Inside the black box: how consistent are global food security crisis assessments?

Abstract: The world relies on assessments by the United Nations-facilitated Integrated Food Security Phase Classification (IPC) to identify where populations are food insecure and to quantify the severity of these crises. IPC sub-national assessments are designed to be comparable over space and time in the 30 plus countries in which they operate. Humanitarian agencies appear to treat these assessments as authoritative and comparable, relying on IPC assessments to allocate more than six billion dollars of aid per year. In this paper, we study whether IPC food insecurity assessments are indeed consistent and comparable across time and space. Analyzing 1,849 IPC assessments covering 742 million people from fifteen countries between 2019 and 2023, we show that IPC assessments face significant challenges related to data availability and food security measurement, resulting from often discordant underlying food security data. We find that the vast majority of assessments are consistent with IPC technical guidance, but that this guidance permits a wide range of assessments for a given set of indicators. We also find evidence that IPC assessments differ in the way they use food security data, often weighing food security indicators differently in different locations. While variation in the way in which assessments use food security indicators can reflect varying contextual factors, we find some evidence that working groups weigh indicators differently across time for the same location. Finally, we show that assessments do not treat closely correlated food security indicators as substitutes, suggesting inconsistency in the treatment of food security indicators across assessments. We discuss implications of these findings for policy and for the use of IPC assessments.

Keywords: humanitarian, food security, measurement, IPC, famine

Inside the black box: how consistent are global food security crisis assessments?

I. Introduction

The Integrated Food Security Phase Classification (IPC) is a globally significant and highly influential tool used to identify and monitor food crises and populations facing acute hunger across the globe. IPC findings provide crucial information for prioritization and allocation of aid funding (United Nations Office for Coordination of Humanitarian Affairs 2024). In 2021, the G7 affirmed that the IPC was “the gold standard for food security and nutrition analysis” (United Kingdom Foreign Commonwealth Development Office 2021, p. 1) and it has been called “the voice of the humanitarian system” (Craze 2024, p. 71). Global aid agencies use IPC assessments to guide the allocation of an estimated six billion dollars in humanitarian aid annually and as of March 2025, its seasonal assessments covered 947 million people (IPC 2025). Despite its global importance in assessing food crises the IPC remains largely unknown beyond the world of humanitarian aid (Craze 2024; Masri et al. 2024).

For the first time, this paper analyzes the consistency of IPC assessments. Given the critical role the IPC plays in global humanitarian information and funding systems, there is considerable value in understanding the degree to which its assessments are consistent with the technical guidance (i.e. results are within prescribed guidelines) and comparable both across locations and within locations over time.

Developed by the FAO Food Security Analysis Unit (FSAU) in Somalia in 2004 (IPC 2008), IPC’s founding objectives included reducing uncertainty in the global systems used to identify and address the incidence and severity of hunger, with the aim of clarifying and improving international crisis assessment and aid allocation (Frankenberger and Verdujin 2011). Pre-IPC, multiple international agencies and governments simultaneously assessed food insecurity within a country. These simultaneous assessments came to different conclusions about the extent and severity of a crisis (Eilerts 2006; Harvey 2005) and this confusion limited and slowed humanitarian response (Eilerts 2006; Harvey 2005). The pre-IPC fragmentation in crisis assessment led to inefficiencies in humanitarian response and resource allocation, as well as a lack of coherence in strategies to address food insecurity (Eilerts 2006; Harvey 2005; IPC 2008).

By 2024, the IPC had become a partnership of at least nineteen international institutions. The institutions and national governments agreed on technical definitions of a continuum of hunger designations ranging from 1-5: minimal, stressed, crisis, emergency, and famine (IPC 2024). These hunger classifications are determined at the district or subdistrict assessment area and are based on specific data requirements. Each hunger designation corresponds to a defined share of the population meeting or exceeding set thresholds on key food security indicators and key contributing factors that influence food insecurity (IPC 2021). IPC categories prescribe different levels of intervention, with urgent action required for classifications 3 and above. Each assessment area includes population estimates by hunger designation, meaning the IPC assesses both the severity of hunger and the number of people affected.

IPC technical working groups rely on detailed technical guidance and a consensus-based process. This combined approach emerged to address several challenges inherent in measuring both the number of people experiencing food insecurity and its severity. Food security is multidimensional and a latent construct, with multiple approaches to measurement (Barrett 2010). Common household food security indicators often provide conflicting assessments of the severity of hunger and may require interpretation that is sensitive to the specific context (Maxwell et al. 2014; Vaitla et al. 2017). In addition, available data may be sparse and dated. Incidents of mass hunger tend to be driven by or are coincident with conflict or weather shocks that can complicate information access, interpretation, and dissemination, meaning populations' experiences can change in the period between data collection and the IPC analysis.

The IPC's approach to classifying food insecurity relies on a structured, collaborative process. At the country level, IPC technical working groups, composed of trained food security analysts, use a 12-step consensus-based process and draw from a detailed set of protocols to review and synthesize evidence from food security indicators and contextual factors (IPC 2021). The protocols guide the interpretation of data on multiple food security indicators, helping to mitigate biases and discrepancies while fostering transparency. According to the IPC, the consensus process is not intended to be formulaic; instead, it allows for contextual interpretation of food security information and other contributing factors. However, the lack of a transparent formula has opened the IPC up to criticisms that it operates as a "black box" and has raised concerns regarding whether the assessment outcomes are accurate and consistent (Craze 2024; Masri et al. 2024).

The IPC places strong emphasis on consistency—seeking to standardize both the evidence used and the way that evidence is interpreted when assessing food insecurity crises. The 2021 IPC

Technical Manual explained that the IPC was designed as “an analytical approach that would be robust and transparent, comparable and applicable across locations, and relevant for decision-making” (IPC 2021 p.4; see also IPC 2010). Consistency implies the reliability and stability of the criteria and methodology used to evaluate food insecurity across countries, time periods, and underlying available data. Consistency ensures that the same standards and processes lead to comparable results, allowing for meaningful cross-regional or longitudinal comparisons of food security conditions. Consistency across IPC assessments is required for comparability: if two places are in urgent need (i.e., IPC classifications of phase 3 or above), users expect that their circumstances and severity are comparable. This comparability underlies efficient distribution of resources over time and space.

Our analysis of IPC consistency focuses on three related questions: (1) are underlying data used by the IPC assessments consistent with *each other*; (2) do technical working groups arrive at severity classifications that are consistent with the technical guidelines; and (3) are IPC food security classifications comparable over time for the same locations, and are they consistent with what research has established about food security indicators and measurement?

First, we analyze the degree to which the core evidentiary set that IPC assessments rely on – the food security data the technical working groups use to make their assessments – present a consistent picture. Internal consistency within the food security data should, in theory, make technical consensus relatively straightforward and lead to assessments that are aligned with IPC guidelines and with each other.

We find that the food security data available to different technical working groups varies over space and time, and the population shares based on the food security indicators themselves frequently suggest conflicting levels of food insecurity. Further, we find that the population estimates derived from different food security measures are only weakly correlated, suggesting that technical working groups – whether implicitly or explicitly – exercise judgement in determining which food security indicators to prioritize given their context. Thus, discrepancies in the underlying data used for assessments highlight the challenges of ensuring consistency across working groups.

Second, we analyze the degree to which the technical working group assessments are consistent with IPC technical guidance and with each other. We find that assessments are consistent with the guidance in the IPC Technical Manual (2021). This means that consensus is operating within designated technical guard rails. We show however that the guidance is very broad: a wide range of assessment outcomes for a given evidentiary set are consistent with IPC guidelines.

Finally, we study how technical working groups weigh evidence from food security indicators across space and time. Results from ordinary least squares regressions show that the implicit weights technical working groups use to compile evidence from a suite of food insecurity indicators are not consistent across working groups. This means that different IPC working groups use the same suite of food security indicators in relatively different ways. Some of this variation makes sense: working groups face different sets of evidence and context matters. Even so, we show that technical working groups do not appear to treat key, closely-correlated food security indicators as substitutes, suggesting some inconsistency in how food security indicators enter into consensus processes.

Our results highlight both strengths and challenges in the consistency of IPC assessments. We find that technical working groups generally adhere to formal IPC protocols. We also show that these protocols allow for considerable interpretive flexibility, meaning that different working groups can reach different, yet technically consistent, conclusions when faced with similar evidence. We find that TWGs are not formulaic in their use of food security data, rarely using simple averages or taking modal values, and we identify differences in how TWGs weigh and interpret indicators. The variation in the availability and coherence of food security data across contexts — combined with differences in how TWG weigh and interpret indicators — further complicates efforts to ensure consistency.

Our results contribute to a growing body of work studying global early warning for food crises and analyzing methods to measure and predict food insecurity. Lentz et al. (under review) analyzed IPC accuracy and found evidence of conservatism in IPC assessments. In an analysis of the degree to which the IPC assessments agree with different treatments of the underlying food insecurity data, they show that IPC technical working groups on average undercount the number of individuals living in circumstances of food crises. The authors conclude that current IPC assessments miss as many as one in four hungry people worldwide (within the assessment catchments). A recent study by Maxwell et al. (2023) attempts to 'ground-truth' IPC classifications directly. Using extensive interviews and observations, they apply the Household Hunger Scale (HHS) to distinguish between households classified as IPC phase 4 (emergency) and phase 5 (famine). They find that distinguishing between a household in emergency versus famine status is extremely difficult, even with considerable on the ground observation of circumstance. This study is one of the few—if not the only—efforts that begins with IPC classifications and works backward to examine the characteristics of households within each classification.

The IPC plays a central role in shaping global humanitarian responses, guiding the allocation of billions of dollars in aid. Reaching consensus in food security assessments is a nuanced and critical task, especially given data limitations and contextual differences. If assessments are inconsistent across contexts, the fairness, credibility, and effectiveness of these decisions may be undermined. Our work underscores the value of greater transparency in how technical working groups navigate data gaps and contextual challenges. We close with recommendations for future research on and careful use of IPC by researchers and decision makers.

II. Literature Review

Social science researchers and policy makers are often concerned with the measurement and analysis of latent constructs central to human welfare including poverty, hunger, agency, and empowerment (Sen 1983; Barrett 2010; Bourguignon 2006). These concepts share a common feature: they are not directly observable and therefore not directly measurable, distinguishing them from readily quantifiable indicators including height, weight, income, prices, and disease incidence.

Given that measures of latent constructs are imperfect proxies for the true unobserved concepts, an active literature focuses on measurement. For example, earlier work asks how to enhance measures of poverty and food security to encompass new dimensions or to incorporate subjective and objective analyses into their construction and interpretation (Alkire and Foster 2011; Ravallion 2016; Charmes and Wieringa 2003; Pradhan and Ravallion 2000). A second area of research proposes new measures for latent constructs. For example, researchers have developed a range of new metrics to assess concepts including water security (Young et al. 2019), women's empowerment (Alkire et al. 2013; Narayanan et al. 2019), and food security (Clapp et al. 2022; Herforth et al. 2020).

It is not possible to directly assess the accuracy of measures of latent constructs, and so scholarship validating new and existing measures generally explores the consistency of measures with one another, whether measures capture the same populations, or whether measures capture the same latent concept (Klasen 2000; Barrett 2010; Maxwell et al. 2014; Gebreyesus et al. 2015;

Headey and Ecker 2017; Bageant et al. 2024). Consistency is akin in these applications to reliability, essential to a measure's credibility and comparability. A poverty measure, for example (Ravallion 2020) is consistent if it classifies two households or subgroups with the same welfare status into the same poverty category (poor or not poor).

While consistency is often a desired trait in measurement, context specificity can complicate that objective. For example, an individual classified as poor using one country's poverty line may not be poor using another's (Foster 1998).

Food security measurement research is at a relatively earlier methodological stage: a range of measures exists, and recent research has begun to establish which among these assess different dimensions of food security and which have some overlap. Vaitla et al. (2017) for example use household survey data to study the relationships among four measures of food insecurity (Food Consumption Score (FCS), Household Dietary Diversity Score (HDDS), Household Hunger Score (HHS), and reduced Coping Strategies Index (rCSI)) (for detailed information on each measure, see FANTA 2018) and to identify latent (unobservable) dimensions of food insecurity. They find that food insecurity dimensions include both a lack of quantity of food consumed, captured by HHS and rCSI, and a lack of diversity of food consumed, captured by FCS and HDDS but the authors also discuss the likely existence of other relevant food security dimensions (see also Clapp et al. 2022). Similarly, Headey and Ecker (2013) classify HDDS and FCS as measures of dietary diversity, distinct from what they term 'experiential measures' of food security, such as rCSI and HHS. Maxwell et al. (2014) describe HDDS and FCS as capturing shortfalls in diet quality and argue that HHS and rCSI are more suited to capturing quantity shortfalls.

Like relative poverty lines, some food security measures may be highly context specific, however. For example, in some contexts, it is important to ensure that fasting-by-choice is not recorded as affirmative to the HHS question: "has any household member gone a whole day and night without eating at all because there was not enough food?" (FANTA 2011, p. 6). Similarly, in regions where few households raise livestock, questions regarding their sales will not provide information about changes in household livelihoods or coping strategies (LCS).

III. IPC Process

The IPC has developed a consensus-based approach to address the challenges of sub-national and intra-annual food insecurity status assessment. The process is designed to help the working group achieve a strong understanding of the underlying data, and to reach agreement. Country technical working groups (TWGs) composed of trained food security analysts engage in this "consensus with technical guard rails" process. Working groups tend to meet several times a year and consist of analysts drawn from government, UN staff and domestic and international NGO staff. The guard rails include training, a step-by-step approach to analysis within the TWGs, the participation of an experienced facilitator, and detailed guidance and reference tables articulated in the IPC technical manual.[1] A 12 step process helps analysts reach consensus regarding (1) the classification of each IPC analysis area in the country and (2) the proportion of population in each phase classification (IPC 2021).[2] A qualified facilitator helps define ground rules for building consensus. In TWG meetings, held either in person or online, analysts work methodically through IPC analysis worksheets to compile and discuss evidence and the IPC protocols. Each country's technical working group undertakes this process for each IPC assessment area (e.g., geographic location) within the country. The number of assessments per country ranges between 10 assessments (in Madagascar) to 179 (in Sudan).

TWGs draw on available food security indicators (FSI) for acute food insecurity assessments.[3] The reference table in the IPC Technical Manual provides guidelines to TWGs about how to interpret FSI. See Figure SI.1 in the online Supplementary Information (SI) for a reproduction from IPC (2021). The TWG also considers information about contributing factors including food production, food and income source, weather information, and humanitarian aid programs. Due to data limitations, we rely on the key outcome indicators (food security indicators) in our analyses.

The technical working group uses the FSIs to first classify the percent of the population in each location currently experiencing each phase. They then classify each location based on the severity stage experienced by the worst (at least) 20th percentile of the population. For example,

[1] Consensus is not always achieved within TWGs, and disagreements may arise. In such cases, the analysis team uses neutral facilitation to address disagreements and to seek agreement at the country level to minimize delay in releasing time sensitive findings (IPC 2021). If agreement cannot be reached, dissenting organizations can choose to disagree with the analysis results, with a range of options existing for expressing less than full agreement with the final results, although cases are rare (IPC 2021). An external quality review may be requested by members of the TWG or supporting partner(s) reflecting the minority view (IPC 2021). Vetting of classification and population estimations is also recommended as good practice for IPC consensus-building (IPC 2021).

[2] Reaching consensus within the TWG involves identifying the analysis parameters, critical analysis, reviewing a comprehensive range of data, discussing the evidence, addressing disagreements through neutral facilitation, seeking agreement at the country level, documenting minority views when necessary, and considering external quality reviews for classifications 4 and 5 (IPC 2021).

[3] The IPC uses additional and separate data for acute malnutrition and chronic food insecurity assessments and are not described here.

if a population is distributed as 50%, 20%, 10%, 10%, and 10% across increasing severity phases from phase IPC 1 to IPC 5, applying the classification threshold of 20% would classify the area as IPC 4, since 20% of the population are categorized as IPC 4 or 5. Similarly, if the distribution is 5%, 5%, 5%, 80%, and 5%, for phases 1-5 respectively, the same classification of IPC 4 would be applied based on the rationale of prioritizing the circumstances of the worst-off 20th percentile. Thus, in these examples, the assessment of IPC 4 reflects both an underlying population of 20% experiencing phase 4 or worse and 85% experiencing phase 4 or worse. For this reason, we focus on population shares in each phase rather than on the phase classifications themselves.

