Possible but requires additional research
Adolescent males (11–17 years)	Possible but requires additional research
Adolescent females (11–17 years)	Possible but requires additional research
Adult males (18–59 years)	Demonstrated
Adult females (18–59 years)	Demonstrated
Pregnant and/or lactating women	Possible but requires additional research
Elderly (≥ 60 years)	Demonstrated

The ease of interpretation of the AHEI-2010 was assessed as low (but potentially high if the cut-off is confirmed by further research).

3.2.5. Criteria for examination of the data collection tools

In the article describing the development of the AHEI-2010, the authors used validated semiquantitative food frequency questionnaires (FFQ) to derive data for the AHEI-2010 *(38)*. Nevertheless, single or multiple quantitative 24-hour dietary recalls can also be (and have been) used to derive data for the AHEI-2010 *(39, 40)*.In addition, some authors have developed short screeners to derive data for the AHEI-2010 *(41, 42)*, but because these tools are very specific to the studies and populations studied, we did not assess them further.

3.2.5.1. Criterion validity

Multiple quantitative 24-hour dietary recalls and FFQs can be considered standard dietary assessment methods that have been extensively validated. Nevertheless, the appropriateness of these methods for use in different population groups and contexts ought to be verified in a validation study in a study population *(43)*.

3.2.5.2. Interviewer burden

According to the Diet, Anthropometry and Physical Activity (DAPA) Measurement Toolkit, single or multiple quantitative 24-hour dietary recalls are associated with high interviewer burden, because these methods require the interviewers to have advanced training, a good understanding of the methods and a reasonable knowledge of how food is acquired, prepared and consumed in the survey areas. By comparison, FFQs are easier and more flexible to administer *(44)*.

3.2.5.3. Respondent burden

According to the DAPA Measurement Toolkit, while respondent burden is relatively low for a single quantitative 24-hour dietary recall, it is higher for multiple quantitative 24-hour dietary recalls due to the number of interviews. Furthermore, this method is dependent on the respondent's ability to recall intake accurately, which can increase its burden. FFQs are associated with a higher respondent burden than single or multiple quantitative 24-hour dietary recalls because they require more time, literacy and knowledge about food. Furthermore, the respondent burden is bound to increase: the greater the number of FFQ items and the longer the recall period (e.g. diet over the past year versus past month), the greater the respondent burden *(44)*.

3.2.5.4. Collection time

According to the DAPA Measurement Toolkit, interviews using a single 24-hour recall method are reported to last from 20 to 30 minutes depending on the respondent's diet *(44)*. According to the National Cancer Institute Dietary Assessment Primer, completing a FFQ aimed at capturing the total dietary intake usually requires 30 to 60 minutes *(45)*.

3.3. Dietary Approaches to Stop Hypertension score (DASH score)

Detailed information related to the development and performance of the DASH score was found in a peer-reviewed article by Fung et al. (2008) *(46)*. We found additional information related to the validity and use of the AHEI-2010 in other populations and contexts in a recent peer-reviewed meta-analysis by Soltani et al. (2020) *(47)*. Requests for clarification and confirmation of certain items of information were sent to the developers by email and responses received.

3.3.1. Rationale and purpose

The DASH score is a measure of diet quality that was developed for the USA context, mainly by a research team from Harvard School of Public Health and Harvard Medical School. According to the developers, "the DASH score was designed to reflect adherence to the DASH-style diet which is high in fruits and vegetables, moderate in low-fat dairy products, and low in animal protein but with substantial amount of plant protein from legumes and nuts, and which substantially reduces both systolic and diastolic blood pressure among hypertensive and normotensive individuals" *(46)*. Subsequently, different approaches have been used to define the DASH-style diet *(47)*, including, among the most used, the modified DASH score developed by Yu et al. (2014) *(48)*, the nutrient-based Mellen index developed by Mellen et al. (2008) *(49)*, and the DASH-like eating plan developed by Cuenca-García et al. (2014) *(50)*. Nevertheless, the most used version is the original DASH score developed by Fung et al. (2008) *(46)*. We concluded that the purpose of the DASH score was to measure at population level adherence to a dietary pattern that has been proven favourable for health in the USA context.

3.3.2. Uses

The developers designed the DASH score to "explore the long-term effect of DASH-style diet on cardiovascular end points" *(46)*. Because the DASH score seems to be a complex and resource-intensive metric (see following sections), all its items depending on quintiles of distribution in the studied population, this metric would be difficult to use for identification, monitoring and surveillance, policy and programme design, and evaluation. We therefore concluded that the DASH score could be used only for research.

3.3.3. Scoring system

The DASH score is composed of eight items related to food groups or nutrients (Table 4). The scoring system is based on quintiles of distribution in the population. For fruits, vegetables, nuts and legumes, low-fat dairy products and whole grains, high intake is desired: the first quintile of consumption is thus assigned 1 point and the fifth quintile assigned 5 points. For sodium, red and processed meats and sweetened beverages, low intake is desired: the lowest quintile is thus given a score of 5 points and the highest quintile 1 point. The component scores have to be summed to obtain the total DASH score which ranges from 8 to 48 (the higher the score the higher adherence to the dietary pattern) *(46)*.

3.3.4. Criteria for examination of the metric

3.3.4.1. Content

The developers mentioned that "the DASH score was designed to reflect adherence to the DASH-style diet which is high in fruits and vegetables, moderate in low-fat dairy products, and low in animal protein but with substantial amount of plant protein from legumes

Table 4. Diet Scoring Criteria for the DASH-style diet and mean[a] intake for Q1 (low consumption) and Q5 (high consumption) in the cohort

Component	Foods	Scoring criteria	Q1, servings/ day	Q5, servings/ day
Fruits	All fruits and fruit juices	Q1=1 point Q2=2 points Q3=3 points Q4=4points Q5=5 points	0.7	4.1
Vegetables	All vegetables except potatoes and legumes		1.1	4.6
Nuts and legumes	Nuts and peanut butter, dried beans, peas, tofu		0.3	1.5
Whole grains	Brown rice, dark breads, cooked cereals, whole grain cereal, other grains, popcorn, wheat germ, bran		0.1	2.4
Low-fat dairy	Skim milk, yogurt, cottage cheese		0.1	2.3
Sodium[b]	Sum of sodium content of all foods in FFQ	Reverse scoring: Q1=5 point Q2=4 points Q3=3 points Q4=2points Q5=1 points	1041 mg	2676 mg
Red and processed meats[b]	Beef, pork, lamb, deli meats, offal, hot dogs, bacon		0.4	1.8
Sweetened beverages[b]	Carbonated and non-carbonated sweetened beverages		0	1.2

[a] Mean of 5 FFQs.
[b] Higher quintiles represent higher intake; however, in constructing the DASH score, high intake and high quintiles received lower scores.
Source: Fung TT et al., 2008 *(46)*.

and nuts" *(46)*. We concluded that the DASH score assesses adherence to a dietary pattern that has been proven favourable for health and thus assesses the subconstruct "favourable dietary pattern".

We assessed the content of the DASH score as holistic.

3.3.4.2. Cross-context equivalence

The DASH-style diet was originally developed for the USA context *(51, 52)*, and the DASH score has been evaluated in this specific context *(46)*. Subsequently, this metric has been used unchanged in other contexts, such as Europe and China *(47)*, although it was not designed, developed and validated to be equivalent across contexts. Nevertheless, because all the items of the DASH score depend on quintiles of distribution in the studied population, we can conclude that the DASH score is at best measurement equivalent in contexts in which it has already been used (mainly in China and the United States, and in Europe).

We assessed the DASH score as measurement equivalent.

3.3.4.3. Validity for initial purpose

In an analysis of longitudinal quantitative dietary data from women aged 30–55 years (Nurses' Health Study) in the USA, the developers found that "women in the top quintile of the DASH score, compared with those in the bottom quintile, had a relative risk of

0.76 (95% confidence interval, 0.67–0.85) and the magnitude of risk difference was similar for nonfatal myocardial infarction and fatal CHD". They found that "the DASH score was also significantly associated with lower risk of stroke (multivariate relative risks across quintiles of the DASH score were 1.0, 0.92, 0.91, 0.89, and 0.82)". They also found that "the DASH score was significantly associated with lower plasma levels of C-reactive protein and interleukin 6, in cross-sectional analysis in a subgroup of women with blood samples" *(46)*.

The initial purpose of the DASH score was to assess adherence to a dietary pattern favourable for reducing both systolic and diastolic blood pressure among hypertensive and normotensive individuals in a clinical trial in the USA. We concluded that the developers of the DASH score used adequate a priori study design (longitudinal design with large sample sizes and multi-adjusted regressions) to demonstrate that this metric indicates adherence to a dietary pattern inversely associated with a risk of major NCDs in two cohorts in the USA.

The validity for initial purpose of the DASH score was assessed as high.

Other elements of validation of the DASH score

In the first article describing the development of the DASH score, neither sensitivity to change, test-retest reliability nor internal consistency were reported as having been evaluated *(46)*. In recent meta-analyses, it was found that diets that score highly on the DASH score were associated with a significant reduction in the risk of all-cause mortality, cardiovascular disease, cancer, type 2 diabetes and neurodegenerative disease in different cohorts (mainly from the USA, but also China and Europe) *(35, 47)*, which confirms that the DASH score assesses adherence to a dietary pattern favourable for health.

3.3.4.4. Population groups covered

All analyses presented for validation of the DASH score are for men and women aged ≥ 18 years (including the elderly) *(35, 46, 47)*. Additional research is needed to explore how the DASH score could be adapted for other population groups such as adolescents and children (Table 5).

Table 5. Population groups and applicability of DASH score

Population group	Applicability
Children (2–10 years)	Possible but requires additional research
Adolescent males (11–17 years)	Possible but requires additional research
Adolescent females (11–17 years)	Possible but requires additional research
Adult males (18–59 years)	Demonstrated
Adult females (18–59 years)	Demonstrated
Pregnant and/or lactating women	Possible but requires additional research
Elderly (≥ 60 years)	Demonstrated

3.3.4.5. Interpretation threshold

No interpretation threshold was proposed in the article describing the development of the DASH score *(46)*. During our email exchange, the developers confirmed that there is no interpretation threshold of the DASH score. We concluded that additional research is needed to determine a cut-off point as the best predictor of a lower risk of NCDs.

3.3.4.6. Ease of computation

The process of computing the DASH score appears to be complex and requires multiple steps as follows: cleaning the data; merging the food composition database; processing dietary data to estimate gram consumption of food, ingredients and nutrients per respondent; assigning food, ingredient and nutrient to the appropriate metric item; summing the quantity of consumption per metric item per respondent; assigning points per respondent; calculating the score; and tabulating the indicator.

The ease of computation of the DASH score was assessed as very low.

3.3.4.7. Ease of interpretation

The DASH score could only be interpreted within a possible range of 8 to 48 (the higher the score the better adherence to the dietary pattern).

The ease of interpretation of the DASH score was assessed as low.

3.3.5. Criteria for examination of the data collection tools

In the article describing the development of the DASH score, the authors used validated semi-quantitative FFQ to derive data for the DASH score *(46)*. However, single or multiple quantitative 24-hour dietary recalls can also be used to derive data for the DASH Score *(53)*. In addition, some authors have developed short screeners to derive data for the DASH Score *(41)*, but because these tools are specific to the studies and populations involved, we did not evaluate them further.

3.3.5.1. Criterion validity

Multiple, quantitative 24-hour dietary recalls and FFQs can be considered to be standard dietary assessment methods that have been extensively validated. Nevertheless, the appropriateness of these methods for use in different population groups and contexts ought to be verified in a validation study in a study population *(43)*. In the case of the FFQ used in the study by Fung et al. (2008) *(46)*, the authors mentioned that previous validation studies revealed good correlations between nutrients assessed by the FFQ and multiple weeks of food records completed over the preceding year.

3.3.5.2. Interviewer burden

According to the DAPA Measurement Toolkit, single or multiple quantitative 24-hour dietary recalls are associated with a high interviewer burden, because these methods require the interviewers to have advanced training, a good understanding of the methods and a reasonable knowledge of how food is acquired, prepared and consumed in the survey areas. In comparison, FFQs are easier and more flexible to administer *(44)*.

3.3.5.3. Respondent burden

According to the DAPA Measurement Toolkit, while respondent burden is considered relatively low for a single quantitative 24-hour dietary recall, it is considered higher for multiple quantitative 24-hour dietary recalls due to the number of interviews. Furthermore, this method depends on the respondent's ability to recall intake accurately, which can increase its burden. FFQs are associated with a higher respondent burden than single or multiple quantitative 24-hour dietary recalls because they require more time, literacy and knowledge about food. Furthermore, the respondent burden is bound to increase: the greater the number of FFQ items and the longer the memory period (e.g. diet over the past year versus past month) the greater the respondent burden *(44)*.

3.3.5.4. Collection time

According to the DAPA Measurement Toolkit, interviews using a single 24-hour recall method are reported to last from 20 to 30 minutes depending on the respondent's diet *(44)*. According to the National Cancer Institute Dietary Assessment Primer, completing a FFQ aimed at capturing the total dietary intake usually requires 30 to 60 minutes *(45)*.

3.4. Diet Quality Index-International (DQI-I)

Detailed information related to the development and performance of the DQI-I was found in the peer-reviewed article by Kim et al. (2003) *(38)*. Requests for clarification and confirmation of some information items were sent to the developers by email and responses received. Based on these responses, we found additional information related to the use of the DQI-I in other populations and contexts in various peer-reviewed articles *(54–63)*.

3.4.1. Rationale and purpose

The DQI-I is a measure of diet quality that was designed for international comparisons by a research team from the University of North Carolina at Chapel Hill. According to the developers, "the DQI-I was designed to provide an overall measure of diet quality of adults aged ≥ 20 years that can be used for international comparisons". They consider that "because the DQI-I focuses on 4 of the sub-constructs of diet quality (i.e. variety, adequacy, moderation and overall balance), it provides a global tool for monitoring healthfulness of diet of both developed and developing countries and for exploring aspects of diet quality related to the nutrition transition" *(38)*. We concluded that the purpose of the DQI-I was to measure at population level four subconstructs of healthy diets ("diversity", "nutrient adequacy", "moderation" and "macronutrient balance") in different settings.

3.4.2. Uses

According to the authors, "the DQI-I is a composite measure of diet quality created to evaluate healthfulness of diet not only within a country for monitoring purposes but also across countries for comparative work". It is also "a global tool for monitoring healthfulness of diet and for exploring aspects of diet quality related to the nutrition transition" *(38)*. While the DQI-I seems to be a complex and resource-intensive metric (see following sections), it was designed for international comparisons when monitoring the healthfulness of diet. We therefore concluded that it could be used for identification, monitoring and surveillance, but also for policy and programme design, evaluation and research.

3.4.3. Scoring system

According to the authors, the DQI-I is composed of 17 items related to food groups or nutrients, distributed in four submetrics: variety, adequacy, moderation and overall balance (Table 6). Scores for each item are summarized in each of the four submetrics and the scores for all four submetrics are summed, resulting in the total DQI-I score which ranges from 0 to 100 (0 being the poorest and 100 being the highest possible score) *(38)*.

3.4.4. Criteria for examination of the metric

3.4.4.1. Content

The authors mentioned that "the DQI-I focuses on 4 major aspects of a high-quality healthy diet: variety, adequacy, moderation and overall balance" *(38)*. We concluded that the DQI-I assesses four subconstructs of the construct of healthy diets: "nutrient adequacy", "moderation", "diversity" and "macronutrient balance".

We assessed the content of the DQI-I as multiple.

