Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Issue: *Paths of Convergence for Agriculture, Health, and Wealth*

Validity and reliability of food security measures

Carlo Cafiero,[1] Hugo R. Melgar-Quiñonez,[2] Terri J. Ballard,[1] and Anne W. Kepple[1]

[1]Statistics Division, United Nations Food and Agriculture Organization, Rome, Italy. [2]McGill Institute for Global Food Security, School of Dietetics and Human Nutrition, McGill University, Montreal, Quebec, Canada

Address for correspondence: Carlo Cafiero, Statistics Division, United Nations Food and Agriculture Organization, via delle Terme di Caracalla 4, Room C-412, 00153 Rome, Italy. Carlo.Cafiero@FAO.org

This paper reviews some of the existing food security indicators, discussing the validity of the underlying concept and the expected reliability of measures under reasonably feasible conditions. The main objective of the paper is to raise awareness on existing trade-offs between different qualities of possible food security measurement tools that must be taken into account when such tools are proposed for practical application, especially for use within an international monitoring framework. The hope is to provide a timely, useful contribution to the process leading to the definition of a food security goal and the associated monitoring framework within the post-2015 Development Agenda.

Keywords: food security; validity of empirical measures; post-2015 Development Agenda

Introduction

Measurement is indisputably an important element of the process through which we advance knowledge. It is indispensable when we need to highlight changes such as the progress toward set targets. To contribute to knowledge and to allow correct assessments, however, measurement should be *valid* and *reliable*, posing two fundamental but distinct problems regarding *what* is being measured and *how* it is done.

Assessing validity and reliability of measures is particularly problematic in the social sciences, where the phenomena we wish to measure are often not directly observable. The definition of what is being measured may become confounded with its measurement to the point that use of different measures would imply adoption of different notions of the phenomenon being analyzed, an issue that could easily go unappreciated by less sophisticated readers. Confusion may thus arise, especially when dealing with complex constructs.

This problem, we argue, has been particularly evident in the academic and political debates on food security definition and measurement over the last half century. Hundreds of indicators have been proposed, ranging from quantification of food supplies at the country level, to more or less detailed characterizations of food consumption at the household or individual level, and including measures of nutritional outcomes with respect to growth and nutritional deficiencies. Some of these indicators are presented as "measures" of food security, others are combined in various ways to produce "indexes," but questions remain regarding whether they can indeed be considered proper measures.

With such questions in mind, in this paper we discuss a few of the existing indicators of food security. After highlighting some of the elements of the debate that are relevant for the definition of a measurable food security construct, we introduce a conceptual framework that distinguishes the validity of a measurement tool from the reliability of the measures it produces. We then apply the proposed framework to indicators belonging to two categories: those based on the concept of food consumption adequacy, and those that treat food insecurity as a condition that can be identified and characterized by experiences and behavioral responses that appear to be common across most cultures. In the first group, we consider the *prevalence of undernourishment* (POU, which is currently used as the official indicator to monitor the target 1C of the first Millennium Development Goal to eradicate extreme poverty and hunger), the *Food*

doi: 10.1111/nyas.12594

Consumption Score (FCS, used by the World Food Programme in their food security and vulnerability assessments), and the *Household Dietary Diversity Score* (HDDS, which has been proposed as perhaps the simplest possible proxy for the adequacy of food consumption at the household level).

In the second group, we include a number of *experience-based food security scales*, starting from the Household Food Security Survey Module (HFSSM, developed by the United States Department of Agriculture and used in the United States and Canada), and continuing with the Household Food Insecurity Access Scale (HFIAS, promoted by the Food and Nutrition Technical Assistance-II (FANTA-II) initiative), the harmonized Latin American and Caribbean Food Security Scale (ELCSA, from the Spanish *Escala Latinoamericana y Caribeña de Seguridad Alimentaria*), and the recently launched Food Insecurity Experience Scale (FIES), developed and adopted by the Food and Agriculture Organization (FAO) in collaboration with the Voices of the Hungry Project.

We conclude, based on our analysis, that the inadequacy of food consumption can be reliably measured at the country level as the percentage of the population that does not have access to enough food to fulfil normal requirements. Proposed metrics of food consumption adequacy at the household or individual levels, unfortunately, do not possess the analytic foundations required to determine their empirical validity as food security measures. In particular, there is no basis to establish the reliability of the measures obtained in practice or to ensure comparability across applications.

Experience-based scales, on the other hand, emerge as very promising tools to measure a valid concept of food insecurity at the household and individual levels. Their use, especially if coupled with other indicators of individual or household socioeconomic condition or nutritional status, can contribute to a better understanding of the determinants and consequences of household and individual food insecurity.

In our discussion, we also draw conclusions on the suitability of these indicators as elements of a monitoring system at the global scale, in the hope of providing a timely, useful contribution to the process leading to the definition of the post-2015 Development Agenda, where a specific goal on food security will likely be included.

Toward the definition of a measurable construct of food security[a]

In the early years, the "food problem" was fundamentally identified with having enough supplies to cover the needs of the population.[1–3] It was only after 1974 that the expression "food security" started to be widely used, even though mostly still referring to adequacy of country-level food supplies. In the 1980s, the focus shifted to looking at the food problem from the perspective of people's access to food, owing to the growing evidence of widespread malnutrition and famines in the developing world, even in periods of relatively abundant food supplies.[4] To mark the difference in these different approaches, the qualified expression "*household* food security" began to be used, stressing the idea that evidence of food insecurity was to be investigated at the household level. In the 1990s, another shift occurred from considering food security essentially a matter of dietary energy adequacy to including economic, social, nutritional, and psychological considerations.[5,6] This evolution in thinking about food security has been described as a process characterized by the occurrence of "important and overlapping paradigm shifts:" (a) "from the global and the national to the household and the individual, (b) from a food-first perspective to a livelihood perspective, and (c) from objective indicators to subjective perception."[7]

The search for indicators has always been present in the debate that has accompanied the continuing evolution in thinking about food security, with a proliferation of definitions and hundreds of indicators being proposed for different uses.[8]

It is perhaps with the World Food Summit (WFS)[9] and, later, the United Nations Millennium Declaration on the Millennium Development Goals (MDG),[10] when globally endorsed, time-bound targets associated with the overarching goal to eradicate hunger were set,[b] that the relevance of measurement

[a]Retracing the history of food security definitions is beyond the scope of this paper, and we focus here only on elements that are relevant for measurement. For a thorough review of the evolving definition of food security, mainly from the institutional and political perspective, see the volume by Shaw;[5] see also Gibson.[6]

[b]The WFS set the target to halve, by the year 2015, the "number of people suffering from hunger." Four years

and its relation with the definition of the object being measured emerged in all its importance. While the goal and the targets were framed with respect to hunger, the choice of indicators revealed the difficulty in operationalizing the concept as a measurable construct. Estimates of the prevalence and number of undernourished individuals, as provided by the FAO,[11] and of the percentage of children under 5 years of age that are underweight, as provided by UNICEF,[12] were adopted as official indicators of the MDG monitoring framework. Neither of the two estimates, however, can be considered a measure of hunger. Rather, they reflect respectively, the extent of insufficiency of food consumption in a country and the presence of symptoms of child malnutrition.

In the extensive debate around food security measurement, including that generated from the 2002 International Scientific Symposium hosted by the FAO,[13] it has clearly emerged that the demand for measures of food security reflecting the various dimensions of availability, access, utilization, and stability (as captured by the definition set forth by the WFS in 1996[c,14]) has not yet been adequately fulfilled. This ongoing debate has been summarized in several recent publications where indicators are reviewed.[15–19]

Most notably, the need has emerged to go beyond the national level when assessing the state of food insecurity in a country, owing to the fact that relevant differences in developmental achievements are seen more within, rather than between, countries. This consideration, along with the increased statistical capacity in many countries, has promoted an increased use of data from population surveys as opposed to national accounts and macroeconomic data. Relying more directly on survey data is justified also as a response to another identified need, namely to improve the way the food access dimension is captured at the individual or household levels. To respond to the new need, efforts have been made in several directions. We focus here on two of these efforts: the improvement of measures of food consumption and the development of direct, experience-based measures of the severity of food insecurity, defined as the inability to obtain adequate food in quantity and quality because of a lack of money or other resources.

Validity and reliability of household food security measures

In choosing the criteria to assess the overall quality of the indicators we review, we are primarily guided by the consideration that individual or household food security metrics should be effective means to monitor progress toward the goal of ensuring food security for all, mindful of the prospect that such a goal will likely be agreed upon by the international community when defining a renewed, post-2015 Development Agenda.[d]

The criteria we apply are those of *validity* and of *reliability* (which in turn includes *accuracy and precision*) in a framework that partly overlaps with the one proposed by Frongillo.[20]

Validity

The terms *validity* and *validation* have been used with different meanings in different occasions, and general consensus may still not exist.[e,21]

Wikipedia provides a very broad definition of validity as follows: "In science and statistics, validity is the extent to which a concept, conclusion or measurement is *well-founded* and corresponds *accurately* to the real world."[22] We believe this definition is too broad and that the issues of whether a measurement is well-founded and whether it is accurate are two

later, the MDG proposed to "halve, between 1990 and 2015, the proportion of people who suffer from hunger" (a much less demanding target in the presence of a fast growing population).

