

The FIES methodology to measure household and individual food security: innovations and challenges

Carlo Cafiero
FAO statistics division

Beirut, 26 June 2019

Motivation

From MDG-1 to SDG-2: much more than continuation of an advocacy campaign

- MDG-1: Eradicate extreme poverty and hunger
 - Target 1.C: To halve the proportion of individuals suffering from hunger in the period between 1990 and 2015
 - Indicator 1.8 Prevalence of underweight children under-five years of age
 - Indicator 1.9 Proportion of population below minimum level of dietary energy consumption
- SDG-2: End hunger, achieve food security and improved nutrition and promote sustainable agriculture
 - Target 2.1: By 2030, end hunger and ensure access by all people, in particular the poor and people in vulnerable situations, including infants, to safe, nutritious and sufficient food all year round
 - Indicator 2.1.1 Prevalence of undernourishment
 - Indicator 2.1.2 Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES)

The SDG monitoring framework

- Broader, more ambitious, and potentially more effective
 - 242 indicators, many targets aim at “zero” or “100%”, leaving no one behind, with national authorities in the driving seat

But...

- More demanding in terms of methods, standards and tools
 - Many new areas of interest, that are still not part of national statistical systems
 - Comparability of the indicators across countries is essential, to ensure meaningful aggregation
 - Indicators must be timely, relevant, scalable and reliable

Measuring *food insecurity* in a relevant, timely, reliable, cross-country comparable way

- MDG indicators did not fully address new demands
 - Malnutrition:
 - Children underweight largely insufficient as it confounds acute and chronic malnutrition
 - New indicators: stunting, wasting+overweight in children, anaemia in WRA, low-birthweight, exclusive breastfeeding, adolescent and adult overweight
 - Food access:
 - PoU: only national level, 2-3 years delay, insufficiently precise to capture very low levels (< 5%)
- Did we have alternatives?
 - FCS, HDDS: lacking a basis for ensuring cross-country comparability
 - HFIAS, ELCSA: lack cross-country comparability
 - HHS: only relevant for severe food insecurity

Innovations:

Measurement in the realm of social sciences

Measurement principles

A **measurement system** consists of:

- a measurement ***tool***,
 - a ***protocol*** for application of the tool to objects, generating numbers to reflect the magnitude of an attribute (*measurand*), and
 - a ***reference scale***, used for calibration.
- Fundamental properties
 - **Validity**
 - measures reflect the magnitude of the attribute being measured (and nothing else).
Ideally measures changes if and only if there is a change in the attribute being measured
 - **Reliability**
 - Precision: small errors
 - Accuracy: no bias

Is measurement legitimate/possible in the social sciences?

- Measurands in social sciences are not physical objects, but rather abstract *conceptual constructs* that are “invented” or “discovered” while attempting at measuring them.
- Measurement is legitimate, to the extent that objects can be *classified, ordered, compared with each other, and/or against a standard of reference*
- There is a difference between **data** and **measures**
“Measurement in the behavioral and social sciences never takes place while data are collected—it always happens after they are collected.” (Van der Linden)

Claim 1: In social sciences (including in food security assessments), **“measuring” is a very abused expression.** Too often, numbers are associated to objects and treated as if they were measures, when in reality, they are not.

Claim 2: Here, measurement can **only be conceived as probabilistic. There will always be residual uncertainty around the measures**

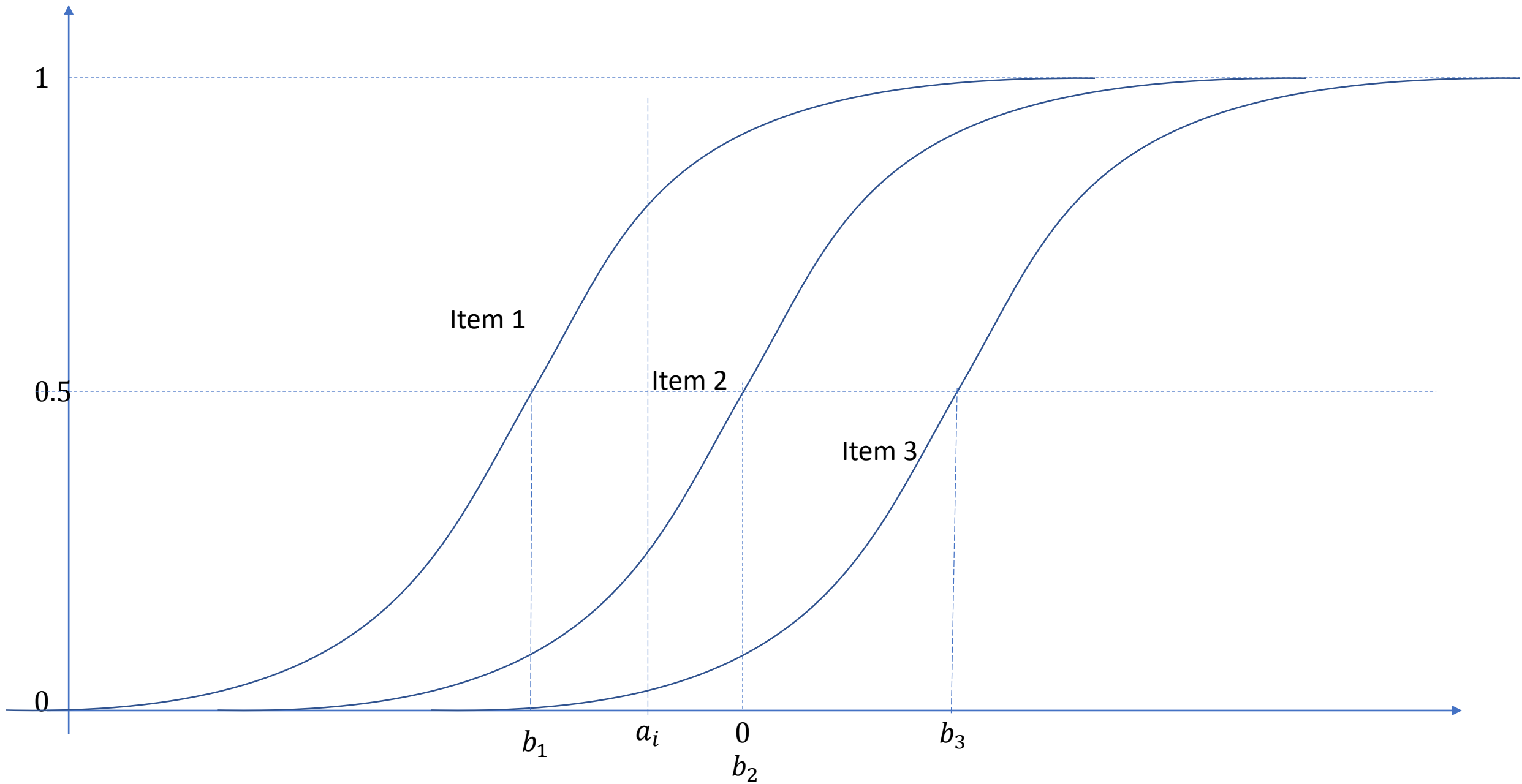
Fundamental measurement in the social science