Before TWGs begin their analysis, the FSI are assigned reliability scores that reflect timeliness and representativeness of the data. In theory, these reliability scores could help technical working groups weigh the range of food security indicators in their assessments. However, in practice these scores are given at the level of the survey as a whole and, in most areas, a single survey collects multiple FSI. Therefore, the reliability scores rarely vary across FSI within an assessment area, and they provide limited actionable information to the TWG about which FSIs to weigh more heavily.

IV. Data

Our sample includes all available technical working group phase classifications and population assessments with underlying food security and livelihood indicators held by the IPC Global Support Unit for the period 2019-2023. TWGs assess food security using up to eight different food security measures, and two livelihood measures.[4] In practice, few TWGs have more than four food security indicators and one livelihood measure available. We focus our analysis on assessments derived from the five most commonly available indicators: FCS, HDDS, rCSI, FCS and LCS.[5] We do not have raw, household food security indicator data, and instead use the data

[4] When and where available, they also incorporate other information, including on contributing factors such as hazards, safe water access, and nutritional status. We do not observe this information, which is unevenly available across and within TWGs. The IPC Technical Manual (2021) refers to these data "second-level outcomes" and "contributing factors" to denote the difference from the required food security indicators, which are "first-level outcomes."

[5] Few TWGs have information on any of the following food security indicators: Quantity: Adequate Energy Intake, Dietary Energy Intake, Household Economy Analysis, Food Insecurity Experience Scale, or Livelihood Change measures. See IPC Technical Manual (2021) for more details on each indicator.

that TWGs use in their assessments: aggregated data on the population that is assigned to each phase for each food security indicator.

Table 1 shows the sets of FSI used in the 1849 sample assessments drawn from 15 countries. The most common food security indicators available to the technical working groups are FCS, rCSI, and LCS.[6] HDDS is least common among the indicators, available in only 898 assessments. The total sample with at least three core FSI indicators is 1849 assessments.

Most TWGs work with fewer than five FSIs. We refer to the sample of area assessments that have data on all five FSIs (n= 599) as Group 1. We analyze three other groupings of data, reflecting the next three largest sample sets in our analyses. The Group 2 sample of area assessments with four indicators has 360 observations; the Group 3 sample with another combination of four indicators has 237 observations; the Group 4 sample using the three most common indicators (rCSI, FCS, and LCS) has 497 observations. We provide the complete list of technical working group analyses in SI Table SI.1.

According to the IPC Technical Manual, different FSI provide information relevant to different severity classifications (see SI Figure SI.1). The rCSI is more helpful when the situation is less severe (phases 1 and 2) but cannot differentiate well amongst phases 3 to 5. Similarly, HDDS and FCS are informative between phases 1 and 4 but cannot differentiate between phases 4 and 5. HHS is the lone indicator in our sample that can reliably distinguish between a phase 4 and phase 5 assessment.

Table 1. Summary of cases and FSI availability in IPC assessments

	FCS	**rCSI**	**HHS**	**LCS**	**HDDS**	**Total Number of Assessments**
Sample A (Number of unique TWG Analyses)	1823 (27)	1812 (27)	1072 (16)	1823 (27)	898 (15)	1849 (27)
Sample B (Number of unique TWG Analyses)	1691 (27)	1690 (27)	990 (16)	1691 (27)	862 (15)	1693 (27)
Disaggregation of Sample B By FSI Availability Group						

[6] We exclude the following from our analyses due to small sample size: three cases of FCS with lower cut-off (L_FCS) and two cases FIES.

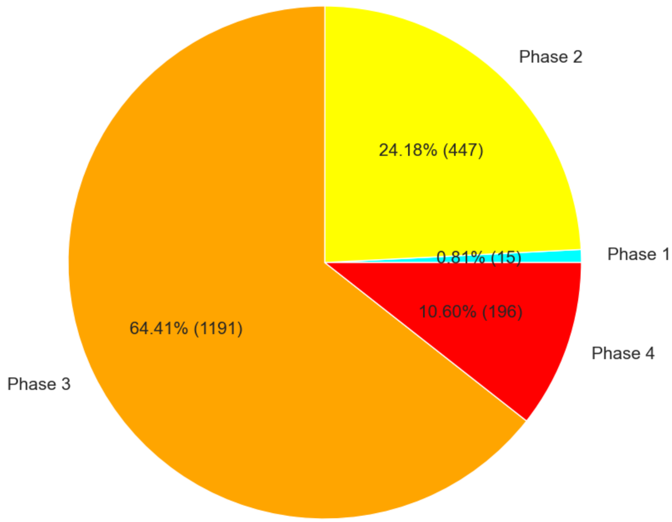
Group 1	FCS-RCSI-HHS-LCS-HDDS	599	599	599	599	599	599
Group 2	FCS-RCSI-HHS-LCS	360	360	360	360	0	360
Group 3	FCS-RCSI-LCS-HDDS	237	237	31	237	237	237
Group 4	FCS-RCSI-LCS	497	497	0	497	26	497

Caption: This table presents the sample size used in our analysis, defined by the relative availability of five FSIs—FCS, rCSI, HHS, LCS, and HDDS. Each TWG's sample is categorized into an FSI group based on whether more than 50% of the assessed areas within a TWG analysis include a given indicator. Sample A represents the total IPC assessments included in this study. Sample B includes cases that are used for regression analyses where all relevant indicators are available within each FSI Availability Group.

Figure 1 presents the relative frequency of the IPC assessments in our sample. Our sample includes no phase 5 assessments (phase 5 indicates famine or catastrophe) and less than 1% of classifications in our data are IPC phase 1 (indicating none or minimal acute food insecurity). Just under two-thirds (64%) of the assessments are IPC 3 (crisis), by far the most common phase in our sample. Phase 3 and above indicate urgent action is required (e.g., food assistance). Given that we observe few classifications in phase 1 and none in phase 5, we caution that our consistency findings should not be generalized to phases 1 or 5.[7]

[7] As described above, our sample is limited to relatively recent data available to the IPC global support unit. The sample may include more populations in IPC phase 3 or above (crisis or worse) than the universe of IPC assessments.

Figure 1. Distribution of consensus-based phase classifications



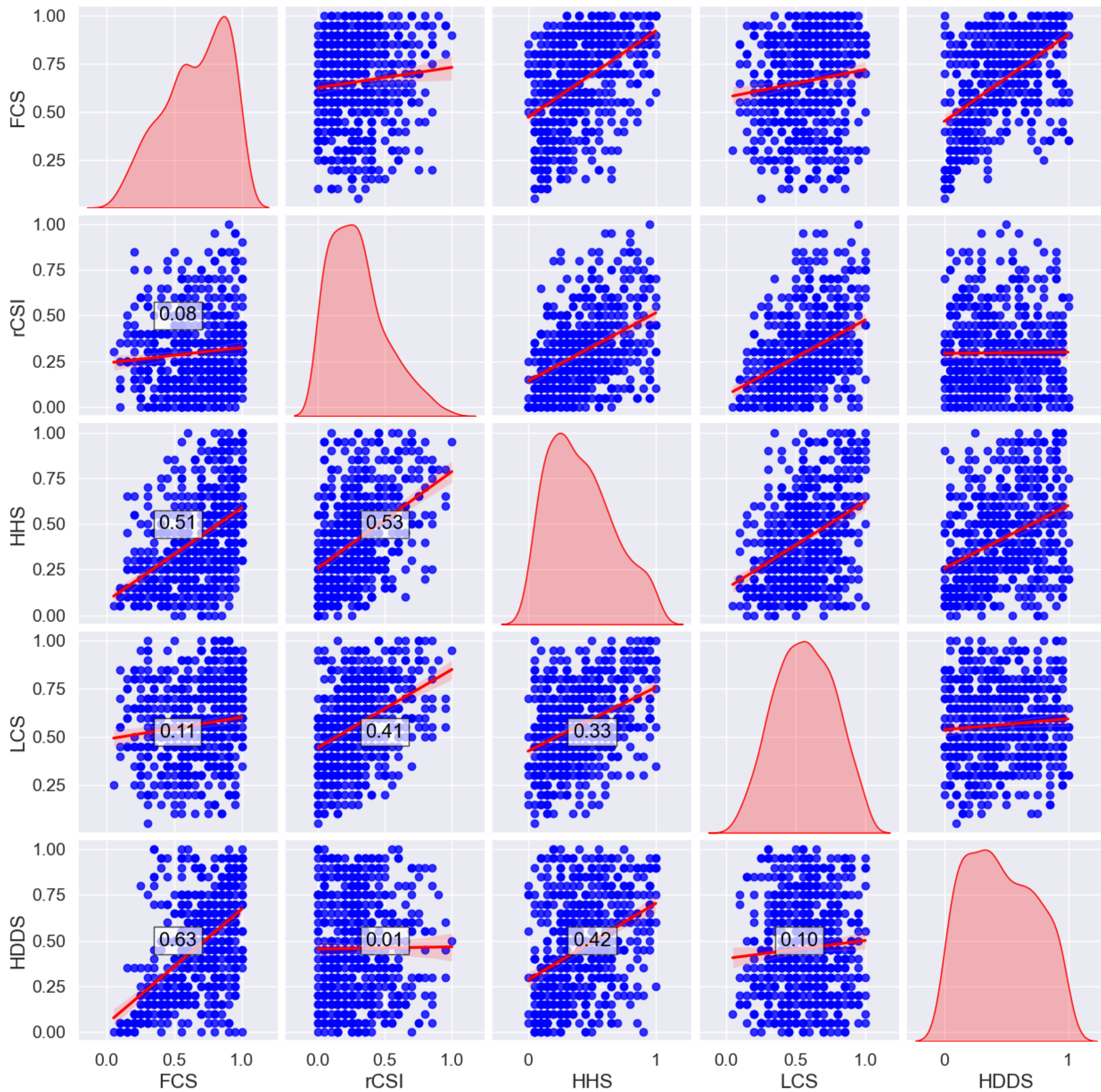
Caption: This figure presents the relative frequency of each phase classification, with counts in parentheses. The total sample includes 1849 assessments.

V. Analysis 1: consistency of underlying data

We begin by establishing how often the multiple FSIs for a given place and time concur in their phase classifications according to the IPC Technical Manual guidelines, and how often they differ. We present a series of descriptive analyses using the underlying FSI information available to the TWGs.

First, we explore how the predictions of the percent of population in urgent need predicted by FSIs vary for the same place and time. Figure 2 illustrates the relationships between pairs of FSI-implied populations in 3+ (crisis or worse). The analysis uses the Group 1 sample (refer to Table 1). We compute kernel density estimates (KDEs) for each variable, present pairwise scatterplots for the FSI, and use ordinary least squares to estimate the slope coefficient for the percent of population deemed to be in urgent need by each pair of FSI. We then generate a series of cubic polynomial lines that are smoothed fits for each food security indicator. The resulting graphs show the relationships between food security indicator-based 3+ population shares (i.e., 3+ is the population experiencing phase 3 crisis levels of food insecurity or worse).[8]

Figure 2. Pairwise Relationships and OLS Regression Coefficients Among FSI Implied 3+ Population (%) for Sample with All-5-FSI



Caption: This figure presents a pairwise scatterplot matrix with regression lines, illustrating the relationships among the share of the population assessed to be in phase 3+ based on Group 1 assessments (those with five FSIs: HHS, rCSI, FCS, LCS, and HDDS, n = 599). The diagonal

[8] All analyses using phase 4+ data or showing other data groups are available upon request.

displays kernel density estimates (KDEs) for each variable, while the lower/upper triangular section includes regression plots with slope coefficients from ordinary least squares (OLS) regressions from each pair annotated at the center of each plot.

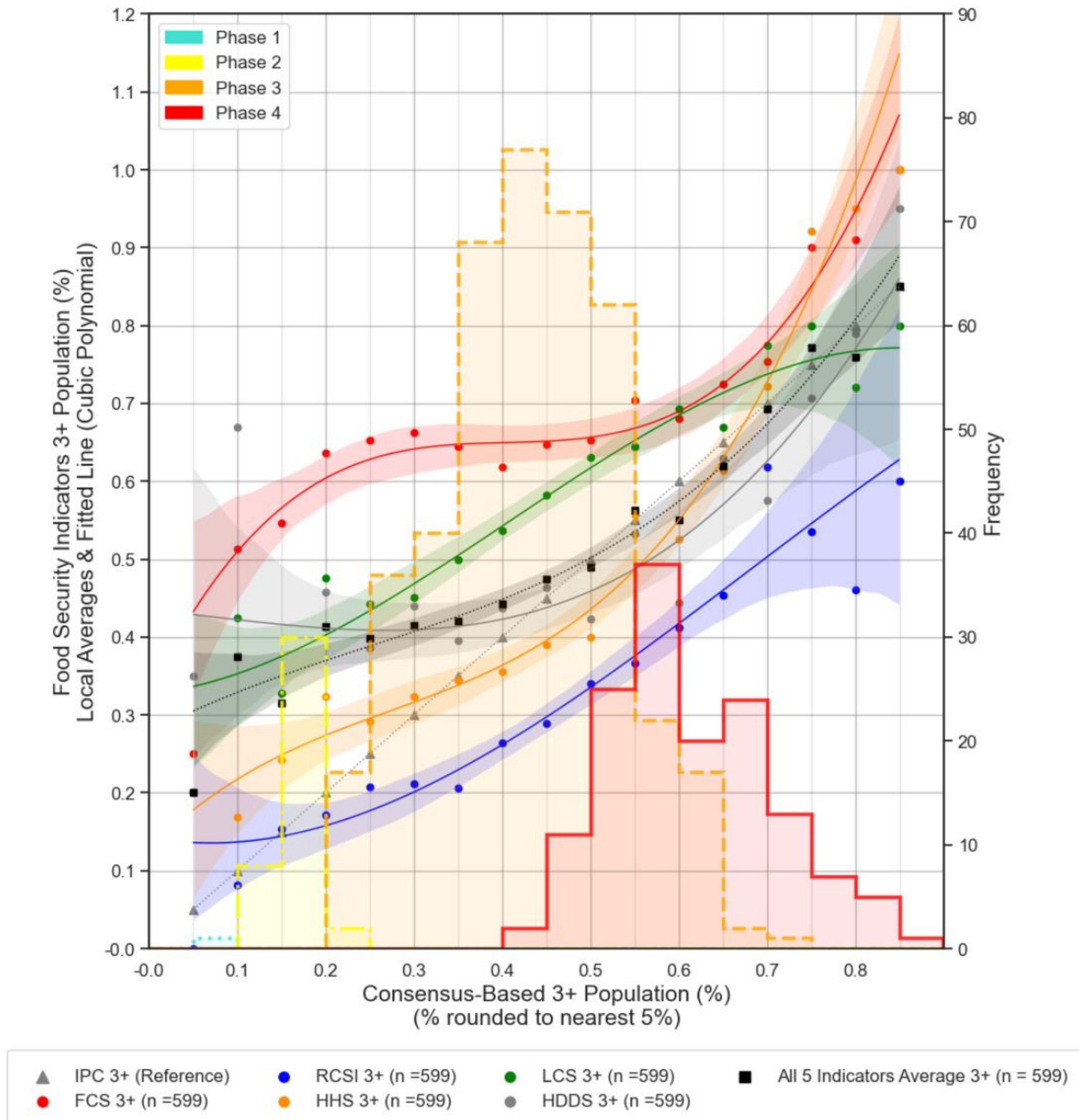
Figure 2 shows the relationship between the percent of population in 3+ across FSI pairs is relatively weak; five of the 10 relationships have correlations of one third or less. The wide variation in FSI data regarding populations in urgent need demonstrates a key challenge for technical working groups. When faced with data suggesting divergent IPC categorizations, these groups must determine how to aggregate the evidence — deciding which food security indicators to prioritize over others.