[c]"Food security, at the individual, household, national, regional and global levels [is achieved] when all people, at all times, have physical and economic access to sufficient, safe and nutritious food to meet their dietary needs and food preferences for an active and healthy life."

[d]At the time this paper was written, the most authoritative proposal for a relevant target on food security to be considered in the post-2015 Development Agenda reads as follows: "By 2030, end hunger and ensure that all people have access to adequate, safe, affordable, and nutritious food all year round." (Zero draft of the introduction and proposed goals and targets on sustainable development for the post-2015 Development Agenda, produced by the Open Working Group on Sustainable Development Goals, available at http://sustainabledevelopment.un.org/content/documents/4523zerodraft.pdf).

[e]The International vocabulary of metrology[22] does not have an entry for the term *validity*.

separate issues. Knowing whether a measurement tool is inadequate because the underlying theoretical conceptual framework is poorly grounded or because of the inaccuracy of the measures it produces, may be useful, for example, in guiding research aimed at improving them.

With reference to food security measurement, researchers often considered *criterion validity* and *concurrent validity*,[23] by suggesting that validation of a measurement tool could be sought by comparing the measures it produces with, respectively, an existing gold standard or with other measures capturing other aspects of the same construct obtained at the same time on the same objects. However, as food security is a nonmaterial, unobservable construct (i.e., a *latent trait*), for which no objective benchmark exists, we cannot strictly rely on criterion validation, while concurrent validation must rely on the *a priori* assumption that the concurring measures are valid. In reviewing validation studies of food security measures, it is difficult to separate judgment of the appropriateness of the underlying conceptualization of food security from the technical properties of the specific measurement tool being validated. To do so, we need to somehow steer away from criterion and concurrent validation as the proper means to assess validity.[23] Thus far, the processes of defining food security as a measurable construct and of validating proposed metrics have been so inextricably linked that judgments on technical characteristics of the indicators have often been confused with judgments on the appropriateness of the underlying concepts. This has led to two opposite and equally dangerous mistakes: concerns regarding the appropriateness of the concept have been presented as limitations of the measures, and acceptance of the appropriateness of the concept has been considered sufficient to claim validity of the indicators.

For these reasons, and by extending a definition accepted in the educational and psychological testing domains, we propose that validity be intended as the degree to which both evidence and theory support the interpretations of measures entailed by proposed uses of the tool.[24–26] This definition has the advantage that both theoretical and empirical considerations have a role in assessing validity. The explicit need to formulate a model that links theory and evidence in a probabilistic setting, moreover, may force analysts to take statistical aspects into proper consideration, which, in our opinion, has been greatly overlooked in most discussions on food security measurement.

Reliability[f]

A measurement tool is reliable if it consistently produces good results. It is a judgment on the overall degree of uncertainty that inevitably surrounds the measures.

Traditionally, reliability has been considered the combination of two distinct qualities, accuracy (or trueness) and precision, respectively related to the two kinds of errors that may affect a measure. The first is a measure of how small the *systematic error* is, while the second is a measure of how small the *accidental* (or *random*) *error* is.

According to Frongillo, "accuracy is assessed by in-depth analysis and by relating the measure to a criterion measure, which may be a more definitive measure, determinant or consequence."[21] However, as the food insecurity construct is unobservable, there is clearly no definitive direct measure in the more traditional sense. Establishing accuracy must therefore confront the issue of also considering how the measure is related to the indirect evidence on the attribute of interest through a theoretical model. The possibility to assess the presence and the extent of systematic measurement error therefore depends on the appropriateness of the model that links changes in the attribute of the object of interest (the measurand)[27] to changes in the observed evidence (the data).

Reliability is important also for another reason, as it implies, for example, invariance of the measure obtained when the tool is applied to the same object under different conditions.[g] In the area of food security measurement, this means that measures have

[f] A note of caution is appropriate here. The most recent edition of the International Vocabulary of Metrology[22] has dropped the term *reliability* and changed the meaning associated with the term *accuracy*, which is defined now as a synonym of what we still call reliability, as the result of the combination of *trueness* and *precision*. We have decided to continue to use the terms reliability instead of accuracy, and accuracy instead of trueness, as we believe they still reflect the most common, albeit imprecise, understanding of the concepts.

[g] As some measurement error is unavoidable, invariance here must be intended in a probabilistic sense, meaning that the expected size and direction of measurement errors are not affected by the object being measured.

no bias and are similarly precise when applied to different populations. It has thus a strict relation to the possibility of meaningfully comparing measures obtained from different population groups within a country or across countries. Lacking means to assess reliability, it is difficult to defend comparability.

Before moving on, it is worth noting that a fundamental underlying theme when discussing measures is feasibility—intended as the possibility that sufficiently reliable and valid measures can be obtained under normal conditions. While feasibility is key, as practically relevant measures should be feasible to produce, we should not make the mistake of considering a measure to be acceptable only because it is feasible.

A review of selected food security indicators

From the perspective of the procedure to establish validity outlined in the previous section, we discuss indicators on the basis of measures of access to food. A distinction is made between two categories of indicators: those that are based on the concept of adequacy of food consumption and those that are founded on the severity of the constraints on food access, also called experience-based food security indicators.[h]

Measures of the adequacy of food consumption

It has often been suggested that the extent of food insecurity in a population could be conceived as the proportion of individuals who do not consume enough food and that it would suffice to "take a large sample of people and determine how many aren't eating enough" to measure it.[28]

While intuitively appealing, such an approach poses several theoretical and practical challenges. To start with, even defining food as a generic measurable entity presents nontrivial problems. Nutrition studies have revealed how not only the quantity, but also the quality and the variety of foods and how they are combined, contribute to determining the overall nutritional value of what is consumed. How to define and measure the food aggregate for the purpose of assessing its adequacy is thus problematic. Different solutions have been adopted, leading to different operational definitions of food. The most common solution is to use the dietary energy content to aggregate the quantities of different foods and to conduct the assessment with respect to the resulting total dietary energy consumption (DEC). Inadequacy is then defined as when habitual food consumption fails to provide the amounts of dietary energy required for an active and healthy life,[29] a choice that has been criticized because energy-rich diets may be deficient of essential nutrients. The energy content of a certain combination of foods, while measurable, is not the best way to capture adequacy with regard to nutritional value.[i]

[h] There are several examples of indexes and scores (including some of the ones we review here) that are commonly—but perhaps too loosely—presented as measures of household or individual food security. Sometimes also explicit reference is made to the multidimensionality of the food security concept, hinting that the proposed indicator might be intended to capture more than one attribute of household food security, including, for example, the utilization dimension. A thorough discussion on the problems with multidimensional measurement is beyond the scope of this paper and we limit ourselves to the analysis of validity of the proposed indicators as measures of the main attribute they have been designed to capture, namely food access.

[i] The use of total dietary energy as a metric to capture the nutritional adequacy of food has traditionally been justified by the consideration that "it would seem improbable that a dietary intake which is sufficient to cover energy requirements will be insufficient to meet protein requirements, except for the possible exception of populations which subsist on starchy roots, tubers, plantains, etc., all of which are extremely low in protein content. On the other hand, if a diet is adequate in protein content but the quantity of food eaten is insufficient to meet energy needs, some of the protein intake will be used as a source of energy and the diet would in practice be deficient both for energy and protein requirements."[35] More recently, the debate has moved beyond energy and protein consideration to include micronutrient adequacy. A meaningful measure of the overall diet that would reflect its nutritional value, especially with reference to micronutrients, would require detailed observation on the quantity and quality of food consumed, as well as how it is prepared. Reference to average consumption of individual foods or nutrients, in this case, is likely to be of limited use. For example, assessing the quantity of citrus fruits and of dark leafy vegetables independently does not provide the

A related problem is how to establish adequacy of the food consumed. This is still an active area of research, aimed at reconciling consistency between nutritional theory and feasibility of data collection. Requirements have been established for individual macronutrients and of their combination, as well as for micronutrients, with reference to a series of parameters linked to sex, age, body mass and composition, physical activity level, physiological status of pregnancy or lactation, and to possibly idiosyncratic levels of metabolic efficiency.[30–32] This means that adequacy of the food consumed by an individual should be ideally assessed against each person's own specific needs. This in turn would call for detailed information on the characteristics listed above to be able to assess the nutritional requirements of individuals. A single threshold could be used for a group only if variability of the requirements within the group is known to be very small, possibly smaller than the measurement errors that can be expected to affect the variable that is used to measure consumption, lest increasing the risk of one of two possible mistakes. Failing to recognize inadequacy when it is there, or identifying inadequacy of consumption when, in reality, it is not there. With the data usually available through household and individual surveys measuring food consumption, it is impossible to reduce the uncertainty on individual nutrient requirements to such negligible size as to prevent misclassification. This is why, for example, nutrient requirements are always specified as ranges of values within a class or population group.[j] The main consequence of this is that the extent of nutrient adequacy in a sample of individuals or in a population can only be properly assessed in probabilistic terms, which in turn requires defining a model to link the probability of inadequacy to the observed food consumption level and the other relevant characteristics of the individual.