- To prove validity there must be a convincing ***model*** that links the observables (the “evidence”, or “data”) to the attribute being measured.
 - Proponents must explain *how* the things we can see are linked to the things we would like to measure
 - The model can only be framed probabilistically
- To assess reliability, one must evaluate the extent, magnitude and direction of measurement “errors”
 - As there is no way to get to the “actual” magnitude of a latent trait by direct observation, reliability can only be determined with reference to the expectations induced by the measurement model

The Rasch model (G. Rasch, 1960)

$$Prob(X_{i,j} = 1) = \frac{\exp(a_i - b_j)}{1 + \exp(a_i - b_j)}$$

- The data (observables), $X_{i,j} \in \{0,1\}$, is the “**response**” of the i^{th} respondent to the j^{th} “**item**” (hence, **Item Response Theory** or **IRT**)
- Both items’ and respondents’ relevant measured attributes (unobservables) can be placed on the same, one-dimensional scale of severity.
- The probability that the subject whose attribute’s position on the underlying scale is a_i , might respond to an item positioned at b_j on the same scale is a (logistic) function of the difference $(a_i - b_j)$



The Rasch model

- The probability to affirm an item is increasing in the distance between the item and the respondent
 - Example:
 - The more “competent” is a student, the more likely it is that she will answer correctly any item
 - “Easier” items will be answered correctly more often than “difficult” ones
- Measures are defined/produced on an *interval scale*, not a ratio one
 - As the probability depends only on the difference between measures, the model is defined up to an arbitrary constant
- Maximum likelihood principles can be applied to estimate the values of a and b , given a set of data $\{x\}$.

Items' severity is revealed by the responses

| | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | RS |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------------|-------------|----|
| 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 5 |
| 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 3 |
| 6 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 5 |
| 7 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 4 |
| ... | | | | | | | | | |
| N | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 6 |
| Avg. | 0.95 | 0.76 | 0.82 | 0.72 | 0.52 | 0.65 | 0.4 | 0.21 | |

Items can be sorted by severity

| | Q1 | Q3 | Q2 | Q4 | Q6 | Q5 | Q7 | Q8 | RS |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------------|-------------|----|
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 5 |
| 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 3 |
| 6 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 5 |
| 7 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 4 |
| ... | | | | | | | | | |
| N | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 6 |
| Avg. | 0.95 | 0.82 | 0.76 | 0.72 | 0.65 | 0.52 | 0.4 | 0.21 | |



Respondents can be located, based on the sorted items

| | Q1 | Q3 | Q2 | Q4 | Q6 | Q5 | Q7 | Q8 | RS |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------------|-------------|----|
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 5 |
| 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 3 |
| 6 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 5 |
| 7 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 4 |
| ... | | | | | | | | | |
| N | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 6 |
| Avg. | 0.95 | 0.82 | 0.76 | 0.72 | 0.65 | 0.52 | 0.4 | 0.21 | |

Evaluating the extent of “fit”

| | Q1 | Q3 | Q2 | Q4 | Q6 | Q5 | Q7 | Q8 | RS |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----|
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 5 |
| 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 3 |
| 6 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 5 |
| 7 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 4 |
| ... | | | | | | | | | |
| N | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 6 |
| | OK | OK | ?? | OK | OK | OK | ?? | OK | |