Table 6. Components of Diet Quality Index-International (DQI-I) and the percentage of the sample in component subcategories in China and the USA[a]

Component	Score	Scoring criteria	China	USA
Variety	0–20 points			
Overall food group variety (meat/poultry/fish/eggs; dairy/beans; grain; fruit; vegetable)	0–15 points	≥1 serving from each food group/day = 15	2.4	23.3
		Any 1 food group missing/day = 12	28.8	41.6
		Any 2 food groups missing/day = 9	43.6	26.9
		Any 3 food groups missing/day = 6	25.0	6.9
		≥4 food groups missing/day = 3	0.3	1.2
		None from any food groups = 0	0.0	0.1
Within-group variety for protein source (meat, poultry, fish, dairy, beans, eggs)	0–5 points	≥3 different sources/day = 5	28.1	68.4
		2 different sources/day = 3	28.6	25.1
		From 1 source/day = 1	27.0	6.1
		None = 0	16.3	0.4
Adequacy	0–40 points			
Vegetable groups[b,c]	0–5 points	≥3–5 servings/day = 5, 0 serving/day = 0		
		≥100%	82.2	42.1
		<100–50%	14.7	37.7
		<50%	3.1	20.2
Fruit groups[b,c]	0–5 points	≥2-4 servings/d = 5, 0 serving/day = 0		
		≥100%	0.4	19.6
		<100-50%	2.4	23.4
		<50%	97.2	57.0
Grain group[b,c]	0–5 points	≥6-11 servings/d = 5, 0 serving/d = 0		
		≥100%	99.1	9.6
		<100-50%	0.7	59.8
		<50%	0.2	30.7
Fibre[b,c]	0–5 points	≥20-30 g/d = 5, 0 g/d = 0		
		≥100%	3.9	13.9
		<100-50%	28.7	52.6
		<50%	67.3	33.5
Protein[b]	0–5 points	≥10% of energy/d = 5, 0% of energy/d = 0		
		≥100%	80.3	95.3
		<100-50%	19.6	4.5
		<50%	0.1	0.1
Iron[b,d]	0–5 points	≥100% RDA (AI)/d = 5, 0% RDA (AI)/d = 0		
		≥100%	68.3	68.9
		<100-50%	30.4	22.5
		<50%	1.3	8.7
Calcium[b]	0–5 points	≥100% AI/d = 5, 0% AI/d = 0		
		≥100%	2.9	16.0
		<100-50%	36.4	44.9
		<50%	60.7	39.1
Vitamin C[b,e]	0–5 points	≥100% RDA (RNI)/d = 5, 0 RDA (RNI)/d = 0		
		≥100%	43.3	44.0
		<100-50%	37.1	27.9
		<50%	19.6	28.1
Moderation	0–30 points			
Total fat	0–6 points	≤20% of total energy/d = 6	33.7	5.5
		>20-30% of total energy/d = 3	31.5	27.4
		>30% of total energy/d = 0	34.9	67.1
Saturated fat	0–6 points	≤7% of total energy/d = 6	57.6	11.4
		>7-10% of total energy/d = 3	24.5	27.2
		>10% of total energy/d = 0	18.0	61.4
Cholesterol	0–6 points	≤300 mg/d = 6	77.2	66.4
		>300-400 mg/d = 3	8.2	14.4
		>400 mg/d = 0	14.6	19.2
Sodium	0–6 points	≤2400 mg/d = 6	9.5	30.9
		>2400-3400 mg/d = 3	9.3	29.9
		>3400 mg/d = 0	81.3	39.2

Component	Score	Scoring criteria	China	USA
Empty calorie foods	0–6 points	≤3% of total energy/d = 6 >3-10% of total energy/d = 3 >10% of total energy/d = 0	94.5 2.8 2.7	63.7 22.6 13.8
Overall balance	0– 10 points			
Macronutrient ratio[f] (carbohydrate: protein:fat)	0–6 points	55 ~ 65:10 ~ 15:15 ~ 25 = 6 52 ~ 68:9 ~ 16:13 ~ 25 = 4 50 ~ 70:8 ~ 17:12 ~ 30 = 2 Otherwise = 0	4.8 14.0 15.6 65.6	1.2 5.1 9.6 84.1
Fatty acid ratio (PUFA:MUFA:SFA)	0–4 points	P/S = 1 ~ 1.5 and M/S = 1 ~ 1.5 = 4 Else if P/S = 0.8 ~ 1.7 and M/S = 0.8 ~ 1.7 = 2 Otherwise = 0	14.5 19.3 66.2	7.1 16.2 76.7

AI, adequate intakes; M/S, ratio of MUFA to SFA intake; MUFA, monounsaturated fatty acids; P/S, ratio of PUFA to SFA intake; RDA, recommended dietary allowance; RNI, recommended nutrient intake; SFA, saturated fatty acids.
[a] Values are the percentages of the sample in subcategories.
[b] Used as a continuous variable.
[c] Based on 7118 kJ (1700 kcal)/9211 kJ (2200 kcal)/11304 kJ (2700 kcal) diet; 1 kcal = 4.1868 kJ.
[d] Scoring system based on the AI value for China and RDA value for the USA.
[e] Scoring system based on the RNI value for China and RDA value for the USA.
[f] Ratio of energy from carbohydrate to protein to fat.
Source: Kim S et al., 2003 *(38)*.

3.4.4.2. Cross-context equivalence

The DQI-I was originally developed and evaluated in two national contexts (China and USA) *(38)*. Since it is based on secondary data analysis, the same construct was measured across contexts, with the content of each item used being interpreted in the same way: this means that the difference in scores between two individuals is the same across contexts as is the definition of zero. During our email exchange, the developers mentioned that the DQI-I had been used in at least 79 other studies, in contexts such as Canada, Greece, Guatemala, Iran (Islamic Republic of), Portugal, Republic of Korea, Sri Lanka, South Africa and Tunisia *(54–63)*.

We assessed the DQI-I as scalar equivalent.

3.4.4.3. Validity for initial purpose

In their analysis of cross-sectional, nationally representative, quantitative dietary data from men and women aged ≥ 20 years in two national contexts (China and USA), the developers found that "as the scores of the DQI-I moved from lower to higher levels, desirable intakes (e.g. fruit and vegetable) increased steadily and intakes of less desirable nutrients and food (e.g. fat, SFA) gradually declined". They also found that "the percentages with intakes less than the US nutrient recommendations, indicating a higher probability of inadequacy in intake, decreased as the DQI-I scores increased in both countries" *(38)*. The initial purpose of the DQI-I is to measure at population-level four subconstructs of healthy diets ("diversity", "nutrient adequacy", "moderation" and "macronutrient balance") in different settings. While the developers have developed and evaluated the metric in two large nationally representative dietary surveys from the China and USA, we concluded that they demonstrated that higher DQI-I scores were more likely to meet nutrient recommendations only for vitamin A, riboflavin, vitamin E, zinc and calcium in simple analysis. This showed that the DQI-I indicates the subconstruct "nutrient adequacy". They also determined trends in the mean intake of several nutrients and foods (fruits and vegetables) across tertiles of DQI-I scores. In this analysis, three nutrients related to moderation were evaluated: energy from fat and saturated fatty acids, and sodium. While higher DQI-I scores were associated with lower energy from fat and saturated fatty acid, there was no association with sodium. We concluded that this demonstrated that the DQI-I barely indicates the subconstruct "moderation". From different studies (see below), we concluded that the ability of the DQI-I to indicate the subconstruct "moderation" was also questionable. To our knowledge, there were no studies investigating whether the DQI-I indicates the subconstructs "diversity" and "macronutrient balance".

The validity for initial purpose of the DQI-I was assessed as medium.

Other validation elements of the DQI-I

In the article describing the development of the DQI-I, "the scores of the metric were examined in terms of how they reflected variation in the individual components on which the index was based" *(38)*, which can be considered to be a way of evaluating the internal consistency of the metric. Sensitivity to change and test-retest reliability have not been reported as being assessed *(38)*.

In studies conducted by teams other than the one that developed the DQI-I, evidence that the DQI-I was consistently associated with lower risks of NCDs was inconsistent. In an analysis of cross-sectional quantitative dietary data from Guatemalan adults aged 25–42 years, it was found that the DQI-I was poorly related to cardiometabolic risk *(61)*. In an analysis of cross-sectional quantitative dietary data from the Tehran Lipid and Glucose Study population, it was found that the DQI-I could not predict BMI and waist circumference in Iranian adults aged ≥ 19 years after 6.7 years of follow-up *(62)*. In an analysis of cross-sectional quantitative dietary data from Iranian women aged 20–50 years, it was found that the DQI-I was inversely and directly associated with serum levels of total cholesterol and high-density lipoprotein cholesterol, but there was no significant association with other cardiovascular risk factors *(63)*.

3.4.4.4. Population groups covered

The analyses presented for validation of the DQI-I are for men and women aged ≥ 20 years (including the elderly) *(38)*. During our email exchange, the developers mentioned that the DQI-I had been used in at least 79 other studies involving various population groups such as children and adolescents *(54–57)* and pregnant women *(58, 59, 64)*. We concluded that a targeted systematic review would be required to assess the validity of the DQI-I in these different studies and determine more precisely whether further research is needed to validate the DQI-I in these population groups (Table 7).

Table 7. Population groups and applicability of DQI-I

Population group	Applicability
Children (2–10 years)	Already used but requires additional research
Adolescent males (11–17 years)	Already used but requires additional research
Adolescent females (11–17 years)	Already used but requires additional research
Adult males (18–59 years)	Demonstrated
Adult females (18–59 years)	Demonstrated
Pregnant and/or lactating women	Already used but requires additional research
Elderly (≥ 60 years)	Demonstrated

3.4.4.5. Interpretation threshold

According to the developers, "the scores of the four submetrics of the DQI-I were dichotomized into good and poor categories using the cut-off point of 60% of full scores. The absolute cut-off point was chosen by the authors over one based on distribution in distinguishing good and poor quality diets to provide a standard that was meaningful for comparative purposes rather than a data-driven criterion determined by specific country's data distribution" *(38)*. Nevertheless, it was not explicitly mentioned whether the 60% cut-off point could be used for the total DQI-I score. During our email exchange, the developers mentioned that the total DQI-I score had been interpreted in various ways in subsequent studies using DQI-I (applying a continuous measure, using tertiles or quartiles, and a 60% cut-off point) and were unable to confirm that the 60% cut-off point is the best threshold to distinguish between good and poor quality diets using the DQI-I. We concluded that additional research is needed to confirm whether the 60% cut-off point is the best threshold to distinguish between good and poor quality diets using the DQI-I.

3.4.4.6. Ease of computation

The process of computing the DQI-I appears to be complex and requires multiple steps as follows: cleaning the data; merging the food composition database; processing dietary data to estimate gram consumption of food, ingredient and nutrients per respondent; assigning food, ingredient and nutrient to the appropriate metric item; summing the quantity of consumption per metric item per respondent; assigning points per respondent; calculating the score; and tabulating the indicator.

The ease of computation of the DQI-I was assessed as very low.

3.4.4.7. Ease of interpretation

The DQI-I can be interpreted by applying a possible range of 0 to 100 (the higher the score the better the diet quality). However, interpretation of its four submetrics dichotomized into good and poor categories using the 60% cut-off point for full scores could be more challenging.

The ease of interpretation of the DQI-I was assessed as very low.

3.4.5. Criteria for examination of the data collection tools

In the article describing the development of the DQI-I, the authors used multiple quantitative 24-hour dietary recalls to derive data for the DQI-I *(38)*. In another article where the relevance of the DQI-I for assessing the quality of the Mediterranean diet was questioned, multiple quantitative 24-hour dietary recalls and FFQs were used to derive data for the DQI-I *(65)*.

3.4.5.1. Criterion validity

Multiple quantitative 24-hour dietary recalls and FFQs can be considered as standard dietary assessment methods that have been extensively validated. Nevertheless, the appropriateness of these methods for use in different population groups and contexts ought to be verified in a validation study in a study population *(43)*.

3.4.5.2. Interviewer burden

According to the DAPA Measurement Toolkit, single or multiple quantitative 24-hour dietary recalls are associated with high interviewer burden, because these methods require the interviewers to have advanced training, good understanding of the methods and a reasonable knowledge of how food is acquired, prepared and consumed in the survey areas. In comparison, FFQs are easier and more flexible to administer *(44)*.

3.4.5.3. Respondent burden

According to the DAPA Measurement Toolkit, while respondent burden is relatively low for a single quantitative 24-hour dietary recall, it is higher for multiple quantitative 24-hour dietary recalls due to the number of interviews. Furthermore, this method depends on the respondent's ability to recall intake accurately which can increase its burden. FFQs are associated with a higher respondent burden than single or multiple quantitative 24-hour dietary recalls because they require more time, literacy and knowledge about food. Furthermore, the respondent burden is bound to increase with a greater number of FFQ items and longer recall period (e.g. diet over the past year versus past month) *(44)*.

3.4.5.4. Collection time

According to the DAPA Measurement Toolkit, interviews using a single 24-hour recall method are reported to last from 20 to 30 minutes depending on the respondent's diet *(44)*. According to the National Cancer Institute Dietary Assessment Primer, completing a FFQ aimed at capturing the total dietary intake usually requires 30 to 60 minutes *(45)*.

3.5. Global Diet Quality Score (GDQS)

Detailed information related to the development and performance of the GDQS was found in peer-reviewed articles published in a supplementary issue of the Journal of Nutrition in 2021 regarding the GDQS *(66–70)*. We found additional information related to the validity and use of the GDQS in data collection options and tabulation guidelines published in 2021 by Intake–Center for Dietary Assessment *(71)*. Requests for clarification and confirmation of certain information were sent to the developers by email and responses received.

3.5.1. Rationale and purpose

The GDQS is a food group-based metric of diet quality that was developed for diverse contexts (HIC and LMIC), by an international research initiative led by a research team at Harvard University, the National Institute of Public Health in Mexico and scientists at Intake–Center for Dietary Assessment. According to the developers, "the GDQS was designed to be appropriate for use among non-pregnant, non-lactating women of reproductive age (15–49 years) in LMICs but has also been shown through secondary data analysis to be valid for use in high-income countries, thereby providing a simple, standardized metric appropriate for population-based measurement of diet quality globally" *(71)*. "The GDQS is based on the Prime Diet Quality Score which was improved to represent the diversity of nutritionally important foods more fully across LMIC globally, and the most up-to-date scientific evidence regarding relations between consumption of different foods and health" *(66)*. We concluded that the purpose of the GDQS was to measure diet quality at population level in diverse settings.

3.5.2. Uses

According to the developers, "GDQS data are intended to be reported and used at the population or sub-group level, not at the individual level, for population-based assessment, target-setting, programme/policy design, cross- or within-country comparison, assessing population-level changes in diet quality, and monitoring and evaluation of programmes and policies that aim to improve diet quality" *(71)*. Because the GDQS was developed as a simple, rapid and feasible standardized metric for diverse context, we concluded that it could be used for identification, monitoring and surveillance, but also for policy and programme design, evaluation and research.

3.5.3. Scoring system

The GDQS is composed of 25 food groups which are considered to be important contributors to nutrient intake and/or NCD risk. Points are assigned based on three or four categories of consumed amounts (defined in g/d) specific to each group (Table 8). There are 16 healthy food groups (more points for higher intake), seven unhealthy food groups (more points for lower intake), and two food groups classified as unhealthy when consumed in excessive amounts. The GDQS is obtained by summing points across all 25 food groups, and ranges from 0 to 49 *(66)*.

The GDQS+ submetric includes the 16 healthy food groups included in the GDQS, is scored with the same categories of consumed amounts used in the GDQS, and ranges from 0 to 32. The GDQS- submetric includes the 9 GDQS food groups classified as unhealthy or unhealthy in excessive amounts, is scored with the same categories of consumed amounts used in the GDQS, and ranges from 0 to 17. The GDQS+ and GDQS- quantify the collective contribution of healthy foods (those that should be consumed in higher amounts) and unhealthy foods (those that should be consumed in lower amounts), respectively *(66)*.

Table 8. GDQS and GDQS submetric food groups and scoring[a]

	Categories of consumed amounts (g/day)				Point values			
Food group	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Food groups included in the GDQS and GDQS+								
Healthy								
Citrus fruits	<24	24–69	>69		0	1	2	
Deep orange fruits	<25	25–123	>123		0	1	2	
Other fruits	<27	27–107	>107		0	1	2	
Dark green leafy vegetables	<13	13–37	>37		0	2	4	
Cruciferous vegetables	<13	13–36	>36		0	0.25	0.5	
Deep orange vegetables	<9	9–45	>45		0	0.25	0.5	
Other vegetables	<23	23–114	>114		0	0.25	0.5	
Legumes	<9	9–42	>42		0	2	4	
Deep orange tubers	<12	12–63	>63		0	0.25	0.5	
Nuts and seeds	<7	7–13	>13		0	2	4	
Whole grains	<8	8–13	>13		0	1	2	
Liquid oils	<2	2–7.5	>7.5		0	1	2	
Fish and shellfish	<14	14–71	>71		0	1	2	
Poultry and game meat	<16	16–44	>44		0	1	2	
Low fat dairy	<33	33–132	>132		0	1	2	
Eggs	<6	6–32	>32		0	1	2	
Food groups included in the GDQS and GDQS-								
Unhealthy in excess amount								
High fat dairy (in milk equivalents)[b]	<35	35–142	>142–734	>734	0	1	2	0
Red meat	<9	9–46	>46		0	1	0	
Unhealthy								
Processed meat	<9	9–30	>30		2	1	0	
Refined grains and baked goods	<7	7–33	>33		2	1	0	

Table 8. Continued

Food group	Categories of consumed amounts (g/day)				Point values			
	1	2	3	4	1	2	3	4
Sweets and ice cream	<13	13–37	>37		2	1	0	
Sugar-sweetened beverages	<57	57–180	>180		2	1	0	
Juice	<36	36–144	>144		2	1	0	
White roots and tubers	<27	27–107	>107		2	1	0	
Purchased deep fried foods	<9	9–45	>45		2	1	0	

GDQS, Global Diet Quality Score; GDQS-, GDQS Negative Submetric; GDQS+, GDQS Positive Submetric.