Another problem relates to the reference period for food consumption measurement. Actual food intake can vary from day to day without serious implications for the human body, which has the ability to store energy and to use it flexibly. It is thus clear that adequacy of food intake cannot be established by considering what is eaten in a day or over a very short period. This also has nontrivial implications for data collection and methods of analysis.

In summary, when defining indicators of food security on the basis of food consumption data, one has to make several choices on various aspects and clearly define the concept of adequacy of food consumption so that it is unambiguously understood before even considering how to measure it. Acceptance of the validity of the concept should then not be confused with considerations of the feasibility of reliable assessments.

In the following section, four indicators of the inadequacy of food consumption will be considered: the FAO's estimates of the POU, indicators of caloric inadequacy based on food consumption data from household surveys, dietary diversity, and the FCS.

The FAO's estimates of prevalence of undernourishment

The first attempts at measuring the incidence of undernutrition by considering the distribution of food consumption within populations were made by the FAO as early as 1974 and published in 1977 for 70 countries, covering 90% of the world's population.[33,34]

The specific indicator, first introduced by Sukhatme in 1961,[35] is an estimate of the prevalence of people with insufficient access to food in a population. It is based on the concept of adequacy of DEC relative to dietary energy requirement (DER) and is calculated using information on availability of food for domestic consumption, food energy needs in the population, and inequality in access to food.

In practice, the distribution of food consumption in a population is represented by a probability distribution for the levels of habitual (i.e., annual averages) daily DEC of the average individual in the population.[k] Such distribution is then used to estimate the extent of food consumption inadequacy by evaluating the probability mass below a certain threshold, which is termed the prevalence of undernourishment.[36] It can be described as equivalent to the probability that a randomly selected

proper evidence on the effective level of iron intake, which improves by combining the two foods.

[j] See, for example, UNU/WHO/FAO,[31] in particular, page 5.

[k] The average or representative individual in the population is a statistical concept that summarizes the distribution of the population by age, gender, body mass, and physical activity levels.

individual in the population is found to be consuming, on a regular basis during the year, an amount of food that is insufficient to provide the energy required for an active and healthy life.

Validity. On the basis of the short description given above, it should be recognized that, contrary to widespread misconceptions, the FAO undernourishment indicator is not simply an assessment of average availability of food against requirements at the national level. Rather, it is a carefully devised statistical method to estimate the prevalence of dietary energy inadequacy in the population on the basis of a valid concept of food deprivation.[37] It cannot be deemed an indicator of household food security, however, because it is intended to be applied to the entire population and not to individual households, even if data collected from households are of crucial importance to inform the measure. Since the POU is composed of information regarding food supply as well as access, it can provide information regarding the extent to which high prevalence rates of undernourishment are due to low levels of overall food availability as opposed to particularly unequal distribution of access to food.

Reliability. Accuracy of POU estimates has been the subject of a long debate.[38–41] The main issue was the choice of the threshold used to evaluate the probability of DEC inadequacy. Critics questioned the use of a single threshold, claiming that it would imply large biases in the estimate.[40,41] It turns out that the arguments advanced in supporting this claim are based on the wrong presumption that estimates are obtained from the analysis of the marginal distribution of DEC as derived from a joint bivariate distribution, requiring observation of both DEC and DER. This criticism has been shown to be unfounded.[38,42]

Even if there is no convincing argument to claim that POU estimates may be systematically biased, questions can be raised on their precision. FAO bases their estimates on three major sources of data: (1) official data on population size and composition; (2) production and trade data from official sources; and (3) data on food acquisition from nationally representative Household Income and Expenditure Surveys (HIES). Indirect methods are used when data are insufficient to provide direct estimates of the needed parameters, as it has often been the case for the parameters describing the inequality in food consumption.[37]

Errors in the estimates may be due to the quality of the data used to inform the estimates of the parameters of the model and to the methods and assumptions made to impute them indirectly.

Population data are used both to calculate food availability per capita and to estimate the range of DER necessary to assess the threshold used for classification. Through their impact on the calculated average dietary energy supply (DES), periodic revisions of official population assessments have thus direct consequences on the POU.

Production and trade data are input into the compilation of food balance sheets (FBSs), which provide the estimate of total availability of food in a country.[42] The net supply of each food source is converted into dietary energy, and an estimate of the per capita DES is obtained by dividing the total DES by the country's population size.[43]

FBSs are routinely produced by the FAO for a large number of countries. Concerns have been raised with respect to the reliability of the underlying official data on production and trade, and of the accuracy of assessment of nonfood uses, especially of stock variations. Such concerns are particularly serious when referred to the individual components of the elements that compose an FBS, or to the overall estimate in any single year, which may be affected by measurement errors that are difficult to quantify. Efforts are made to reduce the potential impact of such errors on the POU by cross-validating the various elements of the balance, for example, by checking the consistency of feed use of crop products against assessments of the size of livestock herds in the country, or by reporting estimates as 3-year averages, thus greatly reducing the impact of inaccurate measures of yearly stock variations.

Another concern regarding the use of FBSs to estimate average DEC relates to the difficulty in accounting for food losses. Food supply, as recorded in the FBS, may overestimate the quantity available for household consumption when food losses at the level of retail distribution are not negligible, as recently suggested.[44]

Feasibility. The parametric approach informing undernourishment estimates is at the root of its feasibility. Series of POU estimates for most countries

in the world have been published yearly by the FAO since 1999. This has been possible only through use of a series of indirect methods to estimate the needed parameters in the absence of direct evidence, which may prove to be quite unreliable. When new evidence has become available, such as, for example, when suitable data from surveys have been made available to revise the indirectly estimated coefficient of variations, the series of POU have been significantly revised, for example, with the release of the 2012 edition of the State of Food Insecurity in the World.

Survey-based indicators

Spurred by the concerns on the reliability of POU estimates and by the rising popularity of using household surveys to inform poverty analyses,[45–47] food security assessments have been increasingly based on data from household surveys.[19,48,49]

In principle, if household surveys collected information on relevant concepts of food consumption and on the characteristics needed to assess the requirements of household members, it would be possible to identify households that do not consume enough food and to measure the extent of food inadequacy at the household level. Which data on household food consumption to collect and how to determine adequacy, however, have proven to be difficult questions to answer. An obvious temptation has been to apply the same concept of dietary energy inadequacy that informs the estimates of the POU at the national level in the analysis of survey data to produce individual- or household-level measures of food consumption inadequacy.

Measures of DEC can be obtained by converting the quantities of food items reported as acquired or consumed by a household or an individual into the corresponding dietary energy. Accumulated evidence suggests, however, that these are not reliable measures of the actual dietary energy intakes of individuals. A recent experiment with the Living Standard Measurement Survey, for example, demonstrates how different designs of the food consumption module can lead to differences of up to 884 kcal even in the average DEC measured over comparable groups of households in a representative survey in Tanzania.[50]

Much work thus still needs to be done to improve quantitative food consumption data collection through large-scale representative surveys in order to inform reliable measures of household food consumption adequacy.[18,51] Even if systematic measurement errors can successfully be reduced, quantitative food data collection will remain an expensive endeavor. For this reason, researchers and agencies over the past two decades have attempted to define indicators that could be informed by easier-to-collect data in the hope that it may still provide relevant information on the adequacy of food consumption.

Dietary diversity score

Dietary diversity has been considered one approach to assess adequacy of food consumption at the household level for use as a food security indicator. Quantitative indicators of dietary diversity, termed dietary diversity scores (DDSs), have been defined as simple counts of the number of food groups consumed in a reference period.[52] (For a discussion of other possible dietary diversity measures, see, among others, Drescher *et al.*[53]).

Validity. There is no doubt that variety is a key element of high-quality diets and that variety or diversity can be meaningfully measured. The relevant issue in this discussion though is whether, and to what extent, variations in the DDS reflect differences in the food security status of an individual or a household and do so in a consistent way.

The HDDS was developed by the Food and Nutrition Technical Assistance (FANTA) Project as a proxy measure of household food access,[54] whereas individual dietary diversity measures reflect nutrient adequacy of the diet, as validated against quantitative 24-h intake nutrient data. Both are defined as how many of a predefined set of different food groups are consumed by an individual or a household over a given reference period, even though the number and composition of the food groups may vary.[55]

The original validation study of household dietary diversity as a food security indicator— leading to development of the HDDS—was a 10-country analysis that showed a strong association between the number of food groups consumed by the household during the previous 24 h and measures of per capita consumption and dietary energy availability, as obtained from data collected on the same households.[56] The HDDS tool could thus be a good, and much less expensive, substitute for more

demanding consumption modules to assess food access problems in a group of households.

