Global Estimates Systematically Undercount Acute Hunger

Abstract

The Integrated Food Security Phase Classification (IPC) system is the global method for classifying food insecurity severity. As of 2023, international agencies and governments use IPC assessments of food crises to allocate more than six billion dollars of humanitarian food assistance annually. Despite concerns that IPC estimates overstate global food insecurity, our analysis of data from approximately one billion people, in 10,000 IPC assessments conducted between 2017-2023, demonstrates the opposite: IPC estimates are prone to understate the extent and severity of crises. Our primary estimates suggest that IPC assessments substantially underestimate the number of acutely hungry people in the world, missing approximately one in five. We find evidence of under-classification around the IPC threshold that determines whether an assessed area is to be classified as 'stressed' as opposed to 'in crisis' - a threshold meant to trigger deployment of humanitarian resources. Contrary to widely held assumptions, our findings suggest that the prevalence and severity of acute hunger is likely significantly higher than current global estimates would indicate.

Introduction

Humanitarian information systems are sets of processes and tools used to identify and anticipate the locations and magnitude of crises around the world. A prominent example, the Integrated Food Security Phase Classification (IPC), was established in 2004 to provide information about current and emerging food insecurity crises: including severity, scope, and size of affected population (1, 2). Acute hunger affects hundreds of millions of people worldwide with long-term consequences for health, development, and security (3,4,5). IPC assessments guide the timely allocation of humanitarian resources to the locations that need them most.

A core IPC objective is to identify emerging crises in time to guide response that prevents their escalation. Between 2017 and 2023, the IPC conducted more than 10,000 sub-national food security assessments across more than 30 countries. Famines are rare findings in those assessments, occurring only twice between 2017 and 2023; near-famine conditions have occurred a handful of times in that same period (6; see also 7). Instead, most of the IPC's work has focused on identifying where food security is deteriorating, to facilitate intervention before conditions degrade to the severity of a famine.

In 2023, an estimated 765 million individuals around the world lacked sufficient food to meet their basic daily caloric needs. Among them, 282 million were classified as experiencing acute food insecurity, circumstances in which "a person's inability to consume adequate food puts their lives or livelihoods in immediate danger" (8, 9,10). Accurate measurement of the magnitude and the location of acute hunger is essential to support effective response. Because acute caloric and nutritional deprivation can have both short and long-term impacts on individual health and wellbeing, the value of timely and reliable assessments of food insecurity is high.

Beyond crisis response, IPC estimates serve as a foundation for major annual United Nations reports on hunger (e.g., 10, 11). They are also cited in funding appeals and response plans (e.g., 12, 13, 14) and are submitted as evidence to the UN Human Rights Council (15) and to the International Court of Justice (16).

A crucial question, therefore, is whether current global assessments accurately measure the populations experiencing food insecurity. This paper provides a quantitative evaluation of these global IPC assessments. Analyzing food security assessments covering 917 million people, (with repeated observations across multiple rounds totaling 2.8 billion people) from the majority of countries characterized by frequent food crises between 2017-2023, we use three related methods to evaluate current estimates: we test for systematic bias around the threshold indicating urgent need for assistance, and we compare official assessments to two counterfactual estimates derived from underlying household survey data.

The results of our analyses indicate that the number of acutely food insecure people estimated by the IPC is likely too low. Our primary analysis of nearly 10,000 area assessments across 33 countries finds that, on average, current assessments miss approximately one in five acutely hungry people. Because true food insecurity status cannot be directly observed, we cannot measure rates of misreporting with certainty. However, our findings show a systematic pattern of conservative (i.e., lower) assessments relative to the underlying data.

We also show that concerns from the international community of donors and policymakers about potential overestimation by the IPC process do not align with our empirical findings. International response already falls short of meeting global need, even based on IPC's conservative estimates: in 2023, the estimated gap between global hunger-related need and available funding ranged from 33% to 53% (11,17). This means that not only are millions of acutely hungry people excluded from IPC estimates, but even those who are counted lack sufficient support.

Our findings underscore the urgency of additional financial and programmatic support to address and alleviate acute hunger worldwide.

The IPC

The IPC is the recognized global standard tool for food security and nutrition assessment at national and sub-national levels (18, 19, 20, 21). The IPC operates as a consortium of 19 partner organizations in more than 30 countries, with plans to expand their assessments by ten additional countries by 2026. Governments and non-governmental organizations (NGOs) worldwide rely on the IPC to monitor and identify acute food insecurity and malnutrition, and to classify famine (1, 2). As of 2023, IPC estimates are widely used to guide the allocation and targeting of approximately six billion dollars' worth of humanitarian assistance per year —by international organizations, donor governments and national NGOs (21). The IPC uses multiple indicators, and a convergence-of-evidence approach to assess the severity and geographic distribution of acute food insecurity sub-nationally in many cases multiple times each year. In this way, the IPC differs from measures like the Prevalence of Undernourishment (PoU), calculated once each year by the FAO as a model-based national-level estimate of chronic undernourishment, primarily used for global monitoring and long-term trend analysis.

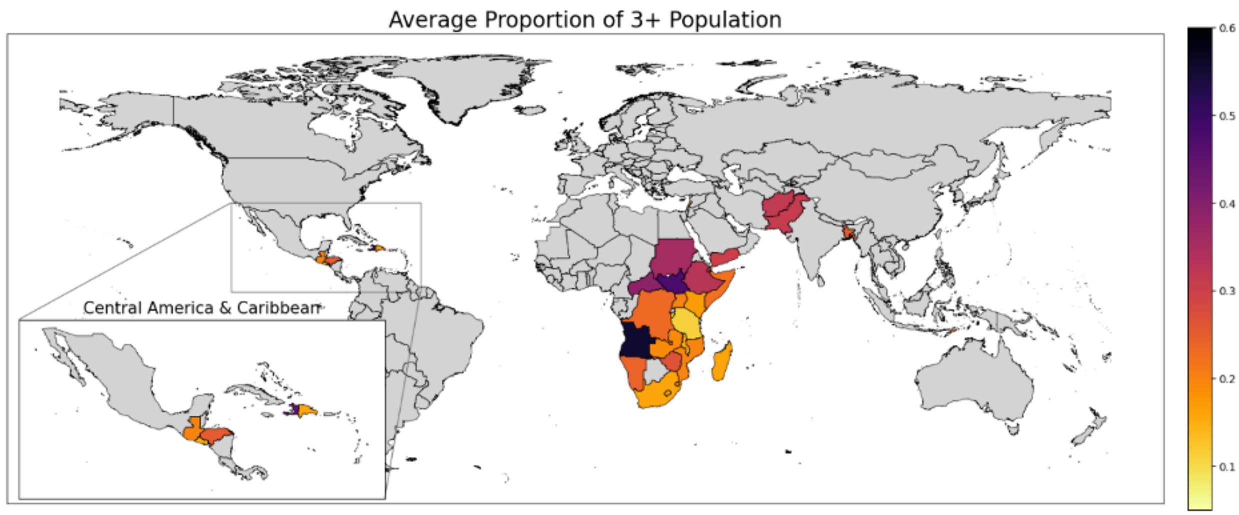
The IPC process assesses a diverse set of information about food insecurity to develop a single severity classification ("phase"), and population estimates in each phase for each sub-national assessment area. These classifications are intended to guide interventions. The assessments are based on IPC partners' survey data – a range of food security measures collected from each sub-national assessment area – as well as other relevant contextual information.

IPC technical working groups (TWGs) have access to different combinations of food security indicators (FSIs), depending on data collected by their partners. Common indicators used by TWGs include measures of dietary quality such as the Food Consumption Score (FCS) and Household Dietary Diversity Score (HDDS) and measures reflecting food insecurity experience and coping, including the Household Hunger Scale (HHS), Livelihood Coping Strategies (LCS) and reduced Coping Strategies Index (rCSI). Correlations vary by FSI pair (22, 23). FSIs capture different aspects of food insecurity, underscoring both the multidimensionality of food insecurity and the challenges faced by analysts facing conflicting data (24, 25, 26). For example, while each FSI has an associated phase classification, the FSIs often point to different rates and severities of food insecurity, thus suggesting a wide range of phases and population estimates for the same location at the same time (26; See SI Table A4 and Figure A1).

To harmonize these discordant indicators, IPC food insecurity assessments rely on a multi-step consensus-based process. Each analyzed country has a TWG of trained food security analysts drawn from United Nations organizations, NGOs, and national and local institutions. TWGs meet between two and four times per year to assess and interpret recent FSI data collected by partners and to draw on other information such as food prices, conflict location and severity, market functioning, and natural disasters and hazards that can contribute to food insecurity. Each TWG uses a ‘convergence of evidence’ process, which consolidates available evidence from a range of sources using a set of predefined protocols (19). TWG members evaluate and debate the available evidence, building a consensus on subnational classifications of (1) severity and (2) populations in each severity class. The consensus process occurs behind closed doors.

For each sub-national assessment area, working group members reach consensus both on the number of people who are food insecure and on the severity of their circumstances. Classifications range from phase 1 (none/minimal), 2 (stressed), 3 (crisis), 4 (emergency), to 5 (catastrophe/famine). A classification of phase 3 or higher (phase 3+) means that households are “in crisis” and "marginally able to meet minimum food needs but only by depleting essential livelihood assets or through crisis coping strategies" or worse (19, p.37), placing them at risk of future deprivation. Any classification at or above 3 implies that “urgent action is required” (19, p.37). IPC classifications are intended to be comparable across locations and over time. See Figure 1 for details about scope and assessment of the current IPC.

Figure1. IPC coverage used in our analysis and geographic distribution of the proportion of population assessed in phase 3+



Caption: The figure presents the average proportion of population in phase 3+ for our sample of IPC assessments in 32 countries, ranging from yellow (low) to black (high). The averages are based on the most current area assessments within each country that were available (2020 or later). The number of areas classified by IPC ranges from 8 to 348 per country. IPC analyses report the 'Tri-National Border Regions' (El Salvador, Guatemala, and Honduras) separately, which we treat as a 33rd "country." Countries not covered by IPC are colored gray. In West Africa, food insecurity is classified using the Cadre Harmonisé, an analytical approach similar to – but distinct from – the IPC process.

Food security assessment is difficult

The IPC's consensus-based methods are in response to a major challenge: food insecurity assessment is inherently difficult, especially in data-scarce environments. One key issue in food insecurity assessment is that, like poverty, food insecurity is multidimensional and as a latent variable is not directly measurable (24, 25). The measurement of food insecurity relies on proxy indicators such as household dietary quantity, dietary quality, and or coping strategies (e.g., skipping meals) (22, 25, 27). Moreover, some proxy indicators require contextually specific adjustments (22, 23, 28). For example, in some contexts the consumption of wild foods is a common strategy among food *secure* households while in others turning to wild foods is a sign of acute distress. Thus, collecting and interpreting food security information, such as on the consumption of wild foods, requires nuance (29).

A second key issue related to food insecurity measurement is that the literature has yet to identify any single indicator or combination of indicators to serve as a reliable proxy for food insecurity. Instead, research suggests that the current suite of food insecurity measures capture different aspects of food insecurity and current measures are often only weakly correlated (22, 23, 28, 30). Thus, multiple FSIs can imply different conclusions about the food insecurity status of the same household at a single point in time.

Further, IPC assessments are often conducted in data-scarce environments including remote areas, conflict zones or highly dynamic humanitarian crises. As a result, timely, reliable information about food insecurity is generally limited for TWGs, who must rely on the technical protocols and the consensus process to interpret and contextualize the data they do have.

The IPC's consensus approach provides a means to address important limitations and constraints related to food insecurity measurement in general but also particular challenges that characterize the contexts in which IPC operates: discordant food insecurity indicators that capture different aspects of food insecurity and that indicate different needs, missing or poor-quality data, and rapidly changing situations on the ground. Even so, the closed-door nature of the consensus approach creates another challenge: the perception of bias.

IPC working group members and users are concerned that the consensus process is biased

Governments and international agencies rely on IPC assessments to allocate humanitarian funds and to prioritize particular regions over others. Yet, donors, the media, and even IPC TWG members themselves have expressed concerns about the IPC assessment process (1, 2, 7, 31, 32, 33). Key informant interviews with over 20 humanitarian decision-makers and analysts conducted by the authors identified a common concern: that IPC population estimates are regularly and substantially inflated, over-stating the magnitude of crises (26).

Key informants raised two primary concerns contributing to the perception that IPC estimates of the number of individuals in crisis are inflated by the TWGs. First, while they recognize the significant data challenges associated with measuring food security, they express concern that such data challenges mean that resulting IPC assessments are not well-grounded. Second, they expressed concern about the IPC consensus process, describing it as “closed-door” and a “black box” (see also 7). One donor commented that working group members have the incentive to inflate the number of people “in crisis”, a technical term in the IPC assessment referring to classifications at IPC phase 3 or higher, and therefore in urgent need of assistance, to crowd-in funding for a country. A former IPC TWG member explained that while there was “general agreement on the classification … Probably [the TWG process] overestimates the population.” A related concern was that vested interests could push consensus toward results that would benefit certain sub-national assessment areas as a form of political patronage, or in order to benefit aid organizations and other actors operating in those locations. Findings from our interviews are consistent with humanitarian financing experts who write, “We encountered deep suspicion among donors that humanitarian organizations have for years been inflating figures in a bid to compensate for funding gaps” (34, p. 1).

Results

We analyze nearly 10,000 sub-national area assessments classifying the food insecurity status of 917 million unique individuals in 33 countries between 2017-2023. The sub area assessments over time include 172 country-by-round assessments, and a total of 2.8 billion people, including populations whose status is assessed in multiple rounds. Each assessment is for an IPC assessment area (a geographic unit often similar to Admin 3 level) at a given time. The assessments are based on underlying FSI data collected by household survey; for a subset of our observations, we observe the FSI data. This sample includes 27 TWG meetings during 2020-2022 for 15 countries, covering 743 million people, including populations assessed in multiple rounds (details of our sample are in methods and SI Tables A1-A4).