[a] Due to the importance of cheese in many food cultures and the significantly different nutrient density of hard cheeses in comparison with other dairy products, we recommend converting consumed masses of hard cheeses to milk equivalents when calculating total consumption of high fat dairy for the purpose of assigning a GDQS consumption category [using cheddar cheese as a typical example, a conversion factor of 6.1 can be computed as the mass of 1 serving of milk (237 mL × 0.95 g/mL = 225 g) divided by an isocaloric mass of cheddar cheese (37 g)].

Source: Bromage S et al., 2021 *(66)*.

3.5.4. Criteria for examination of the metric

3.5.4.1. Content

The developers mentioned that "the GDQS is composed of 25 food groups that are globally important contributors to nutrient intake and/or NCD risk as informed by current nutrition science and epidemiologic literature" *(66)*. While the mention of food groups that are globally important contributors to nutrient intake and/or NCD risk would suggest that the GDQS assesses the subconstructs "nutrient adequacy" and "moderation", we concluded that the GDQS assesses adherence to a dietary pattern where the consumption of 25 food groups is either promoted or limited and thus assesses the subconstruct "favourable dietary pattern".

We assessed the content of the GDQS as holistic.

3.5.4.2. Cross-context equivalence

The GDQS was developed and evaluated in 14 national or subnational contexts (China, Ethiopia, Ghana, India, Kenya, Malawi, Mali, Mexico, Nigeria, Rwanda, Senegal, Uganda, United Republic of Tanzania and the USA). Since it is based on secondary data analysis, the same construct was measured across contexts, with the content of each item used being interpreted in the same way: this means the difference in scores between two individuals is the same across contexts as is the definition of zero.

We assessed the GDQS as scalar equivalent.

3.5.4.3. Validity for initial purpose

In analyses of cross-sectional data from non-pregnant, non-lactating WRA in 12 national or subnational contexts (China, Ethiopia, India, Mexico and 12 Millennium Villages in 10 countries), the developers "considered that the GDQS performed well compared with the MDD-W in capturing nutrient adequacy, and anthropometric and biochemical indicators of undernutrition". They observed "that the GDQS tended to correlate more favorably than the MDD-W with energy-adjusted fiber, folate, iron, protein, saturated fat, and zinc intakes, whereas the MDD-W tended to correlate better with energy-adjusted monounsaturated fat, vitamin A, and vitamin B12 intakes. Performance of the GDQS and MDD-W in predicting overall nutrient inadequacy (mean of probability of adequacy for protein, fiber, calcium, iron, zinc, vitamin A, folate, and vitamin B12) did not differ (except in one dataset where MDD-W outperformed the GDQS). In adjusted regression models, the GDQS and MDD-W were significantly inversely associated with underweight in Ethiopia and India; low mid–upper arm circumference in Ethiopia, India, and the African Millennium Villages; and unassociated with underweight or anemia in China. These metrics were also inversely associated with serum folate deficiency in Ethiopia, associated with higher serum folate concentrations in Mexico, inversely associated with anemia in the African Millennium Villages, and associated with higher hemoglobin concentrations or inversely associated with anemia in Ethiopia" *(66)*.

In analyses of cross-sectional data from NPNL WRA in four national or subnational contexts (China, Ethiopia, India and Mexico), the authors "considered that the GDQS performed comparably or better than the AHEI-2010 in capturing diet-related NCD risk". They observed that "the GDQS significantly outperformed the AHEI-2010 in predicting the metabolic syndrome (MetS) in urban China. In rural China, the AHEI-2010 was positively associated with MetS, whereas the GDQS was marginally associated with lower odds of MetS and significantly associated with lower odds of high waist circumference, hypertension, and low HDL cholesterol. Both the GDQS and AHEI-2010 were positively associated with overweight and high waist circumference in India. Although not predictive of the MetS in urban or rural Mexico, the GDQS was inversely associated with continuous BMI, waist circumference, and LDL cholesterol" *(66)*. In multivariable analyses of longitudinal data from women in the Mexican Teachers' Cohort (25–49 years), the developers found that "a higher GDQS was associated with less gain in weight and in waist circumference, and the GDQS was significantly more strongly associated with weight change than the MDD-W and with waist circumference change than the AHEI-2010" *(67)*. In multivariable analyses of longitudinal data from women in the US Nurses' Health Study II (27–44 years at inception in 1989 although completion of a FFQ every four years made it possible to work on a subsample of women aged over 50 years), the developers found that "a higher GDQS was inversely associated with weight gain and type 2 diabetes risk, and the GDQS performed nearly as well as the AHEI-2010" *(68, 69)*.

The initial purpose of the GDQS is to measure diet quality at population level in diverse settings. We concluded that the developers used an adequate a priori study design (cross-sectional and longitudinal designs with large sample sizes and multi-adjusted analyses) to demonstrate that this metric indicates the subconstruct "nutrient adequacy" in 14 diverse contexts (similar correlation to the mean probability of adequacy as the MDD-W) and subconstruct "moderation" in five diverse contexts (similar or better associations with diet-related NCD risk than the AHEI-2010).

The validity for initial purpose of the GDQS was assessed as high.

<u>Other elements of validation of the GDQS</u>

Sensitivity to change of the GDQS has been evaluated and demonstrated through analyses of longitudinal cohort data in Mexico *(67)* and the USA *(68, 69)*. Test-retest reliability and internal consistency have not been reported as being assessed.

3.5.4.4. Population groups covered

All analyses presented for validation of the GDQS are for NPNL and WRA (in the Chinese dataset women were aged more than 18 years) and many analyses were also done for adult men (China, Ethiopia and 12 Millennium Villages in 10 countries), suggesting that the GDQS is valid for men. During our email exchange, the developers mentioned that they are currently in the "process of validating the GDQS for children 24–59 months, 5–9.9 years, and 10–14.9 years and have some initial validation results" (Table 9). They also mentioned that "elderly are included in the sample of data collected for the primary validation studies carried out in Thailand (under analysis) and US (data collection phase)". They planned to explore the validity of the GDQS for pregnant women and lactating women subgroups but expect to "have some limitations regarding sample sizes".

Table 9. Population groups and applicability of GDQS

Population group	Applicability[a]
Children (2–10 years)	Ongoing research
Adolescent males (11–17 years)	Ongoing research
Adolescent females (11–17 years)	Demonstrated only in 15–17 years old
Adult males (18–59 years)	Demonstrated
Adult females (18–59 years)	Demonstrated only in 18–49 years old
Pregnant and/or lactating women	Ongoing research
Elderly (≥ 60 years)	Ongoing research

[a] As reported by metric developers

3.5.4.5. Interpretation threshold

According to the developers, "GDQS scores ≥23 are associated with a low risk of both nutrient adequacy and NCD risk, scores ≥15 and <23 indicate moderate risk, and scores <15 indicate high risk" *(66)*. During our email exchange, the developers mentioned that "they applied a regression approach to predict continuous or categorical outcomes (related to either nutrient adequacy or NCD risk) in covariate-adjusted models that accounted for potential nonlinearity of these relationships, graphed relationships between GDQS and predicted mean or odds of each outcome, and compared these graphs across outcomes and datasets to identify optimal data-driven cutoffs across datasets".

3.5.4.6. Ease of computation

According to the developers, "the ease of computation of the GDQS appears to be greatly improved when using the GDQS app because the GDQS application includes a global database of 5500 foods from all regions of the world where foods and beverages are pre-classified into corresponding GDQS food groups" *(70)*. In addition, "the process of tabulating GDQS data with data collected using the GDQS app requires only 2 steps [the GDQS tabulation is automated with the GDQS app] while tabulating data collected with a quantitative 24-hour dietary recall survey or a FFQ require 6 and 7 steps, respectively" *(71)*.

The ease of computation of the GDQS score was assessed as low (when using quantitative 24-hour dietary recall or FFQ) and high (when using the GDQS app).

3.5.4.7. Ease of interpretation

The GDQS can be interpreted by applying a possible range of 0 to 49 (the higher the score the better the diet quality) and population-based cut-offs of 15 and 23 to report the percentage of the population at high risk for poor diet quality outcomes (GDQS < 15) and percentage at low risk for poor diet quality outcomes (GDQS ≥ 23).

The ease of interpretation of the GDQS was assessed as high.

3.5.5. Criteria for examination of the data collection tools

According to the developers, various data collection methods can be used to derive data for the GDQS: "quantitative 24-h dietary recall, semi-quantitative or qualitative FFQ, and the GDQS app. Among the standard dietary assessment method, the quantitative 24-h dietary recall is considered as an ideal data source for the GDQS while the semi-quantitative or qualitative FFQs require a certain number of prerequisites (e.g. to comprehensively list the foods commonly consumed by the target population and allow for the quantity of intake of each food to be derived for a 24-hour reference period). The GDQS app is a new and dedicated method to collect population-based GDQS data, with a simplified quantification method using a standard set of 10 3D cubes as visual aids to enable the respondent to easily classify the quantity of consumption per GDQS food group" *(71)*.

3.5.5.1. Criterion validity

To assess the relevance of the simplified quantification method using a standard set of 10 3D cubes, the authors applied secondary analysis to six datasets from different LMICs comparing the GDQS with the original and simplified quantification approaches and demonstrated that were no differences in performance between the two approaches *(70)*. Nevertheless, this is only a theoretical validation and we concluded that additional research in a field setting is needed to confirm the validity of the GDQS app. During our email exchange, the developers mentioned that they had or were undertaking such research by carrying out "an initial field test of the GDQS app in Ethiopia (data not published) and a validation study in Thailand (under analysis)" and that "Intake is currently collecting data for a validation study using the GDQS app (against weighed records) in the USA". In addition, they mentioned that studies were planned to start soon in "Bangladesh, Cameroon, Democratic Republic of the Congo, India, Nepal and Nigeria".

3.5.5.2. Interviewer burden

According to the developers, "the GDQS app does not require interviewers to have specific training or expertise in nutrition or food preparation in a given context to collect high-quality data to tabulate the GDQS" *(70)*. During our email exchange, the developers specified that a "short training is needed for the interviewer to walk through the functions and features of the app and how to use the app, as well as to provide skills for how to conduct the interview". They also specified that the GDSQ app provides "the interviewer with prompts about when more specific details are needed about a food consumed to be able to classify the food into the correct GDQS food group", indicating that investigators do not need knowledge at a higher level than that of a normal consumer. We concluded that the GDQS app requires only short specific training.

3.5.5.3. Respondent burden

According to the developers, "using interviewers to conduct the interviews with the help of the application facilitates data collection in contexts where targeted respondents may have a low level of literacy" *(70)*. Since data from the GDQS app field-tests is not yet available, it is difficult to assess to what extent this method requires specific effort from respondents or is cognitively complex. However, we can assume that the GDQS app would require the same or less specific effort from respondents than a classic quantitative 24-hour dietary recall and would be the same or less cognitively complex. During our email exchange, the developers mentioned that they estimate the GDQS app to require less specific effort from respondents. To support this, they "collected qualitative data from respondents during pre-test in Ethiopia to learn about their perceptions of using the cubes and are in process of preparing a paper to report the results" and they are “collecting feedback about use of the cubes from respondents as part of validation study in the USA”.

3.5.5.4. Collection time

According to the developers, “collecting data with the GDQS app takes an average of 10-20 minutes per respondent, depending on the complexity and diversity of the diet” *(72)*. Because the GDQS app relies on a simplified quantification method, we concluded that a 10- to 20-minute completion time is realistic compared to the 20- to 30-minute completion time of a standard 24-hour recall *(44)*.

3.6. Global Dietary Recommendations (GDR) score

Detailed information related to the development and performance of the GDR score and the DQQ was found in the peer-reviewed article by Herforth et al. (2020) *(73)*. We found additional information in the Diet Quality Questionnaire Indicator Guide *(74)* and a report of the use of the DQQ in 41 countries of diverse contexts (HIC and LMIC) *(75)*, both published in 2022 by the developers of the indicators. Clarification and confirmation of some information items were discussed with the developers in a one-hour exchange meeting on 9 September 2022, as well as in email exchanges.

3.6.1. Rationale and purpose

The GDR score assesses adherence to a dietary pattern respecting WHO global dietary recommendations while the eight food group-based indicators reflect adherence to specific dietary recommendations and All-5 indicates minimal adherence to national and global dietary guidelines (see 3.6.3. for more details). These metrics and the DQQ were developed in tandem for diverse contexts (HIC and LMIC) by Gallup, Harvard Department of Global Health and Population, the Global Alliance for Improved Nutrition and other collaborators. According to the developers, “the GDR score is a measure of the adherence to 11 global dietary recommendations (mostly coming from the WHO Healthy Diet Fact Sheet 2018), which include dietary factors protective against non-communicable diseases” *(74)*; “the 8 food-group-based indicators reflects adherence to individual global recommendations for 1) fruits and vegetables, 2) dietary fiber, 3) free sugars, 4) saturated fat, 5) legumes, 6) nuts and seeds, 7) whole grains, and 8) processed meats” *(74)*; and “the All-5 is an indicator of the minimal adherence to dietary guidelines, defined as consumption of 5 food groups recommended across national and global dietary guidelines” *(74)*. All these indicators “are part of a suite of indicators designed to be highly feasible to collect and calculate and can be used to monitor adherence to dietary recommendations at the population level to provide information about the nature of diet quality in a population and change over time. This suite of indicators was developed simultaneously with a Diet Quality Questionnaire module designed for incorporation into multi-topic surveys such as The Gallup World Poll” *(73)*. We concluded that the purpose of the GDR score, the eight food group-based indicators and All-5, was to measure adherence at population level to all or part of the global dietary recommendations in diverse settings.

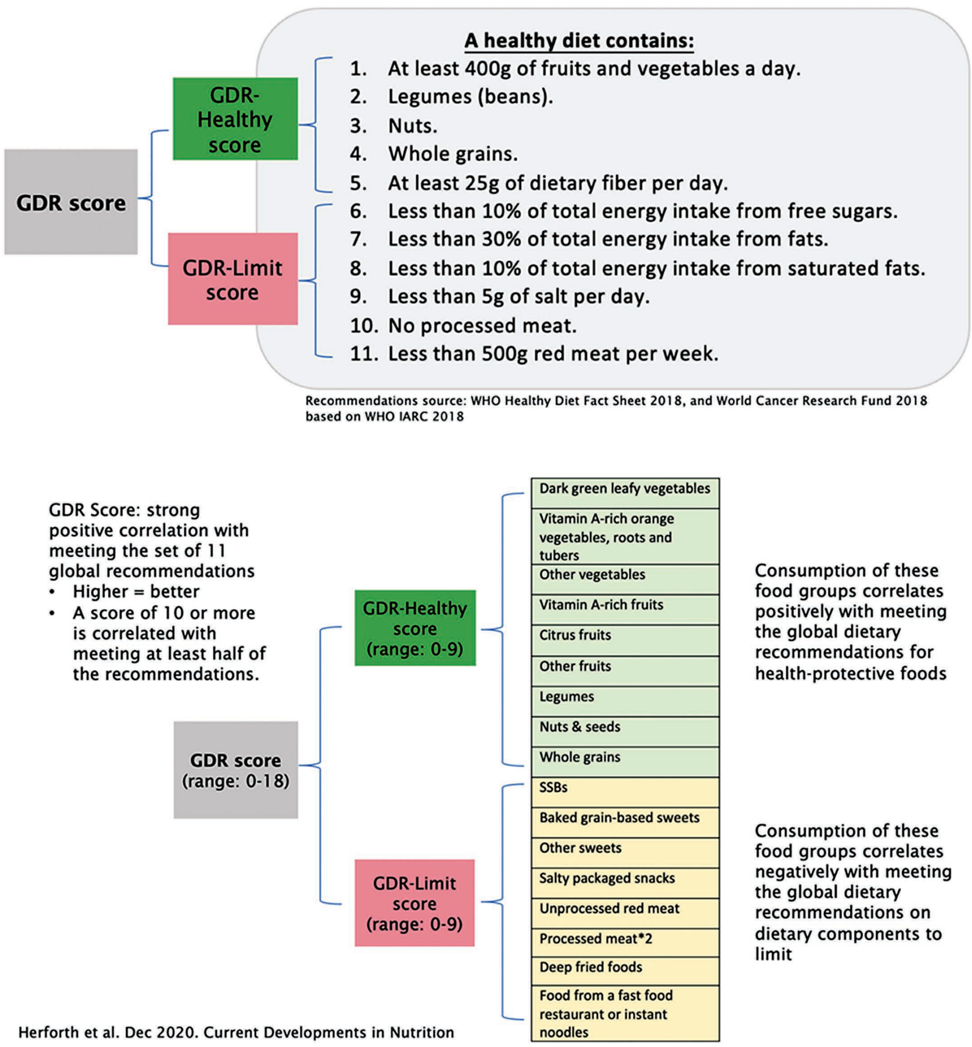
3.6.2. Uses

According to the developers, the intent of the GDR score, the eight food group-based indicators and All-5 "was to be highly feasible to collect and calculate and can be used to monitor adherence to dietary recommendations at the population level, to provide information about the nature of diet quality in a population and change over time" *(73)*. Because the GDR score, the eight food group-based indicators and All-5 were developed as a simple, rapid and feasible standardized metric for diverse context, we concluded that they could be used for identification, monitoring and surveillance, but also for policy and programme design, evaluation and research.