A subsequent validation of the HDDS was included in a three-country analysis comparing use of the FCS and different versions of DDSs, including the HDDS.[57] Both the FCS and the various DDSs were found to be significantly, although not strongly, correlated with measures of calorie consumption as gauged from 7-day household recall data.

An alternative approach has been to validate DDS as proxies for the food access dimension of food security conceptualized more broadly than simple dietary energy adequacy. A number of field studies have found that household dietary diversity is consistently associated with household food security, as measured through the Household Food Insecurity Access Scale (HFIAS), with food expenditure, and with various indicators of socioeconomic status.[58–61]

Cross-analysis of the HDDS with household food security status can indeed help to understand the impact of food insecurity on access to specific foods, thus guiding policy and programming. In a two-round survey across seasons in Central Mozambique, food-insecure households had lower fish consumption than food-secure households at both pre- and postharvest times, and had a 66% reduction in fish consumption in the postharvest time compared to a 36% reduction in food-secure households after severe flooding destroyed fish ponds.[62]

Reliability. Even though DDS values are expressed as integer numbers defined over a finite range, it is still unclear whether they comprise even an ordinal scale of the construct they are meant to capture. Lacking a formal theory that links the number of food groups consumed to levels of either nutrient adequacy or food insecurity, it is difficult to assess how DDS values obtained in different contexts can be accurate and precise indicators of the construct they are intended to capture. There are open questions remaining on the optimal number of food groups to consider or the minimal size of servings needed for consideration in a food group. No solid conclusions can be thus drawn, in general, on the reliability of DDSs as food security measures. The HDDS has been validated against household caloric availability, but no cut-off point has been defined to classify households with low or adequate household diversity.[56]

One of the major attractions of dietary diversity tools is that they can be easily computed with data obtained through relatively inexpensive collection methods as "rapid, user-friendly and easily administered low-cost assessment tools."[56] Ease of computation, however, has a potential cost in that errors—for example, in wrongly classifying food items into food groups—may induce discrete jumps in the diversity score. Reliability of the classification can be improved when food groups are adapted to locally available foods.[56]

A fair conclusion is that, although they cannot be considered comprehensive measures of food security, DDSs somehow reflect energy consumption at the household level. Analyzed together with other food security–related information, they can provide a holistic picture of the food security status and its impact on access to a diverse diet.

Food consumption score

The FCS, which has been used by the WFP to inform Comprehensive Food Security Vulnerability Assessments (CFSVA) and Emergency Food Security Assessments (EFSA), is defined as "the frequency-weighted diet diversity score calculated using the frequency of consumption of different food groups consumed by a household during the 7 days before the survey."[62]

It is computed from data on occurrence and frequency of consumption of food from different food groups, collected through surveys, and on its value used to classify households' food consumption levels as poor, borderline, and acceptable on the basis of the FCS.

Collecting simply the occurrence (yes/no) and the frequency of consumption (how many days during the last week) of a number of food groups was originally thought to be a relatively inexpensive method to obtain information on the nutritional adequacy of food consumption, in an effort to substitute more demanding food intake modules.[63]

Validity. The FCS is an ambitious attempt at defining an original concept of food consumption that would reflect both quantity and quality, on the basis of a parsimonious set of data collected through household surveys. The score aims at summarizing the three aspects reflected in the food

frequency data: dietary diversity, frequency of food group consumption, and nutritional value of food.

In principle, data reduction techniques, such as Exploratory Factor Analysis or Latent Trait Analysis,[64] could have been used to identify an underlying latent trait captured by the food consumption variables, in the hope that such a latent trait could be interpreted as the appropriate food consumption construct. Use of such techniques might have helped in defining ways to combine the observed elementary variables into a single index or score.

Multivariate data analysis techniques have been considered in the construction and interpretation of the FCS, and the Food Consumption Analysis guidelines published by the WFP still list principal component analysis (PCA) of the unweighted 7-day recall food consumption data and cluster analysis of the identified components as very important steps in defining the appropriate classifications. It remains unclear though if and how results of these analyses have been used to define the weighting scheme.[65]

The adopted solution to compute a weighted sum of the truncated frequency of consumption of eight food groups, with weights based on what has been described as "an interpretation by a team of analysts of 'nutrient density,' which is a term used to subjectively describe a food group's quality in terms of caloric density, macro and micro nutrient content, and actual quantities typically eaten,"[58,66] appears to be rather *ad hoc*.

Subsequent validation studies of the FCS[58,67] and published reports[l] do not help in resolving the ambiguity regarding which construct the FCS has been designed to capture.

Validation has been attempted in two ways: by measuring the correlation of household-level FCSs with measures of household dietary energy availability and by comparing classifications of households on the basis of the FCS with those of the same households obtained by using dietary energy inadequacy. As for the first, FCSs revealed correlations with calorie availability in line with those of various definitions of the DDS, which—as already commented—are statistically significant, but not of a magnitude sufficient to claim that the FCS would accurately predict household average DEC, a result confirmed by other studies.[68] It has been suggested that improvements in FCS ability to predict calorie consumption could be obtained by excluding food groups consumed in small quantities,[58] but the recommendation does not appear to have been adopted thus far in refining the FCS methodology, as it seems to be difficult in practice to exclude small quantities.

Regarding the second type of validation, cross-tabulation of food consumption groups based on the FCS with those based on calorie consumption suffer from the drawbacks that are (1) conditional on acceptance of the appropriateness of the thresholds used to classify caloric consumption, and (2) highly dependent on the particular dataset being used.

Similar exercises have been used to fine-tune the thresholds used for classification with the FCS for specific datasets. No thresholds could be found that would yield classifications close to those obtained with caloric consumption thresholds, even for different areas in a country.[69]

All of these attempts have been, thus far, disappointing. It has not yet been possible to identify suitable thresholds, leading to the conclusion that the current FCS needs to be improved and that, in any case, it will be necessary "to go beyond the score" to provide inputs into nutrition-sensitive programming.[69] This raises serious concerns regarding the possibility of creating a globally valid standard classification on the basis of the FCS.

Reliability. According to the WFP, the FCS has been a "reliable indicator of food insecurity in all the CFSVAs and EFSAs where it has been applied."[69] No information is provided, however, to understand how reliability is assessed. As is the case for the simpler HDDS, it is not clear that the FCS comprises even a simple ordinal scale of food consumption adequacy over the numeric range where it is defined. Measures of accuracy or precision would require that the actual scale of adequacy be defined in a way that would allow for an assessment of the likely extent and direction of errors in measuring the underlying construct, something that is simply impossible given the current definition of the indicator and the existing evidence.

[l] A query of WFP's Survey Data Portal (http://nada.vam.wfp.org/index.php/catalog) involving a search for "Food Consumption Score" in the variable descriptions returned 24 surveys (query run on September 14, 2014).

The FCS lacks minimal characteristics, starting from a proper definition of the construct it aims at measuring, to assess precision and reliability of the measures and to ensure comparability across applications. These properties are considered crucial for a measure of food security to be adopted in an international monitoring framework.

Experience-based food insecurity scales

The problems with the analysis of food consumption data, and the need for measures encompassing a broader concept of food insecurity, created the context within which experience-based food insecurity scales emerged.

Ethnographic research carried out in the United States to understand the experience of food insecurity and hunger among low-income households revealed it to be a process initially characterized by uncertainty and anxiety, associated with worry about being able to get enough food. As conditions worsened, it resulted in a decreased amount of stored food in the home, followed by worsening quality and variety of the diet. With increasing severity of the food insecurity condition, the quantity of food consumed per meal was decreased, portion sizes were reduced, and households were forced to skip meals. In the more severe situations, adults and children experienced hunger and did not eat for a whole day or more. Different experiential domains—uncertainty and worry about food, consumption of low-quality food or unbalanced diets, and reductions in the quantity of food consumed—were identified as defining characteristics of household food insecurity.[70–72]

This conceptual understanding of the experience of food insecurity formed the basis to develop the U.S. Household Food Security Survey Module (U.S. HFSSM), which has been applied annually in the United States to monitor the food security situation since 1995,[73] and has served as a model for the creation of many other household experience-based food insecurity scales around the world,[74–77] including for the Food Insecurity Experience Scale (FIES) recently developed at the FAO.[78]

Validity. The substantial innovation brought about by experience-based food insecurity measures is the possibility of capturing the food insecurity situation directly and analyzing it from a behavioral perspective, on the basis of the well-grounded assumptions that (1) food insecurity is a condition that may affect overall welfare in different ways, depending on the type and strength of the constraints faced; (2) food insecurity is characterized by behavioral responses to the different conditions, and (3) both the conditions that characterize it and the overall welfare impact of food insecurity can be located along a scale of severity. This makes it possible to speak legitimately of subjects not only with respect to being food insecure or not, but also as being more or less food insecure than others, depending on the situations they report to be experiencing.

From an analytic point of view, the approach is founded on conceptualization of the severity of food insecurity as a latent attribute that, even though inherently unobservable, can be measured over a one-dimensional scale.