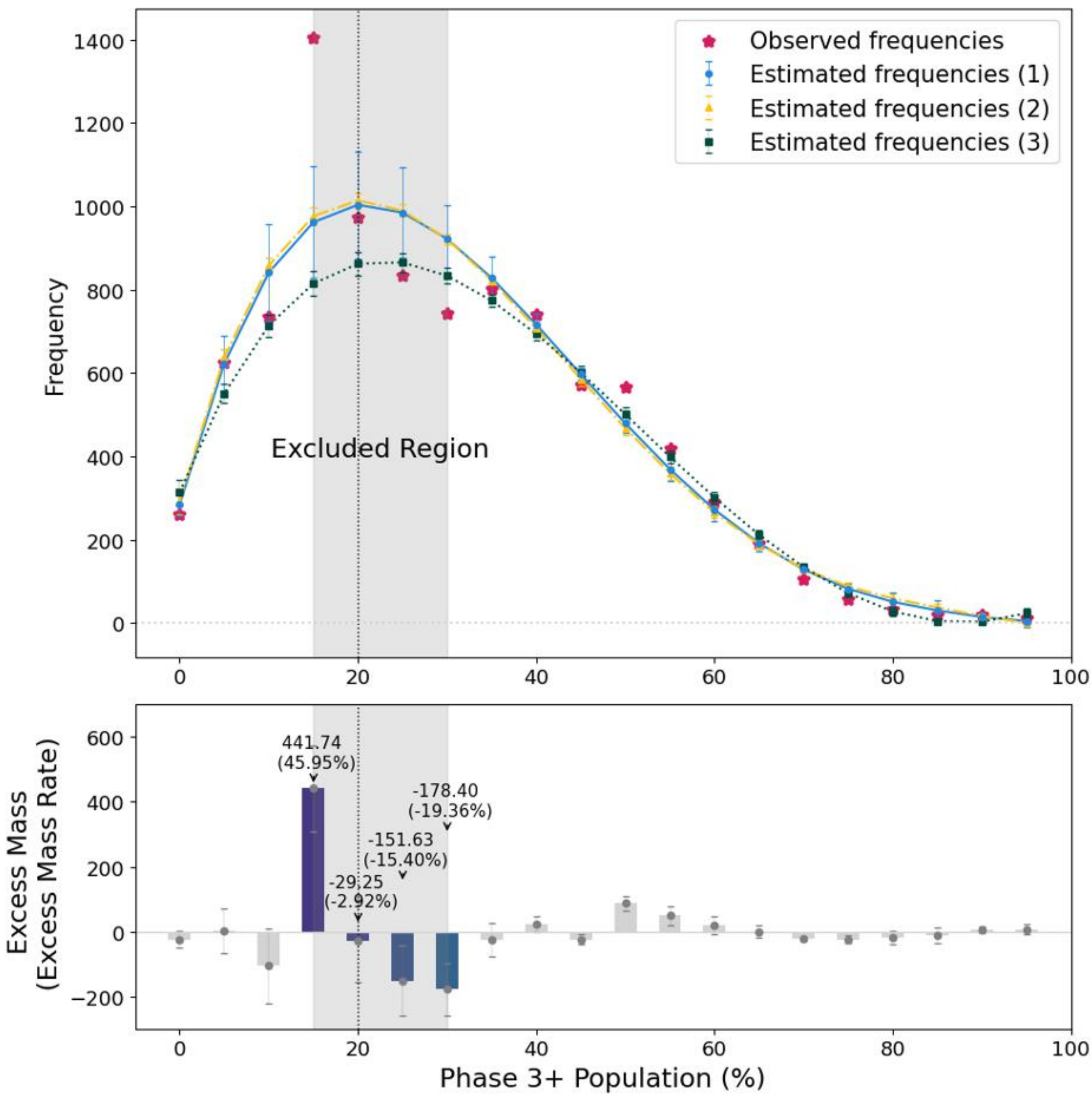
We use three related approaches to evaluate potential bias in the IPC TWG assessments. We first evaluate the distribution of the population percentage in phase 3+ around the 20% threshold. This threshold is significant for study because, according to IPC technical guidance (19), when at least 20% of the population moves from phase 2 to phase 3 or above, this transition signals the urgent need for assistance to donors, governments, and response organizations. Thus, the 20% population threshold is an administratively imposed cutoff that determines whether a location is in phase 2 (stressed) or phase 3 (crisis). We hypothesize that when TWGs are uncertain about the underlying conditions, they may tend to assess an area to be just under this threshold for assistance, resulting in a ‘bunching’ of the population in the distribution right below the threshold. Alternatively, if the TWG wanted to use estimates to increase funding, they would tend to assess an area to be at or just above this threshold, again resulting in bunching. Accordingly, this point in the distribution is where under- or overcounting is most likely to be identified.

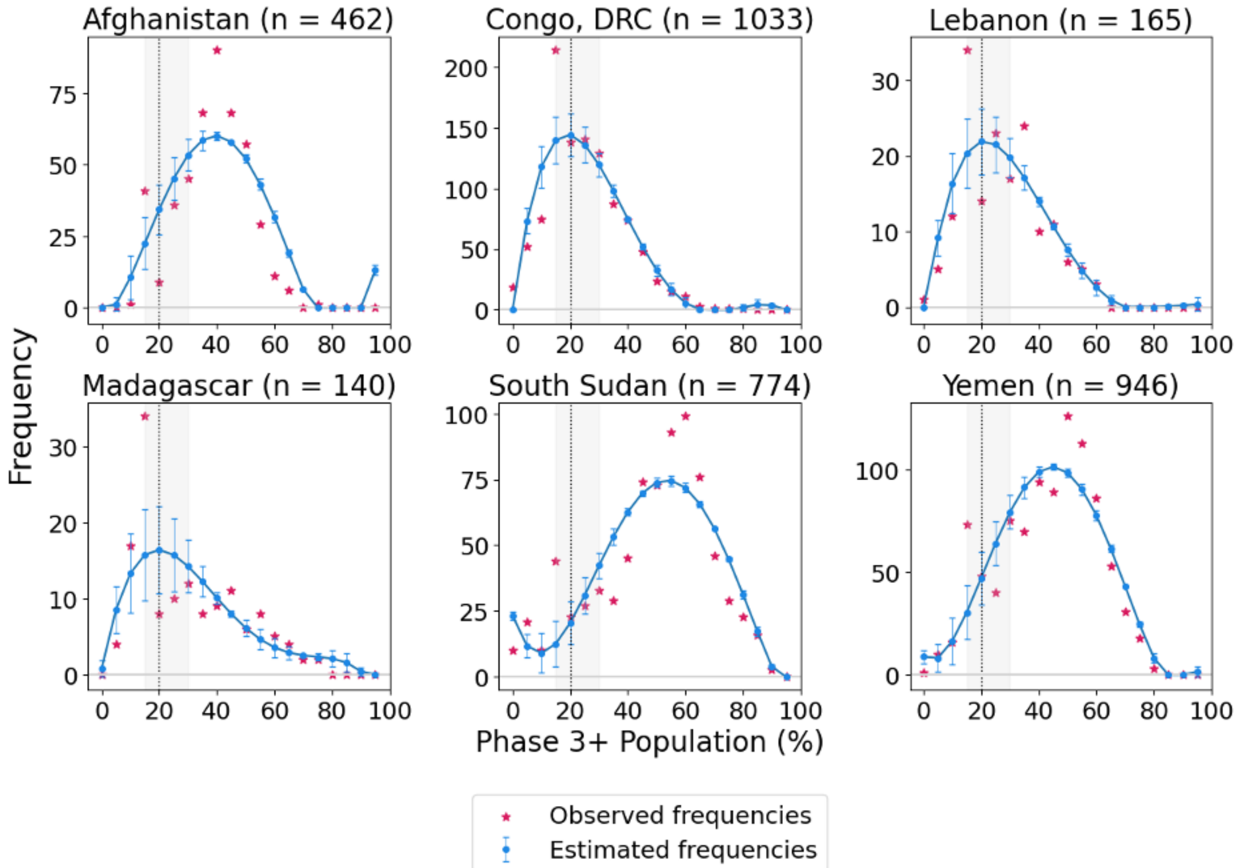
We then compare the assessed population from the IPC convergence process, which we refer to as the “observed” population, to two different counterfactual distributions. These distributions are generated from food security metrics used in the assessment process itself. We use the difference between the observed population and the counterfactual distribution to generate estimates of the population who are undercounted. One counterfactual distribution is generated by using the percent of population from the average of the FSIs, and the other is generated using the worst FSI. The latter approach is motivated by a ‘right to food’ approach – if a household is food insecurity in any dimension, they should be treated as food insecure. This analysis can be treated as an upper threshold on the number of hungry people. Our primary results are generated using the more conservative of these two approaches – the percent of population based on the average of the FSIs. For detail on each of these approaches, see the methods section.

Evidence of undercounting

The results from all three approaches indicate that IPC assessments undercount the number of hungry people relative to the comparators. We begin with the results of the bunching analysis. We evaluate whether the distribution of the IPC assessments of the share of the population in 3+ exhibit bunching by evaluating whether there is an anomalous mass of assessments just above or below the threshold distinguishing a phase 2 classification from a phase 3 classification (i.e., at the 20% threshold) (following 35-37). To make this evaluation, we construct a series of counterfactual distributions that simulate the distribution in the absence of the 20% threshold. We find evidence of bunching just below the threshold compared to an expected, smooth distribution. In other words, we find the TWGs assess a large number of areas to have populations just below the threshold for phase 3+ (upper panel of Figure 2). We illustrate the effect for several countries with very different levels of overall food insecurity (lower panel of Figure 2). We frequently observe an excess mass of assessments just below the phase 3 threshold, and a lower-than-expected number of assessments at or just above the threshold, such as in Lebanon and Madagascar. The results exhibit heterogeneity over countries and time, with evidence of undercounting at the threshold for a range of countries and time periods. Dynamic analyses for one of the countries over time – Afghanistan – are presented in the Supplementary Information (SI Figure A2).

Figure 2. Comparison of observed and estimated distributions of IPC classifications, based on population shares in Phase 3+ (%)





Caption: The top panel generates a smooth counterfactual distribution using data for the full sample (all countries and time periods) and three different exclusion strategies around the threshold. We apply a 4^{th}-degree polynomial fit to data organized into 5% bins, constraining the estimates to be non-negative, while varying the exclusion of bins near the 20% threshold. The first set of estimated frequencies and their 95% confidence intervals, illustrated by the blue solid line with circle markers, is derived by excluding bins in the range [15%, 30%] sequentially, and aggregating coefficients across four different scenarios for each bin. The second set of estimated frequencies and 95% confidence intervals, illustrated by the yellow dot-dashed line and triangle markers, is generated without excluding any bins. The third set of estimated frequencies and 95% confidence intervals, illustrated by the green dotted line and square markers, excludes all bins in the range [15%, 30%] simultaneously and extrapolates across the entire distribution. Red stars illustrate the observed frequencies of IPC assessments, and a vertical black dotted line indicates 20% population threshold, where a sub-national zone being assessed moves from being classified as phase 2 (“stressed”) to being classified as phase 3 (“crisis”) (n =9394).

The middle panel presents the difference between the observed and estimated frequencies at different phase 3+ population bins [15%, 30%] (following 35, 37). These differences are generated from estimated frequencies from the upper panel of Figure 2 (i.e., the blue solid line). Values on each bar are the excess mass, defined as the difference between observed and estimated frequencies. We use the 95% confidence intervals of the estimated frequencies (1) to construct the lower and upper bounds of the excess mass at each bin. Values in parentheses represent the excess mass as a percent of the estimated frequency (i.e., (observed – estimated frequency) / estimated frequency). See SI for excess mass calculations across the entire 3+ population bin. Note we also observe undercounting at the 50% bin, possibly suggesting a bias from TWGs rounding the population estimates to half the population).

The bottom panel shows country specific comparisons of observed and estimated distributions of IPC classifications, based on population shares in Phase 3+ (%) across a selection of six countries. The counterfactual distributions are generated using a 4th degree polynomial following the same approach as the top panel for the first bin-exclusion and aggregation strategy demonstrated in the full sample analysis.

The middle panel of Figure 2 presents the difference in the number of observed versus estimated counterfactual assessments of the percent of population in phase 3+, immediately above and below the 20% threshold, using the pooled sample. It shows that the sum of the missing assessments in three of the distributional bins just across the critical threshold (20%, 25%, 30%) approximately equals the excess number of assessments in the 15% bin. We find evidence that there are 46% more classifications just under the threshold (at 15%) than predicted by the counterfactual: 3% fewer at the 20% threshold, and 15% and 19% less at the two bins just above the threshold (25% and 30% bins respectively). When we statistically compare the number of classifications at each bin (following 35, 37), we find that all of the differences for the bins around the 20% threshold are statistically significantly different from zero with the exception of the 20% bin. Thus, we find strong evidence of bunching just below the 20% threshold, highly suggestive of undercounting.

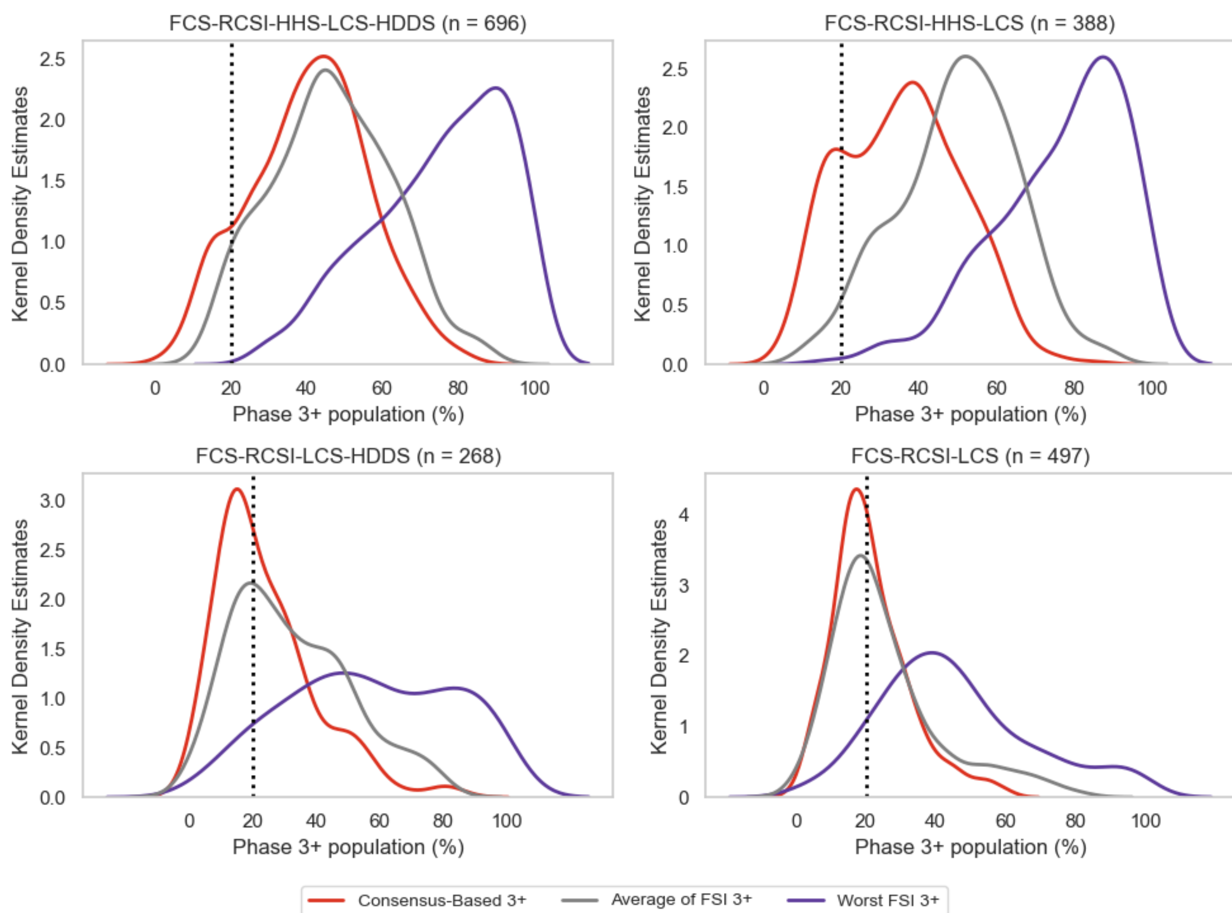
In Figure 2, in the top panel, we observe several peaked bins relative to our counterfactual distribution, including population concentrations at the 40% bin in Afghanistan and the 50% bin in South Sudan and Yemen. These peaks do not appear to result from a shifting of population from adjacent bins and therefore seem to be consistent with a smooth, unimodal distribution. This phenomenon changes the prevalence of population within the classification level but does not change the severity of the classification. In contrast, the pronounced mass point at the 15% bin and associated reduction of mass in the bins just above 20% suggests a non-smooth pattern, likely driven by the presence of the external threshold at 20%, below which populations are not classified as being in “urgent need.”

Second, we present results from our analysis of the full distribution of classifications. This is our preferred method for estimating the number of undercounted people, as it considers the full distribution, not merely assessments around the 20% population threshold. Undercounting may not just take place at the threshold; instead, it may be present in the evaluation of the data in front of the TWGs. We analyze the FSI that the TWGs used in the 1849 assessments where these data are available. Our results from this second analysis method indicate that the observed consensus-based classifications allocate a lower percent of the population to be in phase 3+ than we would predict from the underlying FSIs. TWGs do not have access to the same suite of FSIs. To keep our estimated counterfactual comparable across observations, we conduct separate analyses for each set of available FSIs.