3.6.3. Scoring system

The GDR score is based on the consumption of 17 food groups during the past day and night (Fig. 1). The GDR score is composed of two submetrics (the GDR-Healthy score and the GDR-Limit score, recently renamed NCD-Protect and NCD-Risk) *(75)* and is calculated as follows: GDR-Healthy - GDR-Limit + 9 = GDR score. The GDR score ranges from 0 to 18 and the higher the GDR score, the more recommendations are likely to be met *(73, 74)*. The GDR-Healthy score is based on five global recommendations on nutritious foods for healthy diets (Fig. 1) and reflects adherence to global dietary recommendations on healthy components of the

Fig. 1. Scoring system of the Global Dietary Recommendations Score *(70)*



diet. The nine GDR-Healthy score food groups are summed into a score ranging from 0 to 9, beginning with a score of 0 and adding one point if any food in the group was consumed the previous day or night. A higher score indicates inclusion of more health-promoting foods in the diet *(73, 74)*. The GDR-Limit score is based on six global recommendations on dietary components to limit (see Fig. 1) and reflects adherence to global dietary recommendations on components of the diet to be limited or avoided. The eight GDR-Limit score food groups are summed into a score ranging from 0 to 9, beginning with a score of 0 and adding one point if any food in the group was consumed the previous day or night (except for processed meat where 2 points are added if it is consumed). A higher score indicates higher consumption of foods and drinks to be avoided or limited, and also indicates a higher dietary share of UPF *(73, 74)*.

The eight food group-based indicators were either assessed directly from food group consumption or assessed from multiple food groups *(74)*:

- Whole grain consumption: 1 point is given if whole grains were consumed the previous day or night, assessed by one item of the DDQ. The indicator is expressed as the percentage of the population consuming this food group.

- Pulse consumption: 1 point is given if pulses were consumed the previous day or night, assessed by one item of the DDQ. The indicator is expressed as the percentage of the population consuming this food group.

- Nuts and seeds consumption: 1 point is given if nuts and seeds were consumed the previous day or night, assessed by one item of the DDQ. The indicator is expressed as the percentage of the population consuming this food group.

- Processed meat consumption: 1 point is given if processed meat was consumed the previous day or night, assessed by one item of the DDQ. The indicator is expressed as the percentage of the population consuming this food group.

- WHO-fruits and vegetables score: six items of the DDQ are summed into a score ranging from 0 to 6, beginning with a score of 0 and adding 1 point if any food in the item was consumed the previous day or night. A score of 3 or more indicates the likelihood of consuming at least 400 g fruits and vegetables, which is a global dietary recommendation *(73)*. This indicator is expressed as the percentage of the population with a score of 3 or more. For this indicator, the developers mentioned that this cut-off was not globally validated (ongoing work) *(74)*.

- WHO-fibre score: nine items of the DDQ are summed into a score ranging from 0 to 10, beginning with a score of 0 and adding 1 point if any food in the item was consumed the previous day or night (except for pulses where 2 points are added when consumed). A score of 4 or more indicates the likelihood of consuming at least 25 g fibre, which is a global dietary recommendation *(73)*. This indicator is expressed as the percentage of the population with a score of 4 or more. For this indicator, the developers mentioned that this cut-off was not globally validated (ongoing work) *(74)*.

- WHO-sugar score: five food groups of the DDQ are summed into a score ranging from 0 to 6, beginning with a score of 0 and adding 1 point if any food in the group was consumed the previous day or night (except for soft drinks where 2 points are added when consumed). A score of 2 or more indicates the likelihood of exceeding 10% of dietary energy from free sugars, where limiting free sugar consumption to < 10% of dietary energy is a global dietary recommendation *(73)*. This indicator is expressed as the percentage of the population with a score of 2 or more. For this indicator, the developers mentioned that this cut-off was not globally validated (ongoing work) *(74)*.

- WHO-saturated fat score: eight food groups of the DDQ are summed into a score ranging from -2 to 6, beginning with a score of 0 and adding 1 point if any food in the groups other than sweets, cheese and yogurt, milk, processed meat, unprocessed red meat or fast food was consumed the previous day or night, and subtracting 1 point if any food in the groups poultry or fish and seafood was consumed the previous day or night. A score of 2 or more indicates the likelihood of exceeding 10% of dietary energy from saturated fat, where limiting saturated consumption to < 10% of dietary energy is a global dietary recommendation *(73)*. This indicator is expressed as the percentage of the population with a score of 2 or more. For this indicator, the developers mentioned that this cut-off was not globally validated (ongoing work) *(74)*.

- The All-5 is an indicator of minimal adherence to dietary guidelines, defined as consumption of five food groups recommended across national and global dietary guidelines. Five food groups of the DDQ are summed into a score ranging from 0 to 5, beginning with a score of 0 and adding 1 point if any food in the group was consumed the previous day or night. A score of less than 5 indicates that not all five recommended food groups were consumed (binary score: 1/0). All-5 is expressed as the percentage of the population consuming all five recommended food groups the previous day or night *(74)*.

3.6.4. Criteria for examination of the metric

In this subsection, we examined solely the GDR score because the eight food group-based indicators are metrics that summarized only a single nutrient or food (and are therefore outside the scope of this study), and the All-5 is a measure of the percentage of the population consuming all five food groups *(74)*.

3.6.4.1. Content

The developers mentioned that “the GDR score is a measure of the adherence to 11 global dietary recommendations (mostly coming from the WHO Healthy Diet Fact Sheet 2018), which include dietary factors protective against non-communicable diseases” *(74)*. We concluded that the GDR score assesses adherence to a dietary pattern where the consumption of 17 food groups is either promoted or limited and thus assesses the subconstruct “favourable dietary pattern”.

We assessed the content of the GDR score as holistic.

3.6.4.2. Cross-context equivalence

The GDR score has been developed and evaluated in two national contexts (Brazil and USA). Since it is based on secondary data analysis, the same construct was measured across contexts, with the content of each item used being interpreted in the same way: this means that the difference in scores between two individuals is the same across contexts as is the definition of zero. During our one-hour exchange meeting on 9 September 2022, the developers mentioned that “studies were underway for (1) validating the sentinel foods approach, (2) replicating the initial validation of the GDR score, (3) developing a combined total diet quality indicator, and (4) developing indicators related to environmental impact, in 10 countries: Bangladesh, Brazil, Cameroon, China, Colombia, Ethiopia, Mexico, Netherlands (Kingdom of the), Philippines, USA”.

We assessed the GDR score as scalar equivalent.

3.6.4.3. Validity for initial purpose

To validate the GDR score, the developers used the Healthy Diet Indicator 2020 (HDI-2020) as a standard of diet quality. According to the developers, "the HDI-2020 is an index of WHO global dietary recommendations for the prevention of chronic disease from 1990, 2003 and 2015, that was updated to include current WHO recommendations and other current global recommendations on red and processed meat based on the WHO International Agency for Research on Cancer. Each of the 11 dietary components of the HDI-2020 is equally weighted and expressed as a simple dichotomous score (0/1) for whether each dietary recommendation was met". They also created two HDI-2020 subindexes: one for meeting the recommendations for healthy dietary components and one for meeting the recommendations for dietary components to limit *(73, 74)*.

In analyses of cross-sectional, nationally representative quantitative dietary data from individuals aged ≥ 15 years old (including pregnant and/or lactating women) in the USA (pooled data from the NHANES cycles 2009–2010, 2011–2012 and 2013–2014) and in Brazil (National Dietary Survey 2008–2009), the developers found that "the correlation between the GDR score and the HDI-2020 is 0.55 in Brazil and 0.66 in the USA. The correlations of GDR-Healthy with the HDI-2020 sub-index on healthy foods and of GDR-Limit with the HDI-2020 sub-index on dietary components to limit are of a similar magnitude". They found that "the GDR score is negatively correlated with percentage of energy from UPFs (−0.40 in Brazil, −0.49 in the USA) and has a low negative association with energy intake (−0.15 in Brazil, −0.25 in the USA)" *(73)*.

The initial purpose of the GDR score is to assess adherence to 11 global dietary recommendations in diverse settings. We evaluated that the developers used an adequate a priori study design (cross-sectional designs with large sample sizes, adjusted analyses and receiver operating characteristic (ROC) analysis) to demonstrate that this metric indicates the subconstruct "favourable dietary pattern" (correlation with the HDI-2020) and the subconstruct "moderation" (correlation with percentage of energy from UPFs).

The validity for initial purpose of the GDR score was assessed as high.

Other elements of validation of the GDR score

Sensitivity to change, test-retest reliability and internal consistency have not been reported as being assessed. In analyses of cross-sectional, nationally representative, quantitative dietary data from children and adolescents aged 7–18 years old in China (2011 wave of the China Health and Nutrition Survey (CHNS)), a research group at Peking University explored the associations of the GDR score with overweight and obesity. They found that "the continuous GDR-Limit was positively associated with general obesity (OR = 1.43, 95% CI: 1.17–1.74) and abdominal obesity (OR = 1.22, 95% CI: 1.05–1.43), whereas the continuous overall GDR score was negatively associated with general obesity (OR = 0.85, 95% CI: 0.74–0.97)" *(76)*. Although this study is based on a cross-sectional design, and therefore does not establish relationships with health outcomes, we concluded that this study provides evidence to support the fact that the GDR score indicates the subconstruct "moderation".

During our one-hour exchange meeting on 9 September 2022, the developers mentioned that ”the GDR score was replicated in several other countries within the on-going 10-country study and appears to be a valid indicator of meeting WHO healthy diet recommendations”.

3.6.4.4. Population groups covered

The analyses presented for validation of the GDR score are for males and females aged ≥ 15 years (including the elderly) *(73)*. An additional study based on nationally representative quantitative dietary data in China provides evidence to suggest that the GDR score could be valid for children and adolescents aged 7–18 years old *(76)*. However, additional research is needed to explore how the GDR score could be adapted for this population group and for younger children (Table 10).

Table 10. Population groups and applicability of GDR score

Population group	Applicability[a]
Children (2–10 years)	Partly demonstrated (to be confirmed)
Adolescent males (11–17 years)	Partly demonstrated in 11–14 years old and demonstrated in 15–17 years old
Adolescent females (11–17 years)	Partly demonstrated in 11–14 years old and demonstrated in 15–17 years old
Adult males (18–59 years)	Demonstrated
Adult females (18–59 years)	Demonstrated
Pregnant and/or lactating women	Demonstrated
Elderly (≥ 60 years)	Demonstrated

[a] As reported by metric developers

3.6.4.5. Interpretation threshold

According to the developers, “a GDR score ≥10 is associated with meeting at least 6 out of the 11 global dietary recommendations” *(73)*.

3.6.4.6. Ease of computation

The ease of computation of the GDR score is facilitated by the fact that only data on whether a food group is consumed (or not) is required. The process of tabulating the GDR score using the DQQ involves only the following steps: assigning points per respondent and calculating the score *(74)*.

The ease of computation of the GDR score was assessed as high.

3.6.4.7. Ease of interpretation

The GDR score can be interpreted by applying a possible range of 0 to 18 (the higher the score, the more recommendations are likely to be met and the better the diet quality), and a population-based cut-off of 10 to report the percentage of the population more likely to meet at least 6 out of the 11 global dietary recommendations.

The ease of interpretation of the GDR score was assessed as high.

3.6.5. Criteria for examination of the data collection tools

According to the developers, "the DQQ can be used to derive the GDR score along with a suite of indicators. The DQQ gathers information on consumption (yes/no) of 29 food groups in the previous day" (Table 11). "Food groups are not asked about directly but are rather represented by sentinel foods that are the most frequently consumed items within a food group in a given population. While the DQQ was developed as a tool to rapidly assess diet quality, it does not gather information on all aspects of diet. The DQQ was developed to enable population-level diet quality monitoring and was not developed to assess dietary intake of individuals" *(77)*. "The DQQ has been adapted for over 100 countries" *(78)*.

Table 11. The 29 food groups of the Diet Quality Questionnaire

Plant foods	Animal-source foods	Foods to limit
Foods made from grains	Eggs	Baked sweets
Whole grains	Milk	Other sweets
White roots/tubers	Cheese	Sodas, energy drinks, sports drinks
	Yogurt	Fruit juice and fruit-flavoured drinks
Legumes	Processed meats	
Nuts and seeds	Unprocessed red meat (ruminants)	Sweet tea/coffee/cocoa
	Unprocessed red meat (non-ruminants)	Packaged ultra-processed salty snacks
Vitamin A-rich orange vegetables	Poultry	Instant noodles
Dark green leafy vegetables	Fish and seafood	Deep-fried foods
Other vegetables		Fast foods
Vitamin A-rich fruits		
Citrus fruits		
Other fruits		

Source: Global Diet Quality Project 2022 *(77)*

3.6.5.1. Criterion validity

In a study presented during the Nutrition Conference 2019, the developers designed a preliminary version of the DQQ based on 26 food groups using 24-hour nationally representative dietary intake data from Brazil (Individual Food Intake Survey 2008–2009) and the USA (NHANES 2009–2014). They also ran 82 cognitive interviews in five languages in São Paulo and New York City, in which they compared responses to closed-ended sentinel food questions to open-ended food group questions. The authors concluded that closed-ended questions using sentinel foods capture the vast majority of consumption and are better understood by respondents than open-ended, list-based methods for measuring dietary diversity and other aspects of diet related to NCD risk *(79)*.

In two analyses of cross-sectional, nationally representative, quantitative dietary data from individuals aged ≥ 15 years old and children and adolescents aged 7–18 years old in China (2011 wave of the CHNS), the developers aimed to identify the sentinel foods to adapt the DQQ for China. In both studies they found that "consumption of sentinel foods accounted for over 95% of people who consumed any food item in each food group at national levels" *(76, 80)*, and they further found in one study that "the sentinel foods the DQQ selected captured over 90% of people who consumed each food group in almost every province" *(80)*.

During our one-hour exchange meeting on 9 September 2022, the developers mentioned a study where "the DQQ was validated against a quantitative 24-hour dietary recall in Ethiopia, Solomon Islands and Viet Nam" and they found that "the two methods result in very similar population-based indicators and small differences in consumption of individual food groups" (manuscript under review).

Based on this body of evidence, we concluded that the DQQ is a valid tool to measure adherence of all or part of the global dietary recommendations in diverse settings.

3.6.5.2. Interviewer burden

The developers mentioned in a factsheet written for the 2021 TEAM Consultation "that the DQQ requires no specific interviewer expertise or special training, the main instruction being to read the questions exactly as written, without probing". Because the DQQ gathers simple information on consumption (yes/no) of 29 food groups, we concluded that this method does not require interviewers to have specific training or expertise in nutrition or food preparation in a given context.

3.6.5.3. Respondent burden

According to the developers, "cognitive testing showed that asking about sentinel foods was lower burden for respondents and more likely to be consistently understood, than asking respondents to self-characterize foods into named food groups, which sometimes resulted in misclassification" *(79)*. Because the DQQ gathers simple information on consumption (yes/no) of 29 food groups, we concluded that this method does not require specific effort from respondents and is not cognitively complex.

3.6.5.4. Collection time

According to the developers, "the DQQ takes approximately 5 minutes to implement" *(78)*. During our email exchange, the developers mentioned that this five-minute completion time was based "on roughly 50,000 observations from implementing the DQQ by Gallup". Because the DQQ gathers simple information on consumption (yes/no) of 29 food groups, we concluded that a five-minute completion time is realistic.

3.7. Minimum Dietary Diversity for Women (MDD-W)

Detailed information related to the development and performance of the MDD-W was found in a report published in 2015 by FAO *(81)* and the related peer-reviewed article by Martin-Prével et al. (2017) *(82)*. We found additional information related to the validity and use of the MDD-W in the 2016 and 2021 guides for measurement published by FAO *(83, 84)*. Additional information regarding ongoing work related to the MDD-W was received by email after presentation of this report at the Technical Expert Meeting on Harmonizing and Mainstreaming Measurements of Healthy Diets Globally held at the Rockefeller Foundation Bellagio Center, Italy, from 28 November to 2 December 2022.