Dealing with latent trait measurement significantly changes the perspective of how to look at validity (and reliability) of measurement tools.[79] In this context, a precise causal direction running from the concept to the evidence, from the measurand to the data, defines validity; namely, that changes in the latent trait induce detectable changes in the data used for measurement.[80] When applied to experience-based food security scales, this is equivalent to requiring that, for example, a higher severity of food insecurity increases the probability of occurrence of experiences in those domains, which will therefore be more frequently observed in representative samples of the population.

Validity of a particular experience-based scale as a food security measurement tool thus depends on two conditions: (1) that the severity of food insecurity indeed involves the domains that have been considered in creating the items that compose the scales, and (2) that the occurrence of experiences in those domains can be reliably detected and meaningfully linked, albeit in a probabilistic sense, to the severity of food insecurity.

The existence of fundamental domains of food insecurity experiences as identified by the pioneering ethnographic research in the United States has been extensively validated. Additional ethnographic studies aimed at understanding the experience of hunger from the perspective of the elderly, and of immigrants in the United States and low-income families in Quebec, Canada, revealed similar patterns.[81–85] Although the original ethnographic study was based on a small number of households in a wealthy country, a review conducted years later

of experience-based scales in over 20 countries and settings around the world concluded that these dimensions of the experience of hunger appear to be common across cultures.[78] These analyses suggest that these common domains are robust across cultures, in their broad definition and ranking of severity.[78,86,87]

Verification of the second requirement, namely to theoretically and empirically validate the link between observed experiences and a measure of food insecurity severity, is possible through application of the Rasch model,[88] a probabilistic measurement method pioneered in psychometrics and educational testing that bears fundamental similarities with latent variable analysis and discrete choice models in econometrics.[89–91] As applied to food insecurity, the Rasch model assumes that the probability that a respondent characterized by a certain level of food insecurity would report having experienced the condition described by one of the items can be represented by an increasing function of the distance between the respondent's position on the scale and that of the item.[92] The model further assumes that items are ranked in the same order of severity by all respondents independently of their food insecurity condition, and that the more food insecure a respondent is, the more likely it is that he or she would affirm any given item.

This rather simple logic defines the criteria through which the relative severity both of the items and of the respondents can be estimated and, most importantly, allows evaluating the empirical validity of any given set of items as contributing to the measure of a single, underlying one-dimensional trait. This is obtained by evaluating the degree of consistency between observed patterns of responses and those that would be expected if the restrictions imposed by the theory were true.

The logic behind the analytic procedure used to reach a conclusion on the validity of the measures is as follows: first, as the probability of affirming an item is expected to decrease with the severity of the experience it represents, the order of severity of the various items included in the scale is derived from the shares of respondents that affirm them.[m] Item Response Theory (IRT) in general and the Rasch model in particular provide formal tests of adherence of the data to the theoretical requirements for invariant measurement through the analysis of item misfit statistics.[93] Measures of item *infit* statistics and of reliability can be used to ascertain whether all of the items included in the scale, or at least a subset of them, conform to the requirements of the theory. Virtually all studies reporting applications of experience on the basis of food insecurity measures in a broad range of countries and conditions include an analysis of infit statistics confirming, in the vast majority of the cases, adequacy of the item chosen.[93–96,n] In a recent application of the eight-item FIES questionnaire in 20 countries across Latin America, Africa, and Asia, only one of the items in just one instance revealed an infit value of 1.4, slightly higher than ideal, but still productive for measurement. All other infits were in the range 0.84–1.24, indicating minimal violation of the assumptions that the Rasch model imposes, and confirming that the current version of the FIES is indeed adequate for internationally comparable use.[97]

In conclusion, we can say that there are strong conceptual and empirical bases to claim that the fundamental construct underlying these tools, best described as "resource-constraint food insecurity"[98] is indeed a valid and measurable concept.

Reliability. It has been a rather standard practice to test accuracy of the experience-based scales by studying associations between the measures obtained with the scale and the classifications they produce with those obtained using other variables that are theoretically part of the same construct food insecurity and/or that would vary in an expected way across different levels of food insecurity, when these variables are measured concurrently in the same subjects.

[m] This makes sense, because, unless there is a problem in understanding the questions, the fewer the number of people report having suffered from an experience, the more severe (on average) that experience must be.

[n] Of the various different categories of items that had been tested for use in food security scales, only those related to the experience of having to procure food in socially unacceptable ways were found to be inadequate, as they would not meet the required standards by showing infit values outside the range 0.8–1.2. See Coates *et al.*[77]

National studies conducted in Brazil,[99,100] Mexico,[101] Bolivia, Burkina Faso, and the Philippines,[102] showed an inverse relationship between the severity of household food insecurity and the consumption or purchase of high nutritional quality foods. A study in Albania showed that dietary diversity (number of food groups consumed in a 24-h period) in three vulnerable areas declined with increasing severity of household food insecurity, consistent with the underlying theoretical construct of food insecurity.[103] Similar results had been previously found in rural Mexico.[104] Explorations have been made of associations between biological indicators of nutritional status and experience-based measures of food insecurity. One study conducted with children in Colombia reported increasing probability of being stunted with increasing severity of food insecurity, as measured by an experience-based scale.[105,o] Interestingly, food insecurity classifications obtained with experience-based food security scales have shown correlations with overweight and obesity, especially among women in the United States[106] but also in some developing countries, such as Brazil and Mexico.[107,108]

While these associations may be informative for programmatic purposes, examples of concurrent validation cannot be deemed sufficient evidence of validity or reliability of a food security measurement tool, as we have argued earlier. Food consumption and nutritional status are not prototype gold standards for food security. The experience of food insecurity, even in the absence of food energy deficits or negative effects on nutritional status, is still a serious problem in itself, indicating a violation of the Human Right to Adequate Food and with potentially negative consequences for well-being.[79]

Reliability must thus be assessed in a different way, and IRT provides the means to do so. We have linked the validity of experience-based food security scales to the empirical consistency between the data and the latent trait being measured. Notice that this does not mean, however, that each and every set of responses to an experience-based questionnaire must be fully consistent, since reliability can be assessed as a measure of the degree of empirical consistency. Discrepancies in actual responses in relation to the pattern expected according to the underlying theory (i.e., a respondent who affirms a given item also affirms all less severe items, and upon denying a given item, also denies all the more severe ones) may exist without undermining validity, as they can be attributed either to normal sample variability or to measurement error. IRT models make it possible to assess the size and distribution of such discrepancies to draw conclusions on the reliability of the measures obtained. This is one of the aspects that sets these methods apart from the other methods to measure food security that have been presented in this paper. This has very relevant implications for the possibility of defining a proper global reference standard for food security measurement and the comparability of measures obtained in different countries and different years. Analyzing patterns of responses through the lenses of the theory can guide the choice of which items to include in a global reference scale and which ones to consider common when reporting measures in one country or in 1 year, as compared to such common standards.

This is still a relatively unexplored feature of these methods but results thus far are extremely encouraging. Beginning in 2014, the FAO has inserted the FIES into the Gallup® World Poll (GWP), a branch of Gallup, Inc. conducting nationally representative surveys annually in more than 150 countries. Preliminary analysis of the results obtained from 20 countries revealed measures of Rasch reliability ranging from 0.69 to 0.78, with a median of 0.73 (Rasch is a measure of overall fit of the data to the measurement model, with theoretical values ranging from 0 to 1, with 1 indicating a perfect fit). This is a remarkable result, given that the countries covered, including Russia, Brazil, Indonesia, and Sierra Leone, among others, are very different from each other with respect to food insecurity prevalence.[109] The same data have led the FAO to define a provisional version of a global food insecurity severity reference scale on the basis of the FIES that can be used for equating prevalence of food security at medium and high levels of severity across countries and years. Research is also ongoing to develop

[o]Results of such comparisons are likely to be misleading and difficult to interpret, particularly when food insecurity refers to the condition of adults, or when the timing of the assessments of food insecurity and nutritional status is not properly considered.

methods to equate prevalence levels obtained through other experience-based food security scales used in many countries, such as national scales developed in Brazil and Mexico, and the ELCSA and the HFIAS used in many countries in Latin America, Asia, and Africa.

Conclusions and Recommendations

An overview of the evolution of the international debate regarding the concept of food security has revealed how the problems of understanding, monitoring, and communicating about it have been deeply interlinked.

The measurement issue has received much attention and many indicators have been proposed and applied over the years in recognition of the complexity of food security as defined by the international community. But, as we have argued, this has happened often without due consideration of the theoretical requirements to define proper measures.

Even though the daily practice of fighting food insecurity does not allow for exhaustive and long-term inquiries on validity and reliability, which may be practical only within the most rigorous academic environments, a solid understanding by those proposing the different indicators of the importance of adherence to minimal quality criteria for measurement is imperative.

To shed light on the issue, we have discussed a few of the methods proposed to compile indicators of food security and have critically evaluated the available evidence regarding their validity and reliability. Many proxy indicators of food insecurity have been based on food consumption data. However, feasibility of collecting sufficiently precise data to allow reliable assessment on a regular basis, and with enough disaggregation of the data to be useful in providing meaningful policy guidance, is compromised by the burden it may impose on the resources available to countries for such an endeavor.