In Figure 3, we present the observed distribution of the percent of population assessed to be in phase 3+ for the 1849 assessments with underlying FSIs. The solid red lines present the observed, consensus-based outcomes. The grey line in Figure 3 is our estimated counterfactual distribution of the population assessed to be in phase 3+ which is based on an average of the FSIs for the same locations. This counterfactual represents a scenario in which the TWG categorizes populations by calculating an unweighted average of the available FSIs. This scenario can therefore be thought of as a hypothetical situation in which IPC working group members take an average across the set of available FSI because they do not hold priors about which are most suitable. This counterfactual distribution is noticeably shifted to the right relative to the observed assessments, illustrating that the majority of the IPC assessments categorize a smaller percent of the population to be in phase 3+ than is suggested by a simple FSI average. Regardless of the set of FSI available, we see a mass of the distribution just to the left of the 20% classification threshold between phase 2 and phase 3 (depicted as the dashed black line in Figure 4), further evidence of the 'bunching' phenomenon in IPC assessments discussed above. An alternative estimate using regression-based weights on FSI derived from ordinary least squares regression in SI (Figure A3) shows findings for distributions derived from the estimated coefficients on FSI predicting IPC consensus populations. These findings do not predict the "bump" below 20% threshold, also indicating that the threshold is externally imposed by TWG members and not a function of the underlying FSI information.

Finally, we generate a counterfactual distribution with the highest percent of population assessed in phase 3 + based on the underlying food insecurity indicator that suggests the highest level of food insecurity (illustrated by the purple line in Figure 3). This estimate can be thought of as representing a Right to Food perspective that articulates that all people have the right to adequate, safe and nutritious food (38). Thus, under this perspective, any metric that shows that an area is in urgent need should be counted as hungry. Given that food insecurity is multidimensional (24, 25), we do not report the least severe metric, which would ignore other dimensions of hunger. As expected, the Right to Food analysis suggests that the assessments of the populations in phase 3+ are severely undercounted, with the estimated distributions far to the right of the observed distribution of assessments. This analysis cannot account for contextual interpretations of indicators and therefore may overstate the population in 3+.

Figure 3: Comparisons of distributions of IPC assessments defined by proportion of population in phase 3+ (Observed vs Counterfactual)

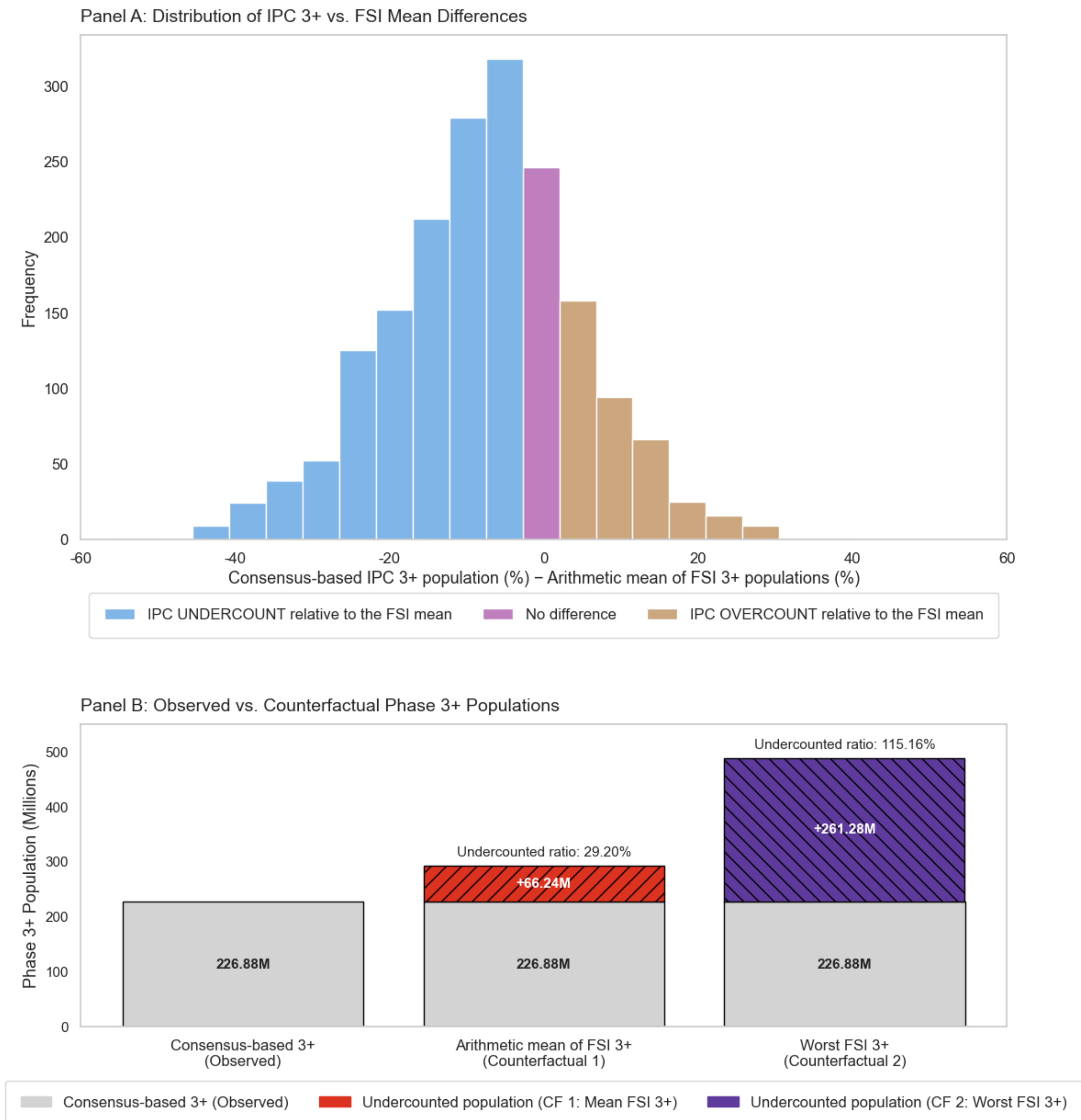


Caption: This figure presents gaussian kernel density estimates for three different distributions of population assessed in phase 3+: the percent of population assessed by the IPC TWGs to be in phase 3+ (red line), the counterfactual distribution of the estimated percent of population assessed to be in phase 3+ using the average of underlying FSIs (gray line), and the counterfactual distribution of the estimated percent of population assessed to be in phase 3+ using the worst underlying FSI 3+ (purple line). Each panel uses a sample based on different combinations of FSIs, (combinations of: food consumption scores (FCS); reduced coping strategies index (RCSI); household hunger scale (HHS); livelihoods coping strategy (LCS); and household dietary diversity score (HDDS)) reflecting the fact that available FSIs tend to vary across assessment areas. The 20% threshold between phase 2 and phase 3 is illustrated by a vertical black dashed line (n = 1849).

We statistically test whether our counterfactual distributions differ from the distribution of the population assessed to be in phase 3+ using a Barrett and Donald (BD) test (39) to ascertain whether one cumulative distribution stochastically dominates the other. We present the results in SI Figure A4. For all groups examined, the test statistics are consistently positive and statistically significant, indicating that the distributions from consensus exhibit lower levels of population in phase 3+, compared to estimates derived from counterfactual scenarios. Consequently, the counterfactual distributions first order stochastically dominate the observed assessed phase 3+ population. This implies that the situations depicted by the counterfactual scenarios are “worse” in terms of representing populations in urgent need, (i.e., in phase 3+).

To visualize the difference between the observed percent of population assessed in phase 3+ and our primary counterfactual – the percent of population assessed in 3+ using the average of the underlying FSIs – we plot a histogram of the difference in the frequency of assessments in Figure 4 (panel a). This difference can be large: in some locations, TWGs assess as many as 40 percentage points fewer people in phase 3+ than what is suggested by the average of the FSIs. Figure 4 also reveals a clear asymmetry: the number of assessments in which the IPC estimates a smaller phase 3+ population than the FSI-based average (blue) far exceeds those indicating possible overcounting (tan). In 14 out of 15 countries, (and 20 out of 22 country-by-round assessments), we find more undercounting than overcounting relative to the mean. We present the number of over and undercounted assessments by country and country-by-round in the SI Figure A5-A65.

Figure 4. Distribution of the difference between consensus-based 3+ population (%) and the arithmetic mean of FSI implied 3+ populations (%) for IPC assessments and number of undercounted populations in phase 3+ using counterfactual distributions estimates by FSI availability group.



Caption: Panel A presents the distribution of observed differences between the share of the population classified in IPC Phase 3+ by TWGs and the arithmetic mean of the FSI-implied Phase 3+ populations, with both rounded to the nearest five percent. Blue bars indicate instances where IPC TWGs classified a smaller share of the population as food insecure (IPC undercount) relative to the FSI mean, tan bars indicate IPC overcounts, and lavender bar at 0 represent no difference.

Panel B compares the observed IPC Phase 3+ population with two counterfactuals constructed from the FSIs: (i) the arithmetic mean of FSIs (Counterfactual 1) and (ii) the worst-case FSI (Counterfactual 2). Gray bars show the observed population, with additional red and purple segments indicating the undercounted populations relative to each counterfactual. Text inside the bars shows observed and undercounted population sizes (in millions), and labels above the bars indicate the undercounted ratio. The sample includes all classifications for which underlying FSIs are available (n = 1,849).

Percent of undercounted people

We use each of the three analyses above to generate estimates of the percent of population in phase 3+ that are undercounted in the IPC assessments.

We first calculate the percent of undercounted people in acute food insecurity crisis using our primary analysis: our comparison of the IPC assessments to counterfactuals derived from the average of the underlying food insecurity indicators. Because the combination of FSIs vary across TWGs, we estimate the percent of undercounting by groups of available FSI. The averaged FSI estimates indicate 293.1 million people in phase 3+, compared to the observed assessments of 226.9 million people. Taking the difference between the two estimates, we find that the true number could be 29.2% more than the observed value (66.2 million / 226.9 million). In other words, the IPC could be missing an average of one in five people who are in urgent need (66.2 million / 293.1 million). Our estimates based on method three – using a counterfactual of the highest percent of population assessed in phase 3 + based on the underlying food insecurity indicator that suggests the highest level of food insecurity – are the largest in magnitude at 115.2% (see Figure 4 panel b). In other words, more than one in five truly hungry people go uncounted in comparison to the averaged measure of food security, while more than one in two go uncounted compared to the worst food security measure.

The bunching analysis cannot provide an estimate of the total number of undercounted 3+ population because it analyzes only the area around the 20 percent threshold. It does, however, allow us to estimate the percentage of undercounting that is attributable to manipulation around the threshold. Estimates of bunching presented in Figure 2 suggest the presence of undercounting in the 20, 25, and 30% bins, where the numbers represent the percent of the population in the area assessed in phase 3+, underscoring the threshold effects of IPC classifications. Our back of the envelope calculations suggest that 16.3 million people are undercounted in the area of the 20% threshold, or about 5.88 percent of the 3+ population in the relevant region of the distribution, 15% to 30% (277.2 million people). Because the samples for the bunching analysis (n=9394) and the comparison to the average FSI (n=1849) differ, we cannot directly compare the two approaches' undercounting estimates. However, if one assumes that the shares of undercounting in each sample are representative of the broader population, then applying 5.88 percent to the smaller sample (n=1849) indicates that approximately 4.3 million people are undercounted out of the 73.7 million people assessed to be between the 15% to 30% bins. These 4.3 million people are 6.5% of the total undercounted population of 66.2 million. Thus, while the threshold effects are important, the majority of the undercounting is happening away from the threshold. We provide details on this calculation in the methods section.

Examining possible mechanisms

The evidence for undercounting is widespread: we found evidence of more undercounting than overcounting in 25 out of 27 country-by-round assessments. The tendency to undercount may arise through several pathways. We use this section to explore four possible mechanisms related to expected humanitarian aid, FSI data quality and inconsistency, unobserved information held by TWGs, and process-based pressures.

First, our estimates do not include humanitarian food assistance (HFA) commitments which TWGs may account for in their assessments. Using the case of Afghanistan, we examine whether our results on undercounting are driven by TWGs accounting for expected humanitarian food assistance. HFA could contribute to downward estimates if a TWG accounts for incoming assistance and adjust the population in 3+ downward in anticipation. We find that the presence of HFA is correlated with a slight decrease in the difference between the consensus-based estimates and the arithmetic mean of the FSI-derived counterfactual estimates, but this effect is statistically significant only at the ten percent level (See SI Figure A7). Thus, for this case, while we cannot make a strong claim that HFA is driving the lower TWG assessments, the possibility cannot be ruled out. However, this finding is limited to one case with sufficient data to permit analysis. As HFA data become more available, it would be worthwhile to expand this investigation to other countries.

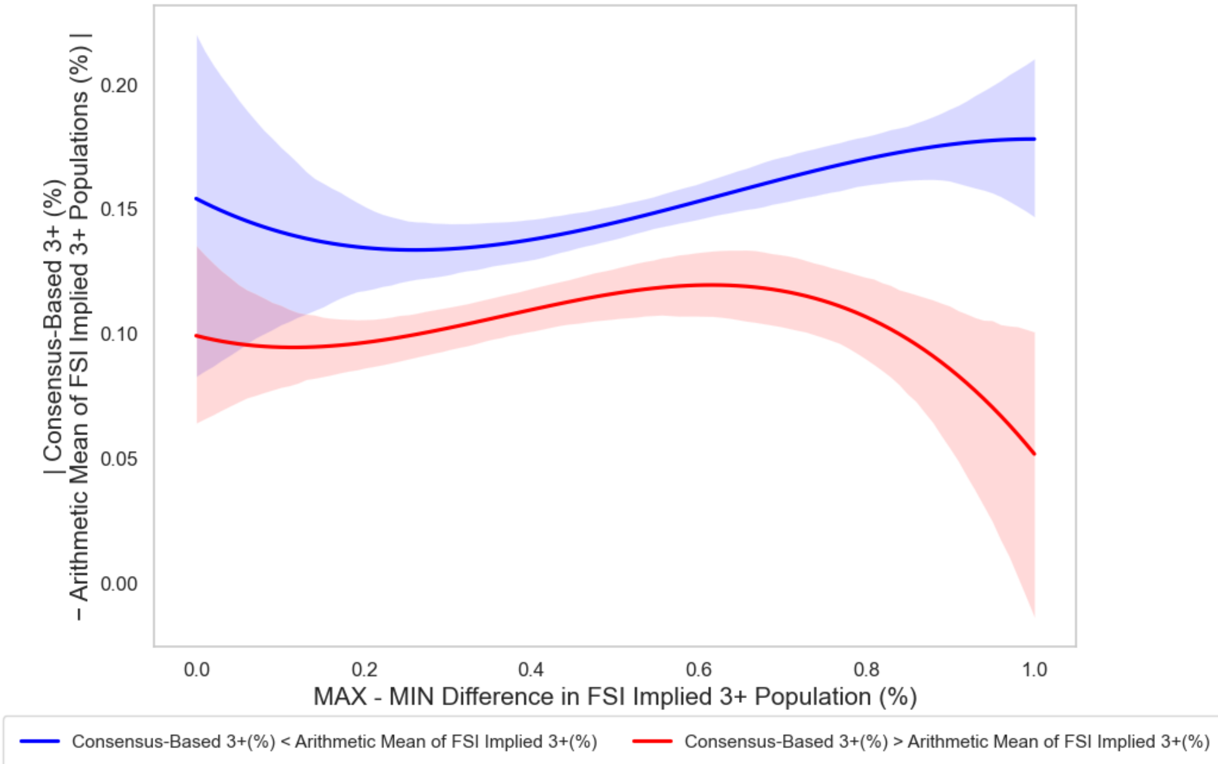
Second, inconsistencies or limitations in the quality of the FSI data available to the TWG may introduce uncertainty, contributing to more conservative classifications. TWG concerns about data quality could contribute to undercounting, with TWGs putting less weight on data with lower reliability scores – potentially discounting or down-weighting data that reflects more severe conditions for example. We do not find evidence that differences in reliability of FSI information influence undercounting, primarily because reliability scores show limited variation within data available to each TWG. While TWGs receive information on the reliability of the surveys used to collect FSI information, specifically on the representativeness and timeliness of the sample, a single survey usually collects all available FSI for each analysis area. As a result, indicators from that survey receive the same reliability score. Therefore, by construction, reliability information cannot help TWGs to determine the quality of underlying FSI.

