Estimating Comparable Prevalence Rates of Food Insecurity through the application of the Food Insecurity Experience Scale

S. Viviani | Food and Agriculture Organization of the UN, ESS Division | Roma | Italy
C. Nwosu | Food and Agriculture Organization of the UN, ESS Division | Roma | Italy
C. Cafiero | Food and Agriculture Organization of the UN, ESS Division | Roma | Italy
M. Nord | Food and Agriculture Organization of the UN, ESS Division | Roma | Italy
DOI: 10.1481/icasVII.2016.b10c

ABSTRACT

This paper presents the methods for estimating comparable food insecurity prevalence rates across countries and time using the Food Insecurity Experience based Scale (FIES) or other similar food insecurity scales. Data are collected though the Gallup World Poll (GWP); however when data collected in national surveys using similar tools is available, those data are used to compute comparable prevalence rates. Results are provided regionally for 2014 and 2015.

Keywords: Food insecurity experience based scales, Rasch model, Voices of the Hungry project, global monitoring

PAPER

Introduction

The FAO Voices of the Hungry (VoH) project has developed methods to measure the access dimension of household or individual food insecurity based on the Food Insecurity Experience Scale (FIES), an 8-item questionnaire that elicits information on individuals' experiences and behaviours resulting from inadequate access to food (see World Summit (2009) for an official definition of this concept and United Nations (2015) to frame the project in the context of the post 2015 Sustainable Development Goals agenda).

The questionnaire was administered in 2014 and 2015 through the Gallup World Poll (GWP) in 147 and 140 countries, respectively. The GWP vehicle ensures continuous global monitoring; current plans are to continue including the FIES survey module (FIES-SM) in the GWP until at least 2018. The FIES-SM was translated into all major languages in each country where it was administered. The samples represent the adult population (15+ years) in each country, and sample sizes vary from 1000 up to 5000 individuals.

In parallel, several independent national surveys have fielded the FIES-SM or other similar food insecurity scales to monitor the food insecurity situation in the country. The samples collected in national surveys are representative at sub-national level, and sample sizes are much wider than in the GWP; therefore, when available, these data are used to calculate the prevalence of food insecurity in the country. The VoH methodology is based on the idea that it is possible to measure the severity of constraints on food access (an unobservable, or latent trait) based on people´s self-reported experiences. This is accomplished through the use of statistical techniques borrowed from Item Response Theory (IRT), commonly used in the educational and psychological testing fields, and in particular the Rasch Model (see Fisher & Molenaar, 1995 and Rasch, 1960). The model assumes that all questions refer to the same latent trait and form a scale of severity on which items (questions) and individuals (respondents) can be mapped. VoH developed a methodological extension of this model to ensure comparability among countries, surveys and from year to year. An R package to implement these methods can be downloaded from the FAO webpage.

The paper is structured as follows. Section 2 describes the FIES questionnaire and its rationale. Section 3 explains the methods to evaluate data quality, comparability across countries and time and scale stability across time. Sections 4 and 5 present some relevant results. Finally, Section 6 draws the main conclusions.

2. The Food Insecurity Experience Scale Survey Module (FIES-SM)

The FIES Survey Module (FIES-SM) is composed of eight questions1 with simple dichotomous responses ("yes"/"no"). Respondents are asked whether, at anytime during a certain reference period, they have worried about their ability to obtain enough food, their household has run out of food, or if they have been forced to compromise the quality or quantity of the food they ate due to limited availability of money or other resources to obtain food. (see FAO, 2013 for a description of the development of the FIES module).

In the version that has been applied globally through the GWP, questions are framed with reference to individuals and have a reference period of 12 months (Table 2-1). This is because the GWP is conducted in different months in different countries and a shorter recall period might result in lack of comparability across surveyed countries due to the possible seasonality of food insecurity.