The need for cost effectiveness and timeliness of reporting has driven the search for innovative ways to obtain reliable measures with universal applicability. For food consumption, attempts have been made at defining relatively simple indicators—the HDDS and the FCS—to be applied at the household level. These are based on concepts of food adequacy that explicitly include dietary quality considerations as captured by dietary diversity, and they have the advantage of being relatively parsimonious with regard to data collection. Unfortunately, the lack of an explicit model linking dietary diversity and food consumption frequency to food consumption adequacy and food security has prevented, thus far, establishing their validity as worldwide comparable measures of food access at the household or individual levels.

Improvements in measures of food consumption adequacy are certainly necessary and worthwhile. However, from the perspective of assessing food insecurity, they will always leave something to be desired. These types of measures are consistent with an approach that relies on proxy measures. An analytic framework to link different aspects of the food consumption experience—such as the quantity and quality of food consumed, or the extent and frequency of episodes of inadequacy—is necessary to interpret them as relevant for the overall food security condition of an individual or household.

The use of experienced-based food insecurity scales, from this perspective, has indeed been a paradigmatic change, both conceptually and in practice. It has provided a clear unifying concept leading to the development of direct measures of food insecurity severity at the household and individual levels. It puts people's experiences and behavioral responses at the core of the definition of what food insecurity means, promoting the shift toward "third generational indicators" that had been invoked long ago.[110,111] Recent experience shows not only that these tools produce valid, cross-country comparable measures, but also that they can be applied in a relatively simple and inexpensive way.

Indicators of the severity of experienced food insecurity can be used within broader concepts of multidimensional poverty, consistently with a focus on unequal access to food and sociocultural aspects of the experience of hunger,[112] and can enhance investigation of associations between food insecurity and various forms of malnutrition, including both under- and overnutrition. Food insecurity can affect health and well-being in many ways, with potentially negative mental and social consequences, in addition to physical well-being, even in the absence of measurable negative effects on nutritional status. One of the unique contributions of experience-based food security measures is that they also capture psychosocial aspects associated with anxiety or uncertainty regarding the ability to procure enough

food, an element of the experience of hunger and food insecurity that other measures do not consider.

It is imperative that those in charge of food security assessments understand in depth the underlying conceptual framework and therefore the pros and cons of the different indicators. Moreover, they need to assure that the framework is understood by those interpreting the outcomes of the assessments and by those designing the food security programs and policies. One of the results of our discussion on validity of current approaches to food security measures is that validating indicators on the basis of their association with other "food security" measures cannot be deemed conclusive, because it is not possible to determine which of the compared measures validates the other. Although an adequate assessment of food insecurity need not be restricted to the indicators described above, it is worrisome that the measurement field is widely populated by an array of indicators that have not been scrutinized to the level required by the importance and magnitude of the phenomenon that we aim to measure and address.[113]

In order to promote food security more effectively, it is imperative to be able to measure it in a more valid and reliable way. The lack of a gold standard should not discourage validation efforts, which would greatly contribute to identifying ways of improving the suitability of the proposed tools and to designing programs that better target those in need. Given the very limited resources available in daily practice, easy-to-apply tools are certainly in high demand. Nevertheless, this reality should not be seen as justification for acquiring and applying tools only because of their low cost.

On the basis of the evidence we have reviewed, we conclude that experience-based food insecurity measures are the most promising tools among those currently available and constitute a valuable instrument that can be applied together with other indicators to better understand the determinants and consequences of household and individual food insecurity. They are founded on a valid concept of food insecurity that covers common domains across cultures and socioeconomic conditions and thus have the potential to form the basis for a valid measure worldwide. Moreover, they promise to effectively combine desirable features of cost effectiveness and rapidity of monitoring with sufficient precision and reliability, and to do so in a way that is applicable worldwide and at the global, national, and subnational levels. The work that the FAO is conducting to promote the FIES will go a long way toward providing the necessary evidence to guide successful practical implementation of a global measurement standard. Research is demonstrating that the FIES, even in its current form, is indeed capable of comparably measuring the prevalence of food insecurity at different levels of severity across populations that differ in their linguistic, cultural, and socioeconomic aspects. Results so far are extremely encouraging and point to the possibility that a global standard to measure household and individual food insecurity is possible. This is a worthy endeavor, and we can anticipate that the development of such a global standard will generate enormous benefits for researchers, analysts, advocates, and policy makers interested in eradicating food insecurity and hunger by providing the necessary tool to monitor progress in a timely and consistent way, both within and across countries.

Conflicts of interest

C.C., T.B., and A.K. are affiliated with the FAO, which is involved in food security measurement.

References

1. FAO. 1946. *World Food Survey*. Food and Agriculture Organization of the United Nations, Washington, DC.
2. FAO. 1952. *Second World Food Survey*. Food and Agriculture Organization of the United Nations, Rome, Italy.
3. FAO. 1963. *Third World Food Survey. Freedom from Hunger Campaign Basic Study No. 11*. Food and Agriculture Organization of the United Nations, Rome, Italy.
4. Sen. 1983. *Poverty and Famines: An Essay on Entitlement and Deprivation*. Oxford University Press.
5. Shaw, D.J. 2007. *World Food Security. A History since 1945*. Palgrave McMillan.
6. Gibson, M. *The Feeding of Nations. Redefining food security for the 21st century*. Boca Raton, FL: CRC Press.
7. Maxwell, S. 1992. Food security: a post-modern perspective. *Food Policy* **21:** 155–170.
8. Hoddinott, J. 1999. *Choosing Outcome Indicators of Household Food Security. Technical Guide No. 7*. Washington, DC: International Food Policy Research Institute.
9. Rome declaration on world food security. http://www.fao.org/docrep/003/w3613e/w3613e00.HTM.
10. U.N. General Assembly. 2001. Road map towards the implementation of the United Nations Millennium Declaration. Report of the Secretary-General. http://millenniumindicators.un.org/unsd/mdg/Resources/Static/Products/SGReports/56_326/a_56_326e.pdf.
11. FAO. 1996. The sixth world food survey.