However, we do observe that within-area inconsistencies among FSI data is associated with conservatism by the TWG. We observe more undercounting when the underlying food insecurity data available to the TWG are noisier, that is, when the indicators exhibit greater range in their suggested phase classification for a given place and time. In Figure 5, we document the presence of more undercounting when a higher fraction of indicators suggests different IPC phases. We show the relative noisiness of the data on the horizontal axis using the difference between the maximum and minimum percent of population in phase 3+ suggested by the FSIs.

The blue line illustrates that underestimates grow larger under “noisy” conditions (i.e., with larger absolute differences), widening the gap between the consensus estimates and the mean. Thus, as the underlying data become more “noisy,” TWGs tend to underestimate hunger (relative to the FSI mean) by larger amounts. In contrast, the red line shows that with more "noise,” overestimates by the TWG decrease. In other words, as we move to the right on the horizontal axis, we see the TWG assessments that overestimate hunger move closer to the arithmetic mean. Thus, these two lines suggest that as FSI data becomes more divergent, TWG predictions become more cautious, assessing a smaller percent of population to be in phase 3+.

We do not find evidence of undercounting in the 4+ category where the classification guidance is more straightforward, given certain FSI do not have cut-offs associated with higher levels of severity (see SI Figure A8). Thus, our findings suggest that greater uncertainty in input data may result in more conservative assessments by the TWGs. Future work could investigate what contributes to greater noise in FSIs in some analysis areas, which could make those areas harder to classify.

Figure 5. Relationship between the difference in Maximum (MAX) and Minimum (MIN) FSI implied 3+ population (%) and the absolute difference between the consensus-based 3+ population (%) and the arithmetic mean of FSI implied population (%).



Caption: This figure presents a third-degree polynomial fitted line with 95% confidence intervals of the observed range in the FSI implied 3+ populations for a given country-by-round level analysis (x-axis) and the absolute value of the difference between the (observed) IPC 3+ population analysis and the arithmetic mean of the FSI implied 3+ populations (y-axis). The data available to the TWG gets "noisier" (in terms of range) as one moves to the right on the x axis. The "residual" in the prediction by the TWG (relative to the arithmetic mean) gets larger as one moves away from the origin on the y axis (n =1849).

We explore a third possible mechanism driving conservativism by the TWGs: TWG members may have context-specific understanding regarding which FSI data are most valuable for their country. It could be the case, for example, that one FSI is wrong in (or inappropriate for) some settings, and therefore including it in the construction of the counterfactual may bias our results. To assess whether any single FSI has an outsize influence on the results, we ran Monte Carlo simulations for four separate groups of indicators using the available FSI dataset. We sequentially omitted one FSI from the average calculation of FSI-implied population estimates and determined whether doing so would significantly affect our findings. The results presented in the SI (Figure A9) suggest that removing any single FSI does not affect our conclusions, and that the distribution of population predicted to be in phase 3+ using the FSI data remains to the right of the assessment. That is, we do not find evidence that any one FSI is driving our results.

Fourth and finally, the convergence process itself could be prone to conservatism, regardless of the data. Key informants voiced concerns regarding both potential overestimation of population figures and the possibility of influential voices advocating for disproportionate assistance to specific areas. While this is possible for a given country, our results do not support these concerns on average. On the other hand, an a priori concern of inflated numbers, paired with limited resources for response, could depress numbers during the TWG process. For example, if TWG members believe that the IPC process tends toward overcounting or are worried that donors believe that there is overcounting, they may pre-emptively undercount. Or, they may err on the side of caution to avoid being accused of crying wolf, especially if circumstances are expected to worsen. Notably, no informant raised the concern that the IPC was too cautious in assessing populations experiencing phase 3 or 4 food insecurity (note: we have only a single phase 5 (famine) in our sample). While we cannot observe the working group deliberations, on average we do not find evidence of systematic overcounting. Our findings should mitigate concerns about TWGs “inflating” population numbers and motivate future research to characterize drivers of the tendency toward undercounting.

Discussion

The IPC is perhaps best known for its assessments of famine, which are rare but catastrophic occurrences attracting considerable media and policy attention. Yet the IPC plays a crucial role in identifying food insecurity more broadly and in marshalling and allocating humanitarian resources to areas in crises to prevent famine from occurring. In this paper, we analyze the IPC’s assessments between 2017 – 2023. We find evidence of a bias towards under-counting in the IPC assessments. Of the 27 country-by-round assessments, undercounting occurs in 25 (SI Figures A5-A6). In locations where we have access to the FSI data underlying the TWG IPC classifications, the TWGs assess 226.9 million people (30.5% of the assessed population) are in phase 3+. At phase 3, urgent action is required to protect lives and livelihoods before households fall into phase 4, where severe hunger puts individuals at risk of losing their livelihoods and lives. Our preferred method, which uses the averaged FSIs underlying the IPC assessments to create counterfactual assessments, suggests the number could be 293.1 million (i.e. 29.2% more people than the number assessed by the IPC to be in phase 3+). Thus, the IPC process may fail to identify more than 1 out of every 5 people in crisis.

Results from the counterfactual grounded in the rights-based approach suggests that the number of acutely food insecure people in the world is even larger at 488.2 million (or 115.2% of the population assessed by the IPC to be in phase 3+). However, given that FSIs are context-specific, using the worst indicator to generate counterfactuals is likely to overstate a crisis. Our bunching analysis also finds evidence of undercounting but is methodologically limited to consider only the population around the 20% threshold.

Our results based on the bunching analysis (Figure 2) also show that the geographic scope of acute hunger is larger than current estimates: 46% of locations currently classified as phase 2 should likely be classified as being in phase 3+, or in “urgent need”. Thus, our findings suggest that many more people in many more locations may be experiencing severe and acute hunger than suggested by the current assessments. Donors, governments, and humanitarian organizations use IPC classifications to guide funding (21). Under-classifying these households puts them at risk of the long-term consequences of hunger: compromised health, livelihoods collapse and decreased future incomes and household welfare (3, 4).

Our results should be viewed in the light of three important limitations. A first, critical constraint of our analysis is that we cannot observe the true experience and therefore the true incidence of food insecurity. We observe the IPC assessments and, for a subset of the IPC assessments in the sample, the underlying food insecurity indicator data. Our counterfactuals are therefore based on assumptions that the food insecurity indicator data and their levels in the reference table sufficiently capture the true incidence of food insecurity. On average, we find undercounting compared to our counterfactuals. In some countries, it is possible that the counterfactual of the averaged FSIs may overestimate the true percent of populations in 3+ because it does not consider important contextual information.

A second limitation is that our bunching analysis only considers locations at the threshold. With this analysis, we cannot identify undercounting that occurs far above the threshold. For example, we are unable to identify locations that may truly have 40% of the population in phase 3+ that are instead assessed to be at 15%. Locations outside of the range of the threshold very likely also have undercounting and the magnitude of that undercounting could be quite severe.

Third, our data only cover the IPC assessments available from 2017 to 2023 and some include only one to two years. While the data include a wide range of countries, our analyses are limited to IPC countries, preventing us from characterizing trends over time and excluding middle-income countries, like India, that may have pockets of severe food insecurity where IPC data are not available. Further, our second and third analyses based on counterfactuals constructed from the FSI data are constrained to the subsample of assessments with available FSI data (1849 out of 9394). Thus, when we calculate the implications for the total percent of undercounted acutely hungry suggested by these analyses, we assume that the percent undercounted in assessments with FSI data are similar to (i.e. a representative sample of) the other assessments where we did not have access to FSI data. We test this assumption (SI Table A5) and find that locations with and without FSI data are similar over a set of observable characteristics, including average phase classification; an exception is that locations with FSI data tend to have higher levels of conflict.

Our methods and results are diagnostic, not prescriptive. Our analysis shows that undercounting may occur frequently and could be significant in scale and scope. But unless we impose strong assumptions on the veracity and underlying true weights of the FSI data, we cannot identify exactly which populations are being undercounted. Further, our distributional analysis relies on the existence of an underlying smooth distribution and cannot identify the exact locations or severities of the populations that are undercounted. Thus, we cannot directly use our approach as an alternative to estimate true rates of food insecurity.

By no means do our results suggest that the IPC consensus process should be abandoned. Given the significant challenges associated with measuring food insecurity, the consensus process plays a valuable role in harmonizing across a wide and dynamically variable information set, while also building buy-in among stakeholders. Even so, better data and more transparency around the data and the process could improve internal and external confidence in the IPC consensus process. For example, the TWG could provide explanations in circumstances in which assessments diverge substantively from assessment levels suggested by underlying FSI data.

A deeper understanding of the relationship among household food security, asset portfolios, and a range of environmental and economic stressors could better inform assessment. Spatially and temporally granular data, including panel data on the food security and livelihood status of households in a range of contexts will improve our understanding of the mapping between FSIs and sub-national classifications used by the TWGs. This understanding could inform TWG deliberations and could help improve the transparency of the convergence process.

As innovative data collection and modeling techniques emerge and develop, so will new opportunities emerge for modeling to complement or even guide – rather than replace – the IPC consensus processes. Machine learning approaches have gained some traction in the food security space (20, 40-43) and these ML-based modelling outcomes could provide valuable support to TWG consensus processes. However, the food security classification and population identification problems may not (yet) be solved by developing and deploying algorithmic solutions. Humanitarian emergencies are often rapidly changing, multi-hazard, cascading crises that occur in data sparse environments, hampering model development and training (43-45).

At least for the near future, the international food insecurity assessment and response communities will continue to face limited and conflicting data. Given that food insecurity assessment will always involve some degree of uncertainty and contextual interpretation, collecting more data is unlikely to fully resolve the problem (46). Thus, the drive for a false level of precision may hinder the appropriate level humanitarian response.

A 2023 report, relying on current IPC population estimates, found a 53% gap between the assessed needs and allocated funding (17) and the scope of global humanitarian funding needs is expected to grow (8, 10). In as much as such funding shortfalls are influenced by a suspicion that current food security estimates are overblown, our evidence indicates that the suspicion may be unfounded.

The number of food insecure individuals worldwide is high and rising (8). Our findings indicate that these figures likely underestimate the actual global population of food insecure people, under-assessing both the scale and the scope of need. More funding could help prevent hunger-related deaths, suffering, and the long-run harm to human welfare. Funding shortfalls combined with prevailing assumptions that the numbers of people in crisis are overstated might contribute to TWG member conservatism and the resulting undercounting. Alleviating the severe funding shortfall may have the additional benefit of reducing undercounting. Reducing the perception of inflated numbers is an important first step.

Methods

1. Interviews

We conducted 21 semi-structured interviews with individuals knowledgeable about the IPC (26). The interview sample was drawn from global staff of the IPC global support unit (GSU): regional coordinators, famine review committee members, the TWG and technical advisory group, as well as other users in IPC partner and donor organizations. These interviews identified issues of concern, helped the study team develop a clearer concept of what can constitute accuracy for IPC outcomes, and also identified possible drivers of (in)accuracy. The Institutional Review Board at University of Texas at Austin reviewed this research.

2. Sample

We use two main samples. The first sample, for our bunching (distributional) analysis, uses information from 9394 assessments from 2017-2023, covering 33 countries, representing 917 million people (the total people assessed, including across multiple rounds, is 2.8 billion) (see SI Table A1, Sample A). IPC assessments occur in countries at risk of high levels of food insecurity; we note that countries in West Africa are classified using the Cadre Harmonisé, an analytical approach similar to, but distinct from, the IPC process. The data in our sample include all assessments with the majority of population data rounded to the nearest 5% bin available from 2017 to 2023 in the IPC's publicly available data platform (47). More than 78% of the assessments occurred in 2020 or later. Assessments include the population share and number of food insecure households in each severity classification (1-5) for each assessment area (often district-level). TWGs allocate percentages of populations to classifications; the majority of TWGs report rounded population estimates to nearest five percent. TWGs can also report population figures generated by multiplying the percentage with the area population estimates. Area classifications of 3+ require at least 20% of the population in an area experiencing food insecurity consistent with phase 3 or above, which undercounts acutely food insecure households who fail to collectively comprise at least 20% of the population (48).

The second sample is a subset of the first. It includes the IPC assessments with the underlying FSI data available (n = 1849). We use this sample to compare the distribution of assessed populations against counterfactuals derived from the underlying FSIs, (SI Table A1. Sample B). Our sample consists of four different combinations of FSIs; the IPC analyses for each data group is listed in Table A4.

3. Bunching (distributional) analysis

We begin by using a bunching analysis to estimate the existence and magnitude of undercounting around a critical assessment threshold: the 20% threshold of population in phase 3+. The 20% phase 3+ population threshold is a critical delineator in IPC assessments, officially distinguishing assessment areas that are stressed from those that are in crisis. When 20% of the population in a location is classified as being in phase 3+, the location moves from being classified as phase 2 ("stressed") to being classified as phase 3 ("crisis"). The phase 3 "crisis" designation is used to identify locations in need of aid; crisis areas receive relatively more attention and more scrutiny, potentially creating incentives to manipulate assessments in this range of the distribution.

To evaluate whether bunching exists, we construct counterfactual distributions that predict the distribution if the threshold were not in place. This approach assumes that the underlying data, when aggregated over a large number of IPC zones and assessment rounds, is smoothly distributed. Thus, in our setting, for example, there is no reason to expect that on average, there would be a mass point in the distribution of true food security at 0.35 relative to 0.4.

To ensure our data are comparable across assessments, we use the sample of assessments where the percentage of population is rounded to the nearest 5% (Sample A in Table 4 above). To estimate this counterfactual distribution, we follow the approach used in 35-37. Chetty et al. (35) examine the effect of a government-imposed tax threshold on the distribution of income by comparing the population earning slightly below the threshold for a higher taxation rate to the population earning slightly above. They identify a non-smooth discontinuity in the distribution, attributing it to the taxable earnings threshold. They compare the observed distribution to a distribution that is smooth across the threshold, assuming that in aggregate, we would not expect to see many more households with income just below that tax threshold than right above.

In our application, we test for the influence of the 20% threshold. We first omit a range of data around the threshold, [-5%, +10%]. Then, we fit multiple polynomials — specifically $3^{rd,}$ 4^{th}, 5^{th}, 6^{th}, and 7^{th} degree — to model the expected distribution of the percentage of population in phase 3+. We use the 4^{th} polynomial results for our main analysis; the others are available in the SI. We run 'p' simulations for each of several scenarios, where in each scenario we strategically exclude one bin at a time within the relevant range (i.e., [15%, 30%]) to generate estimated frequencies. We then aggregated the resulting frequencies across these exclusion scenarios to compute the mean and standard deviation of the estimated frequency for each bin. These final computations form the basis of our counterfactual distribution and provide the standard error for each bin. We compare this counterfactual to the observed distribution to detect any significant concentration or ‘bunching’ near the threshold. Specifically, we compute excess mass at each bin by taking the difference between the observed and estimated frequencies derived from the counterfactual distribution. A positive value indicates a higher-than-expected frequency (i.e., excess mass), while a negative value reflects a lower-than-expected frequency. The excess frequency at 15% serves as an indicator of the degree of bunching. The lower-than-expected frequency at and above the threshold suggests those locations where the population at 3+ is undercounted. We also use the 95% confidence intervals of the estimated frequencies to construct the lower and upper bounds of the excess mass at each bin.