Testing the Rasch model

$$\hat{p}_{i,j} = \frac{\exp(\hat{a}_i - \hat{b}_j)}{1 + \exp(\hat{a}_i - \hat{b}_j)}$$

$$\hat{r}_{i,j} = \text{abs}(x_{i,j} - \hat{p}_{i,j})$$

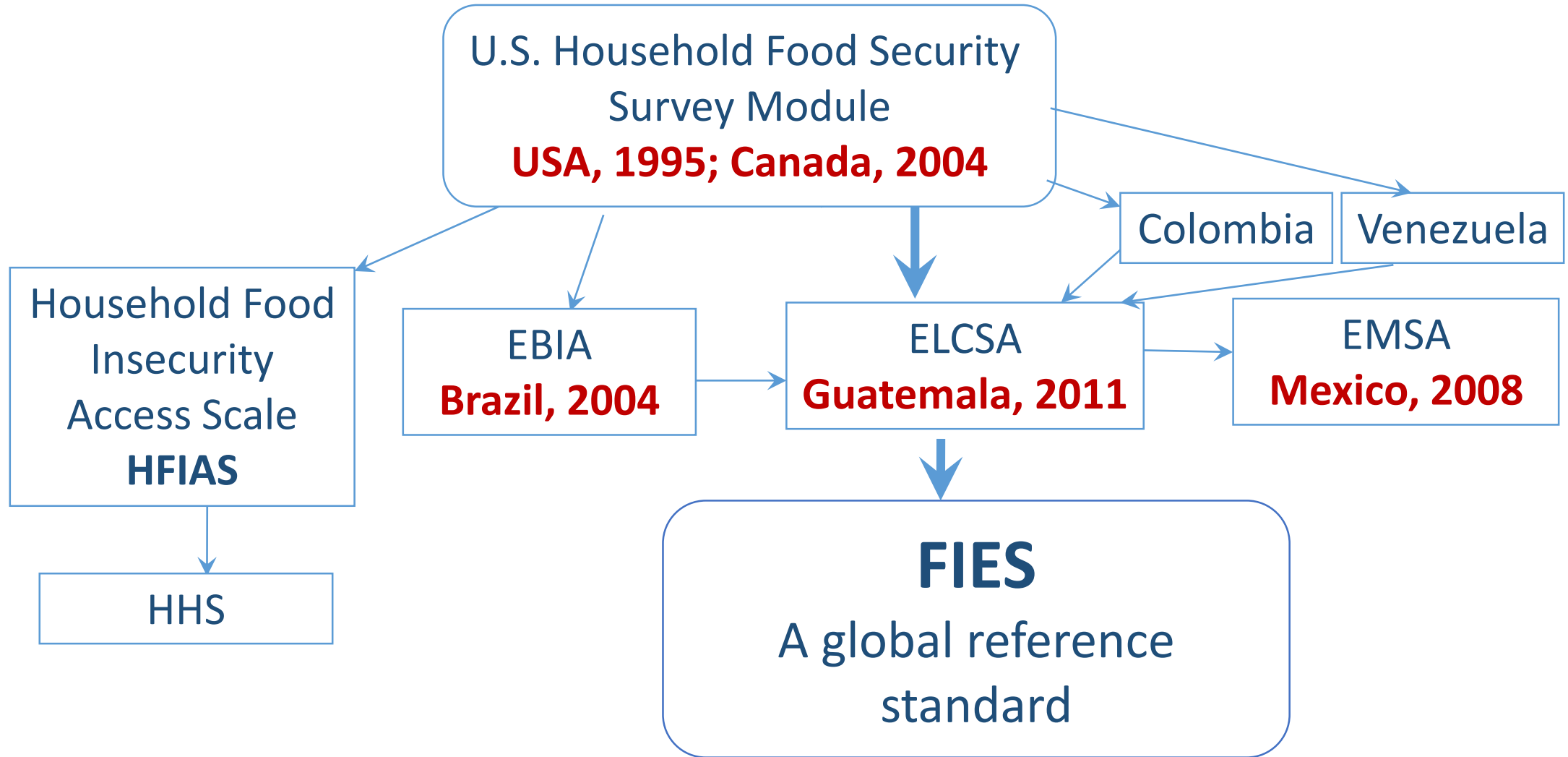
- Analysis of residuals:
 - *Infit* statistics
 - Principal component analysis (PCA)

The Food Insecurity Experience Scale (FIES)

Applied to “food insecurity”

- **Items** are specific conditions or experiences that are typically associated with limited ability to access food: WORRIED, HEALTHY, FEWFOODS, ATELESS, SKIPPED, RUNOUT, HUNGRY, WHOLEDAY (but also: NOTPREFERRED, SHAMEFUL, LOSTWEIGHT)
- **Responses** are the answers to the questions on whether those conditions occurred, during a certain reference period

FIES genealogy



The Food Insecurity Experience Scale survey module

During the last 12 MONTHS, was there a time when, because of a lack of money or other resources:

1. You were worried you would run out of food?
2. You were unable to eat healthy and nutritious food?
3. You ate only a few kinds of foods?
4. You had to skip a meal?
5. You ate less than you thought you should?
6. Your household ran out of food?
7. You were hungry but did not eat?
 - 7.1. If yes, was it rarely, sometimes or often
8. You went without eating for a whole day?
 - 8.1 If yes, was it rarely, sometimes, or often

Objectivity of measures based on experiences and the importance of local adaptation

- While responses to some of the questions included in the FIES questionnaire are conditioned by the respondents' perceptions, or by their subjective conceptualization of what healthy and nutrition food is, the Rasch model based analytic protocol is intended to filter those elements out (they end up in the "residuals")
- A correct characterization of the approach is as a "self reported experiences-based" one
- The information extracted from the dichotomous answers to the 8 FIES questions refers to the factual occurrence of conditions and experiences that are related to food insecurity
- These conditions are intended to be comparable across households or individuals, once controlled for the subjective perception/knowledge of the respondents

Objectivity of measures based on experiences and the importance of local adaptation