12. UNICEF. The state of world children. http://www.unicef.org/sowc/.
13. FAO. 2003. Measurement and assessment of food deprivation and undernutrition. http://www.fao.org/docrep/005/Y4249E/y4249e00.htm.
14. FAO. 1996. *Rome Declaration on World Food Security and World Food Summit Plan of Action.* World Food Summit 13–17 November 1996. Rome, Italy.
15. de Haen, H., S. Klasen & M. Qaim. What do we really know? Metrics for food insecurity and undernutrition. *Food Policy* **36:** 760–769.
16. Coates, J. 2013. Build it back better: deconstructing food security for improved measurement and action. *Global Food Security* **2:** 188–194.
17. Jones, A.D., F.M. Ngure, G. Pelto & S.L. Young. 2013. What are we assessing when we measure food security? A compendium and review of current metrics. *Adv. Nutr.* **4:** 481–505; doi:10.3945/an.113.004119.
18. Carletto, G., A. Zezza & R. Banerjee. 2013. Towards better measurement of household food security: harmonizing indicators and the role of household surveys. *Global Food Security* **2:** 30–40.
19. Perez-Escamilla, R. & A.M. Segall-Correa. 2008. Food insecurity measurement and indicators. *Rev. Nutr. [online]*, **21**(Suppl.): 15s–26s.
20. Frongillo, E. 1999. Validation of measures of food insecurity and hunger. *J. Nutr.* **129:** 506S–509S.
21. International Vocabulary of Metrology—Basic and General Concepts and Associated Terms (VIM), 3rd ed. http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf.
22. http://en.wikipedia.org/wiki/Validity_(statistics). (Italics added).
23. Markus, K.A. & D. Borsboom. 2013. *Frontiers of Test Validity Theory. Measurement, Causation, and Meaning. Routledge Multivariate Application Series. Section 2.5.* New York: Routledge.
24. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. 2009. *Standards for Educational And Psychological Testing.* Washington, DC: American Educational Research Association.
25. Cook, D.A. & T.J. Beckman. 2006. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am. J. Med.* **119:** 166.e7–166.e16.
26. Borsboom, D. 2005. *Measuring the Mind.* Cambridge, UK: Cambridge University Press.
27. International vocabulary of metrology (cit.), definition 2.3, p. 17.
28. Gates, B. 2013. Improving the lives of the poor will take a steady, thoughtful revolution in development data. Development Asia. November.
29. FAO. 2012 *The State of Food Insecurity in the World 2012. Economic Growth is Necessary but not Sufficient to Accelerate Reduction of Hunger and Malnutrition. Technical Note.* http://www.fao.org/fileadmin/templates/es/SOFI_2012/sofi_technical_note.pdf.
30. WHO/FAO/UNU. 2004. *Human Energy Requirements. Report of a Joint FAO/WHO/UNU Expert Consultation, Rome, Italy, 17–24 October 2001.* Rome: FAO.
31. WHO/FAO/UNU. 2007. Protein and Amino Acid Requirements in Human Nutrition. *Report of a joint FAO/WHO/UNU expert consultation (WHO Technical Report Series 935).* Geneva: WHO.
32. Burlingame, B., C. Nishida, R. Uauy & R. Weisell. 2009. Fats and fatty acids in human in nutrition. Joint FAO/WHO Expert Consultation. November 10–14, 2008, Geneva, Switzerland. Special Issue. *Ann. Nutr. Metab.* **55:** 5–300.
33. FAO. 1975. "Population, food supply and agricultural development." In "*The State of Food and Agriculture 1974.*" Chap. 3, pp. 93–196. See in particular the sections titled "Dimensions and causes of hunger and malnutrition," pp. 103–111, and "Indicators of the incidence of undernutrition," pp. 149–151.
34. FAO. 1977. *The fourth world food survey. FAO Statistics Series No. 11 and FAO Food and Nutrition Series No. 10.* Food and Agriculture Organization of the United Nations, Rome, Italy. See in particular "*Part II: Dimensions of malnutrition.*"
35. Sukhatme, P.V. 1961. The World's hunger and future needs in food supplies. *J. R. Stat. Soc. Ser. A (Gen.)* **124:** 463–525.
36. Naiken. 2003. FAO methodology for estimating the prevalence of undernourishment. In *Proceedings of the International Scientific Symposium on Measurement and Assessment of Food Deprivation and Undernutrition*, 26–28 Jun 2002, FAO, Rome, Italy. Agriculture and Economic Development Analysis Division, pp. 7–42. http://www.fao.org/docrep/005/Y4249E/y4249e06.htm#bm06.
37. Cafiero, C. What Do We Really Know About Food Security? National Bureau of Economic Research Working Paper 18861. http://www.nber.org/papers/w18861.
38. Osmani, S.R. 1992. *Nutrition and Poverty. UNU-WIDER Studies in Development Economics.* Oxford: Clarendon Press.
39. Svedberg, P. 1999. 841 million undernourished? *World Develop.* **27:** 2081–2098.
40. Svedberg, P. 2000. *Poverty and Undernutrition. Theory, Measurement and Policy. UNU-WIDER Studies in Development Economics.* Oxford University Press.
41. Naiken, L. 2007. The Probability Distribution Framework for Estimating the Prevalence of Undernourishment: Exploding the Myth of the Bivariate Distribution. FAO Statistics Division Working Paper Series No. ESS/ESSG/009e FAO, Rome, August. http://www.fao.org/fileadmin/templates/ess/documents/food_security_statistics/working_paper_series/WP009e.pdf.
42. http://faostat.fao.org/site/368/default.aspx.
43. FAO. 2001. *Food Balance Sheets. A Handbook.* Rome. http://www.fao.org/docrep/003/x9892e/x9892e00.htm. Food and Agriculture Organization of the United Nations, Rome, Italy.
44. FAO. 2011. Global Food Losses and Food Waste – Extent, Causes and Prevention. Food and Agriculture Organization of the United Nations, Rome.
45. Deaton, A. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* Baltimore: Johns Hopkins University Press.
46. Haughton, D. & J. Haughton. 2011. *Living Standards Analytics Development through the Lens of Household Survey Data.* New York: Springer.

47. Foster, J., S. Set, M. Lokshin & Z. Suraja. 2012. *A Unified Approach to Measuring Poverty and Inequality. Theory and Practice. Streamlined Analysis with ADePT Software*. Washington, DC: World Bank.
48. Smith, L.C., H. Alderman & D. Aduayom. 2006. *food insecurity in sub-Saharan Africa. New estimates from household expenditure surveys. Research Report 146*. International Food Policy Research Institute, Washington, DC.
49. Moltedo, A., N. Troubat, M. Lokshin & Z. Suraja. 2014. *Analyzing Food Security Using Household Survey Data: Streamlined Analysis with ADePT Software*. Washington, DC: World Bank.
50. De Weerdt, J., K. Beegle, J. Friedman & J. Gibson. 2014. *The Challenge of Measuring Hunger. Policy Research Working Paper Series 6736*. The World Bank, Washington, DC.
51. Smith, L.C., O Dupriez & N. Troubat. 2013. Assessment of the reliability and relevance of the food data collected in national household consumption and expenditure surveys. International Household Survey Network, Food and Agriculture Organization of the United Nations and World Bank. February 2014.
52. Kant, A.K., A. Schatzkin, T.B. Haris, *et al.* 1993. Dietary diversity and subsequent mortality in the First National Health and Nutriton Examination Survey Epidemiologic Follow-up Study. *Am. J. Clin. Nutr.* **57:** 434–440.
53. Drescher, L.S., S. Thiele & G.B.M. Mensik. 2007. A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *J. Nutr.* **137:** 647–651.
54. Swindale, A. & P. Bilinsky. 2006. *Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide (v.2)*. Washington, DC: FHI 360/FANTA.
55. FAO. 2010. *Guidelines for Measuring Household and Individual Dietary Diversity*. Rome. http://www.fao.org/docrep/014/i1983e/i1983e00.pdf.
56. Hoddinott, J. & Y. Yohannes. 2002. Dietary diversity as a household food security indicator. Food and Nutrition Technical Assistance Project, Academy for Educational Development, Washington, DC
57. Wiesmann, D., L. Bassett, T. Benson & J. Hoddinott. 2009. *Validation of the World Food Programme's Food Consumption Score and Alternative Indicators of Household Food Security*. IFPRI Discussion Paper 00870. International Food Policy Research Institute (IFPRI), Washington, DC
58. Kennedy, G., M. Razes & T. Ballard. 2014. "Measurement of dietary diversity for monitoring the impact of food based approaches." In *Improving Diets and Nutrition: Food Based Approaches*. B. Thompson & L. Amoroso, Eds. Boston , MA: FAO/CABI.
59. Thorne-Lyman, A.L., N. Valpiani, K. Sun, *et al.* 2010. Household dietary diversity and food expenditures are closely linked in rural Bangladesh, increasing the risk of malnutrition due to the financial crisis. *J. Nutr.* **140:** 182S–188S.
60. Taruvinga, A., V. Muchenje & A. Mushunje. 2013. Determinants of rural household dietary diversity: the case of Amatole and Nyandeni districts, South Africa. *Intl. J. Develop. Sustain.* **2:** 2233–2247.
61. FAO. 2008. Report on use of the household food insecurity access scale and household dietary diversity score in two survey rounds in Manica and Sofala Provinces, Mozambique, 2006–2007. http://www.foodsec.org/fileadmin/user_upload/eufaofsi4dm/docs/moz_diet.pdf.
62. World Food Program-VAM. 2008. Food Consumption Analysis: Calculation and Use of the Food Consumption Score in Food Security Analysis. p. 8. http://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp197216.pdf.
63. Kennedy, G.L., A. Berardo, C. Papavero, *et al.* 2010. Proxy measures of household food consumption for food security assessment and surveillance: comparison of the household dietary diversity and food consumption scores. *Public Health Nutr.* **13:** 2010–2018.
64. Bartholomew, D.J., F. Steele, I. Moustaki & J. Galbraith. 2008. *Analysis of Multivariate Social Science Data*, 2nd ed. Boca Raton, FL: CRC Press.
65. World Food Program-VAM. 2008. Food Consumption Analysis. Calculation and Use of the Food Consumption Score in Food Security Analysis. p. 5.
66. World Food Program-VAM. 2008. Food Consumption Analysis. Calculation and Use of the Food Consumption Score in Food Security Analysis. http://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp197216.pdf. p. 19, footnote 14.
67. Tiwari, S., E. Skoufias & M. Sherpa. 2013. *Shorter, Cheaper, Quicker, Better: Linking Measures of Household Food Security to Nutritional Outcomes in Bangladesh, Nepal, Pakistan, Uganda, and Tanzania. Policy Research Working Paper No. WPS 6584*. World Bank, Washington, DC.
68. Mathiassen, A., S. Montembault & K. Ogden. 2012. Household Food Consumption: looking beyond the score. *Presentation delivered at the International Conference on Dietary and Physical Activity Methods (ICDAM), Rome 17.5.2012, Session B-6 Measuring Dietary Diversity*. http://www.fao.org/fileadmin/templates/nutrition_assessment/Oral_presentations/B6_Mathiassen.pdf.
69. World Food Program-VAM. 2009. Comprehensive Food Security & Vulnerability Analysis Guidelines. 1st ed. http://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp203208.pdf, p. 212.
70. Radimer, K., C.M. Olson & C.C. Campbell. 1990. Development of indicators to assess hunger. *J. Nutr.* **120**(Suppl.): 1544S–1548S.
71. Radimer, K.L., C.M. Olson, J.C. Green, *et al.* 1992. Understanding hunger and developing indicators to assess it in women and children. *J. Nutr. Educ.* **24**(Suppl.): 36S–44S.
72. Radimer, K.L. 2002. Measurement of household food security in the USA and other industrialized countries. *Public Health Nutr.* **5**(6A): 859–864. DOI: http://dx.doi.org/10.1079/PHN2002385.
73. Hamilton, W., J.T. Cook, W.W. Thompson, *et al.* 1997. Household food security in the United States in 1995: technical report. Technical Report prepared for the USDA Food and Consumer Service, Alexandria, VA.
74. Segall-Corrêa, A.M., L. Marín-León & R. Pérez-Escamilla. 2010. "Escala Brasileira de Medida da Insegurança