4. Counterfactuals from FSIs

While the bunching approach narrowly focuses on identifying misclassification around a threshold, to estimate the total number of mis-classified population using the sample with available food security data, we evaluate the full distribution of assessed populations against counterfactuals derived from the underlying FSIs (Sample B in Table A1). We compare the assessments against a counterfactual using the average of the underlying FSI data used by the TWGs. While TWGs are trained to not simply average FSIs, recent research suggests that the implicit weights placed on different indicators vary over time, even within the same country, leading to the same classification being associated with very different levels of FSIs (26). Given such documented inconsistencies, averaging indicators provides a more cautious and transparent approach to developing a metric for comparison.

We then relate our assessments to the distribution of the IPC TWG consensus-based 3+ population estimates. In a third analysis we also use the argument of the "right to food" to classify people as facing acute hunger if any of the underlying FSIs points to households being in phase 3+ (38). Some indicators will miss higher levels of food insecurity (for example, if the household has run out of coping strategies, the coping-strategy related FSIs will not record them as being food insecure). In a rights-based framework, if a household is found to be hungry in any dimension, that is a cause for intervention. This approach can be thought of as an upper bound on the undercounting. However, our primary undercounting results come from the more conservative, second analysis.

We statistically test whether our counterfactual distributions differ from the distribution of the population assessed to be in phase 3+ using a Barrett and Donald (BD) test (39) to compare the empirical cumulative distribution functions of two distributions. Detailed information about the BD test is provided in SI.3(2) and SI Figure A4.

Given that different countries and survey rounds feature varying sets of FSIs, we categorize our sample into four distinct groups by the available FSIs. The results of our analysis were then presented on a group-by-group basis, aligning with the corresponding FSI implied population estimates.

5. Calculating undercounted population

The bunching analysis allows us to estimate the percentage of undercounting that is attributable to manipulation around the threshold. Estimates of bunching presented in Figure 2 suggest the presence of undercounting in the 20, 25, and 30% bins, where the numbers represent the percent of the population in the area assessed in phase 3+. Assuming these 'missing' observations at 20, 25, 30% are instead assessed to be at 15%, we can estimate the total population undercounted just at the threshold by taking the average population of areas assessed at 15% (approximately 0.38 Million), and multiplying that quantity by the number of missing allocations in each of the higher bins, times the implied difference in the percent of population assessed in 3+. Specifically, we calculate the total undercounted population by taking:

$$\begin{aligned} &Total\ undercounted\ 3+\ population \\ &\qquad = \textstyle\sum_{b=\{0.20,0.25,0.30\}} (Average\ population\ of areas\ assessed\ at\ the\ 15\%\ bin) \\ &\qquad * (Insufficient\ Mass_b) \times (b - 0.15) \end{aligned}$$

Our estimate of the undercounted in the area of the threshold is based on several important, conservative assumptions: (1) we assume that there is only an incentive to be conservative relative to the 20% threshold, i.e., bunching found at 15% (below the threshold) comes only from the 20% or nearby bins above the threshold, (2) that the undercounted assessments are assumed to be in regions of 20% to 30% based on a visual inspection of Figure 2a, (3) we assume that the classifications that are overcounted at the 15% level do not have systematically larger or smaller populations than the average assessment at 15%. One limitation of our estimate is that we cannot identify which classifications are truly overcounted at the 15% bin, so we cannot explicitly identify which of those locations have 20%, 25%, or more of the population in phase 3+.

6. Robustness Checks

We conduct a number of analyses to test the consistency of our findings. Our findings are robust to a wide range of functional forms and alternative assumptions. The results of each robustness test are briefly discussed below and presented in more detail in the SI.
For our bunching analyses, we first consider different forms of estimating the counterfactual:

(a) We construct the counterfactual using multiple levels of polynomial and find our results remain substantially unchanged (SI Figure A10).

(b) We test a wider range of bins around the threshold and find that if we include the 10 percent bin, our estimates shrink slightly (from 5.70 to 4.95 percent) (SI Figure A11).

(c) We test whether our results are sensitive to which bin we exclude when constructing the counterfactual and find out results are unchanged (SI Figure A12).

(d) We test whether rounding is driving our results by testing whether we observe bunching for both the unrounded portion of sample A and sample A’ and find evidence for bunching at the 20% threshold for the 3+ throughout (SI Figures A13 - A14).

(e) We test for bunching in a wider range of subsamples and find it across locations and periods (SI Figures A15 - A16).

Data Availability Statement

The University of Texas at Austin's Institutional Review Board approved the study protocol: Study00003367 "Accuracy of Food Security Classifications." To maintain confidentiality, the 21 key informant interviews with humanitarian actors knowledgeable about the IPC are not publicly available.

For our quantitative empirical analyses, we primarily use IPC outcome data, including Phase Classification and Population estimates by IPC regional assessment unit, which is publicly available in the IPC Population Tracking Tool https://www.ipcinfo.org/ipc-country-analysis/population-tracking-tool/en/). Due to privacy restrictions, we cannot share the underlying FSI data used in some of our analyses. These data are owned by each country engaging in IPC assessments and were provided to us, but we were not granted permission to share these data publicly. Therefore, we commit to sharing the data and code necessary to recreate all the materials in the main text and supplementary information, except for the underlying FSI data. Additional data used in the supplementary material are cited in the references.

Code availability statement

We share the code necessary to recreate all the materials or analysis results including figures and tables in the main text and supplementary information through a GitHub repository (https://github.com/mnmx0101/IPC_Paper.git). For any inquiries regarding the code or data, contact Chungmann Kim at (ck24@illinois.edu).

Materials & Correspondence

Correspondence and requests for materials should be addressed to Erin Lentz (erinclentz@utexas.edu).

References

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

2. 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/

3. Hoddinott, John, John A Maluccio, Jere R Behrman, Rafael Flores, and Reynaldo Martorell. 2008. "Effect of a Nutrition Intervention during Early Childhood on Economic Productivity in Guatemalan Adults." *The Lancet* 371 (9610): 411–16. https://doi.org/10.1016/S0140-6736(08)60205-6

4. Ruel, Marie T, Purnima Menon, Jean-Pierre Habicht, Cornelia Loechl, Gilles Bergeron, Gretel Pelto, Mary Arimond, John Maluccio, Lesly Michaud, and Bekele Hankebo. 2008. "Age-Based Preventive Targeting of Food Assistance and Behaviour Change and Communication for Reduction of Childhood Undernutrition in Haiti: A Cluster Randomised Trial." *The Lancet* 371 (9612): 588–95. https://doi.org/10.1016/S0140-6736(08)60271-8

5. Lowcock, Mark. 2018. "A Casement Lecture: Toward a Better System for Humanitarian Financing," 2018. https://cerf.un.org/sites/default/files/resources/cerf_casement-lecture-humanitarian-financing_en.pdf

6. World Peace Foundation. 2025. Contemporary Famines and Near-Famines. August 22. Accessed August 28, 2025. https://worldpeacefoundation.org/wp-content/uploads/2025/01/IPC-FRC-cases-without-agg-mort-301224.pdf

7. Maxwell, Daniel, and Peter Hailey. 2021. "Analysing Famine: The Politics of Information and Analysis in Food Security Crises." *Journal of Humanitarian Affairs* 3 (1): 16–27. https://doi.org/10.7227/JHA.055

8. FAO, IFAD, UNICEF, WFP, and WHO. 2023. *The State of Food Security and Nutrition in the World 2023*. FAO; IFAD; UNICEF; WFP; WHO; https://doi.org/10.4060/cc3017en.

9. World Food Program (WFP). 2023a. "Global Report on Food Crises: Number of People Facing Acute Food Insecurity Rose to 258 Million in 58 Countries in 2022." *WFP*, May 2023. https://www.wfp.org/news/global-report-food-crises-number-people-facing-acute-food-insecurity-rose-258-million-58#:~:text=Acute%20food%20insecurity%20is%20when,the%20Cadre%20Harmonis%C3%A9%20(CH)

10. FSIN and Global Network Against Food Crises. 2024. "Global Report on Food Crises 2024." Rome. www.fsinplatform.org/grfc2024

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

12. United Nations Office for Coordination of Humanitarian Affairs (UN OCHA). 2024b. Sudan: Famine Prevention Plan 2024 - Through Accelerated Mitigative Actions to Health Deterioration of Food Insecurity, Livelihood Impoverishment and Malnutrition. Accessed May 8, 2025. https://www.unocha.org/publications/report/sudan/sudan-famine-prevention-plan-2024-through-accelerated-mitigative-actions-halt-deterioriation-food-insecurity-livelihood-impoverishment-and-malnutrition

13. WFP 2024. News Releases: WFP Response to new IPC Food Security Assessment on Gaza. 25 June. Accessed May 8, 2025. https://www.wfp.org/news/wfp-response-new-ipc-food-security-assessment-gaza

14. WFP 2025. News Releases: Conflict and rising food prices drive Congolese into one of the world's worst food crises according to new IPC data. 27 May. Accessed May 8, 2025. https://www.wfp.org/news/conflict-and-rising-food-prices-drive-congolese-one-worlds-worst-food-crises-according-new-ipc

15. United Nations Human Rights Commission 2020. "There is nothing left for us": starvation as a method of warfare in South Sudan - Conference room paper of the Commission on Human Rights in South Sudan (A/HRC/45/CRP.3). Published 5 October. Accessed May 8, 2025. https://reliefweb.int/report/south-sudan/there-nothing-left-us-starvation-method-warfare-south-sudan-conference-room-paper

16. International Court of Justice. 2024. *Public sitting held on Thursday 16 May 2024 at 3pm, at the Peace Palace, President Sala presiding in the case concerning:* Application of the Convention on the Prevention and Punishment of the Crime of Genocide in the Gaza Strip (South Africa v. Israel). Accessed May 8, 2025. https://www.icj-cij.org/sites/default/files/case-related/192/192-20240516-ora-01-00-bi.pdf

17. Action Against Hunger. 2023. "2023 Hunger Funding Gap Report: What's Needed to Stop the Global Hunger Crisis." Reliefweb. February 22, 2023. https://reliefweb.int/report/world/2023-hunger-funding-gap-report-whats-needed-stop-global-hunger-crisis

18. FCDO, UK. 2021. "G7 Principles Supporting a Vision of Improved Global Food Security Monitoring and Analysis," December 2021. https://www.gov.uk/government/publications/g7-principles-for-improved-global-food-

security-monitoring-and-analysis-for-action/g7-principles-supporting-a-vision-of-improved-global-food-security-monitoring-and-analysis

19. IPC. 2021. “Technical Manual Version 3.1.” https://www.ipcinfo.org/fileadmin/user_upload/ipcinfo/manual/IPC_Technical_Manual_3_Final.pdf

20. Machefer, Melissande, Anne-Claire Thomas, Michele Meroni, Jose Manuel Viega Lopez Pena, Michele Ronco, Christina Corbane, and Felix Rembold 2025. “Potential and limitations of machine learning modeling for forecasting Acute Food Insecurity. *Global Food Security*: 45. https://doi.org/10.1016/j.gfs.2025.100859

21. IPC. 2023. “IPC Global Strategic Programme 2023-2026.” IPC. https://www.ipcinfo.org/fileadmin/user_upload/ipcinfo/docs/IPC_GSP_2023_2026_Press_Release.pdf

22. Maxwell, Daniel, Bapu Vaitla, and Jennifer Coates. 2014. “How Do Indicators of Household Food Insecurity Measure up? An Empirical Comparison from Ethiopia.” *Food Policy* 47 (August): 107–16. https://doi.org/10.1016/j.foodpol.2014.04.003

23. Vaitla, Bapu, Jennifer Coates, Laura Glaeser, Christopher Hillbruner, Preetish Biswal, and Daniel Maxwell. 2017. “The Measurement of Household Food Security: Correlation and Latent Variable Analysis of Alternative Indicators in a Large Multi-Country Dataset.” *Food Policy* 68 (April): 193–205. https://doi.org/10.1016/j.foodpol.2017.02.006

24. Sen, Amartya. 1983. “Poor, Relatively Speaking.” *Oxford Economic Papers* 35 (2): 153–69.

25. Barrett, Christopher B. 2010. “Measuring Food Insecurity.” *Science* 327 (5967): 825–28. https://doi.org/10.1126/science.1182768

26. Lentz, Erin, Kathy Baylis, Hope Michelson, and Chungmann Kim. 2023. “Accuracy Scoping Study: Preliminary Findings on the Accuracy of the Integrated Food Security Phase Classification.”

27. Headey, Derek, and Olivier Ecker. 2013. “Rethinking the Measurement of Food Security: From First Principles to Best Practice.” *Food Security* 5 (3): 327–43. https://doi.org/10.1007/s12571-013-0253-0

28. Lain, Jonathan, Sharad Tandon, and Tara Vishwanath. 2023. “How Much Does the Food Insecurity Experience Scale Overlap with Poor Food Consumption and Monetary Poverty? Evidence from West Africa.” *The World Bank Economic Review*, October, lhad031. https://doi.org/10.1093/wber/lhad031

29. World Food Program (WFP). 2023b. "Livelihood Coping Strategies Indicator for Food Security." https://docs.wfp.org/api/documents/WFP-0000147801/download/

30. Kim, Chungmann, Kathy Baylis, Hope Michelson, and Erin Lentz. 2022. "The Distribution of Multi-Dimensional Food Insecurity over Space and Time." In American Agricultural Economics Association Annual Meetings 2022. https://docs.google.com/presentation/d/1U9sETz4Dh9hYb1h5t3Elku-2-TsYAUIKlFLUEQrjWog/edit#slide=id.p

31. The Economist. 2017. "How Is Famine Declared?" *The Economist*, April 7, 2017. https://www.economist.com/the-economist-explains/2017/04/04/how-is-famine-declared

32. Buchanan-Smith, Margie, Jane Cocking, and Zainab Moallin. 2023. "Independent Review of the Integrated Food Security Phase Classification (IPC) Somalia." HPG commissioned report. London. www.odi.org/en/publications/independent-review-of-the-ipc-in-somalia

33. Enten, François. 2023. "Famines and 'Poor Numbers': How IPC Data Is Communicated through the Media to Trigger Emergency Responses." *Journal of Humanitarian Affairs* 5 (1): 37–51. https://doi.org/10.7227/JHA.101

34. Georgieva, Kristalina, and Sultan Nazrin Shah. 2016. "Coordinating Funding for Humanitarian Emergencies." No. 1 Vol. LIII. United Nations Chronicle. https://www.un.org/en/chronicle/article/coordinating-funding-humanitarian-emergencies

35. Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri. 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *The Quarterly Journal of Economics* 126 (2): 749–804. https://doi.org/10.1093/qje/qjr013

36. Kleven, Henrik Jacobsen. 2016. "Bunching." *Annual Review of Economics* 8 (1): 435–64. https://doi.org/10.1146/annurev-economics-080315-015234.