To investigate the relationships between food security indicators and the resulting consensus outcomes, Figure 3 plots the population estimates suggested by the IPC Technical Manual for each underlying food security indicator against the consensus-based outcomes reached by each TWG. Each line illustrates the population in need of urgent action (IPC 3+) implied by the individual FSIs and should be read using the left-hand axis. The bold black line represents the simple average of the FSI. The resulting consensus outcome is read using the horizontal axis. The shading around each food security outcome indicator line is the 95% confidence interval. The histogram shaded areas present the frequency of cases classified as IPC 3+ and should be read against the right-hand y-axis. At least 20 percent of the population must be in phase 3 or more in order to be classified in phase 3. For example, an IPC 3+ classification for 45% of the population is the most common case, happening about 77 times for phase 3 and 12 times for phase 4. The majority of the phase 3 classifications range between 20% and 60% of the population, evidence that TWGs are asked to classify a wide range of populations.

Figure 3 shows the variation in population shares classified as phase 3+ by each FSI, illustrating the assessment challenge faced by TWGs. The population experiencing acute food insecurity of phase 3 can vary by 45 percentage points, depending on the choice of the FSI. The plot shows that rCSI provides, on average, a lower estimate of the food insecure population relative to FCS and LCS. The HDDS and HHS most closely approximate the averaged food security outcome, as seen by their close tracking of the thick black line. For example, the consensus-based outcome identifies 30% of the population is experiencing food insecurity at a severity level of at least phase 3 (on the x-axis). But applying the IPC Technical Manual to the FSIs points to a range of 20 percent (for rCSI) to 67 percent (for FCS) (left-hand y-axis).

In sum, results indicate that food security data are often discordant and TWGs must make choices about which food security indicators are most relevant for their context.

Figure 3. Consensus-based 3+ population (%) for cases with 5 food security indicators



Caption: The graph presents the 3+ population classified for Group 1 (five food security indicators) (n = 599). The x-axis represents the share of the population the TWG consensus process assesses to be in phase 3+. The y-axis represents the share of the population who would be assigned to IPC 3+ determined only by the food security indicator data (i.e., the population implied by each food security indicator). The 45-degree line in gray depicts the hypothetical situation where the actual outcome from the consensus share in 3+ is equal to the share implied to be 3+ by the food security indicators. The black line represents the arithmetic mean of the population implied to be 3+ by each food security indicator. The colored cubic polynomial lines show smoothed fits for each food security indicator, highlighting variations in the relationship between consensus-based and food security indicator-based 3+ population shares. The shaded areas surrounding each line represent 95% confidence intervals, indicating the level of uncertainty in the estimates. The histograms in the background represent the frequency distribution of cases, differentiated by IPC phase color and should be read against the right-hand y-axis. Green indicates phase 1; yellow indicates phase 2; orange indicates phase 3; and red indicates phase 4.

VI. Analysis 2: Are TWG assessments consistent with technical guidance?

In the preceding section, we demonstrated that food security indicators underlying the IPC assessments are often inconsistent with each other. This creates challenges for TWGs seeking to reach consensus about the severity and number of people experiencing food insecurity. TWGs could take a range of approaches to solving the challenge of conflicting data pointing to several different assessment options. Unpacking the observed TWG outcomes lets us examine if TWGs treat the data in a consistent manner. While there is no single way to establish consistency, we can describe what approaches working groups do and do not appear to use. Specifically, in this section we (1) describe our methods and then examine whether TWG outcomes are (2) are consistent with manual guidance, or (3) use central tendency measures.

1. Method: Assess consistency with the IPC Technical Manual's reference table

While we cannot explicitly measure how closely a working group adheres to IPC guidance, we can study whether the consensus outcomes conform to those predicted by the underlying FSI data. The analysis in this section has three components.

(1) For each assessment, we evaluate the classifications implied by the FSIs as suggested by guidelines in the IPC Technical Manual and we determine whether the assessment is within the range of possible classifications predicted by the underlying FSI data. To put this measure of consistency with the manual in context, we then establish what range of assessments would also conform to the guidelines given the underlying FSI data.

(2) To further assess the consistency of classifications, we examine the extent to which they align with central tendency measures derived from the underlying food security indicator data.

(3) Finally, we compare the modal phase outcomes against the outcomes implied by underlying FSI data. We also compare the observed continuous measure of the percent of population in phase 3+ to the average percent of population in 3+ predicted by the underlying FSIs. Note that IPC analysts are trained to not take simple averages of the population or modal outcomes. For example, if, in the time between the fielding of the FSI survey and the TWG meeting, a large amount of aid or better weather came into the region, one might expect the TWG to assess the area at a lower classification than suggested by the earlier FSIs. Nonetheless these simplifying approaches allow us to compare how much TWG outcomes differ from simple statistical approaches.

2. *Findings: Consensus assessments are consistent with the IPC Technical Manual guidance*

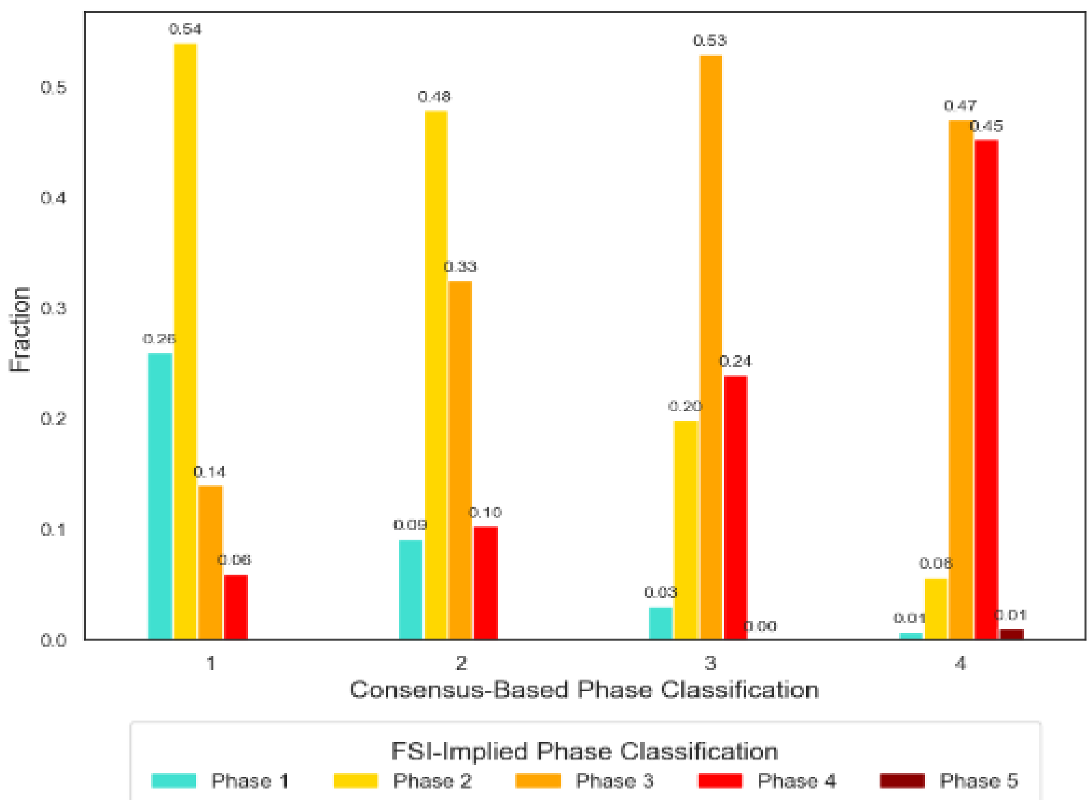
The IPC Technical Manual (2021) provides suggested classifications based on each FSI. We have shown that food security indicators often provide conflicting classifications. We now examine how closely the TWG's assessments (i.e., those we observe) align with the classifications implied by the underlying FSI data they rely on.

If the assessment is within the range suggested by FSI data under the IPC Acute Food Security Reference Table (IPC 2021, p. 37), we classify that assessment as consistent with the technical manual (see Figure SI.1 in the SI for a reproduction of the reference table). Suppose, for example, a working group had three FSI available: the HDDS score is 3-4 food groups, indicating phase 3; the FCS is acceptable but indicates deterioration from typical, indicating phase 2; and the rCSI score is 0-3, indicating phase 1. If the TWG reached a classification of 1, 2, or 3, we reported the observed assessment to be consistent with the manual's guidance.

We find that TWG consensus processes are highly consistent with the classifications indicated by the IPC manual. Only 3% (56/1849) of consensus phases are outside of the range of the FSI-implied outcomes (see SI Table SI.1).

However, our results indicate that it is generally the case that the underlying data and the reference table often suggest a wide range of possible assessments. In Figure 4, we plot for each observed IPC classification the range of other classifications that are possible given the underlying FSI (e.g., phase 1, 2, or 3 in the example above), generating histograms of all possible assessments that are consistent with the manual guidance for each observed phase. Figure 4 allows us to visualize the range of implied outcomes for each set of FSI, revealing a wide range that is often consistent with the manual guidance. For example, when TWGs reach assessments of phase 2, on average, 48% of their food security indicators are consistent with phase 2 levels, while 43% are consistent with phases 3 or 4. Observed phase 4 assessments are nearly evenly split between underlying indicators pointing to phase 3 and phase 4. Combined with the high degree of FSI discordance that we document in the previous section, analysis presented in Figure 4 indicates that the guard rails from the manual guidance are quite broad.

Figure 4. Histogram of all possible counterfactual consistent food insecurity phase classifications



Caption: The 1849 TWG assessments are based on 7428 food security indicators. Each assessment has a range of possible classification phases based on the underlying FSI. The histogram's x-axis shows the range of possible phases for each observed TWG assessment by possible phases. Note: rCSI is generally unable to differentiate between phases 3 to 5 as defined by the IPC manual, with very few exceptions. We assume that if rCSI is a 3 or higher, we assign rCSI as a 3. We recognize that rCSI could also be a 4 but would likely require at least one other FSI consistent with phase 4. Therefore, we rely on other FSI when identifying if the range could include phase 4.

In SI Figure SI.3, we show that only a very small fraction of the classifications has all underlying FSI data supporting the same classification. Most of our data have underlying FSIs that span a range of between two and three phase categories. In other words, often TWGs must classify assessment areas with FSI that point to different classifications.

3. *Findings: TWG phase and population estimates often diverge from the central tendency implied by the underlying FSI data.*

Given the range of possible classifications consistent with technical guidance implied by the underlying FSI data, a reasonable expectation is that technical working groups might rely on a modal aggregation — selecting the most frequently suggested classification — as a decision-making rule.

However, we find that TWG phase classifications do not tend to follow the modal outcome implied by the underlying FSI data. We find that less than half of the consensus-based classifications use the modal outcome suggested by the FSIs (Figure 4. See also SI Table SI.3). Note that this result is not in conflict with the guidance in the technical manual. Analysts are trained to not rely on simple statistical outcomes such as the mode and are instead trained to evaluate the FSI based on local contextual information and on contributing factors. This result does suggest, however, that TWGs are doing something other than a simple algorithmic determination.

We examine the population estimates in Figure 3. We plot the percent of population assessed to be in phase 3+ against the percent of population in phase 3+ suggested by each of the underlying food security indicators for Group 1 data (Figure 3). If the underlying food security indicators suggested the same percent of population is in phase 3+ as the consensus process outcome, those assessments should lie along the 45-degree line (shown in dashed grey). Compared to the consensus outcome, some FSIs tend to predict a larger percent of population in 3+ (e.g. FCS and LCS for almost the entire range) while other FSIs tend to predict smaller percentages (e.g. the rCSI).

As with the phase classifications, the consensus outcomes for population shares in 3+ fall somewhere in the middle of the food security indicators, suggesting that TWGs, on average, don't tend to pick one (or two) preferred indicator(s), nor do they simply average across available indicators to reach consensus outcomes. The average across the FSIs is shown as a thick black line. Again, we find the average departs from the observed consensus 45-degree line, although it tracks more closely as the severity increases.

In sum, multiple possible phase classifications as defined by the IPC manual's reference table are consistent with most TWG assessments. At the same time, the proportion of the population identified as acutely food insecure based on the IPC technical guidance is similarly broad. This highlights both of the challenges faced by TWGs and the considerable discretion the IPC technical manual allows them. Rather than strictly following the modal assessment or mean population estimates implied by the underlying FSI data, TWGs appear to apply a more nuanced judgment that incorporates additional contextual factors.

VII. Analysis 3: is the use of FSI data consistent with one another and across space and time?

The prior section considered whether TWGs treat the data in a manner consistent with the IPC Technical Manual or with assessments based on central tendencies (mean, mode) in the underlying food security indicators.

Given that the working groups are using the food security data in ways that are more nuanced than taking the mean or mode of the assessments implied by the FSI, we turn to an analysis of what they are doing – how they are using the food security data. We introduce and analyze a model of the consensus process to explore how TWGs may implicitly weight different food security indicators. Analysis of these weights help assess whether the use of FSI data is consistent across subgroups, over time within countries, and in relation to existing literature. Additionally, we evaluate the relative influence of FSI data compared to unobservable factors in shaping TWG outcomes.

1. *Method: A simple model of the consensus process*

For determination of population shares, we model consensus of any TWG in any round as a process of weighing the input data, and then adjusting the resulting IPC level and population shares to conform to TWG members' own analyses of the local situation. Thus, we model the IPC's decision-making process as:

$$y_{it} = \beta_{tk}\, FSI_{itk} + \mu_i + \alpha_{it} + \varepsilon_{it}$$

where y_{it} is the IPC analysis of the percent of population in a phase for IPC analysis area, *i*, in round, *t*. The parameters β_{tk} are the weights used by TWG on population shares implied by each food security indicator *k* at time *t*. We are not suggesting that a TWG explicitly allocates weights to each FSI. Rather, the process results in *implicit* weights, which might suggest that TWGs are using FSI information differently. We cannot directly observe how the working group weighs the evidence it has available, but we can infer the degree to which their assessment relies on each indicator. These weights provide some insight into the inner workings and consistency of the convergence process.

Differences in weights across food security indicators may result from TWG member skepticism about the relevance of a particular food security indicator for their country (e.g., the LCS), or concern about the relevance of particular survey questions in their context (e.g., self-reported questions about perceptions of hunger) for example. Thus, we might expect that some weights are common to a specific TWG analysis and round, while others may be common to a location over time.

If TWGs are concerned about the utility and veracity of a specific measure for the country, we would expect the process to result in similar weights on that FSI across the country. TWGs may also adjust the categorization for a specific location for regions that may be particularly vulnerable or have characteristics that make the population-weighted FSI measures less relevant (e.g., refugee camps), represented by μ_i. If the characteristics of these analysis areas do not change (much) over time, we would expect this adjustment to be similar across rounds. TWGs may make adjustments for a specific IPC analysis area in a specific round, α_{it} due to local information that is not captured in the FSI input data. Finally, ε_{it} are random noise. We cannot include location fixed effects since we have only one round of data collection for most countries in our sample.

The analyses we present below use the above model to estimate the weights a TWG places on the population shares at 3+ implied by all FSIs available to the TWG. We use the OLS regressions to estimate the percent of population assessed in phase 3+ as a function of FSIs for several analyses. We first examine weights by each group of FSI data. Since the population shares at 3+ implied by different individual FSIs capture different dimensions of food insecurity, we might expect the weight on each FSI to vary in an expected manner across data groups, and severity classes, and over time and space within a country. For example, given that HDDS and FCS both capture dietary quality (Vaitla et al. 2017; Headley and Ecker 2013), we expect a lower weight on HDDS when a data group includes both FCS and HDDS versus when it includes HDDS alone.

We then compare the weights on the different combinations of available FSIs and by severity level (i.e., phase 3+ or 4+, classified by TWG). Given that some FSIs are more relevant to particular phases, we expect their implicit weights to vary by severity of food crises and thus by phase classification. For example, the IPC manual indicates that rCSI has “non defining characteristics” for (i.e., cannot differentiate between) phases 3-5. We also study how weights vary across countries and over time and space within a country. Phase classifications can mask large variations in the type and extent of food insecurity in a given location (Andree et al. 2020 and Maxwell et al. 2020). We might expect that contextual or cultural factors would make certain indicators more valued in certain locales and that these indicators would remain valued over time for those locales. For example, in locations with basic (low variety) diets, there may be little change in FCS and HDDS, and therefore little information and relatively low weights on these measures in determining food security.