Alimentar e Nutricional – EBIA: Validação, usos e importância para políticas públicas". In *Fome Zero. Uma História Brasileira*, Vol. **3**, pp. 26–43. Brasil: Ministério de Desenvolvimento Social e Combate à Fome.
75. Deitchler, M., T.J. Ballard, A. Swindale & J. Coates. 2010. *Validation of a measure of household hunger for cross-cultural use. Technical Report.* Washington, DC: FANTA-2. Academy for Educational Development, Washington, DC.
76. Pérez-Escamilla, R., H. Melgar-Quiñonez, M. Nord, M.C. *et al.* 2007. Escala Latinoamericana y Caribeña de Seguridad Alimentaria (ELCSA) [Latinamerican and Caribbean Food Security Scale]. *Perspectivas en Nutrición Humana, Colombia* **2007**(Suppl.): 117–134.
77. Coates, J., E.A. Frongillo, B.L. Rogers, *et al.* 2006. Commonalities in the experience of household food insecurity across cultures: what are measures missing? *J. Nutr.* **136:** 1438S.
78. Ballard T.J., A.W. Kepple & C. Cafiero. 2013. The Food Insecurity Experience Scale. Development of a Global Standard for Monitoring Hunger Worldwide. Rome: Food and Agriculture Organization of the United Nations. http://www.fao.org/3/a-as583e.pdf.
79. See Borsboom (cit.), Chapter 3, and Markus and Borsboom (cit.), section 2.5.
80. See Borsboom (cit.), Chapter 6.
81. Wolfe, W.S., C.M. Olson, A. Kendall & E.A. Frongillo Jr. 1996. Understanding food insecurity in the elderly: a conceptual framework. *J. Nutr. Educ.* **28:** 92–100.
82. Wolfe, W.S., E.A. Frongillo & P. Valois. 2003. Understanding the experience of food insecurity by elders suggests ways to improve its measurement. *J. Nutr.* **133:** 2762–2769.
83. Hamelin, A.M., M. Beaudry & J.P. Habicht. 2002. Characterization of household food insecurity in Québec: food and feelings. *Soc. Sci. Med.* **54:** 119–132.
84. Kaiser L.L., H. Melgar-Quiñonez, C. Lamp, *et al.* 2002. Food security and nutritional outcomes of Latino Preschoolers. *J. Am. Diet. Assoc.* **102:** 924–929.
85. Kaiser L.L., H. Melgar-Quiñonez, M.S. Townsend, *et al.* 2003. Food insecurity and food supplies in Latino households with young children. *J. Nutr. Educ. Behav.* **35:** 148–153.
86. Swindale, A. & P. Bilinsky. 2006. Development of a universally applicable household food insecurity measurement tool: process, current status, and outstanding issues. *J. Nutr.* **136:** 1449S–1452S.
87. Melgar-Quiñonez, H., M.C. Alvarez Uribe, Z.Y. Fonseca Centeno, *et al.* 2010. Psychometric characteristics of the Food Security Scale (ELCSA) applied in Colombia, Guatemala y México. *Revista de Segurança Alimentar e Nutricional, Campinas* **17:** 48–60.
88. Rasch, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: Danish Institute for Educational Research. Reprinted with a foreword and afterword by B.D. Wright in 1980 by the University of Chicago Press, and in 1993 by MESA Press, Chicago, IL.
89. McFadden. 1980. Econometric models for probabilistic choice among products, *J. Business* **53:** S13–S29.
90. Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge, UK: Cambridge University Press (for a comprehensive review).
91. Train, K. 2009. *Discrete Choice Methods with Simulation.* 2nd ed. New York: Cambridge University Press.
92. Nord, M. 2014. *Introduction to Item Response Theory applied to Food Security Measurement: Basic Concepts, Parameters, and Statistics.* Technical Paper. Rome: FAO. http://www.fao.org/economic/ess/ess-fs/voices/en.
93. Alvarez, M.C., C. Montoya, A. Estrada, H. Melgar-Quiñonez. 2006. Validación de escala de percepción de la seguridad alimentaria doméstica en Antioquia, Colombia. *Salud Publica Mex* **48:** 474–481.
94. Melgar-Quiñonez, H. & M. Hackett.2008. Measuring household food security: the global experience. *Rev. Nutr.* **21:** 27s–37s.
95. Nord, M. 2012. Assessing potential technical enhancements to the u.s. household food security measures. Economic Research Service of the United States Department of Agriculture Technical Bulletin No. (TB-1936), 130 pp.
96. Faye, O, A. Baschieri, J. Falkingham & K. Muindi. 2011. Hunger and food insecurity in Nairobi's Slums: an assessment using IRT models. *J. Urban Health* **88**(Suppl. 2): 235–255.
97. FAO. 2014. Validity of the Food Insecurity Experience Scale. Preliminary results from application through the Gallup World Poll™. Mimeo. Rome: FAO.
98. Ohls, J., L. Radbill & A. Schirm. 2001. Household food security in the United States, 1995–1997: technical issues and statistical report. Final Report of the Project to Analyze 1996 and 1997 Food Security Data. December 2001, p. 6.
99. Segall-Corrêa, A.M., R. Pérez-Escamilla, L. Marín-León, *et al.* 2008. "Evaluation of household insecurity in Brazil: validity assessment in diverse sociocultural settings." In *Memoria del Primer Concurso de Investigaciones REDSAN*, Santiago, Chile.
100. Panigassi, G., A.M. Segall-Corrêa, L. Marin-León, *et al.* 2008. Food insecurity as an indicator of inequity: analysis of a population survey. *Cadernos de Saude Publica* **24:** 2376.
101. Pérez-Escamilla, R., P. Parás & R. Vianna. 2012. Food security measurement through public opinion polls: the case of ELCSA-Mexico. *Presented at the International Scientific Symposium on "Food and Nutrition Security Information: From valid measurement to effective decision-making," 17–19 January.* Rome: FAO.
102. Melgar-Quinonez, H.R., A.C. Zubieta, B. MkNelly, *et al.* 2006. Household food insecurity and daily per capita food expenditure in Bolivia, Burkina Faso and the Philippines. *J. Nutr.* **136:** 1431S–1437S.
103. United Nations Albania. Baseline Nutrition and Food Security Survey. 2012. Tirana. http://www.unicef.org/albania/Baseline_Nutrition_and_Food_Security_Survey.pdf.
104. Melgar-Quiñonez, H., A.C. Zubieta, E. Valdez, *et al.* 2005. Desarrollo de un instrumento para monitorear inseguridad alimentaria en la Sierra de Manantlán, Jalisco. *Salud Publica Mex.* **47:** 413–422.
105. Hackett, M., H. Melgar-Quiñonez & M.C. Álvarez. 2009. Household food insecurity is associated with

stunting and underweight among preschool children in Antioquia, Colombia. *Pan Am. J. Public Health* **25:** 506–510.
106. Townsend, M.S., J. Peerson, B. Love, C. Achterberg & S.P. Murphy. 2001. Food insecurity is positively related to overweight in women. *J. Nutr.* **131:** 1738–1745.
107. Velásquez-Melendez, G., M.M. Schlüssel, A.S. Brito, A.A. Silva, *et al.* 2011. Mild but not light or severe food insecurity is associated with obesity among Brazilian women. *J. Nutr.* **141:** 898–902.
108. Morales-Ruán, M.C., I. Méndez Gómez-Humarán, T. Teresa Shamah-Levy, *et al.* 2014. La inseguridad alimentaria está asociada con obesidad en mujeres adultas de México. *Salud Publica Mex.* **56:** S54–S61.
109. Reports on the application of the FIES through the Gallup® World Poll™ in 2013 and 2014 in a range of countries. http://www.fao.org/economic/ess/essfs/voices/documents/en/.
110. Barrett, C.B. 2002. "Food Security and Food Assistance Programs". In *Handbook of Agricultural Economics*. B.L. Gardner & G.C. Rausser, Eds. Amsterdam: Elsevier Science.
111. FAO. 2002b. Keynote speech by Eileen Kennedy, International Scientific Symposium, 26–28 June, Rome. Available at: http://www.fao.org/docrep/005/Y4249E/y4249e0c.htm#bm12.
112. Sen. 1981. *Poverty and Famines*. Oxford University Press.
113. Maxwell, D., J. Coates & B. Vaitla. 2013. *How do Different Indicators of Household Food Security Compare? Empirical Evidence from Tigray*. Medford, USA: Feinstein International Center, Tufts University. http://fic.tufts.edu/assets/Different-Indicators-of-HFS.pdf.