37. Allen, Eric J., Patricia M. Dechow, Devin G. Pope, and George Wu. 2017. "Reference-Dependent Preferences: Evidence from Marathon Runners." *Management Science* 63 (6): 1657–72. https://doi.org/10.1287/mnsc.2015.2417

38. Fakhri, Michael. 2023. "UN Special Rapporteur on the Right to Food." Office of the United Nations High Commissioner for Human Rights. https://www.ohchr.org/sites/default/files/documents/issues/food/A_78_202_summary-report.pdf

39. Barrett, Garry F., and Stephen G. Donald. 2003. "Consistent Tests for Stochastic Dominance." *Econometrica* 71 (1): 71–104. https://doi.org/10.1111/1468-0262.00390

40. Lentz, E.C., H. Michelson, K. Baylis, and Y. Zhou. 2019. "A Data-Driven Approach Improves Food Insecurity Crisis Prediction." *World Development* 122 (October): 399–409. https://doi.org/10.1016/j.worlddev.2019.06.008

41. Andree, Bo Pieter Johannes; Chamorro Elizondo, Andres Fernando; Kraay, Aart C.; Spencer, Phoebe Girouard; Wang, Dieter. 2020. Predicting Food Crises (English). Policy Research working paper, no. WPS 9412 Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/304451600783424495/Predicting-Food-Crises

42. Wang, Dieter, Bo Pieter Johannes Andrée, Andres Fernando Chamorro, and Phoebe Girouard Spencer. 2022. "Transitions into and out of Food Insecurity: A Probabilistic Approach with Panel Data Evidence from 15 Countries." *World Development* 159 (November): 106035. https://doi.org/10.1016/j.worlddev.2022.106035

43. Zhou, Yujun, Erin Lentz, Hope Michelson, Chungmann Kim, and Kathy Baylis. 2022. "Machine Learning for Food Security: Principles for Transparency and Usability." *Applied Economic Perspectives and Policy* 44 (2): 893–910. https://doi.org/10.1002/aepp.13214

44. Caldwell, Seth. 2022. "Assessing the Technical Feasibility of Conflict Prediction for Anticipatory Action." https://data.humdata.org/dataset/2048a947-5714-4220-905b-e662cbcd14c8/resource/926e652a-d161-4b16-8598-936f706ee454/download/conflict-prediction-october-2022.pdf?_gl=1*1b76rqg*_ga*MTQ4ODk3OTcxOC4xNzA5NzQ0MTI2*_ga_E60ZNX2F68*MTcwOTc0NDEyNS4xLjEuMTcwOTc0NDE4Ny42MC4wLjA

45. Lentz, Erin C., and Daniel Maxwell. 2022. "How Do Information Problems Constrain Anticipating, Mitigating, and Responding to Crises?" *International Journal of Disaster Risk Reduction* 81 (October): 103242. https://doi.org/10.1016/j.ijdrr.2022.103242

46. Glasman, Joël, and Brendan Lawson. 2023. "Ten Things We Know about Humanitarian Numbers." *Journal of Humanitarian Affairs* 5 (1): 1–10. https://doi.org/10.7227/JHA.098

47. IPC. n.d. "Population Tracking Tool" Accessed December 2, 2023. https://www.ipcinfo.org/ipc-country-analysis/population-tracking-tool/en/

48. Maxwell, Daniel, Abdullahi Khalif, Peter Hailey, and Francesco Checchi. 2020. "Viewpoint: Determining Famine: Multi-Dimensional Analysis for the Twenty-First Century." *Food Policy* 92 (April): 101832. https://doi.org/10.1016/j.foodpol.2020.101832

Global Estimates Systematically Undercount Acute Hunger

Supplementary Information (SI)

SI.1 Sampling process and summary statistics

In Table A1., our data from Sample A and Sample A' come from the IPC Population Tracking Tool (1), a platform that provides the public access to historical IPC population data for over 30 different countries. Our sample includes all assessments available between 2017 and 2023, as of December 2023. In 2017, IPC assessments were available for 17 countries. By 2023, the number of countries undertaking IPC assessments had increased to 33, reflecting the expansion of the IPC to new countries during this period. The initial sample (including Sample A and Sample A') consisted of 10,890 assessments. Of these, 9,394 used a decision rule to round the majority (>50%) of their IPC 3+ population estimates to the nearest 5% bin. For consistency, we analyze this subsample - Sample A in Table A1.

Our data from Sample B is the sample of assessments conducted between 2017 and 2023, which included a portion of the population in 3+, and were available in the IPC's Consolidated Data Tool (CDT), as of December 2023. Data are summarized at the assessment area level, not the household. The CDT is IPC's internal platform for collecting evidence for its consensus outcomes, providing access to Food Security Indicators (FSI), population estimates and information on contributing factors and humanitarian food assistance (HFA), when available. This includes information on the population share corresponding to each phase based on key FSI such as Food Consumption Score (FCS), Household Dietary Diversity Score (HDDS), Household Hunger Scale (HHS), reduced Coping Strategies Index (rCSI), and Livelihood Coping Strategies (LCS). FCS and HDDS reflect dietary quality while rCSI, HHS, and LCS reflect food insecurity experience and coping; correlations vary by FSI pair (2, 3). Each FSI has its own threshold for different phases as defined by the IPC Manual 3.1 (see IPC Technical Manual – protocol 2.2. (4)). We merge the Acute Food Insecurity (AFI) outcomes with underlying FSI outcomes from the CDT based on analysis date and area name.

Disaggregating data from Table A1, Tables A2 and A3 provide country-specific summary statistics, including the count of Technical Working Group (TWG) meetings, phase 3+ population estimates in millions, total population figures, and the number of countries experiencing a phase 3 or higher level of food insecurity. They detail the latest and earliest year-month assessments per country and average phase classifications and phase 3+ populations.

The IPC Technical Manual's Reference Table (4) indicates a classification for each individual FSI. We find that the underlying FSIs do not consistently point to the same classification or to the same population figures, making the convergence of evidence process by the TWGs challenging. Approximately 40% of observations that are deemed to be in phase 2 or 3 by TWGs have underlying FSIs that imply phases of up to 2 levels of difference or population estimates that vary widely across FSIs.

In Table A4, we present summary statistics by county and round by FSI availability group, using the same approaches described for Table A3. For example, in the March 2021 assessment for Afghanistan, the population experiencing acute food insecurity (3+) consistent with rCSI was 15% while for FCS it was 77%. These measures capture coping strategies and dietary quality, underscoring the multidimensionality of food insecurity.

Figure A1 shows pairwise correlations of the share of populations in 3+ by FSI. Vaitla et al. (3) compared several food security indicators across multiple countries. They find FCS and HDDS are well correlated, reflecting that they are often based on the same suite of questions; rCSI and HHS are closely related at the household level. Thus, these pairs of indicators should provide similar information to TWGs and should be somewhat substitutable. We present Pearson correlations of the percent of the population in phase 3+ for pairs of FSIs. In Figure A1, we show results for the sample with five FSIs (HDDS, FCS, HHS, rCSI, and LCS n=600). We find that the highest correlation, 0.53, is between the percents of population in Phase 3+ for HDDS and FCS, two indicators that are correlated by construction. Other correlation pairs are relatively low. We lack access to the raw data and we are therefore not able to determine the drivers of the weak correlations. The weak correlations might be attributable to issues related to data collection, cleaning, or computation of the food security indicators. These processes all occur before the IPC TWG work. It is also possible that relationships among FSIs for populations in Phase 3 or above could be highly context dependent.

In sum, our findings show that the appropriate phase classification and population are not obvious or mechanical, underscoring the motivation for a consensus-based approach.

Table A1: Description of the samples

	Year Coverage	# of countries	# of TWG meetings	# of assessments	# of area assessments in urgent need (3+)	# of total estimated population (M)	# of estimated population in urgent need (M)
Sample A (Full sample) – majority is rounded	2017.01 – 2023.10	33	172	9394	6362	2803.97	730.72
Sample A’ (Full sample) – majority is not-rounded	2017.01 – 2023.03	20	45	1496	490	461.17	70.42
Sample B (Sub-sample with underlying FSIs available)	2020.10 – 2022-11	15	27	1849	1388	743.34	226.88

Caption: The IPC manual allows reporting of unrounded population estimates (%) or rounded estimates to the nearest 5%. Using the IPC population tracker data (as of Dec/02/2023) with 10,890 entries, we segmented our sample into rounded and unrounded samples depending on whether more than 50% of assessments per country and year-month had rounded their proportion of population assessed to be in phase 3+. For example, Afghanistan’s November 2017 assessment had 12 out of 23 assessments that were not rounded and is therefore categorized into Sample A’. Additionally, we created Sample B, which includes data with assessment area -specific food security indicators. Finally, IPC analyses report the 'Tri-National Border Region' (El Salvador, Guatemala, and Honduras) separately, which we treat as a 33rd “country.”

Table A2. Summary statistics by country: Sample A

Country	# of TWG meetings	# of classifications	# of total population (million)	# of 3+ classifications	# of 3+ population (million)	Latest report (Y-M)	Earliest report (Y-M)	IPC phase (mean)	Phase 3+ population (mean)
Afghanistan	11	462	387.30	420	146.33	2023-10	2017-08	3.04	38%
Angola	1	17	2.67	16	1.32	2021-06	2021-06	3.29	55%
Bangladesh	1	17	38.24	13	8.91	2023-03	2023-03	2.76	25%
Burundi	11	88	126.37	10	16.53	2023-09	2017-04	2.10	13%
entral African Republi	9	500	48.06	475	19.09	2023-09	2018-03	3.11	43%
Congo, DRC	7	1033	593.62	668	156.58	2023-07	2018-08	2.69	25%
Djibouti	3	45	3.48	22	0.54	2023-03	2020-10	2.49	23%
Dominican Republic	1	32	10.62	6	1.55	2022-10	2022-10	2.16	16%
El Salvador	3	25	10.64	1	0.90	2020-11	2017-11	1.76	8%
Eswatini	8	57	8.74	31	1.83	2023-06	2017-07	2.49	21%
Ethiopia	4	263	179.09	184	41.84	2021-05	2019-07	2.72	26%
Guatemala	3	67	50.84	34	10.21	2022-03	2018-11	2.48	20%
Haiti	6	184	56.68	175	23.33	2023-08	2018-10	3.11	41%
Honduras	2	27	12.09	19	3.43	2020-12	2018-12	2.70	26%
Kenya	9	219	140.73	79	23.03	2023-07	2019-07	2.27	18%
LAC (Tri-National Border Region)	2	11	0.84	1	0.13	2020-10	2018-02	1.91	8%

Lebanon	3	165	16.51	113	4.45	2023-10	2022-09	2.68	27%
Lesotho	6	60	8.85	29	1.70	2023-07	2019-05	2.48	20%
Madagascar	10	140	39.97	85	11.58	2023-07	2017-08	2.67	30%
Malawi	6	192	112.05	23	12.36	2023-07	2020-07	2.00	11%
Mozambique	8	509	108.24	144	12.51	2023-05	2017-06	2.05	13%
Namibia	5	70	12.66	43	2.46	2023-07	2019-10	2.61	20%
Pakistan	6	149	100.80	137	29.21	2023-04	2018-10	2.96	32%
Somalia	5	1731	84.21	1098	23.61	2023-08	2022-07	2.65	27%
South Africa	1	34	59.13	9	9.34	2020-09	2020-09	2.26	16%
South Sudan	10	774	116.05	689	57.64	2023-09	2017-01	3.16	49%
Sudan	4	744	188.23	510	43.44	2023-06	2020-06	2.74	25%
Tanzania	4	68	19.59	21	2.65	2022-10	2017-07	2.31	14%
Timor-Leste	1	14	1.34	11	0.30	2022-11	2022-11	2.79	24%
Uganda	8	140	98.06	95	8.75	2023-08	2017-01	2.61	24%
Yemen	5	946	79.72	844	35.88	2023-01	2018-12	3.05	43%
Zambia	5	371	50.27	149	7.20	2023-08	2019-05	2.39	17%
Zimbabwe	4	240	38.25	208	12.13	2020-10	2019-02	2.90	31%

Caption: The table provides a country-specific summary statistic of Sample A as described in Table A1. All percentage figures are rounded to zero decimal places.

Table A3. Summary statistics by country: Sample B from Table A1

	# of TWGs meetings	# of assessments	# of total population (million)	# of 3+ classifications	# of 3+ population (million)	Latest report (Y-M)	Earliest report (Y-M)	IPC phase (mean)	IPC 3+ population (mean)	FCS 3+ population (mean)	rCSI 3+ population (mean)	HHS 3+ population (mean)	LCS 3+ population (mean)	HDDS 3+ population (mean)
Afghanistan	4	180	169.89	176	72.51	2022-03	2021-09	3.24	42%	79%	22%	39%	59%	51%
Central African Republic	3	191	15.20	191	6.35	2022-09	2021-04	3.12	44%	46%	28%	79%	76%	.
Congo, DRC	3	529	299.17	346	80.28	2022-08	2021-02	2.67	25%	57%	17%	43%	45%	38%
Djibouti	1	15	1.18	6	0.13	2022-03	2022-03	2.40	19%	34%	5%	.	12%	25%
Ethiopia	1	13	8.81	13	5.43	2021-05	2021-05	3.54	60%	72%	57%	61%	91%	.
Guatemala	1	22	17.36	14	3.93	2022-03	2022-03	2.64	22%	27%	10%	.	43%	.
Haiti	1	32	9.91	32	4.33	2021-09	2021-09	3.16	44%	55%	27%	28%	53%	71%
Kenya	2	46	30.57	10	3.57	2021-09	2021-02	2.09	12%	31%	10%	.	16%	.
Lebanon	1	52	5.36	41	1.99	2022-09	2022-09	2.79	37%	51%	47%	.	72%	94%
Madagascar	3	47	14.16	39	5.05	2022-04	2022-11	2.94	**37%**	73%	49%	**37%**	47%	35%
Mozambique	1	62	14.02	17	1.82	2021-11	2021-11	2.25	20%	22%	16%	.	15%	31%
Pakistan	2	44	33.33	40	8.48	2021-10	2021-03	2.91	26%	61%	12%	.	44%	22%
South Sudan	1	77	12.37	75	6.63	2022-	2022-	3.39	53%	79%	29%	72%	64%	64%

						10	10							
Sudan	2	358	90.98	210	16.44	2022-05	2021-04	2.60	**20%**	**20%**	5%	.	38%	.
Yemen	1	181	21.02	178	9.94	2020-10	2020-10	3.29	50%	46%	38%	33%	61%	25%

Caption: The table provides a country-specific summary statistics of sample B as described in Table A1. All percentage figures are rounded to zero decimal places.

Table A4: Country-round summary statistics by FSI data group: Sample B from Table A1.