- Questions should be easily understood, and elicit a simple “yes” or “no” answer.
 - Questions are not intended to test respondents’ memory on the actual quantity or quality of the food consumed, or their knowledge about nutrition principles
 - The only purpose is to reveal whether, at times during the reference period, they have been **constrained in their ability to access food**
 - If necessary, items **must be adapted to local language and culture** to make sure they elicit the most efficient answer
- The FIES items are some of the most often used ones, but by no means they should be considered an exhaustive list: **more items can and should be added, whenever more precision is needed.**
- Items **must be dropped** if they are found not to conform to Rasch assumptions
- Keeping as many as the FIES “core” items as possible ensures more robust calibration

Estimating the Rasch model

- Using R
 - Download and install the RM.weights package
- Using STATA
 - Use the raschtest + fitstat_ers commands
- Using SPSS
 - Install the EXTRASCH extension (uses the eRm package under R)
 - Install the STATS_R33 extension (<https://www-01.ibm.com/support/docview.wss?uid=swg22008058>)

Relevant results from the analysis of FIES data

- Items' fit statistics
- Correlation among residuals
- Respondents' severity measures estimates and standard errors. Two things must be noted:
 - Measures are defined on an arbitrary scale, centered on the mean of the estimated item parameters
 - Residual uncertainty (standard errors) is quite large, when based on only 8-10 items. However, a probability of being located beyond a given position on the scale of severity can be associated with each raw score, assuming a normal distribution of severity around the mean, with s.d. equal to the estimated s.e.
- Prevalence rates can be obtained as (weighted) average probabilities in a representative sample

Statistical definition of SDG 2.1.2 indicator

$$FI_{mod+sev} = \sum_{i=0}^8 P_{mod+sev}^i \times wN^i / N$$

$$P_{mod+sev}^i = 1 - \Phi(T_{m+s}; \mu = a_i; \sigma = s.e.(a_i))$$

where:

$P_{mod+sev}^i$ = Probability of being food insecure, for the purpose of informing SDG 2.1.2.

wN^i/N = weighted proportion of respondents in a representative sample of the population, with raw score $RS = i$

a_i = Rasch model respondent parameters for a respondent with raw-score $RS = i$

T_{m+s} = international «food insecurity» line

$\Phi(.)$ = Normal Distribution function

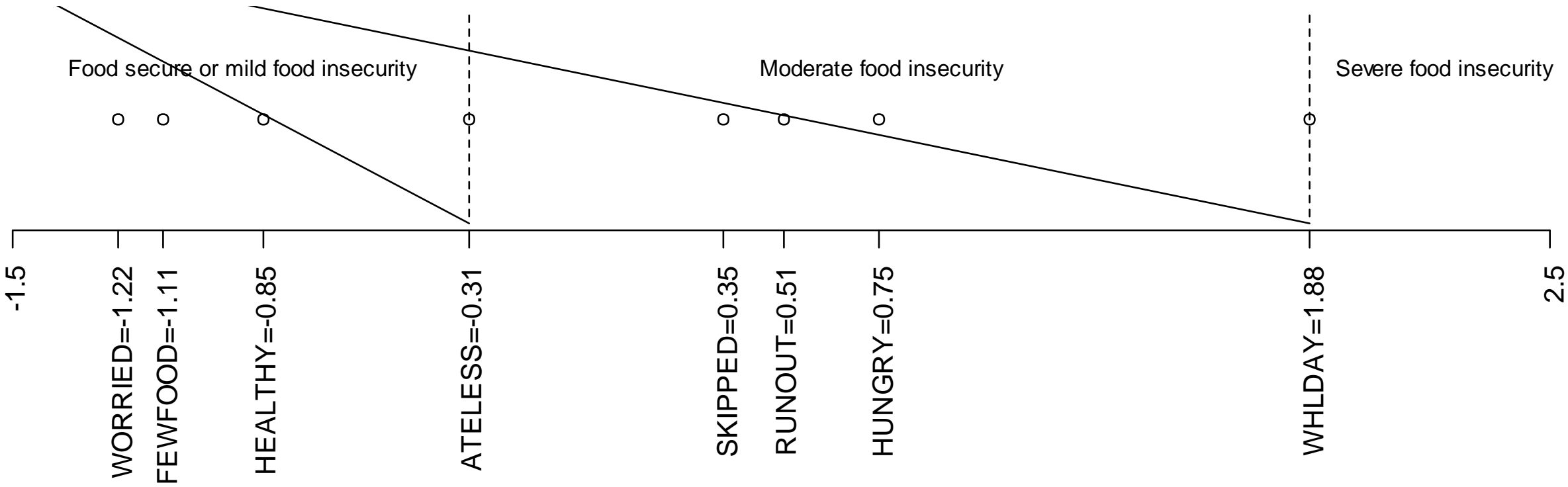
Main challenges

- Estimating the Rasch model, understanding the meaning of the parameters and the role of residual uncertainty
 - FIES-based measures are not sufficiently precise to be used for targeting purposes, yet they provide reliable estimates of the prevalence of food insecurity in a population
- How to calibrate measures against a global reference
 - While not necessary, if the analysis is limited to a single application in a given population, it becomes crucial for comparisons over time and space.