Finally, we use these regressions to analyze how much of the variation in population phase assessments can be explained by the underlying FSI data. We estimate these results by data group for two sets of regressions - one where we force the weights on the FSIs to be fixed over TWGs, and one where we let the weights vary by TWG [or country]. In the second, we interact country fixed effects with FSI to allow for country-specific understandings, and thus weights, of FSI.

2. *Findings: TWGs do not adhere to consistent relationships between FSIs. In some cases, their treatment of particular FSIs does not align with IPC guidance across severity classes. Evidence indicates TWGs treat FSI differently across countries and within the same country over time.*

We first consider the implicit weights on the vector of population shares at 3+ implied by all FSIs available to the TWG.[9] We find the weights vary substantially across different groups of FSI data. Some of these changes are unexpected given previously established relationships among the underlying FSI data.

Figure 5 presents the coefficients from four ordinary least squares regression models, estimating relative weights of different FSIs used by the TWGs in their assessments of populations in 3+. Each quadrant of the figure presents the results for a particular FSI data group. Comparing results within a quadrant provides evidence about relative weighting of indicators within a set of assessments with the same FSI data availability, while comparing across the quadrants provides insight into how the weights on individual indicators change with different sets of available indicators.

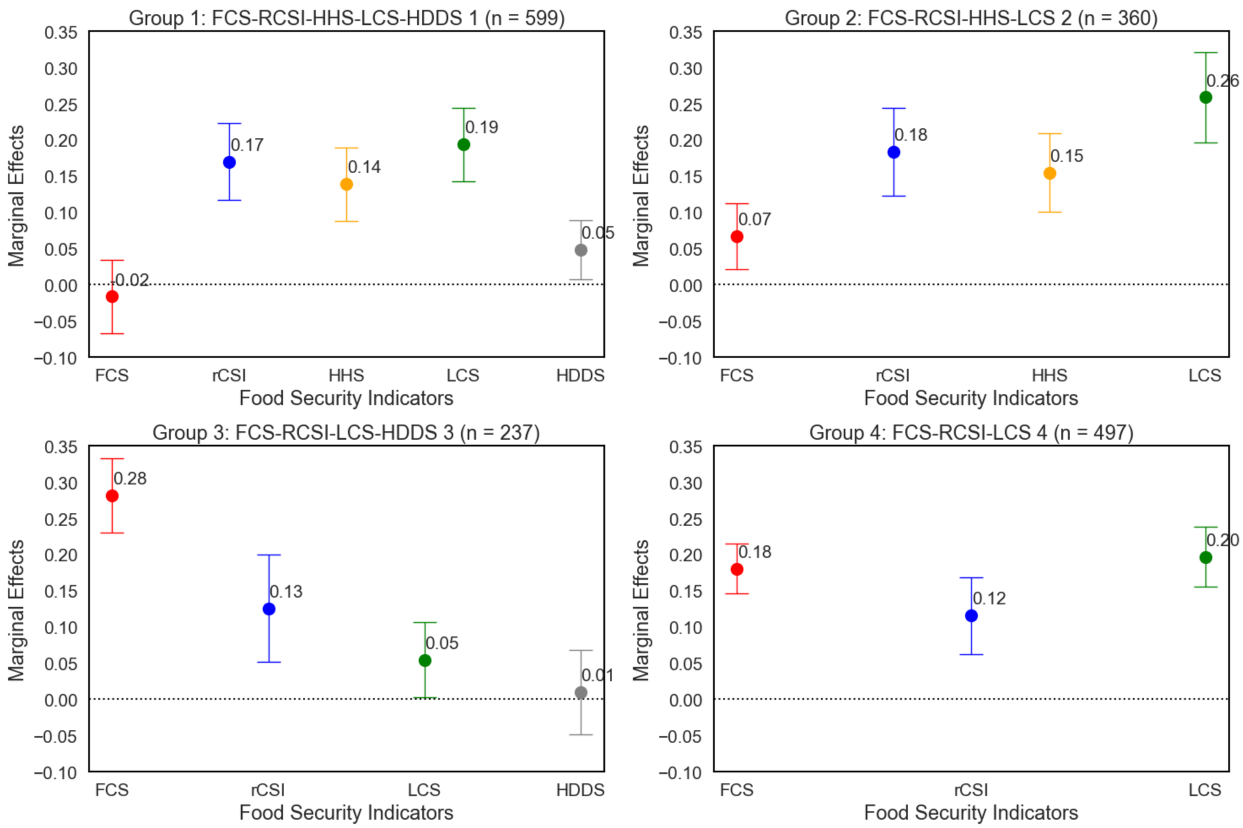
With the exception of FCS in Group 1 and HDDS in Group 3, the weights on each FSI are statistically significantly different from zero, indicating that these FSI contribute positively in TWG assessments. Within each group, most indicator weights are statistically different from one another except for Group 4.

For each data group, the sum of the coefficients on all FSI data is generally less than one. Thus, our model suggests that all of the underlying indicators would need to increase by more than one percentage point (i.e., more population in need) to induce the TWGs to increase the assessed

[9] We can observe a graphical representation of the implicit weights in Figure 3. The horizontal gap between the black line and the 45-degree dotted line can be interpreted as the role of the unobservable consensus process. When the black line is above the 45-degree line, the 3+ population implied by averaging across the food insecurity indicators is higher than the actual 3+ population reached through consensus. When it is below, the 3+ population implied by averaging across the food insecurity indicators is lower than the observed 3+ population reached through consensus. Notably, the implicit weights on the FSIs seem to vary by severity. As more of the population is in IPC phase 3 or above (crisis or worse), the averaged (thick black) line more closely tracks with the consensus (dotted) line, suggesting that the consensus approach to classifying more of the population more closely approximates taking an average than when classifying a smaller population in need. There are fewer classifications, and therefore more noise at the extremes across all indicators.

population in 3+ by one percentage point. This indicates TWGs are, if anything, under-sensitive to FSI data, and rely on unobserved information in their assessment of populations in 3+.

Figure 5: OLS model coefficients (i.e., weights) by combinations of available FSI: 3+ phase classification



Caption: This figure presents the coefficients from four OLS models analyzing the consensus-based 3+ shares. The four quadrants depict the coefficient and 95% confidence intervals for each FSI for each data availability group. Each model is specified as *y = f(FSI 3+)* where *y* represents the consensus-based 3+ population shares and FSI 3+ is a vector of 3+ population shares implied by FSIs.

By comparing the weights **across the data groups** in Figure 5, we can compare whether TWGs treat FSI information as substitutes. We do not see evidence of TWGs using FSIs in a consistent manner across data groups. Nor do we see that they are using measures that tend to correlate strongly as substitutes.

Compared to Group 1 TWGs, Group 2 TWGs do not have HDDS. Relative to Group 1, in Group 2, all four indicators (i.e., FCS, rCSI, HHS, and LCS) increase in importance. Prior research has established strong positive correlations between the two dietary diversity indicators, FCS and HDDS (they are often based on the same suite of questions) (Maxwell et al. 2014; Vaitla et al. 2017). These indicators should provide similar information to TWGs. This might suggest that TWGs might reasonably treat them as substitutes and with a higher weight on FCS while other FSI remain unchanged in Group 2 relative to Group 1. While we observe a higher weight on FCS in Group 2, we also see higher weights on experiential indicators (HHS, LCS, and rCSI), which are more correlated with one another than with HDDS. Similarly, moving from Group 1 to 3, we should see more weight placed on rCSI as we lose HHS as a measure given that rCSI and HHS are closely related at the household level (FANTA 2018). Instead, in Group 3, we observe more emphasis placed on FCS, which is not as well correlated with HHS and therefore not an obvious substitute for HHS.

In the SI (see Figure 6 and SI Figure SI.4), we also show that weights vary substantially between phase 3+ and phase 4+ assessments, and in some cases, in ways not consistent with IPC technical guidelines. For one data group, TWGs appear to place considerably more weight on rCSI for 4+ than for 3+. This is inconsistent with manual guidance (see SI Figure SI.1), which states that rCSI cannot be used to differentiate among phases 3–5.

We have established that implicit weights vary by group of available FSI data and phase classification. Variation in the way that assessments use available evidence could be consistent with TWGs down weighting or upweighting FSI based on their appropriateness for the local context. When we examine implicit weights on FSI by country, we find variation (SI Figure SI.5). We recognize that some TWGs may deem one FSI to be less relevant in some countries or districts than in others, which could explain some of our findings. However, we might expect to see the treatment of FSI persist across time within location.

In the SI (Figures SI.6 - SI.7), we show that weights vary by country and across rounds within the same country for three countries with multiple rounds. **We find evidence that working groups weigh indicators differently across time for the same location.** It does not appear that TWGs within countries have consistently ‘favored’ (or ‘disfavored’) FSI metrics, which could be indicative of a need for spatially contextualized indicators. Instead, we find that TWGs place different (implicit) weights on the same indicators across analyses.

Thus, the consensus process does not appear consistent in terms of how indicators are treated across time or space. Differences in weights may reflect a well-functioning convergence process. For example, some TWG participants might bring to the analysis unobservable information that may be correlated with a specific FSI, implying that a specific population share for that FSI should receive higher or lower weight. Similarly, some TWG participants may have specialized knowledge of certain areas or vested interests, leading them to advocate for phase adjustments—either upward or downward—relative to predictions from our TWG-level regression models. However, such differences could also raise concerns about inconsistency across space and time, especially if weighting patterns appear arbitrary or deviate from the guidance in the IPC Technical Manual.

3. The FSI explain about half of the variation in the predicted population assessments

Finally, we explore how much of the variation in the shares of the population assessed to be in phase 3+ are explained by the FSIs. In Table SI.3 and Figure SI.8, we report the R-squared from three estimates (population in 3+ estimated as a function of FSI information, then plus country-level fixed effects, and then plus interaction terms) across our four data groups. The R-squared measures how much variation is explained by the variables included in each model, shedding light on the relative importance of the FSI compared to other unobserved drivers of the assessments. Results reveal a significant amount of unexplained variation across the four data groups: 35–44% of the explained variance is attributed to unobservable factors. See SI Figure SI.8. This suggests that, beyond FSI-implied population estimates and country-specific context, additional, unobservable factors outlined in the IPC analytical framework (see SI Figure SI.1) play a key role in shaping consensus outcomes and explain around half of the variation in TWG outcomes.

VI. Discussion and Conclusion

The IPC, recently described in the media as “the tiny watchdog,” plays an important role in sounding global alarms about food insecurity and famine (Masri et al. 2024). It is heavily cited in major global reports on hunger, such as the United Nations-backed Global Report on Food Crises (FSIN and Global Network Against Food Crises 2024) and it is attracting increasing interest from policymakers, journalists, and researchers (Craze 2024; Masri et al. 2024). As the IPC Technical Manual (2021) highlights, the objective of the IPC is to provide “the essential information needed in a wide range of contexts in consistent, comparable and accountable ways” (p. 3). Given current resource constraints in aid, applying inconsistent thresholds for defining crises across space and time may have adverse consequences for those most in need. In this paper, we therefore assess the consistency of IPC assessments.

First our findings show TWGs work with highly discordant food security indicators, some of which they need to interpret contextually. Reaching consensus based on these data is therefore a significant undertaking. Second, we find that TWGs generate outcomes that are reliably consistent with the IPC Technical Manual but that IPC guidance allows for a broad range of classifications. Given that the IPC guidelines accommodate a range of assessment options, we analyze how TWGs are using the data available to them. We show TWGs do not appear to implement more mechanistic approaches based on central tendencies in the data, such as taking mean or mode. Third, using regression techniques, we examine the consistency of the estimated weights on FSIs. We estimate the weights on each FSI for the percent of population in IPC phase 3 or above (crisis or worse) by (a) the set of available FSIs and (b) country and analysis round. We also compare findings across levels of severity. We find that weights vary substantially over all of analyzed dimensions, and in ways that are not predictable given the established relationships among the underlying FSIs. The fact that the estimated weights vary is not in and of itself an indication of inconsistency of the process but does suggest that TWGs treat food security indicators differently across multiple dimensions.

This finding – that TWGs are highly consistent with guidelines but less consistent in their use of food security indicators across different data groups, time periods, and severity classes – creates an opportunity for the IPC to better document when and why TWGs consider certain indicators to be more or less useful than others. This would improve the legibility of the TWG process and help readers understand when FSIs may have contextually specific interpretations or when certain FSIs are more trusted than others. In turn, this transparency could contribute to increased trust in the results. With the data available, we cannot identify the causes for the inconsistency in use of FSI or if the inconsistency we document is of material concern. The inconsistency could reflect limitations of our models; we might be missing critical information observed by the TWG. In fact, consensus-based processes of the TWGs may be *working*, with the inconsistencies we document reflecting that TWGs adjust outcomes due to unobservable data and information. Or the consensus process may not be working optimally; biases, political capture, the loudest voice in the room, or other unobservable factors may influence outcomes. Future work that systematically incorporates data on within-TWG contextualizing factors could more comprehensively model the TWG process, would allow us to identify the influence of FSIs and the convergence process on TWG outcomes with more precision. This remains an important area for future work and underscores a tension between comparability and contextualization. Any reforms to improve transparency and consistency, however, should be designed in ways that avoid adding cost or delay to the IPC process, given the importance of timeliness and the reality of limited resources.

Our findings suggest that TWGs apply the data inconsistently. This may contribute to the conservative bias observed in food insecurity classifications and documented in Lentz et al. (under review). Variable weighting of indicators by context could systematically push assessments toward more cautious estimates. Alternatively, both the inconsistency in data use and the conservative bias may stem from a common underlying factor, such as risk aversion or institutional incentives. Further understanding these dynamics is critical for improving the transparency and reliability of IPC assessments.

Our analysis has several limitations. First, any study of the IPC must contend with a lack of an observable truth. That is, we cannot assess "consistency" in the purest sense. Rather, we use the IPC technical guidance and underlying food security indicators to examine TWG fidelity to process. Other approaches may be insightful as well. With our approach, we can observe and identify differences between our estimated outcomes and the outcomes of the consensus process. Where we observe differences, we cannot conclusively determine the drivers of the observed differences. Data limitations preclude us from accounting for contextualizing, contributing factors in our analysis. Contributing factors are important pieces of evidence for TWG analysis. These factors could (and should) shape TWG outcomes. Future research on the role of contributing factors in arriving at consensus outcomes is needed. Therefore, we underscore that observed differences do not necessarily mean inaccurate or wrong consensus-based outcomes.

Second, our findings take the available data and the IPC Technical Manual reference table as given. The IPC has stewardship over the consensus process but not over the data inputs. In practice, this means that while we have access to the aggregated FSI data, we do not observe the underlying analyses or data collection. Thus, any errors in data inputs that enter TWG analyses will also enter into ours.

Third, our analysis reflects the data available. Our findings may not apply to countries or time periods outside of our sample. We lack many classifications in IPC 1 and do not observe any IPC 5. Therefore, our findings should not be generalized to IPC 1 or 5 classifications. Future work incorporating data from more countries and more time periods within countries could add external validity and allow for additional time-series analyses.

Finally, our analysis does not allow us to observe how TWGs arrived at their classifications. Hunger is political. de Waal has argued that the IPC operates on the assumption of benevolence of actors involved (2024). However, members of TWG, donors, or state actors may seek to influence the IPC process and outcomes (Maxwell and Hailey 2021). News reports identify a few such instances, particularly in cases of extreme hunger where states are parties to conflict (Craze 2024; Masri et al. 2024). We do not observe if or when these cases occurred; if they do, they may be outsized contributors to the divergence between the observed outcomes and our estimations.