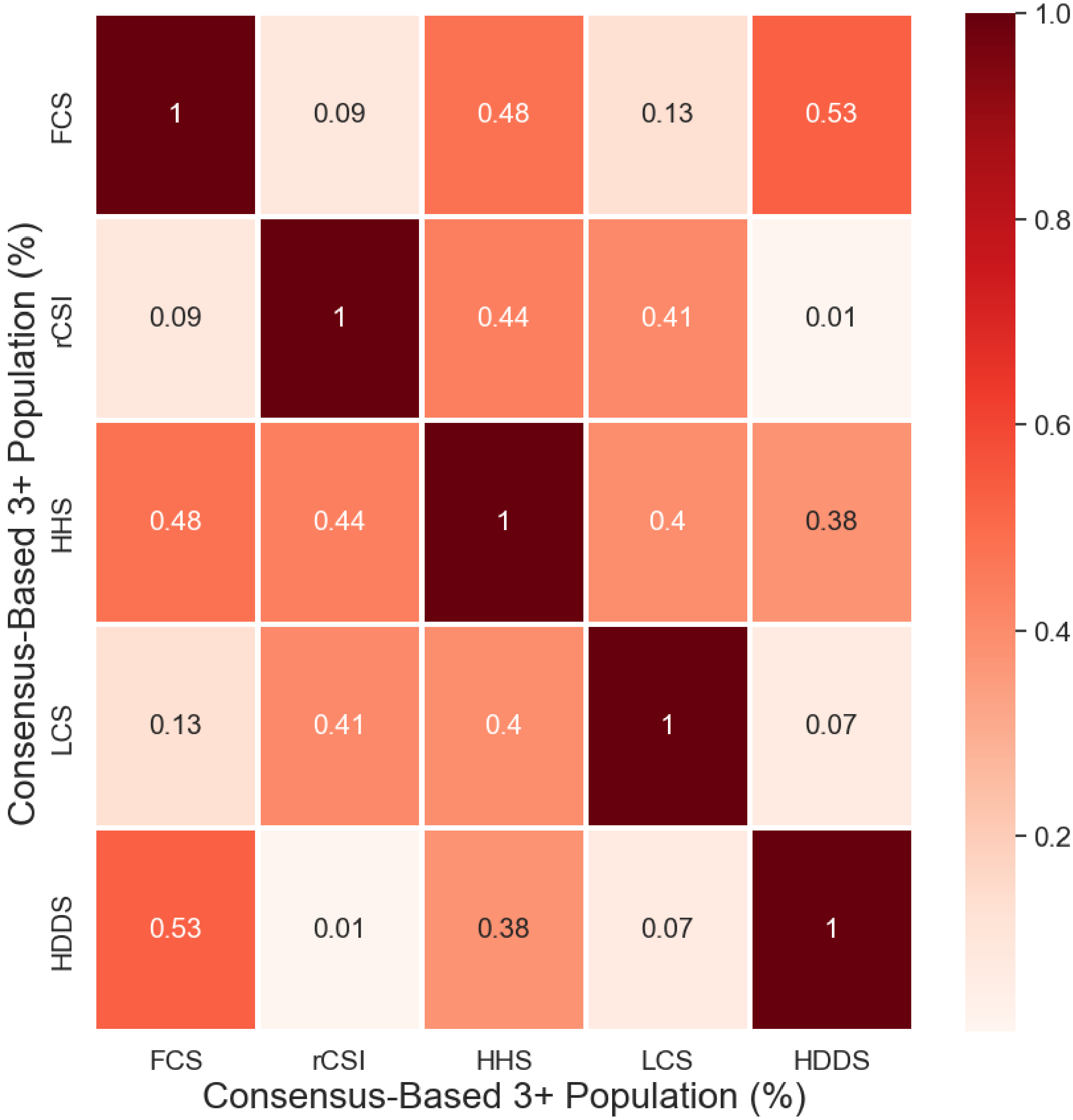
TWGs meetings	# of assessments	# of total population (million)	Latest report (Y-M)	Earliest report (Y-M)	IPC phase (mean)	# of 3+ population (million)	IPC 3+ population (mean)	FCS 3+ population (mean)	rCSI 3+ population (mean)	HHS 3+ population (mean)	LCS 3+ population (mean)	HDDS 3+ population (mean)
Group1: FCS-RCSI-HHS-LCS-HDDS												
Afghanistan | 2021-03	45	34.68	2021-03	2021-03	3.02	12.03	34%	77%	15%	31%	52%	33%
Afghanistan | 2021-09	45	45.64	2021-09	2021-09	3.47	21.52	47%	75%	25%	39%	58%	69%
Afghanistan | 2022-03	45	42.43	2022-03	2022-03	3.40	20.07	47%	90%	26%	52%	68%	56%
Afghanistan | 2022-09	45	47.13	2022-09	2022-09	3.07	18.89	40%	76%	21%	34%	59%	43%
Congo, DRC | 2021-09	179	102.26	2021-09	2021-09	2.71	27.02	26%	60%	18%	41%	47%	47%
Haiti | 2021-09	32	9.91	2021-09	2021-09	3.16	3.93	44%	55%	27%	28%	53%	71%
Madagascar | 2021-05	10	2.68	2021-05	2021-05	3.30	1.14	44%	76%	55%	62%	73%	19%
Madagascar | 2022-04	16	5.25	2022-04	2022-04	2.75	1.68	33%	70%	59%	38%	47%	44%
Madagascar | 2022-11	21	6.23	2022-11	2022-11	2.90	2.23	36%	75%	39%	24%	35%	35%
South Sudan | 2022-10	77	12.37	2022-10	2022-10	3.39	6.63	53%	78%	29%	72%	64%	64%
Yemen | 2020-10	181	21.02	2020-10	2020-10	3.29	9.94	50%	46%	38%	34%	61%	24%

Group2: FCS-RCSI-HHS-LCS												
Central African Republic | 2021-04	65	4.74	2021-04	2021-04	3.05	1.93	40%	45%	10%	78%	74%	.
Central African Republic | 2021-09	62	4.71	2021-09	2021-09	3.11	1.96	45%	45%	41%	80%	75%	.
Central African Republic | 2022-09	64	5.74	2022-09	2022-09	3.19	2.46	47%	47%	34%	81%	79%	.
Congo, DRC | 2022-08	184	102.30	2022-08	2022-08	2.66	26.23	25%	61%	16%	43%	43%	.
Ethiopia | 2021-05	13	8.81	2021-05	2021-05	3.54	5.43	60%	72%	57%	61%	92%	.
Group3: FCS-RCSI-LCS-HDDS												
Congo, DRC | 2021-02	166	94.62	2021-02	2021-02	2.65	27.03	26%	50%	17%	48%	46%	31%
Djibouti | 2022-03	15	1.18	2022-03	2022-03	2.40	0.13	19%	33%	4%	.	12%	24%
Mozambique | 2021-11	62	14.02	2021-11	2021-11	2.25	1.82	20%	22%	16%	.	15%	31%
Pakistan | 2021-10	25	18.59	2021-10	2021-10	2.88	4.66	26%	58%	13%	.	49%	22%
Group4: FCS-RCSI-LCS												
Guatemala | 2022-03	22	17.36	2022-03	2022-03	2.64	3.93	22%	26%	10%	.	43%	.
Kenya | 2021-02	23	15.41	2021-02	2021-02	1.96	1.43	10%	30%	5%	.	16%	.
Kenya | 2021-09	23	15.15	2021-09	2021-09	2.22	2.15	15%	32%	13%	.	16%	.

Lebanon | 2022-09	52	5.36	2022-09	2022-09	2.79	1.99	37%	51%	47%	.	72%	94%
Pakistan | 2021-03	19	14.75	2021-03	2021-03	2.95	3.82	26%	65%	12%	.	37%	.
Sudan | 2021-04	179	44.92	2021-04	2021-04	2.46	7.15	17%	17%	4%	.	33%	.
Sudan | 2022-05	179	46.06	2022-05	2022-05	2.73	9.28	23%	23%	4%	.	43%	.

Caption: This table presents summary statistics for Sample B from Table A1, disaggregated by the relative availability of five Food Security Indicators (FSIs)—FCS, rCSI, HHS, LCS, and HDDS—across countries and rounds used in the TWG analysis. Each TWG analysis is assigned to an FSI group based on whether more than 50% of assessed areas include a given indicator. For example, Congo, DRC (2021-02) and Lebanon (2022-09) are categorized into Group 3: FCS–rCSI–LCS–HDDS and Group 4: FCS–rCSI–LCS, respectively. Mean values for HHS and HDDS are still reported even when they are not part of the group, provided that some subsample (less than 50%) includes records for those indicators.

Figure A1. Correlation Matrix: FSI implied 3+ population (%) for sample with data group 1 (All-5-FSI)



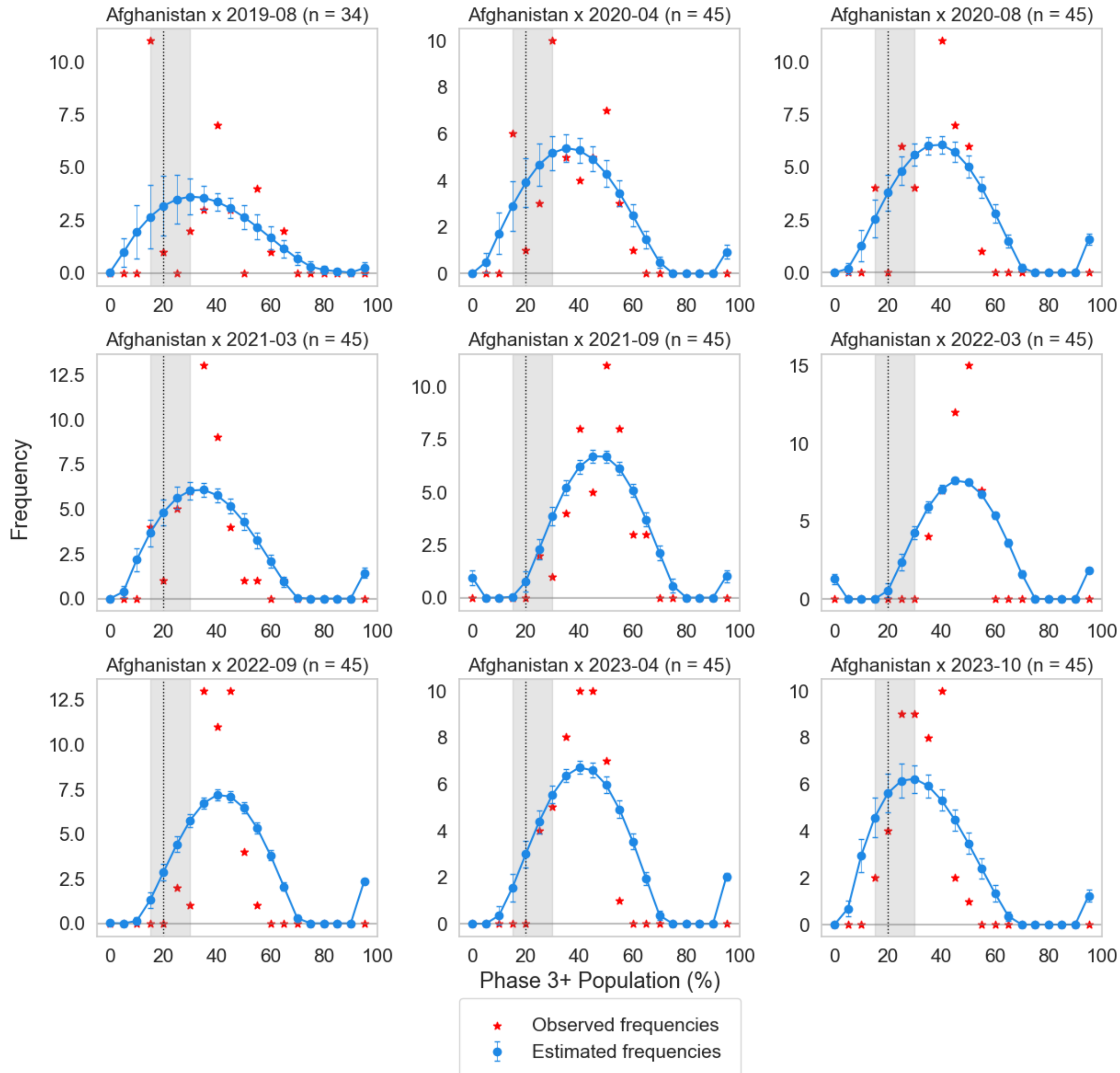
Caption: This figure presents a Pearson correlation matrix illustrating the relationships among the share of the population assessed to be in phase 3+ based on various FSI such as HHS, rCSI, FCS, LCS, and HDDS. The analysis specifically focuses on the sample of analyses for data group 1, which has all 5 FSIs (All-5-FSI sample, n = 600).

SI.2 Subsample bunching analysis: Afghanistan

The bottom panel in Figure 2 in the main text extends our analysis beyond the pooled sample (top panel in Figure 2), showing that the bunching is prevalent across multiple countries. This exhibit distribution patterns similar to those observed in our pooled sample (the top panel in Figure 2, underscoring the consistency of the bunching in diverse contexts.

We further extended our analysis to examine consistency in findings over several rounds within a country, Afghanistan in Figure A2. This figure shows that in Afghanistan, bunching occurs more frequently prior to September of 2021 (2021-09), likely in part reflecting a deteriorating situation following the Islamic Republic takeover in 2021.

Figure A2. Round specific observed and estimated distributions of IPC assessments defined by the proportion of population in phase 3+ for Afghanistan



Caption: This figure shows a comparative distributional analysis of the proportion of population in phase 3+ by assessment round in Afghanistan. Each subplot combines observed frequencies of 3+ population (red stars) and estimated frequencies (4th polynomial fitting by 5% bin) with 95% confidence intervals. We employ the same bin exclusion strategy used to calculate estimated frequencies 1 from Figure 2 in the main text. We enforce a non-negativity constraint to prevent frequency estimates from becoming negative (n = 394).

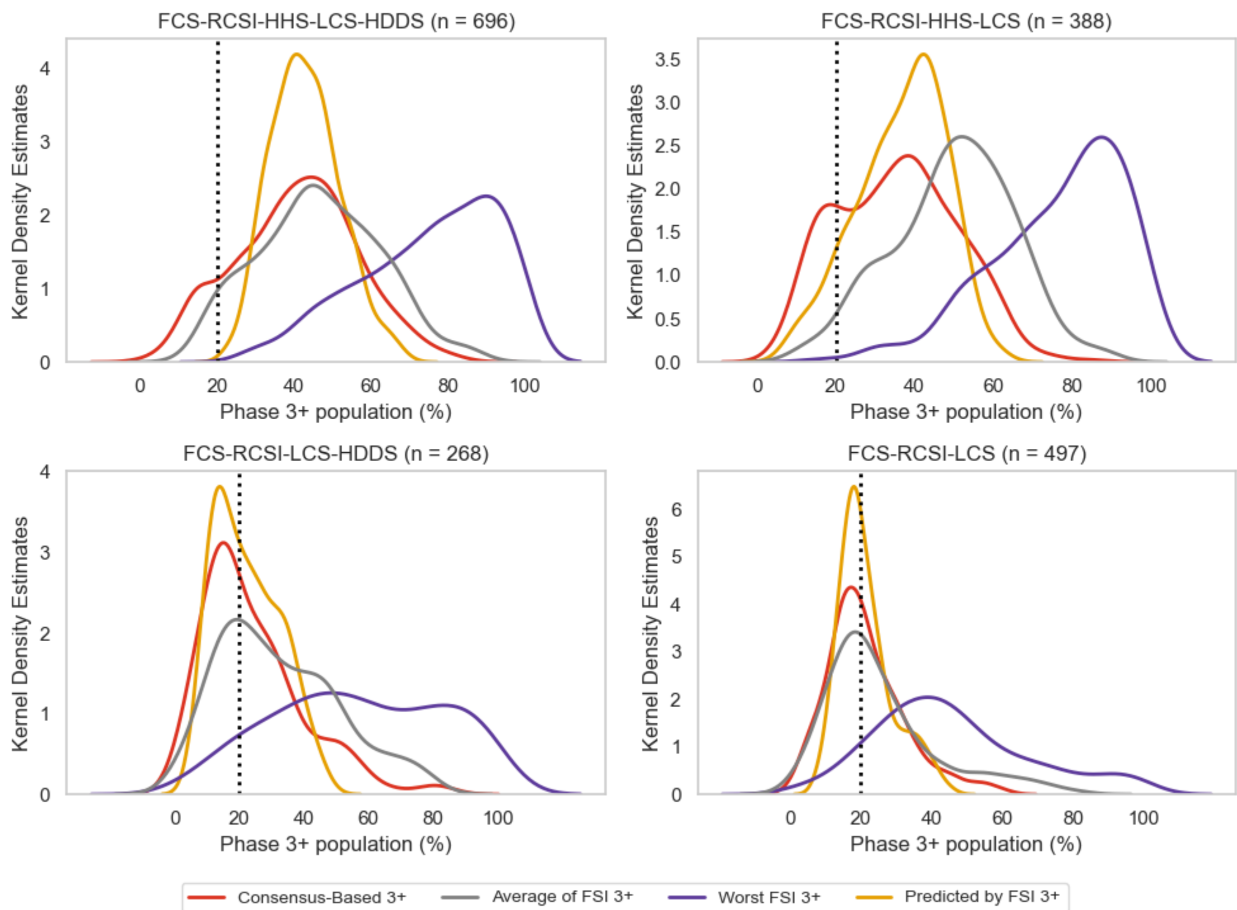
Note: The smaller number of samples or regions in August 2019 is due to TWG's implementation of a distinction between urban and non-urban areas in certain regions starting from 2020. These 11 regions include 'Baghlan', 'Balkh', 'Faryab', 'Helmand', 'Hirat', 'Jawzjan', 'Kabul', 'Kandahar', 'Kunduz', 'Nangarhar', and 'Takhar'.