How to ensure worldwide comparability of SDG 2.1.2, by setting a common threshold

- As measured obtained from application of the FIES or a similar experience-based food security measurement tool are set on an arbitrary scale, there is a need to *calibrate* them against a *reference scale*
- The FIES global reference scale has been established by FAO based on FIES data collected in 147 countries in 2014, and validated for robustness against data collected in 2015 and 2016
- Conventional thresholds have been set to correspond to the severity of the ATELESS item (separating “mild” from “moderate”) and of the WHLDAY item (separating “moderate” from “severe” food insecurity)

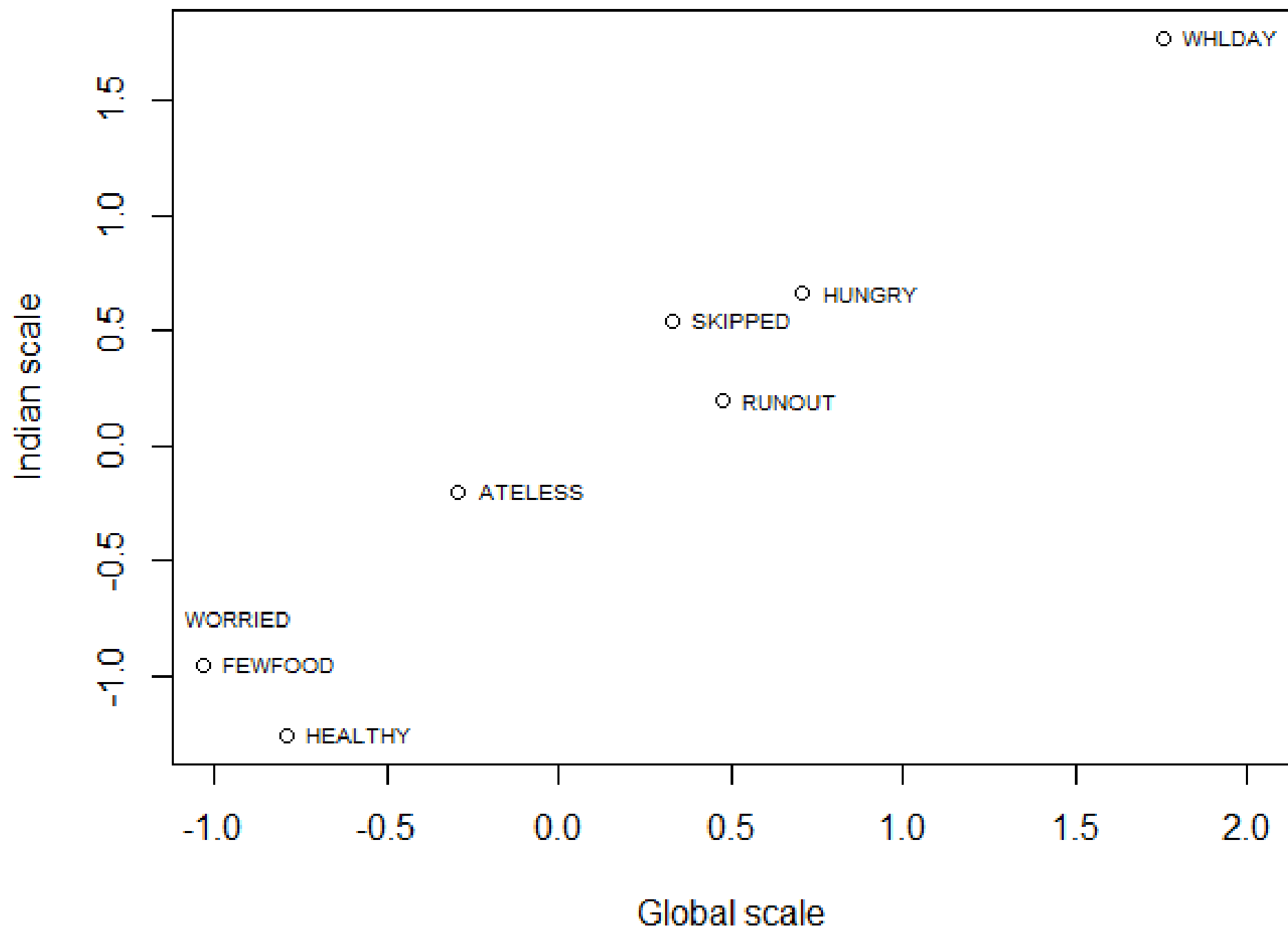
The global FIES reference scale: a benchmark for calibration