3.7.1. Rationale and purpose

The MDD-W is a food group-based dietary diversity indicator for WRA that was developed for use in LMIC by the Women's Dietary Diversity Project (WDDP) Study Group, which is an international consortium. The MDD-W is based on the Women's Dietary Diversity Score *(85)*, a simple nine food group dietary diversity score resulting from the work led by the WDDP to examine the relationship between food group diversity and micronutrient adequacy in the diets of WRA, using data from nine datasets from six countries: Bangladesh (2x), Burkina Faso (2x), Mali, Mozambique, Philippines, Uganda (2x) with dietary intake data from multiple 24-hour recalls *(86)*. According to the developers, "the MDD-W was designed to be a food group diversity dichotomous indicator of whether or not NPNL and lactating WRA have consumed at least five out of ten defined food groups the previous day or night. The proportion of NPNL and lactating WRA who reach this minimum in a population can be used as a proxy indicator for higher micronutrient adequacy (summarized across 11 micronutrients: vitamin A, thiamine, riboflavin, niacin, vitamin B-6, folate, vitamin B-12, vitamin C, calcium, iron and zinc), one of the sub-constructs of diet quality which is deemed to be rather important in the context of LMICs and vulnerable people" *(83, 84)*. We concluded that the purpose of the MDD-W was to measure at population level a minimal level of dietary diversity that can be used as a proxy indicator for higher micronutrient adequacy in LMICs contexts.

3.7.2. Uses

According to the developers, appropriate uses of the MDD-W are: "as a proxy to describe micronutrient adequacy at the population level, description of the proportion of women consuming each food group, tracking changes in diets of groups of women over time, comparing diets of groups of women in different settings, sociodemographic and other characteristics of interest, integration in national and subnational assessments, and evaluating impact of interventions at scale, programmes and policies" *(83, 84)*. Because the MDD-W was developed as a simple, rapid and feasible standardized metric for diverse context, we concluded that it could be used for identification, monitoring and surveillance, but also for policy and programme design, evaluation and research.

3.7.3. Scoring system

The MDD-W is composed of 10 mutually exclusive food groups (no food or ingredient is placed in more than one food group) that have been determined to bear a stronger relationship to micronutrient adequacy than other candidate indicators with different groupings (Table 12) *(81)*. The 10 MDD-W food groups are summed into a score ranging from 0 to 10, beginning with a score of 0 and adding 1 point if any food in the group was consumed the previous day or night. The MDD-W is then tabulated, coding yes or no for each woman scoring ≥ 5, and then calculating the proportion of women scoring ≥ 5 *(83, 84)*.

Table 12. Assigning questionnaire rows to the 10 MDD-W food groups

Questionnaire rows	Food groups
A. Foods made from grains **B.** White roots and tubers or plantains	**1.** Grains, white roots and tubers, plantains
C. Pulses (beans, peas and lentils) **D.** Nuts and seeds **E.** Milk **F.** Milk products	**2.** Pulses (beans, peas and lentils) **3.** Nuts and seeds **4.** Milk and milk products
G. Organ meats **H.** Red flesh mammal's meat **I.** Processed meat **J.** Poultry and other white meats **K.** Fish and seafood	**5.** Meat, poultry and fish
L. Eggs	**6.** Eggs
M. Dark green leafy vegetables	**7.** Dark green leafy vegetables
N. Vitamin A-rich vegetables, roots and tubers **O.** Vitamin A-rich fruits	**8.** Other Vitamin A-rich fruits and vegetables
P. Other vegetables	**9.** Other vegetables
Q. Other fruits	**10.** Other fruits

Source: FAO, 2021 *(83)*.

3.7.4. Criteria for examination of the metric

3.7.4.1. Content

The developers mentioned that "the MDD-W is a simple food-group diversity indicator" *(82)*. We concluded that the MDD-W assesses the subconstruct "diversity".

We assessed the content of the MDD-W as specific.

3.7.4.2. Cross-context equivalence

The MDD-W was developed and evaluated in nine subnational contexts (two rural sites in Bangladesh, urban and rural Burkina Faso, urban Mali, rural Mozambique, peri-urban Philippines, and rural and urban Uganda). Since it is based on secondary data analysis, the same construct was measured across contexts with the content of each item used being interpreted in the same way: this means that the difference in scores between two individuals is the same across contexts as is the definition of zero.

We assessed the MDD-W as scalar equivalent.

3.7.4.3. Validity for initial purpose

In analyses of cross-sectional data from NPNL and lactating WRA in nine subnational contexts (two rural sites in Bangladesh, urban and rural Burkina Faso, urban Mali, rural Mozambique, peri-urban Philippines, and rural and urban Uganda), the developers considered "that the proportion of NPNL and lactating WRA who reach 5 out of 10 food group in a population can be used as a proxy indicator for higher micronutrient adequacy (summarized across 11 micronutrients: vitamin A, thiamine, riboflavin, niacin, vitamin B-6, folate, vitamin B-12, vitamin C, calcium, iron and zinc)". They observed "that the continuous

MDD-W score was significantly correlated to mean probability of adequacy (MPA) at each site. In models that controlled for energy intake, correlations were attenuated in most sites but remained significant. For lactating women, area under the curve (AUC) ranged from 0.56 to 0.90. In sensitivity and specificity analyses, they evaluated indicator performance in detecting MPAs > 0.60 with the food-group cutoff of ≥ 5 food groups, because this cutoff provided the best balance between sensitivity and specificity across sites. Sensitivity and specificity results also showed moderate performance, as well as positive predictive values" *(82)*. The initial purpose of the MDD-W is to assess a minimal level of dietary diversity that can be used as a proxy indicator for higher micronutrient adequacy in LMIC context. We concluded that the developers used an adequate a priori study design (cross-sectional designs with large sample sizes, adjusted analyses and ROC analysis) to demonstrate that this metric indicates the subconstruct "nutrient adequacy" in nine diverse contexts (systematic positive correlation with mean of probability of adequacy).

The validity for initial purpose of the MDD-W was assessed as high.

Other elements of validation of the MDD-W

Sensitivity to change of the MDD-W has been evaluated and demonstrated through analyses of longitudinal cohort data in Burkina Faso, Malawi and Zambia *(87, 88)*. Test-retest reliability and internal consistency have not been reported as being assessed. During our email exchange, FAO project officer confirmed the following "FAO's Food and Nutrition Division, with support from the Federal Ministry for Economic Cooperation and Development (BMZ) through GIZ – Knowledge for nutrition (K4N), has embarked on a project (2021-2023) to advance and expand uptake of the MDD-W. As part of this project results of MDD-W will be compared to other aspects of dietary quality to glean new insights into the association between minimum dietary diversity and intakes of fat, sugar, calories, amongst other, utilizing micro-data from individual food consumption surveys available in the FAO/WHO GIFT platform. Additionally, exploratory analysis to further study the potential misclassification errors related to the 15 grams minimum intake threshold will be undertaken."

3.7.4.4. Population groups covered

All analyses presented for validation of the MDD-W are for NPNL and lactating WRA. While one study demonstrated that a simple 10-food group dietary diversity score using the same food group classification as the MDD-W was consistently associated with micronutrient adequacy for all age and sex groups (two years and older) in China and Mexico *(89)*, additional research is needed. During our email exchange, the FAO project officer mentioned further elements of the ongoing project that involve "on-going work to test the validity of minimum dietary diversity as a proxy indicator for micronutrient adequacy in populations other than women of reproductive age (e.g. adolescents) utilizing micro-data from individual food consumption surveys available in the FAO/WHO GIFT platform". The table above summarizes the current status of population groups for which minimum dietary diversity is applicable (Table 13).

Table 13. Population groups and applicability of MDD-W

Population group	Applicability[a]
Children (2–10 years)	Partly demonstrated (to be confirmed)
Adolescent males (11–17 years)	Partly demonstrated (to be confirmed). Research ongoing and results expected in 2023.
Adolescent females (11–17 years)	Demonstrated only in 15–17 years old. Research ongoing and results expected in 2023.
Adult males (18–59 years)	Partly demonstrated (to be confirmed)
Adult females (18–59 years)	Demonstrated only in 18–49 years old
Pregnant and/or lactating women	Demonstrated in lactating women
Elderly (≥ 60 years)	Partly demonstrated (to be confirmed)

[a] As reported by metric developers

3.7.4.5. Interpretation threshold

The authors have demonstrated that the MDD-W is a dichotomous indicator of whether or not NPNL and lactating WRA have consumed at least five out of 10 defined food groups the previous day or night and can be used as a proxy indicator for higher micronutrient adequacy *(82)*.

3.7.4.6. Ease of computation

The ease of computation of the MDD-W varies according to data collection methods used. According to the developers, "in both cases, the ease of computation of the MDD-W is facilitated by the fact that only data on whether a food group is consumed (or not) is required. The process of tabulating MDD-W data with data collected using the open recall method requires more steps (including a step where the interviewer has to assign food and ingredient to the appropriate food group) than tabulating data collected with the list-based method" *(83)*.

The ease of computation of the MDD-W was assessed as medium (when using open recall method) and high (when using the list-based method or the DQQ).

3.7.4.7. Ease of interpretation

According to the developers, the basic interpretation of the indicator is "X% of women achieved minimum dietary diversity, and they are more likely to have higher (more adequate) micronutrient intakes than the X% of women who did not" *(83)*.

The ease of interpretation of the MDD-W was assessed as high.

3.7.5. Criteria for examination of the data collection tools

The data collection method used to derive data for the MDD-W is a qualitative 24-hour dietary recall of food groups. According to the developers, "two approaches to administer such a questionnaire can be used: the open recall and the list-based. Both approaches require preparatory work to identify the most frequently consumed local/national foods and common dishes, and to classify each food into the correct food groups. Both methods require the respondent to report what was eaten during the previous 24-hours before the interview. Each, however, entails a specific interview protocol and adapted questionnaire. The open recall approach follows the principles of the multiple pass method that is recommended for quantitative 24h recalls but without the last pass (i.e., quantification). For the list-based method, the interviewer solicits yes or no answers to questions from the respondents regarding food groups with a limited number of food items (sentinel), concerning respondent's food consumption over the previous day or night" *(83)*. The DQQ is a list-based method for collecting the MDD-W without the 15 g minimum threshold criterion *(90)*. It is also possible to use quantitative 24-hour recall to derive data for the MDD-W. Nevertheless, the next subsections concern only list-based and open-recall methods and not all possible data collection methods, which may influence several of the assessment criteria noted here.

3.7.5.1. Criterion validity

Agreement of open recall and list-based methods with weighed food records (WFR) for achieving MDD-W was assessed using data from NPNL and lactating WRA in Cambodia, Ethiopia and Zambia. For the pooled sample, proportions achieving MDD-W from both methods were compared to the WFR proportion. According to the developers, "significant differences were found in proportions achieving MDD-W between both methods and WFR, open recall and list-based methods over reporting women achieving MDD-W by 10 and 16 % points, respectively" *(91)*. While both list-based and open recall methods may overestimate dietary diversity, these methods were considered by the authors to remain useful as long as users understand the reasons leading to overestimations *(83)*. We concluded that both list-based and open recall methods are valid tools to measure the MDD-W in LMIC.

3.7.5.2. Interviewer burden

According to the developers, "when comparing the two methods, the open recall method is intuitive for interviewers, but requires longer training time and requires interviewers to have a good understanding of the questionnaire's objectives and a reasonable knowledge of how food is acquired, prepared and consumed in the survey areas, which is time and resource intensive. On the opposite side, the list-based method requires shorter interviewer training time and requires fewer requirements for capacity of interviewers" *(83)*. Because the list-based method gathers simple information on consumption (yes/no) of food groups, we evaluated that this method does not require interviewers to have specific training or expertise in nutrition or food preparation in a given context. On the other hand, we concluded that the open recall method requires a short period of specific training and some expertise in nutrition or food preparation in a given context.

3.7.5.3. Respondent burden

According to the developers, "when comparing the two methods, the open recall method is intuitive for respondents while the list-based method requires respondents to mentally take apart mixed dishes and to mentally move back and forth in time as foods are mentioned" *(83)*. Because the list-based method gathers simple information on consumption (yes/no) of food groups, we evaluated that this method does not require specific effort from respondents and is not cognitively complex. We concluded that the open recall method does not require specific effort from respondents and is not cognitively complex.

3.7.5.4. Collection time

According to the developers, "interviews using the list-based method are reported to last fewer than 6 minutes while those using the open recall method are longer" (time not reported) *(83)*. Because the list-based method gathers simple information on consumption (yes/no) of food groups, we concluded that a six-minute completion time is realistic. According to the DAPA Measurement Toolkit, interviews using a single 24-hour recall method (similar to the open recall method but including quantification which adds significant time to the recall) are reported to last from 20 to 30 minutes depending on the respondent's diet *(44)*. Because the open recall method is non-quantitative, we can assume that interviews last no longer than 20 minutes and are likely to last around 10-15 minutes.

3.8.
Nova UPF score

Detailed information related to the development and performance of the Nova UPF score was found in the article by Costa et al. (2021) *(92)*. We found additional information related to the validity and use of the Nova UPF score in the factsheet written for the 2021 TEAM Consultation. Requests for clarification and confirmation of some items of information were sent to the developers by email and responses received.

3.8.1. Rationale and purpose

The Nova UPF score is a measure of the consumption of UPF that was developed in the Brazilian context by a research team led by Carlos Monteiro, based at the University of Sao Paulo. The concept of UPF, and much of the recommendation to avoid them or at least minimize their consumption, was developed by the same team *(93)*. Numerous recent studies have demonstrated that the dietary share of UPF is associated with the overall quality of contemporaneous diets and the risk of most diet-related NCDs. A meta-analysis of nationally representative dietary surveys in 13 LMICs found that increased UPF consumption negatively affects the overall nutritional quality of diets *(94)*. Systematic reviews and meta-analyses found that an increased UPF consumption was associated with higher risk of obesity, diabetes, CVD, cerebrovascular disease, depression and all-cause mortality *(95–97)*. According to the developers, "the Nova score was designed to measure the previous day's consumption of 23 subgroups of UPF commonly consumed in Brazil in adults and reflect the dietary share of UPF. The Nova score is based on a previous instrument addressing questions on the previous-day dietary intake of a list of 13 subgroups of UPF (answered with "yes" or "no") developed by the same authors for the Surveillance System for Risk and Protective Factors for Chronic Diseases by Telephone Survey in Brazil" *(92)*. We concluded that the purpose of the Nova UPF score was to measure at population level the consumption of UPF in order to reflect the dietary share of UPF in the Brazilian context.

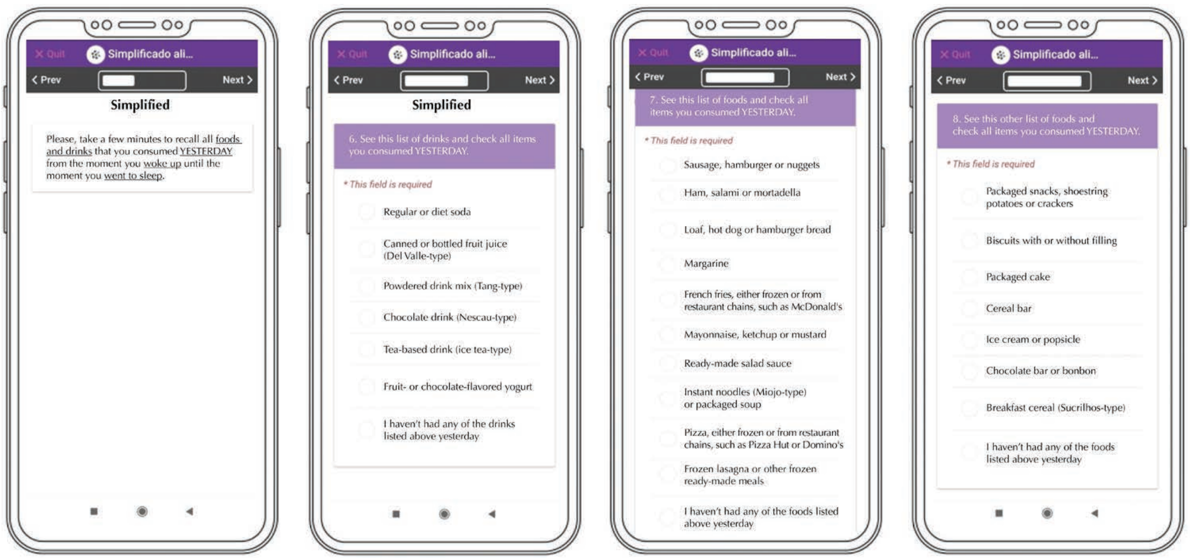
3.8.2. Uses

According to the developers, the Nova UPF score was developed "for monitoring the consumption of UPF by the Brazilian adult population" and other populations across the world (ongoing work) *(92)*. Because the Nova UPF score was developed as a simple, rapid and feasible standardized metric for Brazil but also other countries, we concluded that it could be used for identification, monitoring and surveillance, but also for policy and programme design, evaluation and research.