For users of the IPC, we offer the following insights.[10] As many regular users of IPC are aware, two classifications of a "3" do not necessarily indicate equivalent levels of food insecurity. An assessment of a "3" can indicate that 20% of the population is in IPC 3+ or that 80% is in IPC 3+ (Maxwell et al. 2020). Users should be careful about treating "3" as having specific and consistent meaning. In many cases, the population estimates side-by-side with classifications will be more informative than relying on classifications alone. For food security modelers may wish to account for population estimates, be cautious regarding the external validity of their findings, and be careful about drawing policy prescriptions from their findings. At the same time, food insecurity consistent with a classification of a 3 could result more from inadequate quality in one place (e.g., driven by HDDS and FCS) or from a collapse in coping strategies (e.g., driven by rCSI and LCS) elsewhere. This means that while both populations might experience acute food insecurity, the exact nature of and causes of the experience might differ. Policy implications of addressing food

[10] While our focus was on IPC current status assessments, IPC TWGs also produce projections. Research on other humanitarian systems have compared projections to current assessments as a means of assessing accuracy (Choularton and Krishnamurthy 2019). We warn that IPC users should not assume projections of a given period will be correlated with current status for the same period for two reasons. First, if projections crowd in humanitarian food assistance (as they are intended to do), the projections will not be correlated with the realized current status. Second, in complex humanitarian emergencies in particular, the drivers of food insecurity are dynamic and expecting projections to map to realized current status may be unrealistic. In this way, evaluating the accuracy of IPC projections presents challenges similar to evaluating the accuracy of other early warning programs including those working to reduce deforestation (Finer et al. 2014), or to mitigate the spread of disease (WHO 2011).

insecurity levels of 20% versus 80% clearly differ as do responses addressing deteriorating dietary diversity versus loss of coping strategies.

Given that the nature of food insecurity itself is multidimensional and complex, flexibility in IPC assessments to include contextual factors is one of its strengths. Our results show that IPC TWGs face often discordant food security indicators, yet nearly always still make classifications that are highly consistent with technical guidance. However, the flexibility of the IPC also has important implications for users of IPC information. Future work on further understanding the relationships between food security indicators and guidance regarding which indicators are contextually appropriate to TWGs could offer important next steps.

I. References

Alkire, S., & Foster, J. 2011. Understandings and misunderstandings of multidimensional poverty measurement. *Journal of Economic Inequality*, 9(2), 289–314.

Alkire, S., Meinzen-Dick, R., Peterman, A., Quisumbing, A., Seymour, G. & Vaz, A., 2013. The women's empowerment in agriculture index. *World Development*, 52, 71-91. https://doi.org/10.1016/j.worlddev.2013.06.007

Andree, B. P. J., Chamorro, A., Kraay, A., Spencer, P., & Wang, D. 2020. *Predicting Food Crises* No. 9412; Policy Research Working Paper. The World Bank. https://doi.org/10.1596/1813-9450-9412

Barrett, C. B. 2010. Measuring food insecurity. *Science*, 327(5967), 825-828.

Bageant, L., Lentz, E., Narayanan, S., Jensen, N., and Watson Lepariyo, W. 2024 How do women's empowerment metrics measure up? A comparative analysis. *Food Policy.* 129. https://doi.org/10.1016/j.foodpol.2024.102764

Bourguignon, F. 2006. From income to endowments: the difficult task of expanding the income poverty paradigm. *Poverty and inequality*, 76-102.

Charmes, J., & Wieringa, S. 2003. Measuring women's empowerment: an assessment of the gender-related development index and the gender empowerment measure. *Journal of human development*, 4(3), 419-435.

Choularton, R. J. and Krishnamurthy, P. K. 2019. How Accurate Is Food Security Early Warning? Evaluation of FEWS NET Accuracy in Ethiopia. *Food Security* 11(2). April: 333–44, https://doi.org/10.1007/s12571-019-00909-y.

Clapp, J., Moseley, W. G., Burlingame, B., & Termine, P. 2022. The case for a six-dimensional food security framework. *Food Policy*, *106*, 102164.

Craze, J. 2024. Taxonomies of Hunger. *New Left Review:* 148. July - August.

Eilerts, G. 2006. Niger 2005: Not a famine, but something much worse. *Humanitarian Practice Network.* Vol. 33. April 18. Visited February 5, 2025. https://odihpn.org/publication/niger-2005-not-a-famine-but-something-much-worse/

FANTA. 2011. HHS Indicator Guide. Accessed February 10, 2025. https://www.fantaproject.org/sites/default/files/resources/HHS-Indicator-Guide-Aug2011.pdf

FANTA 2018. FANTA III: Tools. Accessed February 24, 2025. https://www.fantaproject.org/tools

Finer, M., Novoa, S., Weisse, M. J., Petersen, R., Mascaro, J., Souto, T., Stearns, F., and Martinez, R. G. 2018. Combating deforestation: From satellite to intervention. *Science* 360, no. 6395: 1303-1305.

Foster, J. E. 1998. Absolute versus relative poverty. *The American economic review*, *88*(2), 335-341.

Frankenberger, T. R. and Verdujin, R. 2011. *Final Report*, 90. Accessed January 21, 2025. https://www.ipcinfo.org/fileadmin/user_upload/ipcinfo/docs/1_IPC_Glob_Proj_Eval_04_11_Report.pdf

FSIN and Global Network Against Food Crises. 2024. *GRFC 2024*. Rome. Accessed November 4, 2024. https://www.fsinplatform.org/report/global-report-food-crises-2024/

Gebreyesus, S. H., Lunde, T., Mariam, D. H., Woldehanna, T., & Lindtjørn, B. 2015. Is the adapted Household Food Insecurity Access Scale (HFIAS) developed internationally to measure food insecurity valid in urban and rural households of Ethiopia? *BMC nutrition*, *1*, 1-10.

Harvey, P. 2005. Niger: Shared responsibility for a slow response. ODI Global. Explainer. August 9. Visited February 5, 2025. https://odi.org/en/insights/niger-shared-responsibility-for-a-slow-response/

Headey, D., & Ecker, O. 2013. Rethinking the measurement of food security: from first principles to best practice. *Food security*, *5*, 327-343.

Herforth, A. W., Wiesmann, D., Martínez-Steele, E., Andrade, G., & Monteiro, C. A. 2020. Introducing a suite of low-burden diet quality indicators that reflect healthy diet patterns at population level. *Current developments in nutrition*, 4(12), nzaa168.

IPC Global Partners. 2008. *Integrated Food Security Phase Classification Technical Manual*. Version 1.1. FAO. Rome. Accessed January 22, 2025. https://www.fao.org/4/i0275e/i0275e.pdf

IPC Global Partners. 2021. Integrated Food Security Phase Classification Technical Manual Version 3.1. Evidence and Standards for Better Food Security and Nutrition Decisions. Rome. Accessed July 31, 2023. https://www.ipcinfo.org/ipcinfo-website/resources/ipc-manual/en/

IPC. 2023. IPC Global Strategic Programme: 2023-2026. January. Accessed July 7, 2023. https://www.ipcinfo.org/fileadmin/user_upload/ipcinfo/docs/IPC_GSP_2023_2026.pdf

IPC 2024. IPC Partnership and Governance. Visited November 4, 2024. https://www.ipcinfo.org/ipcinfo-website/ipc-global-partnership-and-programme/partnership/en/

IPC 2025. IPC Global Strategic Partnership and Programme. Visited February 5, 2025. https://www.ipcinfo.org/ipcinfo-website/ipc-global-partnership-and-programme/en/

Klasen, Stephen. 2000. Measuring poverty and deprivation in South Africa. *Review of income and wealth* 46 (1): 33-58.

Lentz, E., Baylis, K., Kim, C. and Michelson, H. *under review*. Global Assessment of acutely hungry misses one in four.

Masri, L., Nelson, D., Michael, M., Stecklow, S., McNeill, R., Dowdell, J., and Lesser, B. 2024. An elaborate global system exists to prevent famine. Its failing. *Reuters.* December 4. Accessed February 5, 2025. https://www.reuters.com/investigates/special-report/famine-response-overview/

Maxwell, D., Vaitla, B., and Coates, J. 2014. How do indicators of household food insecurity measure up? An empirical comparison from Ethiopia. *Food policy*, *47*, 107-116.

Maxwell, D. Khalif, A., Hailey, P., and Checchi, F. 2020. "Viewpoint: Determining famine: Multi-dimensional analysis for the twenty-first century." *Food Policy,* vol 92. https://doi.org/10.1016/j.foodpol.2020.101832.

Maxwell, D. and Hailey, P. 2021. Analysing Famine: The Politics of Information and Analysis in Food Security Crises. *Journal of Humanitarian Affairs:* 3(1). https://doi.org/10.7227/JHA.055

Maxwell, Daniel, Guhad Adan, Peter Hailey, Matthew Day, Stephen BJ Odhiambo, Lilian Kaindi, James Njiru, Aishwarya Venkat, and Anastasia Marshak. 2023. Using the household hunger scale to improve analysis and classification of severe food insecurity in famine-risk conditions: Evidence from three countries. *Food Policy* 118: 102449.

Narayanan S, Lentz E, Fontana M, De A, Kulkarni B. 2019. Developing the Women's Empowerment in Nutrition Index in Two States of India. *Food Policy.* Dec; 89:101780.

Pradhan, M., and Ravallion, M. 2000. Measuring poverty using qualitative perceptions of consumption adequacy. *Review of Economics and Statistics*, *82*(3), 462-471.

Ravallion, M. 2016. Toward better global poverty measures. The Journal of Economic Inequality, 14, 227-248.

Ravallion, M. 2020. On measuring global poverty. *Annual Review of Economics*, *12*(1), 167-188.

Sen, A. 1983. Poor, relatively speaking. *Oxford Economic Papers*, 35(2), 153–169.

United Kingdom Foreign Commonwealth and Development Office, 2021. G7 Principles Supporting a Vision of Improved Global Food Security Monitoring and Analysis. Accessed February 24, 2025. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1039534/g7-principles- supporting-vision-improved-global-food-security-monitoring-and-analysis.pdf

United Nations Office for Coordination of Humanitarian Affairs (UN OCHA). 2024. Global Humanitarian Overview: 2024. Accessed February 5, 2025. https://www.unocha.org/publications/report/world/global-humanitarian-overview-2024-enarfres

Vaitla, B., Coates, J., Glaeser, L., Hilbruner, C., Biswal, P., and Maxwell, D. 2017. The measurement of household food security: Correlation and late variable analysis of alternative indicators in a larger multi-country dataset. *Food Policy.* 68: 193-205.

World Health Organization. 2011. *Early warning surveillance and response in emergencies: report of the second WHO technical workshop,* 10-11 May 2011, World Health Organization, Geneva, Switzerland (No. WHO/HSE/GAR/DCE/2011.2). World Health Organization.

Young, S. L., Boateng, G. O., Jamaluddine, Z., Miller, J. D., Frongillo, E. A., Neilands, T. B., ... & Stoler, J. 2019. The Household Water InSecurity Experiences (HWISE) Scale: development and validation of a household water insecurity measure for low-income and middle-income countries. *BMJ global health*, 4(5), e001750.

Inside the black box: how consistent are global food security crisis assessments? Supplementary Information

Figure SI.1: IPC Acute Food Insecurity Reference Table from IPC Technical Manual version 3.1 (b37-39, 2021)

Figure 27: The IPC Acute Food Insecurity Reference Table (Tool 3)

Purpose: to guide convergence of evidence by using generally accepted international standards and cut-offs. The classification is intended to guide decision-making aiming at short-term improvements in food security.

Phase name and description	Phase 1 None/Minimal	Phase 2 Stressed	Phase 3 Crisis	Phase 4 Emergency	Phase 5 Catastrophe/ Famine
	Households are able to meet essential food and non-food needs without engaging in atypical and unsustainable strategies to access food and income.	Households have minimally adequate food consumption but are unable to afford some essential non-food expenditures without engaging in stress-coping strategies.	Households either: • Have food consumption gaps that are reflected by high or above-usual acute malnutrition; or • Are marginally able to meet minimum food needs but only by depleting essential livelihood assets or through crisis-coping strategies.	Households either: • Have large food consumption gaps which are reflected in very high acute malnutrition and excess mortality; or • Are able to mitigate large food consumption gaps but only by employing emergency livelihood strategies and asset liquidation.	Households have an extreme lack of food and/or other basic needs even after full employment of coping strategies. Starvation, death, destitution and extremely critical acute malnutrition levels are evident. (For famine Classification, an area needs to have extreme critical levels of acute malnutrition and mortality)
Priority response objectives	Action required to build resilience and for disaster risk reduction	Action required for disaster risk reduction and to protect livelihoods	**Urgent action required to:** Protect livelihoods and reduce food consumption gaps	Save lives and livelihoods	Revert/prevent widespread death and total collapse of livelihoods

First-level outcomes refer to characteristics of food consumption and livelihood change. Thresholds that correspond as closely as possible to the Phase description are included for each indicator. Although cut-offs are based on applied research and presented as global reference, correlation between indicators is often somewhat limited and findings need to be contextualized. The area is classified in the most severe Phase that affects at least 20% of the population.

Food security first-level outcomes	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
Food consumption (focus on energy intake)	**Quantity: Adequate energy intake**	**Quantity: Minimally Adequate**	**Quantity: Moderately Inadequate** – Moderate deficits	**Quantity: Very Inadequate** – Large deficits	**Quantity: Extremely Inadequate** – Very large deficits
	Dietary energy intake*: Adequate (avg. 2,350 kcal pp/day) and stable	**Dietary energy intake:** Minimally adequate (avg. 2,100 kcal pp/day)	**Dietary energy intake:** Food gap (below avg. 2,100 kcal pp/day)	**Dietary energy intake:** Large food gaps well below 2,100 kcal pp/day	**Dietary energy intake:** Extreme food gap
	Household Dietary Diversity Score*: 5-12 food groups and stable	**Household Dietary Diversity Score:** 5 FG but deterioration ≥1 FG from typical	**Household Dietary Diversity Score:** 3-4 FG	**Household Dietary Diversity Score:** 0-2 FG (NDC to differentiate P4 and 5)	**Household Dietary Diversity Score:** 0-2 FG
	Food Consumption Score*: Acceptable and stable	**Food Consumption Score:** Acceptable but deterioration from typical	**Food Consumption Score:** Borderline	**Food Consumption Score:** Poor (NDC to differentiate P4 and 5)	**Food Consumption Score:** Poor (NDC to differentiate P4 and 5)
	Household Hunger Scale*: 0 (none)	**Household Hunger Scale:** 1 (slight)	**Household Hunger Scale:** 2-3 (moderate)	**Household Hunger Scale:** 4 (severe)	**Household Hunger Scale:** 5-6 (severe)
	Reduced Coping Strategies Index*: 0-3	**Reduced Coping Strategies Index:** 4-18	**Reduced Coping Strategies Index:** ≥ 19 (non-defining characteristics (NDC) to differentiate P3, 4 and 5)	**Reduced Coping Strategies Index:** ≥ 19 (NDC to differentiate P3, 4 and 5)	**Reduced Coping Strategies Index:** ≥ 19 (NDC to differentiate P3, 4 and 5)
	Household Economy Analysis*: No livelihood protection deficit	**Household Economy Analysis:** Small or moderate livelihood protection deficit <80%	**Household Economy Analysis:** Livelihood protection deficit ≥80%; or survival deficit < 20%	**Household Economy Analysis:** Survival deficit ≥20% but <50%	**Household Economy Analysis:** Survival deficit ≥50%
	Food Insecurity Experience Scale*: (FIES 30 days recall): < -0.58	**FIES:** Between -0.58 and 0.36	**FIES:** > 0.36 (NDC to differentiate between Phases 3, 4 and 5)	**FIES:** > 0.36 (NDC to differentiate between Phases 3, 4 and 5)	**FIES:** > 0.36 (NDC to differentiate between Phases 3, 4 and 5)
Livelihood change (assets & strategies)	**Livelihood change:** Sustainable livelihood strategies and assets	**Livelihood change:** Stressed strategies and/or assets; reduced ability to invest in livelihoods	**Livelihood change:** Accelerated depletion/erosion of strategies and/or assets	**Livelihood change:** Extreme depletion/ liquidation of strategies and assets	**Livelihood change:** Near complete collapse of strategies and assets
	Livelihood coping strategies*: No stress, crisis or emergency coping observed	**Livelihood coping strategies:** Stress strategies are the most severe strategies used by the household in the past 30 days	**Livelihood coping strategies:** Crisis strategies are the most severe strategies used by the household in the past 30 days	**Livelihood coping strategies:** Emergency strategies are the most severe strategies used by the household in the past 30 days	**Livelihood coping strategies:** Near exhaustion of coping capacity

Second-level outcomes refer to area-level estimations of nutritional status and mortality that are especially useful for identifying more severe phases when food gaps are expected to impact malnutrition and mortality. For both nutrition and mortality area outcomes, household food consumption deficits should be an explanatory factor in order for that evidence to be used in support of the classification.