Table

Questions in the Food Insecurity Experience Scale Survey Module for Individuals (FIES SM-I) as fielded in the 2014 GWP		
	Now I would like to ask you some questions about food. During the last 12 MONTHS, was there a time when… :	(label)
(Q1)	**… you were worried you would not have enough food to eat because of a lack of money or other resources?**	(WORRIED)
(Q2)	**… you were unable to eat healthy and nutritious food because of a lack of money or other resources?**	(HEALTHY)
(Q3)	**… you ate only a few kinds of foods because of a lack of money or other resources?**	(FEWFOODS)
(Q4)	**… you had to skip a meal because there was not enough money or other resources to get food?**	(SKIPPED)
(Q5)	**… you ate less than you thought you should because of a lack of money or other resources?**	(ATELESS)
(Q6)	**… your household ran out of food because of a lack of money or other resources?**	(RANOUT)
(Q7)	**… you were hungry but did not eat because there was not enough money or other resources for food?**	(HUNGRY)
(Q8)	**… you went without eating for a whole day because of a lack of money or other resources?**	(WHLDAY)

3. Data and methods

As outlined in Section 1, data analysed by the FAO VoH project are obtained through the GWP vehicle, with some exceptions where national data are available (USA, Mexico, Brazil and Guatemala). These national data were collected using tools that are similar to the FIES (see for instance Coleman-Jensen et al, 2015; IBGE, 2014; Tarasuk et al, 2014; Villagómez-Ornelas et al, 2014) and through equating procedure, it is possible to produce results that are comparable to the GWP results. In this section, we describe methods to test data quality, scale stability between years and comparability across countries and time. The data frame of reference is the GWP data, as national data require case-by-case considerations, and will not be described here.

3.1 Quality of Gallup World Poll FIES data

This section describes the methods to assess the quality of the FIES data. Data are analysed using the one-parameter logistic model (or Rasch model) which is used to estimate item (question) and respondent severity parameters along the same latent trait. This paper will not present a detailed description of the modeling framework applied to the FIES data, but further details can be found in FAO (2016), Section 4. Quality assessment statistics derived by the application of the Rasch model are briefly described below:

- **Missing responses** refer to cases where a respondent has stated that they do not know or have altogether refused to answer any of the FIES questions. Such cases are not considered in the analysis, but distribution of missing responses is reported as a general data quality indicator.

- **Item *Infit* and *Outfit* Statistics**: the FIES-SM was administered in close to 200 languages, in various countries throughout the world. Therefore, properly assessing the fit of all the items to the measurement model is crucial. The *infit* and *outfit* statistics assess the performance of the items included in the scale; that is the strength and consistency of the association of each item with the underlying latent trait. One of the Rasch model assumptions is that all items discriminate equally. Thus, ideally, all *infit* statistics should be 1.0;however, infit values in the range of 0.7 to 1.3 are considered acceptable. Similar standards may be applied to item *outfit* statistics. *Outfits* greater than 2.0 should be flagged for further assessment since outfit statistics are sensitive to even a few cases with highly improbable response patterns.

- **Rasch Reliability**: the standard Rasch reliability statistic is the proportion of total variance of severity of food insecurity in the non-extreme sample (omitting extreme raw scores 0 and 8) that is explained by the measure. A modified version of Rasch reliability that weights each raw score equally (rather than by the proportion of the sample in each raw score) is highly correlated with overall model fit across surveys

using the same scale. This "Rasch reliability (equally weighted)" statistic is used as a proxy for model fit by VoH (see Nord, 2014).

▪ **Conditional Independence of items**: the Rasch model assumes that responses to any two items are correlated only because of their mutual association with the underlying latent trait. This means that we do not want two questions that ask about essentially the same behavior or condition caused by food insecurity (Nord, 2014). In general, conditional correlations greater than 0.30 between more than two pairs of items are flagged to identify the presence of a second dimensions in the data.