3.8.3. Scoring system

The Nova UPF score is composed of 23 UPF subgroups: six subgroups of beverages, 10 subgroups of products that replace or accompany meals, and seven subgroups of products often consumed as snacks (Fig. 2). The 23 UPF subgroups are summed into a score ranging from 0 to 23, beginning with a score of 0 and adding one point if any food in the group was consumed the previous day *(92)*.

Fig. 2. Scoring system of the Nova UPF score *(92)*



3.8.4. Criteria for examination of the metric

3.8.4.1. Content

The authors mentioned that "the Nova score was designed to reflect the dietary share of UPF in Brazil" *(92)*. We concluded that the Nova UPF score assesses the consumption of UPF, a group of foods that should be consumed in moderation, and thus assesses the subconstruct "moderation".

We assessed the content of the Nova UPF score as specific.

3.8.4.2. Cross-context equivalence

To date, the Nova UPF score has been evaluated in one subnational context only (University of São Paulo, Brazil). However, other adaptation and validation studies of the NOVA screener in Ecuador, India and Senegal are ongoing *(98)*. During our email exchange, the developers mentioned that "final versions of the adapted tools for Senegal, India and Ecuador included 23, 24 and 27 UPF subgroups, respectively. Although the number of subgroups varied across countries, the final versions of the screeners are very similar and differences in the final scores for each context can be dealt with during statistical analysis (e.g. comparing quintiles of the Nova score or Z-scores across countries)". The developers also mentioned that "the Nova screener is being adapted and validated for the Pacific Islands, Portugal, and Slovakia". Because the Nova UPF score will require specific statistical analysis to make it comparable across countries, we assessed this metric as measurement equivalent.

We assessed the NOVA score as measurement equivalent.

3.8.4.3. Validity for initial purpose

In an analysis of cross-sectional data from 300 men and women aged 18 years or older in Brazil (users of two health centres at the University of São Paulo serving employees and students of the University of São Paulo), the developers considered "that the Nova score for the consumption of UPF (obtained with an electronic self-report questionnaire) shows a good potential in reflecting the dietary share of UPF (obtained with a 24-hour dietary recall) in Brazil". They observed "that the average percentage of dietary share of UPF, calculated based on the 24-hour dietary recall, increases linearly and significantly with the increase in the Nova UPF score for the consumption for UPF". They also observed "a substantial agreement between the classification based on the fifths of the dietary share of UPF and of the Nova score (Pabak index of 0.67)". Finally, they showed "that the prevalence of relatively high consumption of UPF (defined as a Nova UPF score ≥ 5 and UPF participation in the total caloric intake ≥ 49.6%) linearly decreases with increasing age" *(92)*. During our email exchange, the developers mentioned that "high agreements between dietary share of UPF and of the Nova UPF score (Pabak > 0.70) were also observed in the validation studies of the UPF score in Senegal, India and Ecuador" (manuscript under preparation). They also mentioned that "a prospective dose-response association between the Nova UPF score and 15-month weight gain was found in more than 10,000 participants of the NutriNet Brasil cohort study" (manuscript under preparation).

The initial purpose of the Nova UPF score is to reflect the dietary share of UPF in Brazil. We concluded that the developers used an adequate a priori study design, while relying on a modest sample size, to demonstrate that this metric indicates the subconstruct "moderation" in this context. Additional studies (manuscripts under preparation) could confirm that this metric indicates the subconstruct "moderation" in Brazil (using longitudinal data) but also in LMIC contexts.

The validity for the initial purpose of the Nova UPF score was assessed as high.

Other elements of validation of the Nova UPF score

Sensitivity to change, test-retest reliability and internal consistency have not been reported as being assessed.

3.8.4.4. Population groups covered

The analyses presented for validation of the Nova UPF score are for men and women aged ≥ 18 years (including the elderly) *(92)*.During our email exchange, the developers mentioned that "two research groups are currently adapting and validating the Nova-UPF screener with a focus on children in Brazil" (Table 14).

Table 14. Population groups and applicability of Nova UPF score

Population group	Applicability[a]
Children (2–10 years)	Ongoing research
Adolescent males (11–17 years)	Possible but requires additional research
Adolescent females (11–17 years)	Possible but requires additional research
Adult males (18–59 years)	Demonstrated
Adult females (18–59 years)	Demonstrated
Pregnant and/or lactating women	Possible but requires additional research
Elderly (≥ 60 years)	Demonstrated

[a] As reported by metric developers

3.8.4.5. Interpretation threshold

The developers mentioned in a factsheet written for the 2021 TEAM Consultation that "the Nova UPF score in one specific population can be expressed as an average or as the percentage of relatively high scores, such as equal or higher than 5". During our email exchange, the developers mentioned that "preliminary findings from Brazil and Ecuador show that the 20% highest UPF consumers in each of these countries obtained a Nova score of 5 or above". While more work would be needed to confirm the relevance of this threshold, the authors considered that "a threshold of 5+ could be interpreted as high UPF in LMIC" and they "will be in a better position to share further information once we run final analysis, including data from India and Senegal". Like the developers, we concluded that additional research is needed to confirm whether a cut-off of 5 is the best predictor for identifying a high or low dietary share of UPF in diverse settings.

3.8.4.6. Ease of computation

The process of computing the Nova UPF score appears to be simple when using the NOVA screener, and involves only the following steps: assigning points per respondent and calculating the score.

The ease of computation of the Nova UPF score was assessed as high.

3.8.4.7. Ease of interpretation

The Nova UPF score can be interpreted by applying a possible range of 0 to 23 (the higher the score the higher the share of UPF) and a potential cut-off of 5 to report the percentage of the population with a high dietary share of UPF (to be confirmed).

The ease of interpretation of the Nova UPF score was assessed as low (but potentially high if the cut-off is confirmed by further research).

3.8.5. Criteria for examination of the data collection tools

According to the developers, "the Nova screener provides the Nova score for the consumption of UPF. It was developed to include UPF subgroups with greater participation in the daily energy intake, estimated by the national food consumption survey conducted in 2008–2009 Household Budget Survey of the Brazilian Institute of Geography and Statistics. The questions addressing the intake of each of the 23 subgroups of UPF are presented on 3 categories: beverages (6 subgroups); products that replace or accompany meals (10 subgroups); and products often consumed as snacks (7 subgroups). This questionnaire can be uploaded into mobile phones, tablets, or computers with the Epicollect5 Data Collection® software, which stores participants' answers as a database" *(92)*.

3.8.5.1. Criterion validity

Based on the information provided in section 3.8.4.3., we concluded that the NOVA screener is a valid tool to measure UPF consumption in order to reflect the dietary share of UPF in Brazil.

3.8.5.2. Interviewer burden

Because the NOVA screener gathers simple information on consumption (yes/no) of 23 subgroups of UPF, we evaluated that this method does not require interviewers to have specific training or expertise in nutrition or food preparation in a given context.

3.8.5.3. Respondent burden

Because the NOVA screener gathers simple information on consumption (yes/no) of 23 subgroups of UPF, we concluded that this method does not require specific effort from respondents and is not cognitively complex.

3.8.5.4. Collection time

According to the developers, "interviews using the Nova screener are reported to last 3 minutes" *(92)*. Because the NOVA screener gathers simple information on consumption (yes/no) of 23 subgroups of UPF, we concluded that a three-minute completion time is realistic.

4. Evaluation of healthy diet metrics

This section presents a critical and comparative assessment of the selected healthy diet measures for use as global and national monitoring indicators. More precisely, we assessed whether the selected healthy diet metrics were valid for identification, as well as for surveillance and monitoring – uses which usually require standardized approaches that are simple, rapid and feasible, and allow for comparability over time and across contexts, thereby providing a population-level snapshot of the healthfulness of diets. We therefore considered the following elements as requisites for valid metrics to be used in identification, as well as in surveillance and monitoring.

4.1. Coverage of the healthy diet construct

This element takes up our assessment of the content of the metric based on the number of subconstructs of the construct of healthy diets assessed by the metric and defined by three levels (specific, large and holistic). Our hypothesis is that the more subconstructs of the construct of healthy diets are assessed the better (other than for the subconstruct "favourable dietary pattern"). We evaluated the coverage of the healthy diet concept by three levels (low, medium, high).

The NOVA score and the MDD-W were rated low because they assess specific subconstructs ("moderation" and "diversity", respectively). The DQI-I was rated medium because it assesses multiple subconstructs. The AHEI-2010, DASH score, GDQS and GDR score were rated high because they assess the subconstruct "favourable dietary pattern".

4.2. Cross-context equivalence

This element takes up our assessment of the cross-context equivalence defined by four types of equivalence: construct, item, measurement and scalar. Our hypothesis is that scalar equivalence is the most desirable for global monitoring indicators (high) while measurement equivalence is less desirable (medium).

The AHEI-2010, DASH score and Nova UPF score were rated medium while the DQI-I, GDQS, GDR score and MDD-W were rated high.

4.3. Ease of computation

This element takes up our assessment of the ease of computation based on the number of steps needed to compute the metric, and was defined by four levels (very low, low, medium and high). Our hypothesis is that the higher the ease of computation the better. We used the same four levels.

The AHEI-2010, DASH score and DQI-I were rated very low while the GDQS, GDR score, Nova UPF score and MDD-W were rated high. Depending on the collections tool used, the GDQS could also be rated low and MDD-W medium.

4.4. Ease of interpretation

This element takes up our assessment of the ease of interpretation based on the ability to categorize the monitored population as having (or not) a healthy diet and was defined by two levels (low and high). Our hypothesis is that the higher the ease of interpretation the better. We used the same two levels.

The AHEI-2010, DASH score, DQI-I and NOVA score were rated low while the GDQS, GDR score and MDD-W were rated high. Depending on confirmation by ongoing research about the cut-off of 5 to report the percentage of the population with a high dietary share of UPF, the Nova UPF score could be rated high.

4.5. Ease of collection

This element takes up our assessment of the three criteria for examination of the collection tool (interviewer burden, respondent burden and collection time).

The AHEI-2010, DASH score and DQI-I were rated low because they rely on standard tools such as quantitative 24-hour dietary recalls and FFQ which are reported to last from 20 to 60 minutes and associated either with a high interviewer burden (quantitative 24-hour dietary recalls) or high respondent burden (FFQ).

The GDQS was rated medium because it relies on a dedicated application (GDQS app) – a semi-quantitative 24-hour dietary recall with a simplified quantification method – that reduces interview time (now lasting from 10 to 20 minutes) and interviewer burden compared to classic quantitative 24-hour dietary recall. Nevertheless, interviewers need to undergo a short training programme on the use of the GDQS app. The GDQS app would require a priori less specific effort from respondents than classic quantitative 24-hour dietary recall and would be a priori less cognitively complex. Field research is under way to confirm or not these characteristics.

The GDR score, MDD-W and Nova UPF score were rated high because they rely on a dedicated questionnaire that gathers simple information on consumption (yes/no) of a limited number of food groups: this allows for a short interview time (from 3 to 6 minutes) and does not require interviewers to have specific training. These questionnaires also do not require specific effort from respondents and are not cognitively complex.

4.6. Fitness for purpose of healthy diet metrics for global and national monitoring

The result of our critical and comparative assessment of selected healthy diet measures for use as global and national monitoring indicators is summarized in Table 15. From this evaluation, there are four healthy diet metrics that seem currently to be the most suitable for global and national monitoring: the GDR score, GDQS, MDD-W and Nova UPF score.

However, these metrics have some weaknesses that should be highlighted. The main current weakness of the GDR score lies in the fact that the metric has only been validated in two countries (Brazil and USA) and that other validity studies in LMICs should be

Table 15. Summary of the assessment of selected healthy diet measures for use as global and national monitoring indicators

	AHEI-2010	DASH	DQI-I	GDQS	GDR score	MDD-W	Nova UPF score
Coverage of the healthy diet construct	High	High	Medium	High	High	Low	Low
Cross-context equivalence	Medium	Medium	High	High	High	High	Medium
Ease of computation	Very low	Very low	Very low	High[a]	High	High[b]	High
Ease of interpretation	Low	Low	Low	High	High	High	Low (High)[c]
Ease of collection	Low	Low	Low	Medium[c]	High	High	High

[a] High when the GDQS app is used to collect data, low when using quantitative 24-hour dietary recall or FFQ;
[b] High when using the list-based method or the DQQ to collect data, medium when using open recall method;
[c] To be confirmed in ongoing research.

considered. However, its developers have planned to carry out validation studies in 10 countries in very diverse contexts: Bangladesh, Brazil, Cameroon, China, Colombia, Ethiopia, Mexico, Netherlands (Kingdom of the), Philippines and USA.

The main current weaknesses of the GDQS are its relative length and complexity in collecting quantitative dietary information, and the lack of available data on its use on the field. However, its developers have planned to carry out primary data collection with the GDQS app and to validate GDQS data collected with the app with data collected using weighed records.

The main weakness of the MDD-W and Nova UPF score is that each indicate only one of the subconstructs of the construct of healthy diets.

5. Limitations

Some limitations of our assessment of the validity, usefulness and fitness for purpose of existing healthy diet metrics for use as global and national monitoring indicators ought to be mentioned.

Because this work was not intended to be exhaustive, and because our selection was restricted by our inclusion criteria and the time dedicated to this report, we may have missed some relevant healthy diet metrics. However, our selection of indicators was based on those already included in the May 2021 technical consultation on measuring healthy diets *(2)*, which covered the most advanced and relevant healthy diet metrics for use as global and national monitoring indicators.

Another limitation of this work concerns the choice of criteria for examination of the metric, criteria for examination of the data collection tool and criteria for evaluating the appropriateness of the healthy eating metric for global and national monitoring. The inclusion of other criteria, as well as the choice of another evaluation system for each of the selected criteria, could have resulted in a different assessment. Nevertheless, we believe that this risk was limited since our evaluation system is based on a proven framework *(29-31)*. Furthermore, the results obtained from our benchmarking are substantially similar to those proposed by the developers of the GDQS *(72)*.

This report primarily focused on a suitability assessment of healthy diet metrics for use as global monitoring indicators: it was assumed that global monitoring indicators can be easily adopted as national monitoring indicators. However, national surveillance may require indicators that provide a more sophisticated understanding of the dietary health of various subgroups within a country. Thus, although our assessment may have some validity for evaluating the adequacy of healthy eating measures as national surveillance indicators, it may be necessary to reconsider indicator selection by adding those relevant at the national level and adapting the assessment criteria accordingly.

6. Knowledge gaps and priority needs

6.1. Knowledge gaps

- All the selected metrics and their threshold (where present) are relevant only to WRA or the adult and elderly population. It remains largely unknown whether they can be used for other population groups such as pregnant or lactating women, adolescents and children. In addition, their validation occurred in a limited number of contexts (except for the GDQS and MDD-W).
- Very few of the developers of the selected metrics reported having evaluated sensitivity to change (AHEI-2010, GDQS and MDD-W) or internal consistency (DQI-I), and none reported test-retest reliability. The lack of testing for sensitivity to change and test-retest reliability might, for instance, question their suitability for tracking population-level changes over time and measuring the impact of any kind of policy or intervention.
- Validation of dietary measures has most often has been done by estimating solely a correlation coefficient between criterion and test measures, whereas regression methods (often bivariate regression) are preferable for quantifying validity *(6)*. By regressing a criterion measure on a test measure, the absolute error represented by the standard deviation of the residuals (i.e. the root mean square error) quantifies in meaningful units how close the test measure on average predicts the criterion measure, thereby allowing assessment of whether the paired measures are sufficiently and expediently close for the intended purpose(s). Furthermore, correlation coefficients (which estimate relative and not absolute error) are confounded by the underlying variation in the sample, and therefore cannot be easily applied to other samples in which the underlying variation may differ. Regression has the advantage of producing a slope and intercept. If the units of the criterion and test measures are the same, then a test measure that is valid (i.e. accurate and reliable) may result in an intercept of 0 and a slope of 1. Even if the units are not the same, we would expect a valid test measure to have points falling close to the predicted straight line. A standard residual plot (which is analogous to a Bland-Altman plot but instead uses the predicted values as the x-axis) can reveal aspects of this relationship. Regression analysis is also preferred because the slope and intercept can be used to empirically investigate cross-context equivalence. For measurement equivalence, we would expect the slopes from different contexts (e.g. countries) to be about the same. For scalar equivalence, we would expect both the slopes and intercepts from different contexts to be about the same. Cross-context equivalence of each metric was rated in this report primarily depending on how the metric was constructed and whether it has been used in multiple countries rather than through empirical demonstration of equivalence. Methods commonly used with questionnaires to establish cross-context equivalence by estimating item loadings (i.e. multigroup confirmatory

factor analysis, differential item functioning or alignment) may or may not be applicable to healthy diet metrics *(99)*.