Food security second-level outcomes	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
Nutritional status*: **Global Acute Malnutrition (GAM) based on Weight-for-Height Z-score (WHZ)***	**Acceptable:** <5%	**Alert:** 5-9.9%	**Serious:** 10-14.9% or > than usual	**Critical:** 15-29.9% or > much greater than usual	**Extremely Critical:** ≥30%
Nutritional status*: **Global Acute Malnutrition based on Mid-Upper Arm Circumference (MUAC)***		<5%	5-9.9%	10-14.9%	≥15%
Nutritional status*: **Body Mass Index (BMI) <18.5***	<5%	5-9.9%	10-19.9%, 1.5 x greater than baseline	20-39.9%	≥ 40%
Mortality	**Crude Death Rate*:** <0.5/10,000/day	**Crude Death Rate:** <0.5/10,000/day	**Crude Death Rate:** 0.5-0.99/10,000/day	**Crude Death Rate:** 1-1.99/10,000/day OR >2x reference	**Crude Death Rate:** ≥2/10,000/day
	Under-five Death Rate*: <1/10,000/day	**Under-five Death Rate:** <1/10,000/day	**Under-five Death Rate:** 1-1.99/10,000/day	**Under-five Death Rate:** 2-3.99/10,000/day	**Under-five Death Rate:** ≥4/10,000/day

For contributing factors, specific indicators and thresholds for different phases need to be determined and analysed according to the livelihood context; however, some general descriptions for contributing factors are provided below.

Food security contributing factors	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
Food availability, access, utilization, and stability	Adequate to meet short-term food consumption requirements	Borderline adequate to meet food consumption requirements	Inadequate to meet food consumption requirements	Very inadequate to meet food consumption requirements	Extremely inadequate to meet food consumption requirements
	Safe water* ≥15 litres pp/day	Safe water marginally ≥15 litres pp/day	Safe water > 7.5 to 15 litres pp/day	Safe water >3 to <7.5 litres pp/day	Safe water <3 litres pp/day
Hazards and vulnerability	None or minimal effects of hazards and vulnerability on livelihoods and food consumption.	Effects of hazards and vulnerability stress livelihoods and food consumption.	Effects of hazards and vulnerability result in loss of assets and/or significant food consumption deficits.	Effects of hazards and vulnerability result in large loss of livelihood assets and/ or extreme food consumption deficits.	Effects of hazards and vulnerability result in near complete collapse of livelihood assets and/or near complete food consumption deficits.

Notes:

i. **Adequate dietary energy intake** relates to the condition of regularly consuming, over a significant period of time, an amount of food that provides the dietary energy needed to cover the requirements for an active and healthy life. Dietary energy intake is used as a convention and convenience to assess the average energy requirements for a population group. Characteristics that affect requirements include gender, age, body size, body composition and physical activity level as well as unknown factors that produce variations among individuals, as defined by the World Health Organization (WHO, 1985). The energy cut-offs included in the IPC Acute Food Insecurity Reference Table are not intended to be used for empirical assessment of percentage of the population consuming adequate/inadequate amounts of food, but rather, the indicator acts as a reference for food consumption, and the cut-off of 2,100 kcal/day is associated with the Household Economy Analysis (HEA) survival deficit cut-off and borderline FCS. The selected dietary energy requirements are based on average requirements for an average individual (BMI of 21 - 22), engaged in normally active life (physical activity level, or physical activity level = 1.75) for Phase 1, with an average of 2,350 Kcal/day, and in a sedentary lifestyle (physical activity level=1.55) for Phase 2 (FAO, WHO and United Nations University, 2004) with an average of 2,100 Kcal/day.

ii. **The Household Dietary Diversity Score** (HDDS) is an indicator developed by Food and Nutrition Technical Assistance (FANTA) and promoted by FAO. It aims to reflect the economic ability of a household to access a variety of foods and is based on households' self-reporting of the number of food groups consumed in the previous 24 hours. IPC cut-offs have been prepared for HDDS with 12 food groups, based on the FANTA/FEWS NET Household Food Consumption Indicator Study (2015).

iii. The **Food Consumption Score** (FCS) is a WFP corporate indicator collected in all assessments and monitoring activities. The FCS is a composite score based on self-reported information on nine consumed food groups and food frequency (number of days food groups were consumed during the past seven days), weighted by the ascribed relative nutritional importance of different food groups. Based on standard thresholds, households are classified into one of three food consumption groups: poor, borderline, or acceptable, with scores of ≤21, 28 and 35, respectively, except in situations of high oil and sugar consumption, for which the cut-offs used for the same groups are ≤28, 35 and 42, respectively. These same groupings are used as cut-offs for different phases in the IPC Acute Food Insecurity Reference Table.

iv. The **Household Hunger Scale** (HHS) is an indicator developed by FANTA. It assesses whether households have experienced problems of food access in the preceding 30 days, as reported by the households themselves. The HHS assesses the food consumption strategies adopted by households facing a lack of access to food. The cut-offs for the HHS are based on the FANTA (2015) Household Food Consumption Indicator Study report, and the alignment with the Acute Food Insecurity Reference Table phase descriptions.

v. The **reduced Coping Strategies Index** (rCSI) developed by CARE International is an experience based indicator collecting information on household use and the frequency of five different food-based coping strategies over the past 7 days. It is thought to be most useful in early onset crises when households change their food consumption patterns to respond to shocks, but not in protracted emergencies when households are likely to have already exhausted some coping mechanisms. The rCSI cut-offs are based on FANTA (2015) and the validation conducted by WFP.

vi. The **Household Economy Analysis** (HEA) is a livelihoods-based framework founded on the analysis of people in different social and economic circumstances. In particular, the HEA examines the self-reporting of information on: (i) how people access the food and cash they need; (ii) their assets, the opportunities available to them, and the constraints they face; and (iii) the options open to them in times of crisis. Two thresholds define basic needs in the HEA: the Survival Threshold and the Livelihoods Protection Threshold. The HEA Survival Threshold represents the most basic needs, including minimum food energy requirements (calorie requirements), the costs associated with food preparation and consumption if associated inputs are purchased (such as salt, firewood or kerosene), as well as expenditure on water for human consumption. All HEAs should consider the extent of reversible coping that is possible. HEA deficits are presented with cut-offs that reflect the expected situation in terms of livelihood stress and food gaps, as explained in IPC phase descriptions.

vii. **FIES** cut-offs are common, normalized thresholds developed specifically for use with the FIES 30 day recall in the IPC Acute Food Insecurity Reference Table. These thresholds do not correspond to those defined for use of FIES in the context of SDG monitoring and in the IPC Chronic Food Insecurity Reference Table, which are different and based on a 12 month recall period. The threshold that identifies "moderate" food insecurity in the context of SDG monitoring is less severe than the one that identifies IPC Acute Phase 3 or worse. While the standard FIES including 8 questions (i.e. 8 items) does not include cut-offs to differentiate between Phases 3, 4 and 5, an extended version of the FIES has been created and preliminary findings indicate that this extended version might be able to better differentiate between Phase 3, Phase 4 and Phase 5. Use of available FIES extended data for analyses, this should be carefully applied as indirect evidence and only with support from the IPC Global Support Unit.

viii. **Livelihood Coping Strategies** (LCS) is an indicator developed by WFP and is derived from a series of questions regarding the household's experience with livelihood stress and asset depletion due to lack of food or lack of money to buy food during the 30 days prior to the survey. The module needs to be adapted based on local context, both in terms of the strategies selected for data collection and the severity assigned to each strategy during analysis. For IPC Acute Food Insecurity, this indicator needs to be carefully analysed together with evidence on acute events and their impact on the food security pillars (availability, access, utilization and stability). This indicator may have limited use in severe protracted crises, since households may have engaged in and exhausted specific activities prior to the recall period. Analysts should also consider that less vulnerable households may be more capable of changing livelihood strategies and asset levels, and thus may have a higher score, not because they are facing more severe food insecurity, but because they are more capable of responding to shocks (e.g. wealthier households are likely to have more savings, better access to loans, and more animals to sell than poorer households). For the purpose of IPC Acute Food Insecurity classification, analysts should identify the most severe level of coping used by households. IPC cut-offs are based on groupings of strategies, i.e. stress, crisis and emergency strategies, depending on the strategies' sustainability and potential negative impact on future livelihoods and food security of the household.

ix. **Nutritional status and mortality** are used to support the classification of acute food insecurity due to the expected linkages between severity of food deprivation and acute malnutrition and mortality. Household food consumption deficits must be a likely explanatory factor of acute malnutrition and mortality in order for this evidence to be used to support a phase classification. For example, elevated malnutrition due to disease outbreak or lack of access to health care should not be used as evidence for an IPC Acute Food Insecurity Analysis if it is determined to not likely be related to food consumption deficits. Similarly, excessive mortality rates due to trauma-related deaths should not be used as evidence for Acute Food Insecurity Phase classification. A complementary IPC for Acute Malnutrition has been developed to inform decision-makers of the severity and likely drivers of acute malnutrition.

x. **Global acute malnutrition based on weight-for-height Z-score (GAM based on WHZ) is defined as** the percentage of children under five who are below -2 standard deviations of the median of weight for height (<-2 WHZ) or in the presence of oedema. Cut-offs are derived from WHO guidance, as well as from the Review of Nutrition and Mortality Indicators for the IPC study (2009).

xi. **Global Acute Malnutrition based on mid-upper arm circumference (GAM based on MUAC)** is defined as the percentage of children under five who have readings below 125 mm or the presence of oedema. Although GAM based on MUAC is a common measure of acute malnutrition, especially in emergency settings when the IPC Acute Food Insecurity classification is most relevant, global thresholds have not been developed. Evidence on GAM based on MUAC is included in the IPC so that evidence use is maximized, especially in emergency settings. The IPC acknowledges that concordance between MUAC and WHZ varies depending on context and is usually around 40–50 percent. The MUAC thresholds endorsed by the IPC have been developed based on extensive research by the Centers for Disease Control and Prevention and the JRC on the specificity and applicability of MUAC for the detection of GAM prevalence at the population level. MUAC thresholds can only be used in conjunction with the other contextual information by taking into account the immediate causes of acute malnutrition and the locally understood relationship between MUAC and WHZ prevalence, and by using the convergence of evidence approach.

xii. The **Body Mass Index** (BMI) measures central body mass and is an indicator of weight in relation to height. BMI is typically measured on non-pregnant women between 15 and 49 years of age. The IPC thresholds are based on the percentage of people with scores of <18.5. The thresholds use the WHO reference cut-offs that have been adopted by the IPC.

xiii. The **crude death rate** (CDR) is an indicator that accounts for all the deaths that have occurred per day per 10,000 people over a given recall period (often 90 days) in an area or community. According to the IPC Acute Food Insecurity Analysis, the CDR should not include trauma-related deaths, but should include deaths due to unknown causes. IPC cut-offs are based on WHO guidance, as well as on the Review of Nutrition and Mortality Indicators for the IPC study (2009).

xiv. The **under-five death rate** (U5DR) refers to all deaths of children under five (up to 59 months) per 10,000 children under five per day over a given recall period (often 90 days) in an area or community. The U5DR is typically around twice that of the CDR. The U5DR should not include trauma-related deaths. The under-5 mortality rate (i.e. the probability of dying between birth and the fifth birthday per 1,000 live births) can be used in order to understand the indicative U5DR, if the conditions between the collection of data for the under 5 mortality rate and the current situation have not changed.

xv. **Access to safe water** of ≥15 litres per person per day and further severity cut-offs per day per person for other phases are based on Sphere guidance for emergency situations. However, exact information on water quantity is rarely available outside camp settings or other situations where access to water is monitored.

Non-defining characteristic (NDC) is included for some indicators in the IPC Acute Food Insecurity Reference Table when no cut-offs were identified to differentiate between some Phases. For example, given that a "poor" FCS is indicative of Phases 4 and 5 (since it is an NDC to differentiate between Phases 4 and 5), the proportion of households with a "poor" score should be indicative of the proportion of households in Phases 4 and 5.

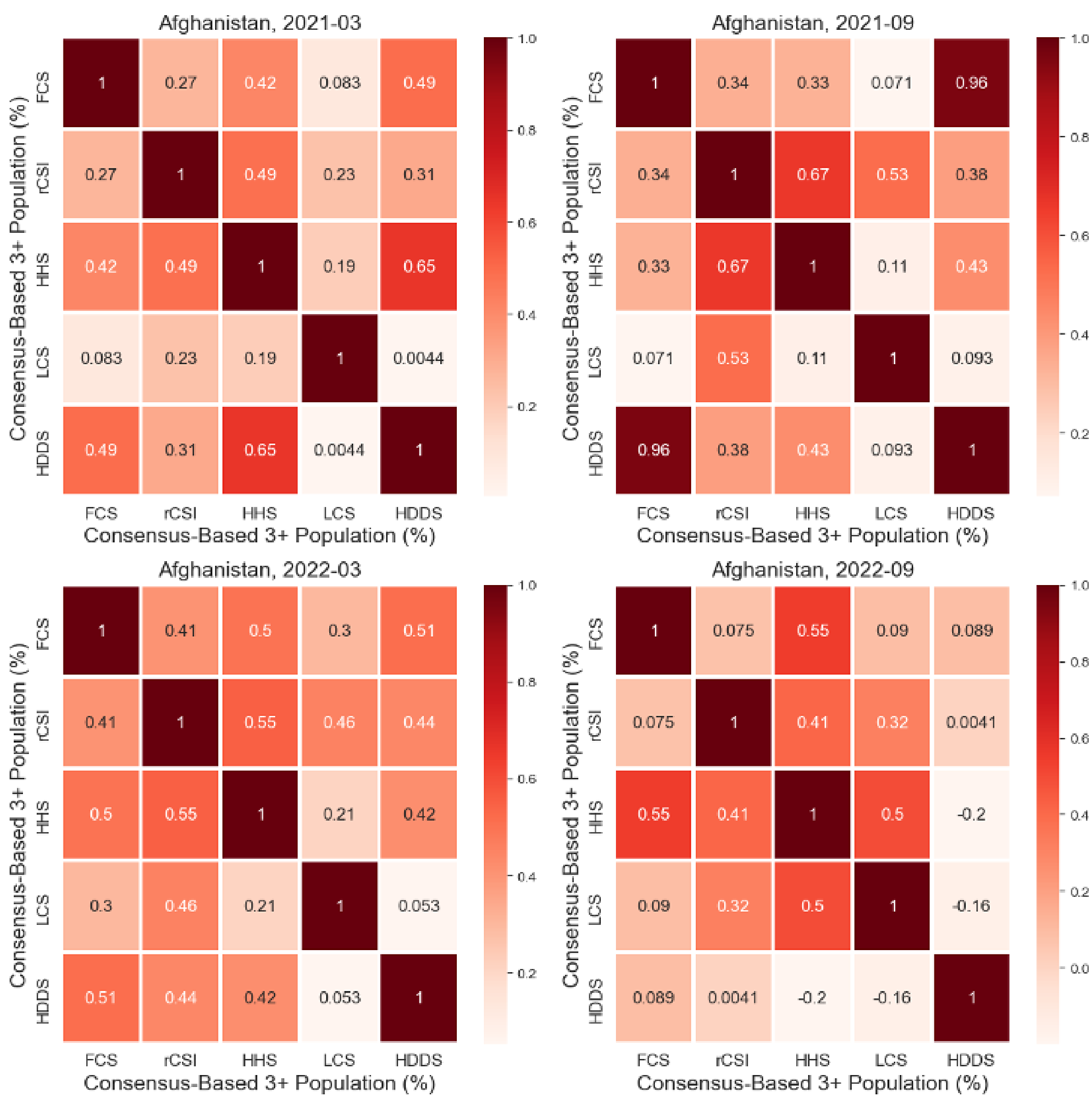
Table SI.1. “Outside of the range” count

Country	Round	“Outside of the Range” Count	Observation
Afghanistan	2021-03	0	45
Afghanistan	2021-09	0	45
Afghanistan	2022-03	0	45
Afghanistan	2022-09	0	45
CAR	2021-04	0	65
CAR	2021-09	1	62
CAR	2022-09	0	64
DRC	2021-02	3	166
DRC	2021-09	3	179
DRC	2022-08	5	184
Djibouti	2022-03	0	15
Ethiopia	2021-05	0	13
Guatemala	2022-03	0	22
Haiti	2021-09	0	31
Kenya	2021-02	2	23
Kenya	2021-09	1	23
Lebanon	2022-09	10	52
Madagascar	2021-05	2	10
Madagascar	2022-04	2	16
Madagascar	2022-11	0	21
Mozambique	2021-11	7	63
Pakistan	2021-03	0	19
Pakistan	2021-10	1	25
South Sudan	2022-10	1	77
Sudan	2021-04	7	179
Sudan	2022-05	4	179
Yemen	2020-10	7	181

We find that the correlation coefficients change substantially over time even within the same country. For example, the correlation between the percent of population in IPC phase 3 or above (crisis or worse) assessed by FCS and rCSI varies from 0.41 in March of 2022, to 0.075 six months later. This discordance in the underlying indicators underscores the challenges of reaching consensus and the importance of the consensus process.