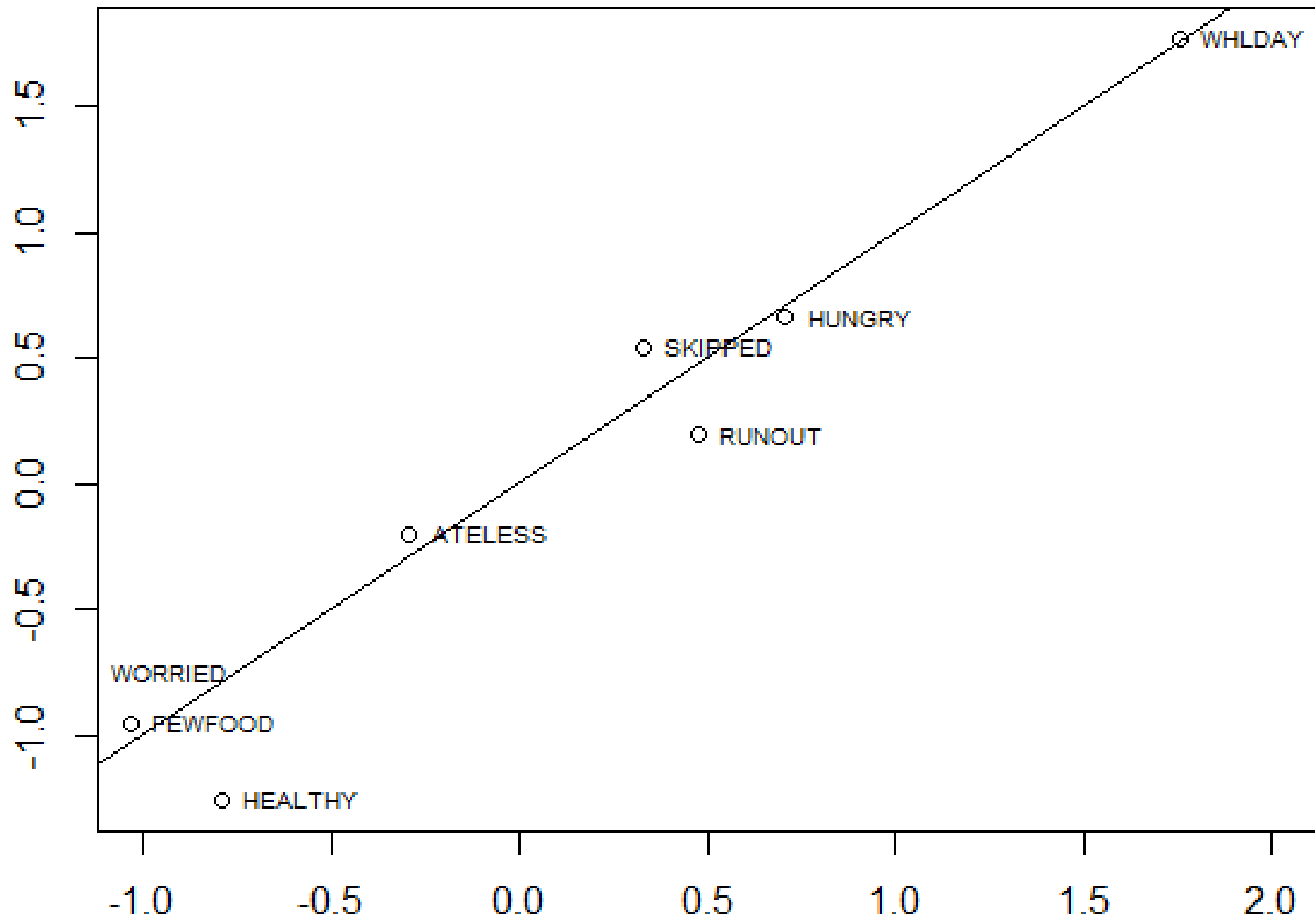
Example: data collected by FAO in
India, through the Gallup World
Poll

FIES analysis protocol to compute SDG 2.1.2

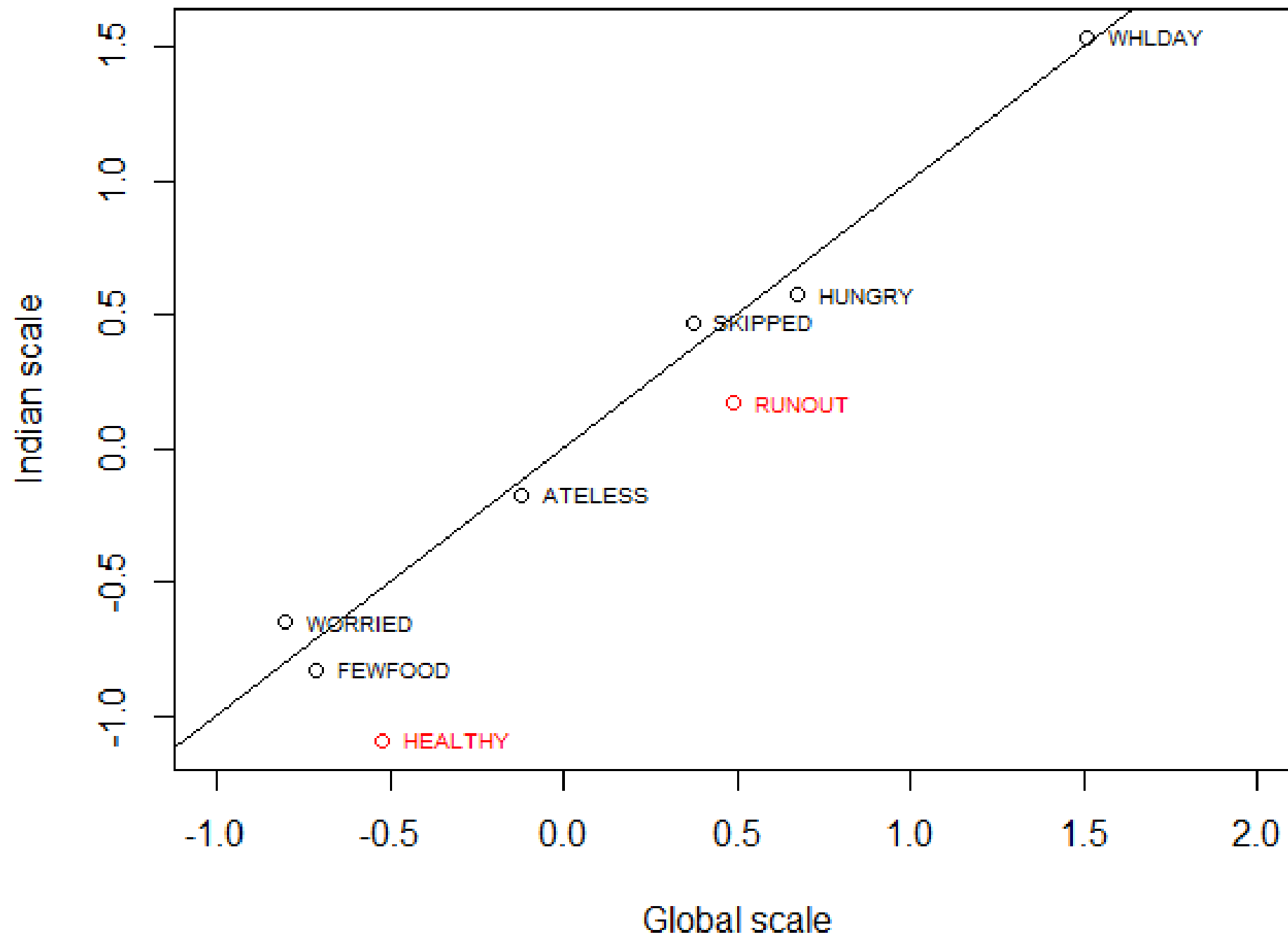
1. Recode answers, so that Yes = 1, No = 0 and everything else is NA
2. Estimate the Rasch model
3. Validate the quality of each item by looking at the INFIT statistics
 1. If $INFIT > 1.3$, drop the item and re-estimate the model
4. Plot the severity of the items against the FIES global standard, and identify the common items.
5. Adjust the global standard so that common items have the same mean and standard deviation on the two scales
6. Estimate the probability to be beyond the threshold $ATELESS_{adjusted}$ for each Raw Score
7. Compute SDG 2.1.2 as the weighted sum of the probability for each raw score, using the number of cases with each raw score as weights