- None of the selected metrics assesses nor indicates the subconstruct "food safety", whether specifically or through a more holistic subconstruct such as "favourable dietary pattern". Furthermore, and to the best of our knowledge, no diet quality metric is able to assess nor indicate the subconstruct "food safety". Indicators of dietary microbiological and chemical safety may be better defined and monitored in a framework other than that applicable to healthy diet indicators.
- It is very difficult to find detailed or even simple estimates of the costs associated with the local adaptation, collection and computation of the indicators.
- We have not found evidence of the trade-offs implied when standardizing collection tools in different contexts. In other words, what information do we lose when using a standardized tool and what does this mean for the metric? Nevertheless, it should be noted that most standardized tools have been contextualized before being administered.
- The limited number of countries/contexts where metrics have been either studied or tested.

6.2. Priority needs

- **Explore the relative validity, reliability and sensitivity of the GDR score and GDQS.** The GDR score and GDQS were identified as the most appropriate indicators for global monitoring and for some national monitoring purposes. Their main difference lies however in the way food group consumption is collected: the GDR score requires simple information on consumption (yes/no) while GDQS requires quantification of consumption. We therefore recommend assessing the relative validity of these two indicators to understand the extent to which quantifying consumption improves metric performance.
- **Explore the suitability of the metrics for other population groups.** The suitability of the selected metrics for other population groups such as pregnant or lactating women, adolescents and children, and in specific contexts, needs to be further explored. Our discussions with the developers of the GDQS, GDR score, MDD-W and Nova UPF score revealed that this work is either in progress or planned.
- **Demonstrate cross-context equivalence empirically.** The ability of the selected metrics to provide an assessment that is cross-context equivalent (e.g. across countries or population groups) needs to be established empirically.
- **Improve collection tools.** Although some data collection tools have been successful in reducing interviewer and respondent burden, there is a need to further improve them to facilitate diet quantification and allow for a more accurate assessment of the healthfulness of diets.

- **Document costs of data collection.** We recommend that the developers of healthy diet metrics document more fully the total cost of the different data collection tools in different contexts, including costs associated with local adaptation of the tools.
- **Explore new metrics.** Although we identified healthy diet metrics that are suitable for global and national monitoring, we believe that we should keep building consensus on healthy metrics that were not included in this report (see p. 7). For example, developing a consensus on an indicator based on the EAT-*Lancet* Planetary Health Diet would make it possible to integrate the environmental sustainability and healthy diet dimensions. Similarly, consensus should also be sought around an indicator that would make it possible to integrate the affordability and healthy diet dimensions.

7. Conclusion

The main objective of this report was to scientifically assess the validity, usefulness, and fitness-for purpose of existing healthy diet metrics for use as global and national monitoring indicators. We have identified seven healthy diet measures that have been critically and comparatively evaluated: the AHEI-2010, the DASH score, the DQI-I, the GDQS, the GDR score, the MDD-W, and the Nova UPF score.

From this assessment, there are four healthy diet metrics which seem to be currently the most suitable for global and national monitoring: the GDR score, GDQS, MDD-W, and Nova UPF score. However, these metrics have certain weaknesses that should be highlighted. Although the GDR score and GDQS were identified as promising indicators of a healthy diet, additional studies are needed to confirm their validity in various contexts and their equivalence across contexts. While the GDR score is a simple, rapid, and feasible population-based metric, it is qualitative and based only on the consumption of food groups. On the other hand, the GDQS is a population-based metric based on semi-quantitative dietary data that involves data collection using the GDQS app and a novel approach for estimating portion size. This data collection and analysis takes longer than the GDR score takes to collect and compute, and the portion size estimation approach still needs to be validated for accuracy. Both metrics require further validation in different demographic groups and contexts. Although the MDD-W is already widely used in large multitopic surveys such as those conducted in the Demographic and Health Survey programme, it assesses only "dietary diversity", which is one of the subconstructs of the construct of healthy diets. The Nova UPF score assesses only "moderation", which is one of the subconstructs of the construct of healthy diets.

The main knowledge gaps identified in this report are the population groups for which the metrics could be used and validity of the metrics for predicting health outcomes. These metrics were often developed for WRA or adult population and their ability to be used for other population groups such as pregnant or lactating women, adolescents and children largely remains unknown. It is a priority to further explore how these indicators could be used for more population groups and settings. This is a priority for the developers of GDQS, GDR score, MDD-W, and Nova UPF score who are currently conducting or planning to conduct such studies. Finally, all of these metrics need to be rigorously assessed for their cross-context equivalence, which is an important measurement attribute for global monitoring.

References

1. The Seventh Meeting of the WHO-UNICEF Technical Expert Advisory group on nutrition Monitoring (TEAM). Meeting report, 5–6 February 2019 Geneva: World Health Organization; 2019 (https://www.who.int/publications/m/item/the-seventh-meeting-of-the-who-unicef-technical-expert-advisory-group-on-nutrition-monitoring-(team), accessed 30 January 2023).
2. Report of the technical consultation on measuring healthy diets: concepts, methods and metrics. Virtual meeting, 18–20 May 2021 Geneva: World Health Organization; 2022 (https://www.who.int/publications-detail-redirect/9789240040274, accessed 30 January 2023).
3. FAO and WHO. Sustainable healthy diets – Guiding principles. Rome: Food and Agriculture Organization of the United Nations/World Health Organization; 2019 (https://www.who.int/publications-detail-redirect/9789241516648, accessed 30 January 2023).
4. Seligman HK, Levi R, Adebiyi VO, Coleman-Jensen A, Guthrie JF, and Frongillo EA. Assessing and monitoring nutrition security to promote healthy dietary intake and outcomes in the United States. Annual Review of Nutrition. 2023;43:5.1–5.21. https://doi.org/10.1146/annurev-nutr-062222-023359.
5. US Department of Agriculture and US Department of Health and Human Services. Dietary Guidelines for Americans, 2020-2025. 9th Edition, Washington DC: US Government Printing Office, 2020 (https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf, accessed 30 January 2023).
6. Frongillo EA, Baranowski T, Subar AF, Tooze JA, Kirkpatrick SI. Establishing Validity and Cross-Context Equivalence of Measures and Indicators. J Acad Nutr Diet. 2019; 119(11): 1817–1830. doi: 10.1016/j.jand.2018.09.005.
7. Miller V, Webb P, Micha R, Mozaffarian D and Global Dietary Database. Defining diet quality: a synthesis of dietary quality metrics and their validity for the double burden of malnutrition. Lancet Planet Health. 2020; 4(8): e352–e370. doi: 10.1016/S2542-5196(20)30162-5.
8. Dwyer J, Picciano MF, Raiten DJ. Future Directions for the Integrated CSFII-NHANES: What We Eat in America–NHANES. J Nutr. 2003; 133(2): 576S-581S. doi: 10.1093/jn/133.2.576S.
9. Ahluwalia N, Andreeva VA, Kesse-Guyot E, Hercberg S. Dietary patterns, inflammation and the metabolic syndrome. Diabetes Metab. 2013; 39(2): 99–110. doi.org/10.1016/j.diabet.2012.08.007
10. Hosseinpoor, AR, Bergen N, Schlotheuber A. Promoting health equity: WHO health inequality monitoring at global and national levels. Glob Health Action. 2015; 8: 29034. doi: 10.3402/gha.v8.29034.
11. Trijsburg L, Talsma EF, de Vries JHM, Kennedy G, Kuijsten A, Brouwer ID. Diet quality indices for research in low- and middle-income countries: a systematic review. Nutr Rev. 2019;25;77(8):515–40. doi: 10.1093/nutrit/nuz017.
12. Madden PJ, Yoder MD. Program Evaluation: Food Stamps and Commodity Distribution in Rural Areas of Central Pennsylvania. University Park: Penn State University Press; 1971.
13. Foote JA, Murphy SP, Wilkens LR, Basiotis PP, Carlson A. Dietary variety increases the probability of nutrient adequacy among adults. J Nutr. 2004; 134, 1779–1785.
14. Verger EO, Mariotti F, Holmes BA, Paineau D, Huneau JF. Evaluation of a diet quality index based on the probability of adequate nutrient intake (PANDiet) using national French and US dietary surveys. PLoS ONE. 2012; 7 (8): e42155. doi:10.1371/journal.pone.0042155.
15. FAO. Food composition challenges. International Network of Food Data Systems (INFOODS). Rome: Food and Agriculture Organization of the United Nations; 2022 (https://www.fao.org/infoods/infoods/food-composition-challenges/en/, accessed 30 January 2023).
16. Kennedy ET, Ohls J, Carlson S, Fleming K. The Healthy Eating Index: Design and Applications. J Am Diet Assoc. 1995; 95(10): 1103–1108. doi: 10.1016/S0002-8223(95)00300-2.
17. Guenther PM, Reedy J, Krebs-Smith SM, Reeve BB, Basiotis PP. Development and Evaluation of the Healthy Eating Index-2005: Technical Report. J Am Diet Assoc. 2008; 108(11): 1896-1901. doi: 10.1016/j.jada.2008.08.016.
18. Guenther PM, Casavale KO, Reedy J, Kirkpatrick SI, Hiza HA, Kuczynski KJ et al. Update of the Healthy Eating Index: HEI-2010. J Acad Nutr Diet. 2013;113(4):569-80. doi: 10.1016/j.jand.2012.12.016.
19. Krebs-Smith SM, Pannucci TE, Subar AF, Kirkpatrick SI, Lerman JL, Tooze JA et al. Update of the Healthy Eating Index: HEI-2015. J Acad Nutr Diet. 2018;118(9):1591-1602. doi: 10.1016/j.jand.2018.05.021.

20. Chaltiel D, Adjibade M, Deschamps V, Touvier M, Hercberg S, Julia C et al. Programme National Nutrition Santé - guidelines score 2 (PNNS-GS2): development and validation of a diet quality score reflecting the 2017 French dietary guidelines. Br J Nutr. 2019;14;122(3):331-342. doi: 10.1017/S0007114519001181.
21. Estaquio C, Kesse-Guyot E, Deschamps V, Bertrais S, Dauchet L, Galan P et al. Adherence to the French Programme National Nutrition Santé Guideline Score is associated with better nutrient intake and nutritional status. J Am Diet Assoc. 2009;109(6):1031-41. doi: 10.1016/j.jada.2009.03.012.
22. Hutchins-Wiese, HL, Bales CW, Starr KNP. Mediterranean diet scoring systems: understanding the evolution and applications for Mediterranean and non-Mediterranean countries. Br J Nutr. 2022; 128(7): 1371-1392. doi:10.1017/S0007114521002476.
23. Willett W, Rockström J, Loken B, Springmann M, Lang T, Vermeulen S et al. Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. Lancet. 2019;393(10170): 447–492. doi: 10.1016/S0140-6736(18)31788-4.
24. Knuppel A, Papier K, Key TJ, Travis RC. EAT-Lancet score and major health outcomes: the EPIC-Oxford study. Lancet. 2019; 394 (10194): 213–214. doi: 10.1016/S0140-6736(19)31236-X.
25. Trijsburg L et al. Method for the Development of WISH, a Globally Applicable Index for Healthy Diets from Sustainable Food Systems. Nutrients. 2020; 13(1): 93. doi: 10.3390/nu13010093.
26. Cacau LT et al. Development and Validation of an Index Based on EAT-Lancet Recommendations: The Planetary Health Diet Index. Nutrients. 2021; 13(5): 1698. doi: 10.3390/nu13051698.
27. Stubbendorff A et al. Development of an EAT-Lancet index and its relation to mortality in a Swedish population. Am J Clin Nutr. 2022; 115(3): 705–716. doi: 10.1093/ajcn/nqab369.
28. Jolles BM, Buchbinder R, Beaton DE. A study compared nine patient-specific indices for musculoskeletal disorders. J Clin Epidemiol. 2005; 58(8): 791–801. doi.org/10.1016/j.jclinepi.2005.01.012
29. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. J Clin Epidemiol. 2011; 64(4): 366–379. doi: 10.1016/j.jclinepi.2010.04.005.
30. Yuen EYN, Thomson M, Gardiner H. Measuring Nutrition and Food Literacy in Adults: A Systematic Review and Appraisal of Existing Measurement Tools. Health Lit Res Pract. 2018; 2(3): e134–e160. doi: 10.3928/24748307-20180625-01.
31. Ribas SA et al. Cross-cultural measurement equivalence of the Healthy Eating Index adapted version for children aged 1–2 years. Br J Nutr. 2021; 126(5): 782–789. doi: 10.1017/S0007114520004729.
32. Coates J et al. A program needs-driven approach to selecting dietary assessment methods for decision-making in food fortification programs. Food Nutr Bull. 2021; 33(3 Suppl): S146-156. doi: 10.1177/15648265120333S202.
33. Chiuve SE et al. Alternative dietary indices both strongly predict risk of chronic disease. J Nutr.2021; 142(6): 1009–1018 (2012). doi: 10.3945/jn.111.157222.
34. Schwingshackl L, Bogensberger B, Hoffmann G. Diet Quality as Assessed by the Healthy Eating Index, Alternative Healthy Eating Index, Dietary Approaches to Stop Hypertension Score, and Health Outcomes: An Updated Systematic Review and Meta-Analysis of Cohort Studies. J Acad Nutr Diet. 2018; 118(1): 74-100.e11. doi: 10.1016/j.jand.2017.08.024.
35. Wang DD et al. Trends in Dietary Quality Among Adults in the United States, 1999 through 2010. JAMA Intern Med. 2014; 174(10): 1587–1595. doi: 10.1001/jamainternmed.2014.3422.
36. Lafrenière J et al. Development and validation of a Brief Diet Quality Assessment Tool in the French-speaking adults from Quebec. Int J Behav Nutr Phys Act. 2019; 16(1): 61. doi: 10.1186/s12966-019-0821-6.
37. Kim S, Haines PS, Siega-Riz AM, Popkin BM. The Diet Quality Index-International (DQI-I) provides an effective tool for cross-national comparison of diet quality as illustrated by China and the United States. J Nutr. 2003; 133(11): 3476–3484. doi.org/10.1093/jn/133.11.3476.
38. Al-Ibrahim AA, Jackson RT. Healthy eating index versus Alternative healthy index in relation to diabetes status and health markers in U.S. adults: NHANES 2007–2010. Nutr J. 2019; 18(1): 26. doi: 10.1186/s12937-019-0450-6.
39. Procter-Gray E et al. Comparison of Dietary Quality Assessment Using Food Frequency Questionnaire and 24-hour-recalls in Older Men and Women. AIMS Public Health. 2017; 4(4): 326–346. doi: 10.3934/publichealth.2017.4.326.
40. Whitton C, Ho JCY, Rebello SA, van Dam RM. Relative validity and reproducibility of dietary quality scores from a short diet screener in a multi-ethnic Asian population. Public Health Nutr. 2018; **21(15):** 2735–2743. doi: 10.1017/S1368980018001830.