We chose Afghanistan because it has multiple TWG analyses, allowing us to examine trends over time. However, we also note that this period coincides with a highly dynamic situation (i.e., the Islamic Republic's return to power in the fall of 2021), which could account for the lack of consistency.

Figure SI.2. Correlation Matrix: FSI implied 3+ population (%) for Afghanistan by round



Caption: This figure presents a Pearson correlation matrix illustrating the relationships among the share of the population assessed to be in IPC 3+ based on various FSI such as HHS, rCSI, FCS, LCS, and HDDS. The analysis specifically focuses on the Afghanistan sample, and the results are presented by round. (n = 180)

Note: In the case of the (AFG: 2022-09) sample, 20 out of 45 analysis areas were lacking HDDS implied population estimates, leading to a notable difference in sample size compared to other time periods.

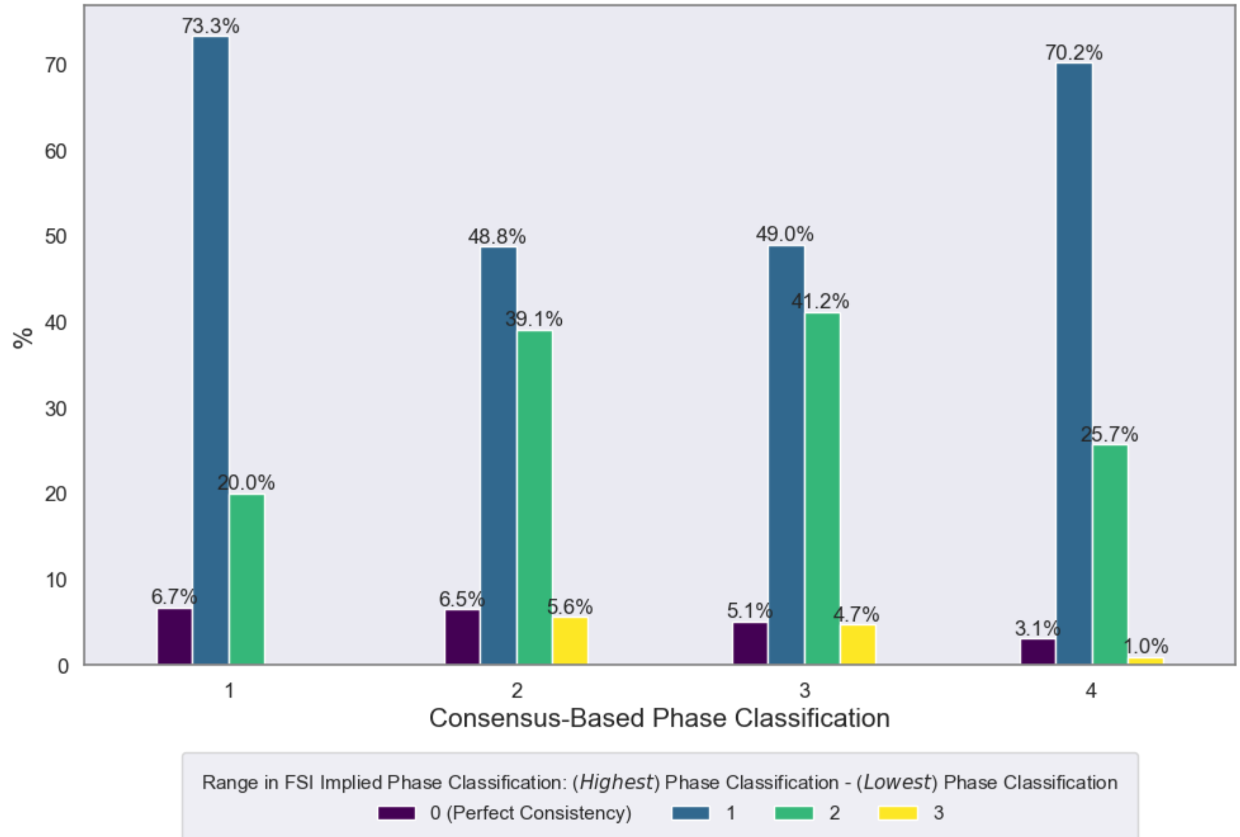
Table SI.2. Share of consistent FSIs with consensus outcome and % of FSIs implying the majority class by consensus outcome

Consensus-Based Phase Classification	% of FSIs Consistent with Consensus-Based Phase Classification
1	27.78%
2	48.39%
3	52.03%
4	44.68%

Caption: In this table, the column labeled “% of FSIs consistent with consensus-based phase classification” presents the share of FSIs in agreement with the consensus-based phase classification (n = 1849).

In Figure SI.3, we illustrate the range of categories suggested by each FSI for each phase classification in our data, measured as the highest suggested phase minus the lowest suggested phase. Only a very small fraction of the classifications have all underlying FSI data supporting the same classification (shown as a range of ‘0’ in the figure). Most of our data have underlying FSIs that span a range of between two and three phase categories.

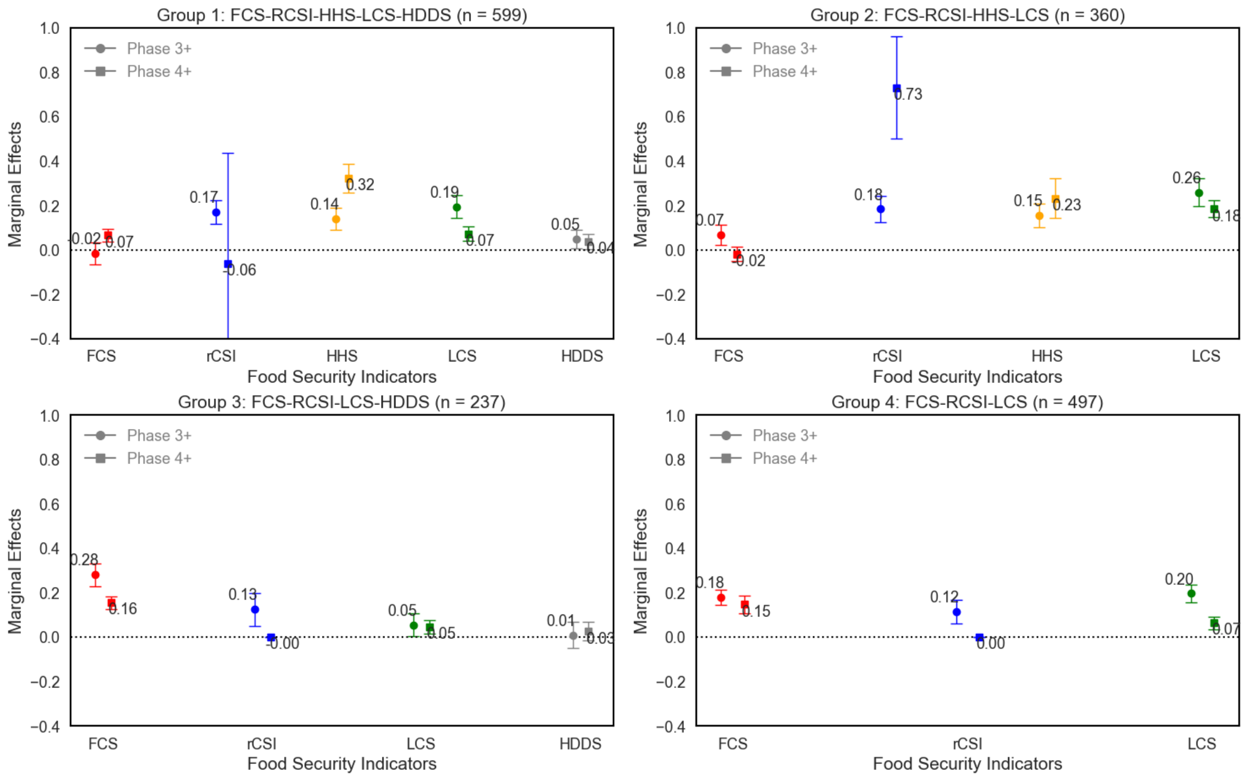
Figure SI.3: Distribution of range in FSI implied phase classification



Caption: The figure shows the range in FSI-based classifications for a given assessment area at a given point in time. The range is computed as the difference between the implied highest FSI classification minus the implied lowest FSI classification. The x-axis is the TWG consensus outcome. 0 indicates perfect consistency, meaning all food security indicators point to the same phase classification (n = 1849).

In SI Figure SI.4, we show Figure 6 overlayed with results from ordinary least squares (OLS) models regressing the consensus-based population in 4+ on the share of the population in 4+ according to each of the three to five FSI indicators, facilitating comparisons of the relative weights across severity class within FSI data availability groups.

Figure SI.4: OLS model coefficients (or weights) by data group for both 3+ and 4+ population shares

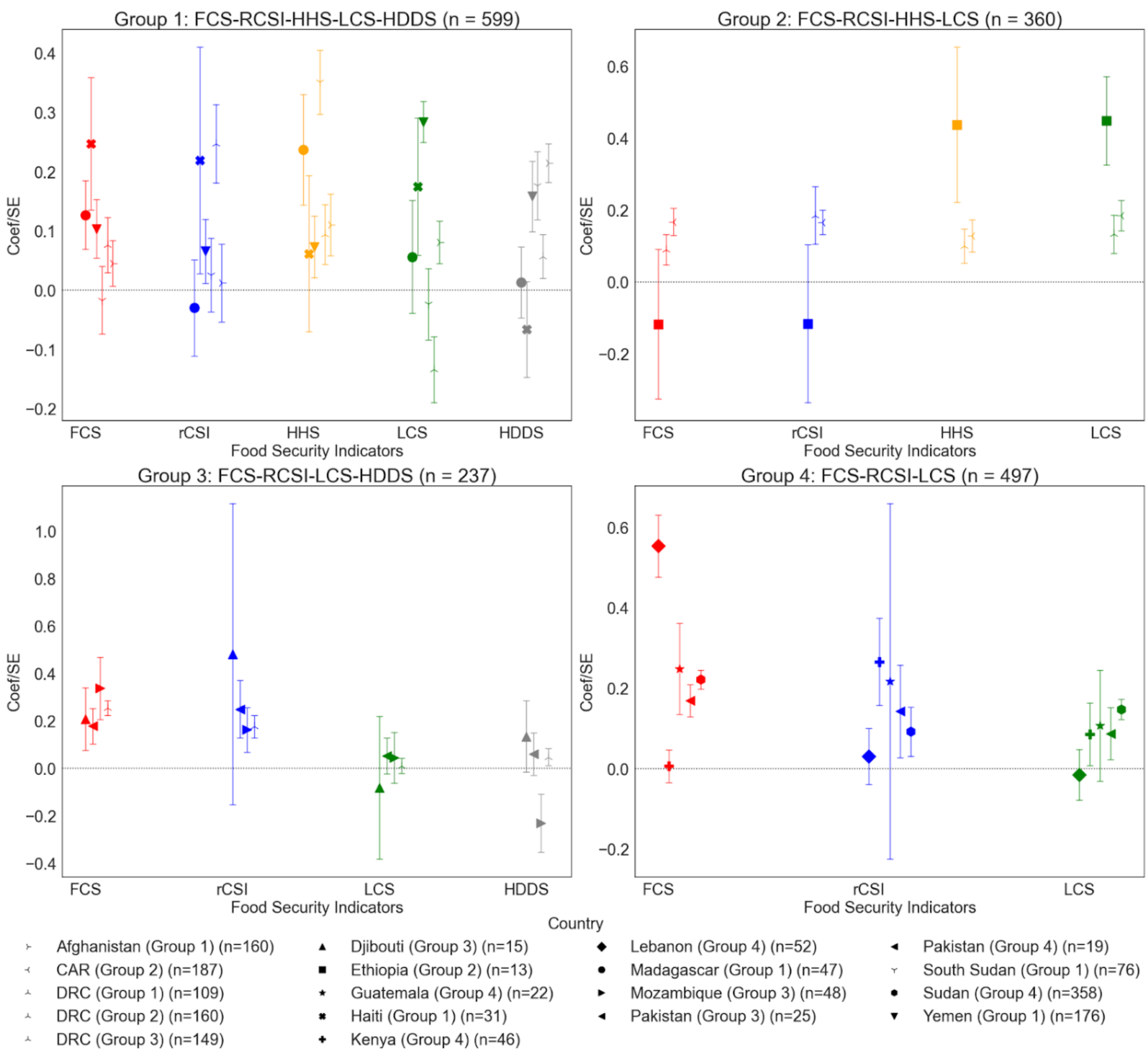


Caption: The figure presents the coefficients from four ordinary least squares (OLS) models analyzing the consensus-based phase 3+ and phase 4+ population shares (represented by circles and squares, respectively). The independent variables consist of a vector of population shares at phase 3+/4+ as implied by all FSIs available to the TWG. Each quadrant corresponds to a different data availability group, displaying the marginal effects and their 95% confidence intervals for each FSI. Note: 3+ and 4+ models are specified as consensus-based phase 3+ population share = f(FSI Implied 3+ population share) and phase 4+ population share = f(FSI Implied 4+ population share), respectively.

Figure SI.4 shows that implicit weights on the FSI indicators change when the severity changes. For 4+ population estimates, rCSI is not statistically significant in most cases, except in Group 2. In Group 2, TWGs appear to place considerably more weight on rCSI for 4+ than for 3+. This is inconsistent with the Reference Table (see SI A), which states that rCSI cannot be used to differentiate among phases 3–5. Therefore, we should not expect the importance of rCSI to increase as severity increases. In contrast, we might expect HHS, which the Manual indicates can differentiate between phases 3 and 4, to be important for TWGs. Our estimates support this, as the importance of HHS increases as severity rises from 3+ to 4+ in Groups 1 and 2.

In Figure SI.5, we turn to estimates of implicit weights by country. Figure SI.5 show coefficients on FSI by data group and country for estimates of consensus based 3+ populations. We find that the weights on FSI are variable across countries within data groups. For example, Ethiopia in data Group 2, weighs HHS and LCS much higher than other group 2 countries of CAR and one year of DRC.