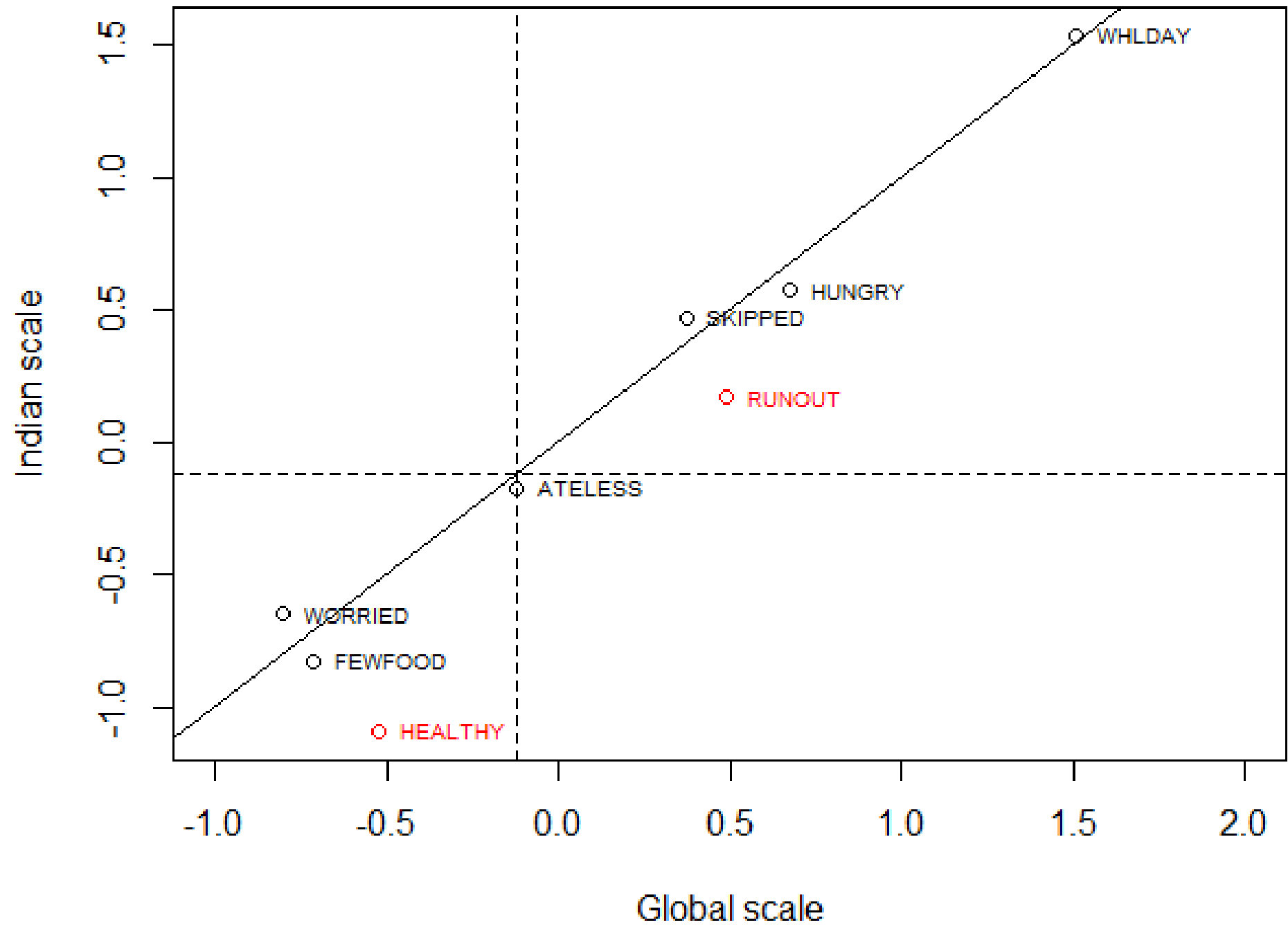


Indian scale

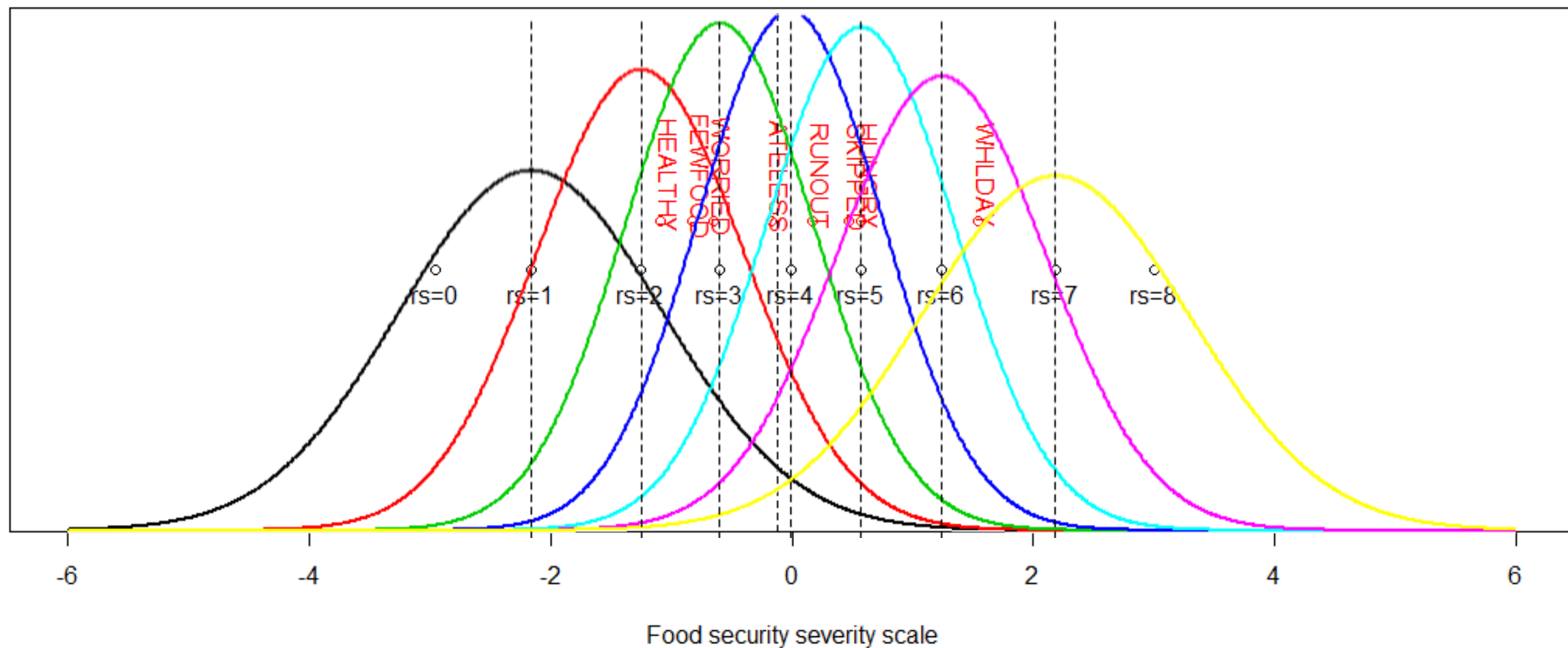


Global scale





FIES scale in India, FAO data 2014-2017



INFIT TABLE BY YEAR

| | 2014 | 2015 | 2016 | 2017 | 2018 | 2014-17 |
|---------|------|------|------|------|------|---------|
| WORRIED | 1.16 | 1.32 | 1.18 | 1.16 | 1.20 | 1.17 |
| HEALTHY | 1.07 | 1.08 | 1.03 | 1.02 | 1.29 | 1.06 |
| FEWFOOD | 1.01 | 1.05 | 1.02 | 1.02 | 1.03 | 1.02 |
| SKIPPED | 0.81 | 0.76 | 0.83 | 0.83 | 0.84 | 0.82 |
| ATELESS | 0.90 | 0.98 | 0.97 | 0.93 | 0.90 | 0.97 |
| RUNOUT | 0.79 | 0.73 | 0.86 | 0.85 | 0.75 | 0.80 |
| HUNGRY | 0.80 | 0.74 | 0.90 | 0.81 | 0.76 | 0.82 |
| WHLDAY | 1.24 | 1.25 | 1.21 | 1.17 | 1.07 | 1.25 |

| Raw score | Average | St. Dev. | Prob. FI |
|-----------|---------|----------|----------|
| 0 | -2.96 | 1.48 | 0.00 |
| 1 | -2.16 | 1.10 | 0.03 |
| 2 | -1.25 | 0.86 | 0.09 |
| 3 | -0.60 | 0.78 | 0.27 |
| 4 | -0.01 | 0.76 | 0.56 |
| 5 | 0.57 | 0.78 | 0.81 |
| 6 | 1.24 | 0.87 | 0.94 |
| 7 | 2.19 | 1.11 | 0.98 |
| 8 | 3.00 | 1.48 | 0.98 |

Challenges:

forcing people to move beyond their “comfort zone” ...

Conclusions and challenges

- Proper use of the FIES requires a significant level of analytic sophistication
- Obtaining the measures requires more than simple arithmetic on the data
- Still many food security analysts, statisticians, econometricians, do not fully understand the subtle implications of the Rasch model
- As there is a component of “art” in the calibration process, who should be invested with the responsibility to decide which is the preferred option when there are competing alternative equating solution?

Discussion points

- Very few countries in the Arab region have provided their consent to FAO to report on SDG 2.1.2.
- The main reasons are because FAO estimates were not based on official national data, and because of limited familiarity with the FIES methodology
- The solution is:
 - Develop statistical capacity of Arab countries' institutions on using the FIES
 - Include the FIES survey module in a national survey. Already happened in Bahrain, Saudi Arabia, Sudan, Lebanon, planned in Morocco and other countries?