3.2 The updated global standard: how to ensure comparability across countries and time

Two potentially competing objectives of the Voices of the Hungry (VoH) project are to calculate prevalence rates of food insecurity that are comparable across countries and, for each country, comparable over time. The VoH report on the 2014 data (FAO, 2016) described the procedure for estimating prevalence rates that are comparable across countries. In order to calculate prevalence rates of food insecurity that are comparable over time in each country, it is desirable to have a fixed global standard scale and a fixed national scale for each country. Otherwise, if either of these changed over time, it would be difficult to distinguish real changes in national prevalence rates from changes due to random fluctuations in either the global standard or the country scale. Moreover, it is desirable that the global standard and national scales be quite precise. However single-year sample sizes of non-extreme cases (i.e., omitting cases that denied all items or affirmed all items) in the GWP are too small in most countries to provide sufficiently precise parameter estimates (cases with extreme responses are omitted when estimating the Rasch model through conditional maximum likelihood, as per Fisher and Molenaar, 1995). To overcome the small sample-size problem, three years of data will be used to calculate the final global standard and the final scale for each country. Thus, from year 3 onward, for each country, the national scales will be fixed. Prevalence rates will change only as the proportions in each raw score change. For 2014 and 2015, provisional national prevalence estimates will be based on a provisional global reference scale and provisional national scales. This will necessitate revising each country's prevalence estimates for 2014 in 2015, and for both 2014 and 2015 in 2016. Operationalizing this plan is straightforward. In the present paper, each country's scale is calculated based on aggregated data from 2014 and 2015 (when available, as a few countries did not have data collected in both years). Therefore, the calculation of country-specific item and raw score parameters, the global standard and thresholds on each country's scale are calculated using the pooled 2014-2015 data. Prevalence rates for each year are calculated by applying the same proportions-by-raw-score to each year's distribution of respondents across raw scores. From here on, we will refer to the country-data obtained by pooling data from 2014 and 2015 as "*pooled data*".

3.3 Scale stability between 2014 and 2015

The stability of the scale within the same country over time indicates that the meaning of the questions has been perceived in the same way by respondents over time. To test the stability of the scale we compare item severity parameters estimated for each country between 2014 and 2015. If, for one country, the difference between item parameter estimates is large enough and significant, we define the corresponding items as "unique". This implies that for that country those items have functioned differently between the two years. This may be due to the sampling framework, to a different interpretation of the items due to language adaptation or to poor fit of the Rasch model in one of the two years. If stability of the scale is observed in the large majority of countries, the VoH strategy is to consider the severity of the items estimated for the pooled data to calculate the global standard, estimate person parameters, measurement error and prevalence rates, and therefore to produce comparable results across countries and over time. The VoH approach to assess year-to-year stability of the scale considers three factors: absolute difference in item parameters, margin of error of item parameters, and the possibility that one or two items are unique but the rest are good.

4. Results

4.1 Quality of Gallup World Poll FIES data

In this section, we describe the results on the performance of the FIES in the 2015 round of the GWP, and in the pooled 2014-2015 data. A brief comparison with 2014 results is also included, but not described in detail (as already done in FAO, 2016). The **percentage of cases with any missing responses** is relatively low in the 2015 and pooled data. This percentage is higher than 5% for 13% of countries in 2014, 18% for 2015 and 14% of countries in the pooled data (results not shown). Table 1 includes a **summary of item infit statistics** for the 2015 and the pooled GWP samples. It can be noted that the infits improve considerably using the pooled data compared to 2015 and 2014 (this latter comparison is not shown). This result supports the use of pooled data to estimate item and person parameters rather than a single-year.

Table 1. *Summary of item infit statistics*

	2015	*Pooled*	*2015*	*Pooled*	*2015*	*Pooled*	*2015*	*Pooled*	*2015*	*Pooled*
Item	Infit 0.8 to 1.2 (% of cases)	Infit 0.8 to 1.2 (% of cases)	Infit 0.7 to 1.3 (% of cases)	Infit 0.7 to 1.3 (% of cases)	Mean infit	Mean infit	Minimum infit	Minimum infit	Maximum infit	Maximum infit
WORRIED	74.2	83.7	89.1	95.2	1.101	1.112	0.6	0.824	1.549	1.549
HEALTHY	90.6	93.2	94.5	98.6	1.003	1.01	0.647	0.651	1.525	1.327
FEWFOODS	89.1	93.2	97.7	97.3	0.977	0.971	0.733	0.688	1.528	1.344
SKIPPED	88.3	92.5	95.3	99.3	0.932	0.923	0.523	0.614	1.257	1.639
ATELESS	78.9	86.4	93.8	98.6	0.904	0.893	0.608	0.6	1.344	1.086
RANOUT	75.8	87.8	92.2	100	0.905	0.915	0.595	0.72	1.414	1.244
HUNGRY	50.8	61.2	75	93.2	0.797	0.834	0.4	0.435	1.263	1.266
WHLDAY	66.4	66.7	83.6	89.1	1.182	1.171	0.836	0.957	2.199	1.855

Regarding the **outfit statistics**, the comparison between the single-year results and those based on the pooled data leads to less discrepant values (not shown). This is in line with the idea that possible anomalous response patterns need to be detected using single-year data. Mean **conditional correlation** is very low considering both the pooled and the year-by-year data (results not shown). Maximum conditional correlation instead improves considerably using the pooled data (Table 2 shows a comparison with 2015).