41. Kotecki JE, Greene MA, Khubchandani J, Kandiah J. A Brief Dietary Screening and Counseling Tool for Health Educators. Am J Health Educ. 2021; 52(3): 111–116. doi: 10.1080/19325037.2021.1902886
42. Gibson RS. Principles of Nutritional Assessment. Oxford: Oxford University Press; 2005.
43. Medical Research Council. DAPA Measurement Toolkit. Cambridge: National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre; 2018 (https://www.measurement-toolkit.org/diet/subjective-methods/introduction, accessed 30 January 2023).
44. National Cancer Institute. Food Frequency Questionnaire at a Glance | Dietary Assessment Primer. Washington DA: NCI/Division of Cancer Control & Population Sciences; no date. (https://dietassessmentprimer.cancer.gov/profiles/questionnaire/index.html, accessed 30 January 2023).
45. Fung TT et al. Adherence to a DASH-Style Diet and Risk of Coronary Heart Disease and Stroke in Women. Arch Intern Med. 2008; 168(7): 713–720. doi: 10.1001/archinte.168.7.713.
46. Soltani S, Arablou T, Jayedi A, Salehi-Abargouei A. Adherence to the dietary approaches to stop hypertension (DASH) diet in relation to all-cause and cause-specific mortality: a systematic review and dose-response meta-analysis of prospective cohort studies. Nutr J. 2020; 19(1): 37. doi: 10.1186/s12937-020-00554-8.
47. Yu D et al. Adherence to dietary guidelines and mortality: a report from prospective cohort studies of 134,000 Chinese adults in urban Shanghai. Am J Clin Nutr. 2014; 100(2): 693–700. doi: 10.3945/ajcn.113.079194.
48. Mellen PB, Gao SK, Vitolins MZ, Goff DC. Deteriorating dietary habits among adults with hypertension: DASH dietary accordance, NHANES 1988-1994 and 1999-2004. Arch Intern Med. 2008; 168(3): 308–314. doi: 10.1001/archinternmed.2007.119.
49. Cuenca-García M et al. Dietary indices, cardiovascular risk factors and mortality in middle-aged adults: findings from the Aerobics Center Longitudinal Study. Ann Epidemiol. 2014; 24(4): 297-303.e2. doi: 10.1016/j.annepidem.2014.01.007.
50. Appel LJ et al. A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. N. Engl. J. Med. 1997; 336(16): 1117–1124. doi: 10.1056/NEJM199704173361601.
51. Sacks FM et al. Rationale and design of the Dietary Approaches to Stop Hypertension trial (DASH). A multicenter controlled-feeding study of dietary patterns to lower blood pressure. Ann Epidemiol. 1995; 5(2): 108–118. doi: 10.1016/1047-2797(94)00055-x.
52. Reyes-García A, López-Olmedo N, Basto-Abreu A, Shamah-Levy T, Barrientos-Gutierrez T. Adherence to the DASH diet by hypertension status in Mexican men and women: A cross-sectional study. Prev Med Rep. 2022; 27: 101803. doi: 10.1016/j.pmedr.2022.101803.
53. Silva AL, Teles J, Fragoso I. Health-related quality of life of Portuguese children and adolescents according to diet quality and food intake. Qual Life Res. 2020; 29(8): 2197–2207. doi: 10.1007/s11136-020-02475-9.
54. Williams J et al. Diet quality of adolescents in rural Sri Lanka based on the Diet Quality Index-International: findings from the 'Integrating Nutrition Promotion and Rural Development' project. Public Health Nutr. 2019; 22(10): 1735–1744. doi: 10.1017/S1368980019000430.
55. Kalmpourtzidou A et al. Diet Quality: A Neglected Parameter in Children With Food Allergies. A Cross-Sectional Study. Front Pediatr. 2021; 9: 658778. doi: 10.3389/fped.2021.658778.
56. Colapinto CK, Fitzgerald A, Taper LJ, Veugelers PJ. Children's preference for large portions: prevalence, determinants, and consequences. J Am Diet Assoc. 2007; 107(7): 1183–1190. doi.org/10.1016/j.jada.2007.04.012.
57. Shin M-K, Kim Y-S, Kim J-H, Kim S-H, Kim Y. Dietary Patterns and Their Associations with the Diet Quality Index-International (DQI-I) in Korean Women with Gestational Diabetes Mellitus. Clin Nutr Res. 2015; 4(4): 216–224. doi: 10.7762/cnr.2015.4.4.216.
58. Conradie C et al. A Priori and a Posteriori Dietary Patterns among Pregnant Women in Johannesburg, South Africa: The NuPED Study. Nutrients. 2021; 13(2): 565. doi: 10.3390/nu13020565.
59. Abassi MM et al. Gender inequalities in diet quality and their socioeconomic patterning in a nutrition transition context in the Middle East and North Africa: a cross-sectional study in Tunisia. Nutr J. 2019; 18(1): 18. doi: 10.1186/s12937-019-0442-6.
60. Gregory CO, McCullough ML, Ramirez-Zea M, Stein AD. Diet scores and cardio-metabolic risk factors among Guatemalan young adults. Br J Nutr. 2009; 101(12): 1805–1811. doi: 10.1017/S000711450813584X.
61. Asghari G et al. The association between diet quality indices and obesity: Tehran Lipid and Glucose Study. Arch Iran Med. 2012; 15(10): 599–605.
62. Zamani B et al. Dietary Quality Index and Cardiometabolic Risk Factors among Adult Women. Iran J Public Health. 2021; 50(8): 1713–1721. doi: 10.18502/ijph.v50i8.6819.

63. Tsoi KY et al. Evaluation of dietary pattern in early pregnancy using the FIGO Nutrition Checklist compared to a food frequency questionnaire. Int J Gynaecol Obstet. 2020; 151(Suppl 1): 37–44. doi: 10.1002/ijgo.13324.
64. Tur JA, Romaguera D, Pons A. The Diet Quality Index-International (DQI-I): is it a useful tool to evaluate the quality of the Mediterranean diet? Br. J. Nutr. 2005; 93(3): 369–376. doi: 10.1079/bjn20041363.
65. Bromage S et al. Development and Validation of a Novel Food-Based Global Diet Quality Score (GDQS). J Nutr. 2021; 151(12 Suppl2): 75S-92S. doi: 10.1093/jn/nxab244.
66. Angulo E et al. Changes in the Global Diet Quality Score, Weight, and Waist Circumference in Mexican Women. J Nutr. 2021; 151(12 Suppl 2): 152S-161S. doi: 10.1093/jn/nxab171.
67. Fung TT et al. Higher Global Diet Quality Score Is Inversely Associated with Risk of Type 2 Diabetes in US Women. J Nutr. 2021; 151(12 Suppl 2): 168S-175S. doi: 10.1093/jn/nxab195.
68. Fung TT et al. Higher Global Diet Quality Score Is Associated with Less 4-Year Weight Gain in US Women. J Nutr. 2021: 151(12 Suppl 2): 162S-167S. doi: 10.1093/jn/nxab170.
69. Moursi M et al. There's an App for That: Development of an Application to Operationalize the Global Diet Quality Score. J Nutr. 2021; 151(12 Suppl 2): 176S-184S. doi: 10.1093/jn/nxab196.
70. Intake–Center for Dietary Assessment. The Global Diet Quality Score: Data Collection Options and Tabulation Guidelines. Washington, DC: Intake–Center for Dietary Assessment/FHI Solutions; 2021 (https://www.intake.org/sites/default/files/2021-04/GDQS%20Overview%20Document%20-%20April%202021.pdf, accessed 30 January 2023).
71. Global Diet Quality Score (GDQS) Toolkit. Washington, DC: Intake–Center for Dietary Assessment; 2022 (https://www.intake.org/news/global-diet-quality-score-gdqs-toolkit, accessed 30 January 2023).
72. Herforth AW, Wiesmann D, Martínez-Steele E, Andrade G, Monteiro CA. Introducing a Suite of Low-Burden Diet Quality Indicators That Reflect Healthy Diet Patterns at Population Level. Curr Dev Nutr. 2020; 4(12): nzaa168. doi: 10.1093/cdn/nzaa168
73. Global Diet Quality Project. DQQ Indicator Guide, Version 7; 2022. [Google Docs] (https://docs.google.com/document/d/1RGwNititbYqSrGpUXhQs7mHWgLIgOxAlN5N0AOZN3qc/edit?usp=sharing&usp=embed_facebook, accessed 30 January 2023).
74. Global Diet Quality Project. Measuring what the world eats: Insights from a new approach; 2022. [Google Docs] (https://doi.org/10.36072/dqq2022, accessed 30 January 2023).
75. Wang H, Herforth AW, Xi B, Zou Z. Validation of the Diet Quality Questionnaire in Chinese Children and Adolescents and Relationship with Pediatric Overweight and Obesity. Nutrients. 2022; 14(17): 3551. doi: 10.3390/nu14173551.
76. Global Diet Quality Project. How to use the DQQ. Diet QUality.org; 2022. [Google Docs] (https://drive.google.com/file/d/1EPZ6WKiRm9nRluwx-j191SsOigKgnzHp/view?usp=sharing&usp=embed_facebook, accessed 30 January 2023).
77. Global Diet Quality Project - DQQ Tools & Data. Global Diet Quality Project; 2022 (https://www.dietquality.org/tools, accessed 30 January 2023).
78. Herforth A et al. Development of a Diet Quality Questionnaire for Improved Measurement of Dietary Diversity and Other Diet Quality Indicators (P13-018-19). Curr Dev Nutr. 2019; 3(Suppl 1): nzz036. (2019). doi.org/10.1093/cdn/nzz036.P13-018-19
79. Ma S, Herforth AW, Vogliano C, Zou Z. Most Commonly-Consumed Food Items by Food Group, and by Province, in China: Implications for Diet Quality Monitoring. Nutrients. 2022; 14(19): 1754 (2022). doi: 10.3390/nu14091754.
80. Martin-Prével Y et al. Moving forward on choosing a standard operational indicator of women's dietary diversity. Technical report. Rome: Food and Agriculture Organization of the United Nations; 2015. DOI:10.13140/RG.2.1.4695.7529
81. Women's Dietary Diversity Project (WDDP) Study Group. Development of a Dichotomous Indicator for Population-Level Assessment of Dietary Diversity in Women of Reproductive Age. Curr Dev Nutr. 2017; 1(12): cdn.117.001701. doi: 10.3945/cdn.117.001701.
82. FAO. Minimum dietary diversity for women. An updated guide for measurement: from collection to action. Rome: Food and Agriculture Organization of the United Nations; 2021 (https://www.fao.org/documents/card/en/c/cb3434en, accessed 30 January 2023).
83. FAO and FHI 360. Minimum Dietary Diversity for Women: A Guide for Measurement. Rome: Food and Agriculture Organization of the United Nations; 2016 (https://www.fantaproject.org/monitoring-and-evaluation/minimum-dietary-diversity-women-indicator-mddw, accessed 30 January 2023).
84. FAO. Guidelines for measuring household and individual dietary diversity. Rome: Food and Agriculture Organization of the United Nations; 2011 (https://www.fao.org/3/i1983e/i1983e00.pdf, accessed 30 January 2023).

85. Arimond M et al. Simple food group diversity indicators predict micronutrient adequacy of women's diets in 5 diverse, resource-poor settings. J. Nutr. 2010; 140(11): 2059S–69S. doi: 10.3945/jn.110.123414.

86. Lourme-Ruiz A et al. Seasonal variability of women's dietary diversity and food supply: a cohort study in rural Burkina Faso. Public Health Nutr. 2022; 25(9): 2475–2487. doi: 10.1017/S1368980021004171.

87. Hanley-Cook GT et al. Seasonality and Day-to-Day Variability of Dietary Diversity: Longitudinal Study of Pregnant Women Enrolled in a Randomized Controlled Efficacy Trial in Rural Burkina Faso. J Nutr. 2022; 152(9): 2145–2154. doi: 10.1093/jn/nxac104.

88. Arimond M et al. Food group diversity and nutrient adequacy: Dietary diversity as a proxy for micronutrient adequacy for different age and sex groups in Mexico and China. Geneva: GAIN Discussion Paper no. 9; 2021. https://doi.org/10.36072/dp.9.

89. Betül TM et al. The DQQ is a Valid Tool to Collect Population-Level Food Group Consumption Data: A Study Among Women in Ethiopia, Viet Nam, and Solomon Islands. J Nutr. 2023; 153 (1): 340–351. doi: 10.1016/j.tjnut.2022.12.014.

90. Hanley-Cook GT et al. Minimum Dietary Diversity for Women of Reproductive Age (MDD-W) Data Collection: Validity of the List-Based and Open Recall Methods as Compared to Weighed Food Record. Nutrients. 2020; 12(7): 2039. doi: 10.3390/nu12072039.

91. Dos Santos Costa C et al. Nova score for the consumption of ultra-processed foods: description and performance evaluation in Brazil. Rev Saude Publica. 2021; 55: 13). doi: 10.11606/s1518-8787.2021055003588.

92. Monteiro CA et al. The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. Public Health Nutr. 2018; **21(1):** 5–17. doi: 10.1017/S1368980017000234.

93. Martini D, Godos J, Bonaccio M, Vitaglione P, Grosso G. Ultra-Processed Foods and Nutritional Dietary Profile: A Meta-Analysis of Nationally Representative Samples. Nutrients. 2021; **13(10):** 3390. doi: 10.3390/nu13103390.

94. Pagliai G et al. Consumption of ultra-processed foods and health status: a systematic review and meta-analysis. Br J Nutr. 2021; 125(3): 308–318. doi: 10.1017/S0007114520002688.

95. Delpino FM et al. Ultra-processed food and risk of type 2 diabetes: a systematic review and meta-analysis of longitudinal studies. Int J Epidemiol. 2022; 51(4): 1120–1141. doi: 10.1093/ije/dyab247.

96. Suksatan W et al. Ultra-Processed Food Consumption and Adult Mortality Risk: A Systematic Review and Dose-Response Meta-Analysis of 207,291 Participants. Nutrients. 2021; 14(1): 174. doi: 10.3390/nu14010174.

97. Innovative Methods and Metrics for Agriculture and Nutrition Actions (IMMANA). A validation study of a dietary assessment instrument capturing ultra-processed food consumption in multiple countries. Sao Paulo: ANH Academy; 2021 (https://www.anh-academy.org/immana/grants/grants-round-3/validation-study-dietary-assessment-instrument-capturing-ultra-processed-food, accessed 30 January 2023).

98. Young SL, Bethancourt HJ, Ritter ZR, Frongillo EA. The Individual Water Insecurity Experiences (IWISE) Scale: Reliability, equivalence, and validity of an individual-level measure of water security. BMJ Global Health. 2021; 6:e006460. doi:10.1136/bmjgh-2021-006460.

Annexes

Annex 1.
List of participants

Technical Expert Meeting on Harmonizing and Mainstreaming Measurement of Healthy Diets Globally

Bellagio, Italy, 28 November–2 December 2022

Victor Aguayo, Director, Child Nutrition and Development, UNICEF, New York, USA

Jane Badham, Managing Director, JB Consultancy, Johannesburg, South Africa

Mia Blakstad, Senior Associate, Food Initiative, The Rockefeller Foundation, New York, USA

Francesco Branca, Director, Department of Nutrition and Food Safety, WHO, Geneva, Switzerland

Jennifer Coates, Associate Professor, Food and Nutrition Policy and Programs, Tufts University Friedman School of Nutrition Science and Policy, Boston, MA, USA

Megan Deitchler, Director, Intake, Center for Dietary Assessment, FHI 360, Washington, DC, USA

Umi Fahmida, Senior Researcher, South East Asian Ministers of Education Organization (SEAMEO) Regional Center for Food and Nutrition (RECFON), Jakarta, Indonesia

Sara Farley, Vice President, Food Initiative, The Rockefeller Foundation, New York, USA

Edward Frongillo, Director, Global Health Initiatives, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

Chika Hayashi, Senior Advisor, Monitoring and Statistics and Unit Chief, Nutrition Data, UNICEF, New York, USA

Anna Herforth, Research Associate, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA

Bridget Holmes, Nutrition and Food Systems Officer and Nutrition Assessment Team Leader, Food and Nutrition Division, FAO, Rome, Italy

Amos Laar, Professor, School of Public Health, University of Ghana, Accra, Ghana

Carlos Monteiro, Professor, Department of Nutrition, Universidade Estadual de São Paulo, São Paulo, Brazil

Lynnette Neufeld, Director, Food and Nutrition Division, FAO, Rome, Italy

Marie Ruel, Director, Poverty, Health and Nutrition Division, International Food Policy Research Institute, Washington, DC, USA

Kuntal Saha, Technical Officer, Monitoring Nutritional Status and Food Safety Events, Department of Nutrition and Food Safety, WHO, Geneva, Switzerland

Shelly Sundberg, Senior Program Officer, Nutrition, Bill & Melinda Gates Foundation, Seattle, Washington, USA

Emorn Udomkesmalee, Senior Advisor, Institute of Nutrition, Mahidol University, Bangkok, Thailand

Eric Verger, Research Fellow, Institut de Recherche pour le Développement (IRD), Montpellier, France

Chris Vogliano, Technical Advisor, Food Systems, USAID Advancing Nutrition, Arlington, VA, USA

Annex 2.
List of metric developers contacted for this report

- Emails were exchanged from 9 September to 4 November 2022 with Dr Teresa Fung as the referent of the AHEI-2010 and DASH score development team.
- Emails were exchanged from 9 to 20 September 2022 with Dr Soowon Kim as the referent of the DQI-I development team.
- Emails were exchanged from 9 September to 11 November 2022 with Dr Megan Deitchler as the referent of the GDQS development team.
- A videoconference was organized on 9 September 2022 with Dr Anna Herforth, Dr Gina Kennedy and Andrew Rzepa as referents of the GDR score development team, and emails were exchanged from 8 to 13 November 2022 with Dr Anna Herforth.
- Emails were exchanged on 14 December 2022 with Dr Bridget Holmes as the referent of the MDD-W development team.
- Emails were exchanged from 9 September to 11 November 2022 with Professor Carlos Augusto Monteiro as the referent of the Nova score development team.

Annex 3. Declaration of Interests

All external experts submitted to WHO a declaration of interest disclosing potential conflicts of interest that might affect, or might reasonably be perceived to affect, their objectivity and independence in relation to the subject matter of this technical report. WHO reviewed each of those and had concluded that none could give rise to a potential or reasonably perceived conflict of interest related to the content covered in the report.

For more information, please contact:

FAO
Food and Nutrition Division Food and Agricultural Organization of the United Nations
Viale delle Terme di Caracalla, 00153 Rome, Italy
Email: nutrition@fao.org
www. https://www.fao.org/nutrition/en/

UNICEF
Division of Data, Analytics, Planning and Monitoring
UNICEF
3 UN Plaza, New York, NY, 10017, USA
Email: data@unicef.org
www.data.unicef.org/nutrition

WHO
Department of Nutrition and Food Safety
World Health Organization
20 Avenue Appia, 1211 Geneva 27, Switzerland
Email: nutrition@who.int
www.who.int/nutrition

9789240072138
9 789240 072138