Figure SI.5: Coefficients (with standard error bars) on FSI by data group and country for estimates of consensus based 3+ populations

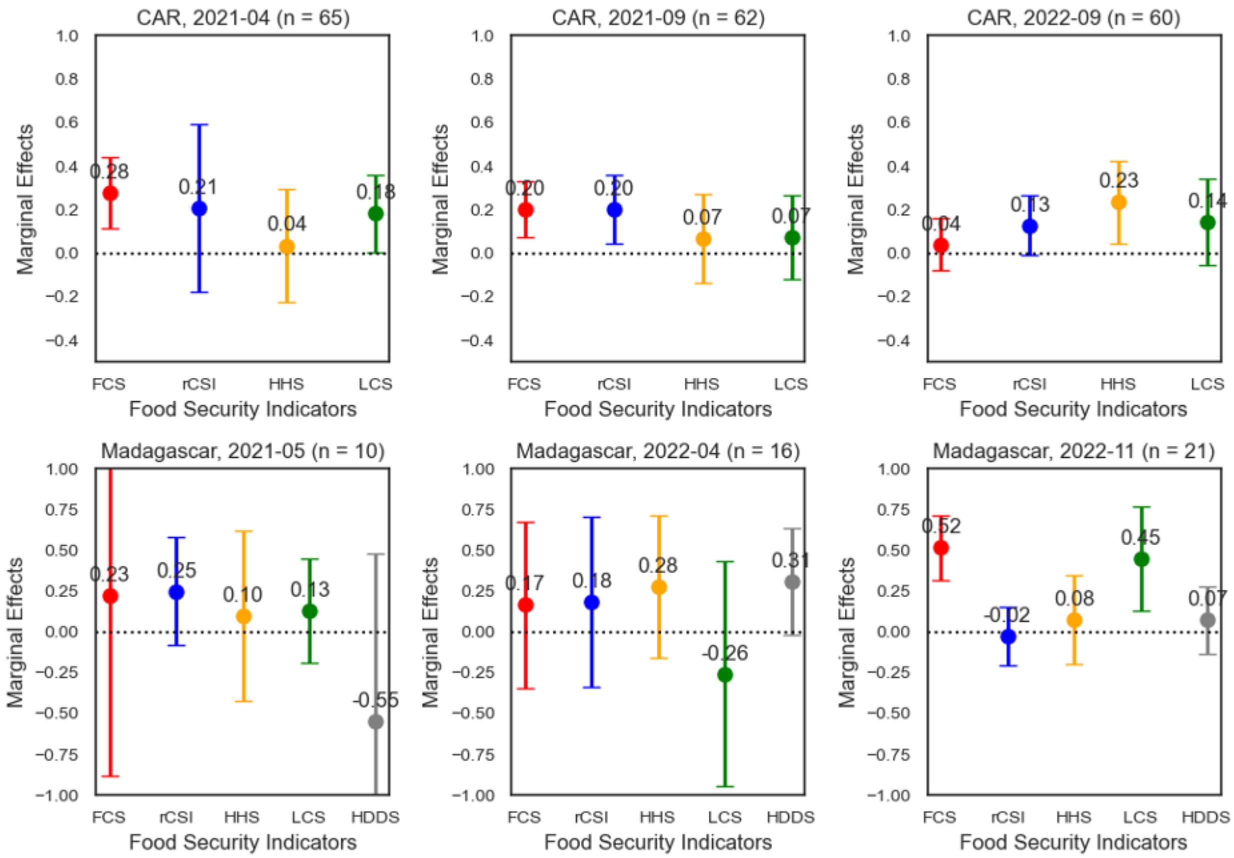


Caption: This figure presents the coefficients from four OLS models analyzing the consensus-based 3+ shares by country. The four quadrants depict the coefficient and 95% confidence intervals for each FSI for each data availability group. Each model is specified as *y = f(FSI 3+)* where *y* represents the consensus-based 3+ population shares and FSI 3+ is a vector of 3+ population shares implied by FSIs.

We estimate the population assessed as 3+ as a function of the percent of population suggested by the underlying FSIs for two cases with multiple rounds of TWG-level data: Central African Republic (CAR) and Madagascar (Figure SI.6) and Afghanistan (Figure SI.7). This allows us to examine whether weights are most constant within TWGs. For example, if certain indicators are not relevant in a particular country, we might expect that TWGs within that country would consistently downplay them across all analyses.

Figures SI.6-SI.7 present findings that the weights on FSI population shares vary over time for CAR, Afghanistan, and Madagascar respectively. It does not appear that TWGs within countries have consistently ‘favored’ (or ‘disfavored’) FSI metrics. Instead, we find that TWGs within CAR place different (implicit) weights on the same indicators across analyses. For example, if LCS is a poor measure for a given locale, we might expect that LCS has a consistently lower weight across TWGs meetings for that country. We do not see that degree of consistency of weights across TWG within a country in the limited cases available. This variation could be a result of changing TWG membership across analyses and therefore lack of consistent agreement on which FSI are most informative for the specific context.

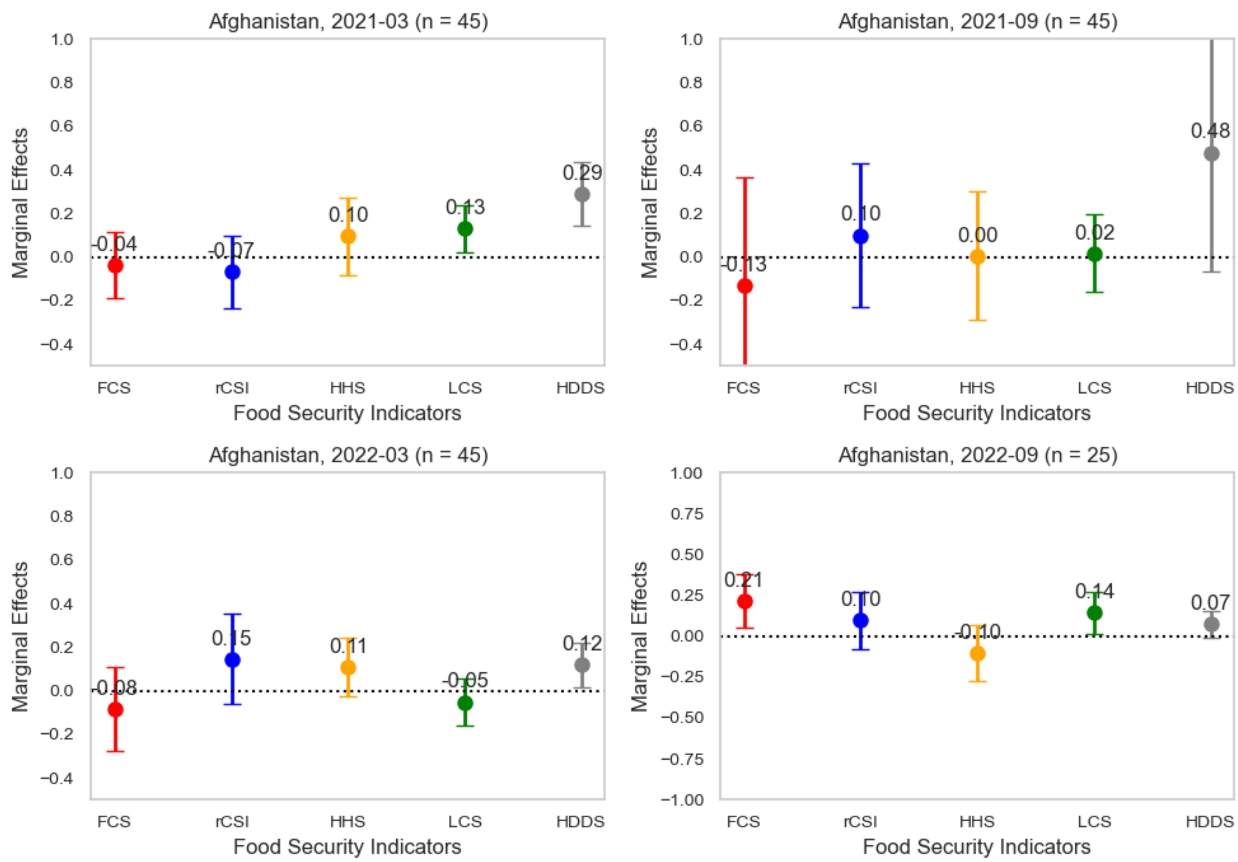
Figure SI.6: Coefficients on FSIs for consensus-based 3+ population by round: Central African Republic and Madagascar



Caption: This figure presents the coefficients from six ordinary least squares (OLS) models regressing the consensus-based population in 3+ on the share of the population in 3+ according to each of the five FSI indicators. Each regression is run for each round of analysis conducted in the Central African Republic (in April 2021, September 2021, and September 2022) (n = 191) and Madagascar (in May 2021, April 2022, and November 2022) (n=47). Lines represent the 95% confidence interval. Coefficients are compared to zero, across FSI indicators for a given analysis, and across analyses for a given FSI. Note: Each regression model is specified as y = f(FSI 3+ (%)) where y is consensus-based 3+ population (%).

Figure SI.7 presents findings that the weights on FSI population shares vary over time for Afghanistan. In both Madagascar and Afghanistan, consistent with findings for CAR, it does not appear that TWGs within countries have consistently ‘favored’ (or ‘disfavored’) FSI metrics. Instead, we find that TWGs within Afghanistan and Madagascar place different (implicit) weights on the same indicators across analyses.

Figure SI.7: Coefficients on FSIs for consensus-based 3+ population by round: *Afghanistan*



Caption: This figure presents the coefficients from four ordinary least squares (OLS) models regressing the consensus-based population in 3+ on the share of the population in 3+ according to each of the five FSI indicators. Each regression is run for each round of analysis conducted in Afghanistan (in March 2021, September 2021, March 2022, and September 2022) (n=180). Lines represent the 95% confidence interval. Coefficients can be compared to zero, across FSI indicators for a given analysis, and across analyses for a given FSI. Note: Each regression model is specified as y = f(FSI 3+ (%)) where y is consensus-based 3+ population (%).

In Table SI.3, we report the R-squared from three estimates across our four data groups. The R-squared measures how much variation is explained by the variables included in each model. R-squared results are reported for model 1 (column 2), which estimates population in 3+ as a function of the underlying FSI information only, model 2 (third column) with underlying FSI information and country-level fixed effects and model 3 (fourth column) for FSI and country fixed effects and their interaction. The final model's use of interaction terms captures country-specific weights on FSI.

Table SI.3: Explained variation (R-squared) in consensus-based population by data groups.

FSI Group	Model 1 FSI-Only R^2	Model 2 FSI and Country R^2	Model 3 FSI, Country and FSI x Country R^2
Group 1: FCS-RCSI-HHS-LCS-HDDS	0.371	0.589	0.649
Group 2: FCS-RCSI-HHS-LCS	0.589	0.636	0.641
Group 3: FCS-RCSI-LCS-HDDS	0.558	0.576	0.605
Group 4: FCS-RCSI-LCS	0.480	0.519	0.565

Note: Model 1 is specified as y = f(FSI 3+ (%)) where y is consensus-based 3+ population (%). Model 2 is specified as *y = f(FSI 3+(%) + country FE)*, where y is consensus-based 3+ population (%). Model 3 regression is specified as *y = f(FSI 3+(%) + country FE + (FSI 3+(%) * country FE))*, where y is consensus-based 3+ population (%).

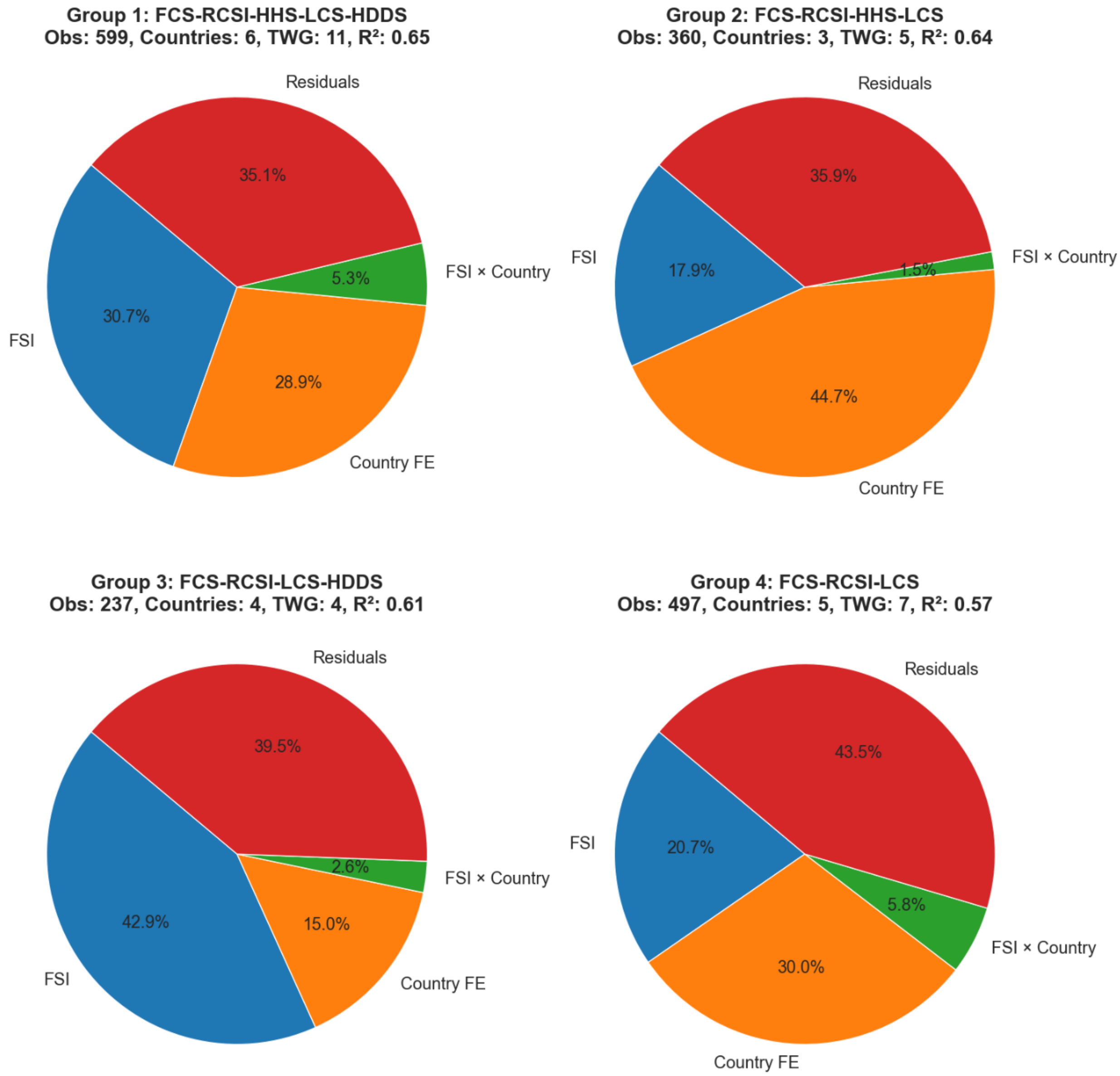
Model 1 shows that underlying FSI data explain between 37 and 59 percent of the variation in population assessments. This indicates that a sizable portion of the variation in TWG assessments (41 to 63 percent) is not based on observable FSI information. Given that FSI data are the primary data source for assessments, this may appear low, underscoring that contextual and contributing factors and consensus process may play an important role. While we expect that TWGs would use information beyond the FSIs to categorize the percent of population in urgent need, we might expect that the information would be similarly important across the groups of data availability. When we allow the different countries to weigh FSIs differently to account for country-specific interpretations of FSI information, we find it does explain much more variation than a country fixed effects model. Model 2 explains between 52 and 64 percent and Model 3 explains between 57 and 65 percent of the variation in the assessments is driven by the FSIs. This suggests that 35-43 percent is driven by changes in how a TWG assesses the same information from different locations.

An ANOVA decomposition of the variance for model specification 3 quantifies the variance explained by food security indicators (FSIs), country fixed effects, and their interactions. Including the interaction term allows us to control for country-wide context-specific interpretations of FSI data. Results reveal a significant amount of unexplained variation across the four data groups:

35–44% of the explained variance is attributed to unobservable factors. This suggests that, beyond FSI-implied population estimates and country-specific context, additional, unobservable factors outlined in the IPC analytical framework (see SI Figure SI.1) play a key role in shaping consensus outcomes.

In the SI Figure SI.8, we present an ANOVA decomposition of explained sources of variation for Model 3, which quantifies the variance explained by food security indicators (FSIs), country fixed effects, and their interactions. We also account for residual variance, highlighting unexplained variation. This analysis allows us to quantify the roles played by each observable aspect of the consensus process. Importantly, including the interaction term allows us to control for country-wide context-specific interpretations of FSI data.

Figure SI.8: ANOVA decomposition of explained source of variation by data group.



Caption: This figure presents the variance decomposition results from ordinary least squares (OLS) regression models estimating the share of the population classified as phase 3+, based on model specification 3 in table 4. Each pie chart represents a different data availability group, displaying the proportion of variance explained by food security indicators (FSI), country fixed effects, interaction effects between FSIs and country identifiers (FSI × Country), and residual variance.