The mean **reliability** (equally weighted across raw scores) for the pooled data is 0.74 and ranges between 0.68 and 0.83. This result is consistent with the year-by-year analyses. In summary, the pooled data leads to improved item performance (and therefore the estimation of person parameters and prevalence rates) in terms of infits and residual correlation. This ensures that prevalence rates calculated according to parameters estimated on the pooled data are not only comparable, but also more reliable than those obtained using year-by-year parameters.

Table 2. *Maximum residual correlation (numbers in red refer to the 2015 samples while numbers in black refer to the pooled data)*

	WORRIED	HEALTHY	FEWFOOD	SKIPPED	ATELESS	RUNOUT	HUNGRY	WHLDAY
WORRIED		0.47	0.3	0.35	0.3	0.86	0.67	0.2
HEALTHY	0.31		0.61	0.33	0.54	0.21	0.26	0.18
FEWFOOD	0.25	0.57		0.28	0.36	0.28	0.21	0.08
SKIPPED	0.2	0.15	0.25		0.67	0.41	0.48	0.5
ATELESS	0.22	0.26	0.29	0.44		0.45	0.56	0.14
RUNOUT	0.19	0.16	0.18	0.35	0.33		0.72	0.33
HUNGRY	0.18	0.17	0.2	0.46	0.38	0.54		0.71
WHLDAY	0.06	0.13	0.04	0.18	0.14	0.25	0.42	

4.2 Comparability across countries and time

The updated global standard based on the pooled data (not shown) is only slightly different from the one calculated using 2014 data only (see FAO, 2016). The provisional thresholds for moderate or severe, and severe food insecurity considering the updated global standard change as follows: the threshold for the moderate or severe level is less severe (-0.29 versus -0.25 using the 2014 data only) while the threshold for the severe level is more severe (1.85 versus 1.83 using the 2014 data only). Items are considered common when their severity in a country differs from the one on the global standard by less than 0.35 units on the global reference scale. In 93 percent of the countries, a set of at least 5 common items was identified, thus allowing a robust equating procedure to be carried out. This result is in line with results of equating based on the 2014 data only.

4.3 Scale stability between 2014 and 2015

The stability of the scale between 2014 and 2015 was analyzed for countries where FIES data in the GWP were available for both years. Of 136 countries considered in the analysis, 99 (73.9%) have all common items between 2014 and 2015, 26 (19.4%) have only one unique item, 7 have two unique items and 2 have three unique items. No country has more than 3 uncommon items. The stability of the scale can be considered, overall, very good. It is worth noting that the instability observed in the GWP may be mostly

due to differences in samples (possibly primarily different languages), and may reinforce the VoH plan to base the global reference scale on three years of data and revise prevalence estimates accordingly. A threee-year GWP sample is likely to be more representative of the population of interest.

5. Prevalence rates at country and regional level

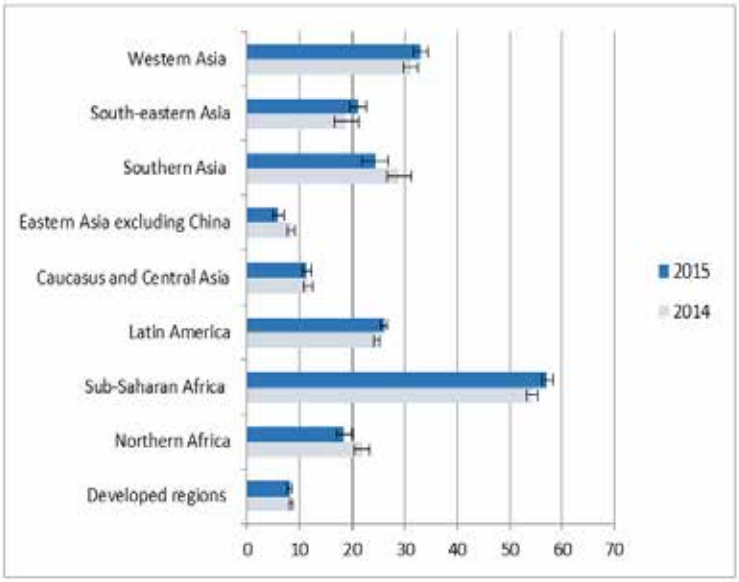
Table 3 shows the distribution of countries for different classes of food insecurity at moderate or severe and severe levels for 2014 and 2015. This distribution does not show particular changes in the two years, neither for the moderate or severe level nor for the severe level. Figure 2 represents the food insecurity prevalence rates calculated at moderate and severe level for the FAO regions, comparing 2014 and 2015. Margins of error are also included. It can be observed that the majority of the regions do not show a significant change infood insecurity prevalence rate over time.

Table 3. *Distribution of countries for different classes of food insecurity*

	Moderate or severe		Severe	
Range (%)	% of cases 2014*	% of cases 2015*	% of cases 2014	% of cases 2015
<5	8.2	8.6	17.0	15.7
5-14.99	32.7	35	33.3	32.9
15-24.99	17.7	12.1	14.3	15
25-50	21.8	22.9	16.3	17.9
>50	19.7	21.4	19.0	18.6

* Prevalence rates for 2014 and 2015 are calculated using the equating with the updated 2014-2015 global standard.

Figure 2 - Prevalence of food insecurity (moderate+severe) by FAO region, 2014 and 2015



6. Conclusions

The FIES provides a quick and reliable tool to estimate the prevalence of food insecurity at different levels of severity. The application of the FIES-SM in the GWP data in 2014 and 2015 has shown how comparability of the results is reached not only between countries, but also over time (using the pooled data approach). By pooling the data from 2014 and 2015 to estimate the Rasch model parameters corresponding to which prevalence rates are calculated, the reliability of the measure is increased. A three-year based analysis (using the 2016 GWP data) will provide a robust global standard for monitoring the evolution of food insecurity worldwide.

References

Coleman-Jensen A., Rabbitt M., Gregory C. and Singh A. (2015) Household Food Security in the United States in 2014, Washington, DC: U.S. Department of Agriculture, Economic Research Service, Economic Research Report, No. ERR-194

FAO (2013) The Food Insecurity Experience Scale: Development of a Global Standard for Monitoring Hunger Worldwide, Rome, FAO FAO (2016) Methods for estimating comparable rates of food insecurity experienced by adults throughout the world, Rome, FAO

Fischer G. H. and Molenaar I. W. (1995) Rasch Models. Foundations, Recent Developments, and Applications, New York: Verlag Springer

Instituto Brasileiro de Geografia e Estatística – IBGE (2014) Pesquisa Nacional por Amostra de Domicílios: Segurança Alimentar 2013, Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística

Nord, M. (2014). Introduction to Item Response Theory Applied to Food Security Measurement: Basic Concepts, Parameters and Statistics. Rome, FAO (available at http://www.fao.org/3/a-i3946e.pdf)

Rasch, G. (1960) Probabilistic models for some intelligence and attainment tests, Copenhagen: The Danish Institute of Educational Research

Tarasuk V., Mitchell A., and Dachner N. (2014) Household food insecurity in Canada, 2012 Toronto: Research to identify policy options to reduce food insecurity (PROOF). Retrieved from http://nutritionalsciences.lamp.utoronto.ca/

United Nations (2015) United Nations 2015 Transforming our world: the 2030 Agenda for Sustainable Development, Resolution adopted by the General Assembly on 25 September 2015

Villagómez-Ornelas P., Hernández-López P., Carrasco-Enríquez B., Barrios-Sánchez K., Pérez-Escamilla R., Melgar-Quiñónez H. (2014) Statistical validity of the Mexican Food Security Scale and the Latin American and Caribbean Food Security, Scale Salud Publica Mex, 2014:56 suppl 1:S5-S11

World Summit on Food Security (2009) Declaration of the World Summit on Food Security, Rome, 16-18