SI.3 Analysis using food security indicators (FSI)

1) Alternative FSI-based population estimates

Our preferred approach is to compare the IPC population outcomes to the population estimates derived from the arithmetic mean of the five FSIs. This approach assumes each FSI is equally valuable for TWGs. In this section, we construct a counterfactual using an OLS regression in which the IPC 3+ population share is modeled as a linear function of the available FSIs. Predicted values are then computed using the estimated coefficients. By visual inspection, we see that the line representing the population predicted by FSI 3+ values has lower variance than either the consensus-based 3+ or average of FSI 3+ estimates. This is as one would expect with regression to the mean. Across the four data groups, consistent with findings for the average of the FSI 3+ estimates, we also do not observe the bunching pattern near the 20% threshold that we observe for the consensus-based 3+ estimates. The absence of bunching in model-based estimates further suggests that the consensus process is sensitive to the administratively imposed 20% threshold that serves as a key criterion for classifying areas as being in urgent need of assistance.

Figure A3. Comparisons of distributions of IPC assessments defined by proportion of population in phase 3+ (observed vs counterfactuals)



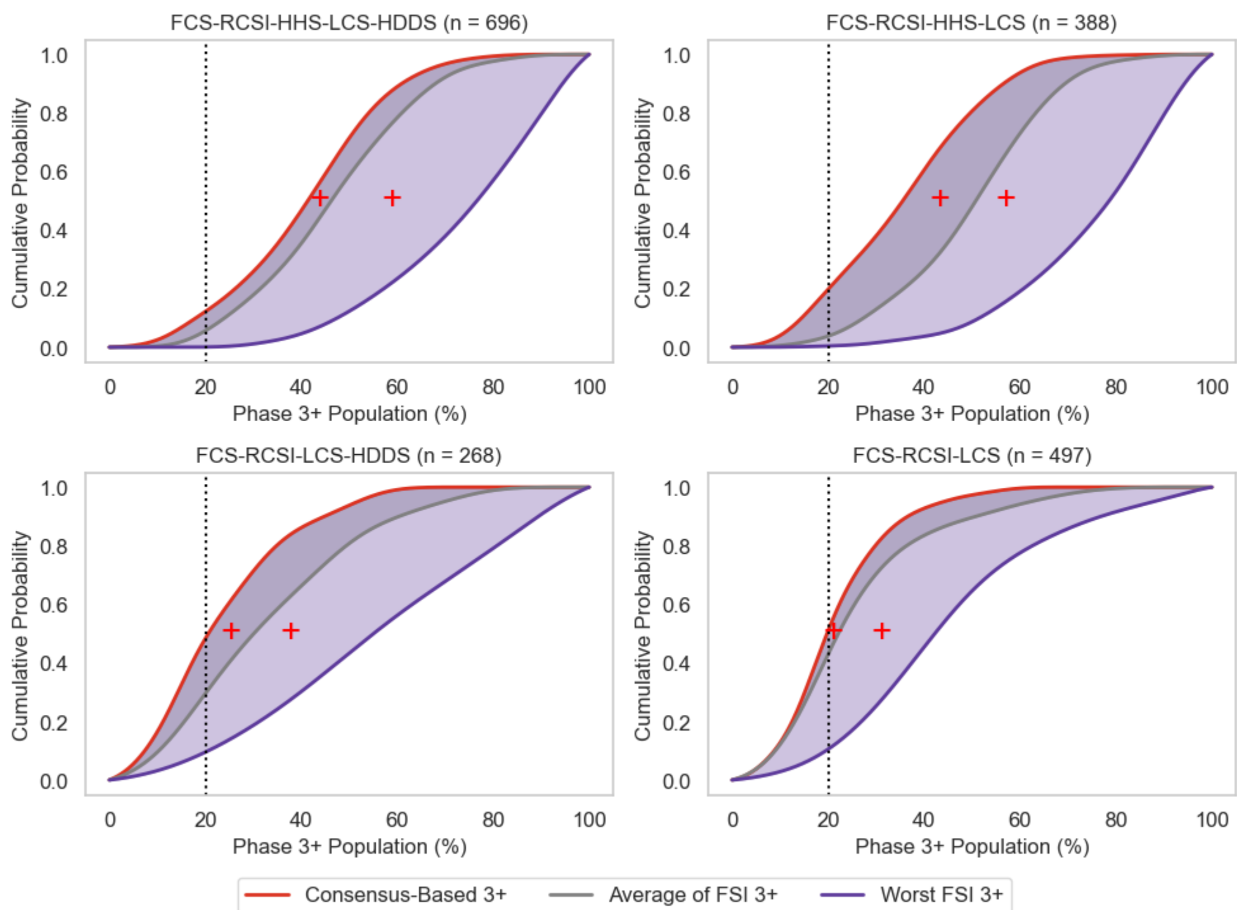
Caption: This figure presents gaussian kernel density estimates for four different distributions of population assessed in phase 3+: the percent of population assessed by the IPC technical working groups to be in phase 3+ (red line); the counterfactual distribution of the estimated percent of population assessed to be in phase 3+ using the average of underlying food security indicators (FSI) (gray line); the counterfactual distribution of the estimated percent of population assessed to be in phase 3+ using the worst underlying FSI 3+ (purple line); and a counterfactual estimate generated using an OLS regression model that predicts IPC Phase 3+ population percentages as a function of the available FSIs (yellow line). Each panel uses a sample based on different combinations of FSIs, (combinations of: food consumption scores (FCS); reduced coping strategies index (RCSI); household hunger scale (HHS); livelihoods coping strategy (LCS); and household dietary diversity score (HDDS)) reflecting the fact that available FSIs tend to vary across assessment areas. The 20% threshold between phase 2 and phase 3 is illustrated by a vertical black dashed line (n = 1849).

2) Stochastic dominance and comparison of distributions of population assessed in phase 3+

We employ the Barrett and Donald (BD) test to examine first-order stochastic dominance between consensus-based and counterfactual 3+ population distributions (5). This non-parametric test method allows for the assessment of stochastic dominance without the need to make specific distributional assumptions (6). We first calculate the empirical cumulative distribution functions (ECDFs for the different distributions). The test statistic is then derived as the supremum of the absolute difference between these ECDFs across the entire support of the distributions. Critical values for the test are estimated using bootstrap resampling techniques, where pseudo-samples are generated from the pooled original samples to construct the empirical distribution of the test statistic under the null hypothesis. The null hypothesis of no stochastic dominance is rejected if the observed test statistic exceeds the critical value, indicating that one distribution first-order stochastically dominates the other.

To visually compare the distributions of the population estimated using the arithmetic mean of FSIs and the worst food security level suggested by the FSIs, we plot the distributions of each in Figure A4. As can be observed, both the estimated distributions using the underlying FSI data are to the right of the observed distribution, suggesting that a larger fraction of people would be assessed as being food insecure under these alternative measures.

Figure A4: Comparisons of cumulative distributions of IPC assessments defined by proportion of population in phase 3+ (observed vs counterfactual)

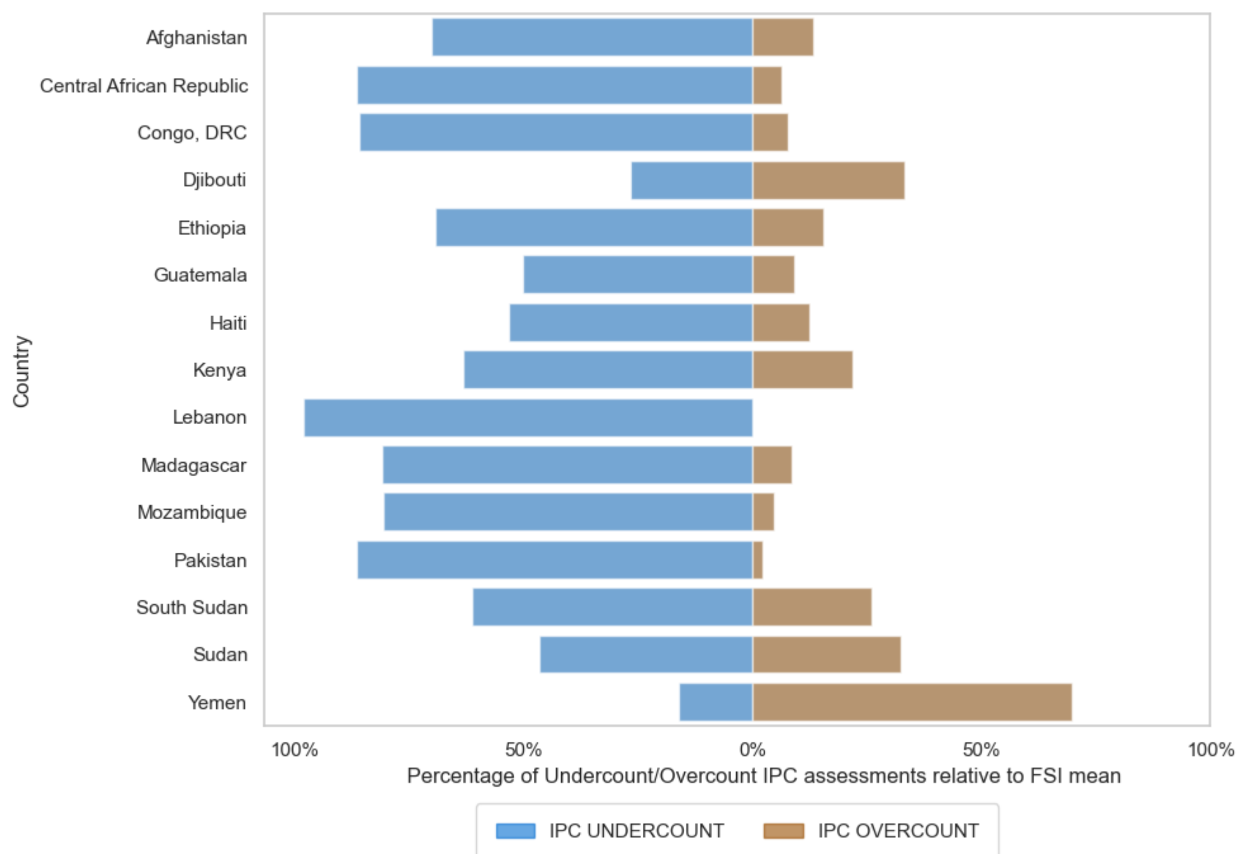


Caption: The 2x2 figure illustrates the cumulative distributions of IPC classifications derived from gaussian kernel density estimates across four unique groups of assessments categorized by the set of available FSIs, comparing consensus-based 3+ population to two hypothetical scenarios: one using the arithmetic mean of FSI 3+ and the other using worst-case FSI 3+ population. The "+" signs indicate significant disparities between the consensus-based CDF and those from alternative scenarios as determined by Barrett and Donald's method for testing for first-order stochastic dominance, suggesting the counterfactual distributions first-order stochastically dominate the consensus-based distribution. These differences are statistically significant at the 1% level (n = 1849).

3) IPC Undercount and Overcount relative to FSI mean

In Figure A5, we illustrate the number of assessments that area reflect overcounting versus undercounting by country. As can be seen, overcounting is more prevalent than undercounting in most countries, with the exception of Djibouti and Yemen. Figure A6 presents the same data by country and round.

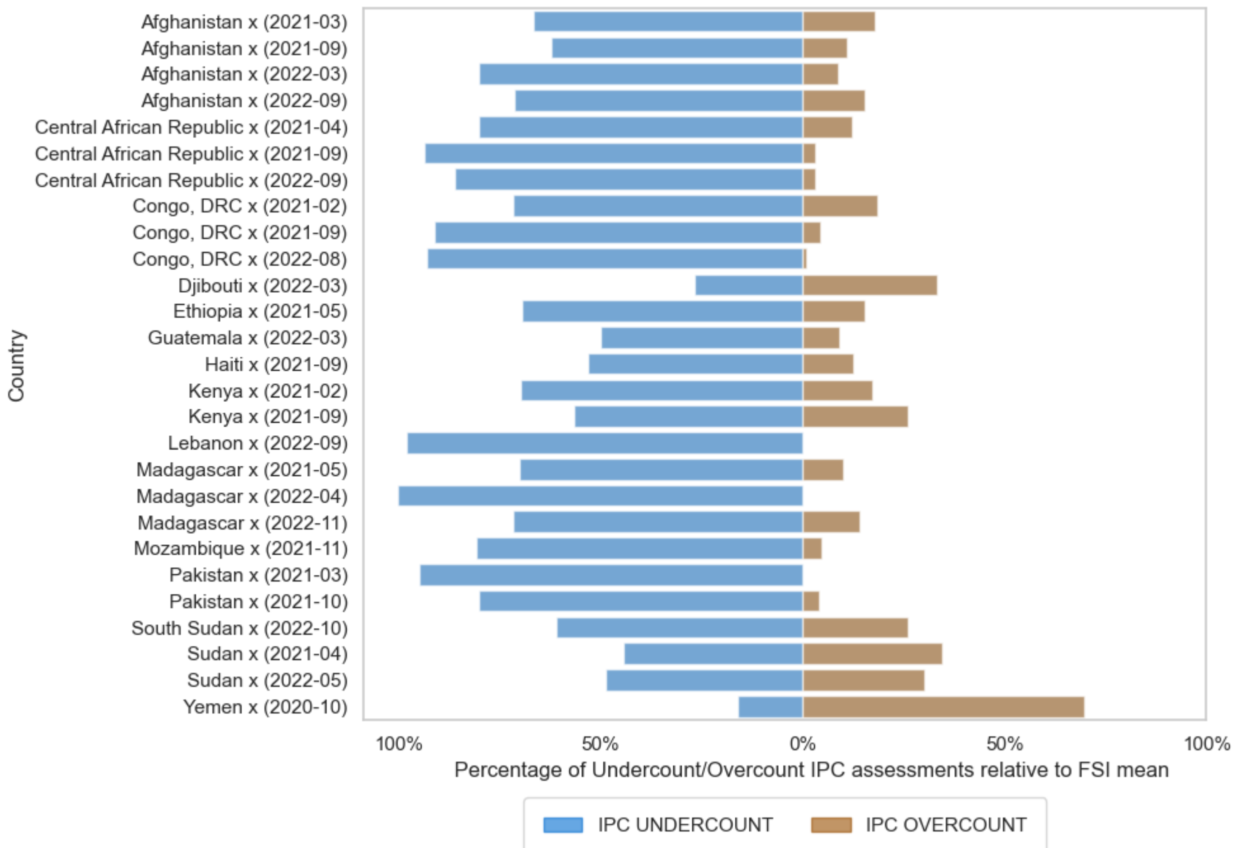
Figure A5. IPC Undercount and Overcount relative to FSI mean by country



Caption: This figure presents a horizontal bar chart comparing the percentage of IPC assessments that are either undercounted or overcounted relative to the FSI mean for Phase 3+ population estimates across different countries and rounds. Each country is listed on the y-axis. Blue bars, indicating IPC undercounts, extend leftward to show the percentage of IPC assessments for consensus-based 3+ population estimates that fall below the FSI 3+ mean. Conversely, the tan bars represent IPC overcounts, stretching rightward to show the percentage of IPC assessments for Phase 3+ estimates that exceed the

FSI mean. The zero percent mark serves as the baseline from which deviations are measured. Note, we exclude one round of assessments from Haiti (2022-09) because it has a unique combination of FSI data.

Figure A6. IPC Undercount and Overcount relative to FSI mean by country-by-round



Caption: This figure presents a horizontal bar chart comparing the percentage of IPC assessments that are either undercounted or overcounted relative to the FSI mean for Phase 3+ population estimates across different countries and rounds. Each country is listed on the y-axis. Blue bars, indicating IPC undercounts, extend leftward to show the percentage of IPC assessments for consensus-based 3+ population estimates that fall below the FSI 3+ mean. Conversely, the tan bars represent IPC overcounts, stretching rightward to show the percentage of IPC assessments for Phase 3+ estimates that exceed the FSI mean. The zero percent mark serves as the baseline from which deviations are measured.

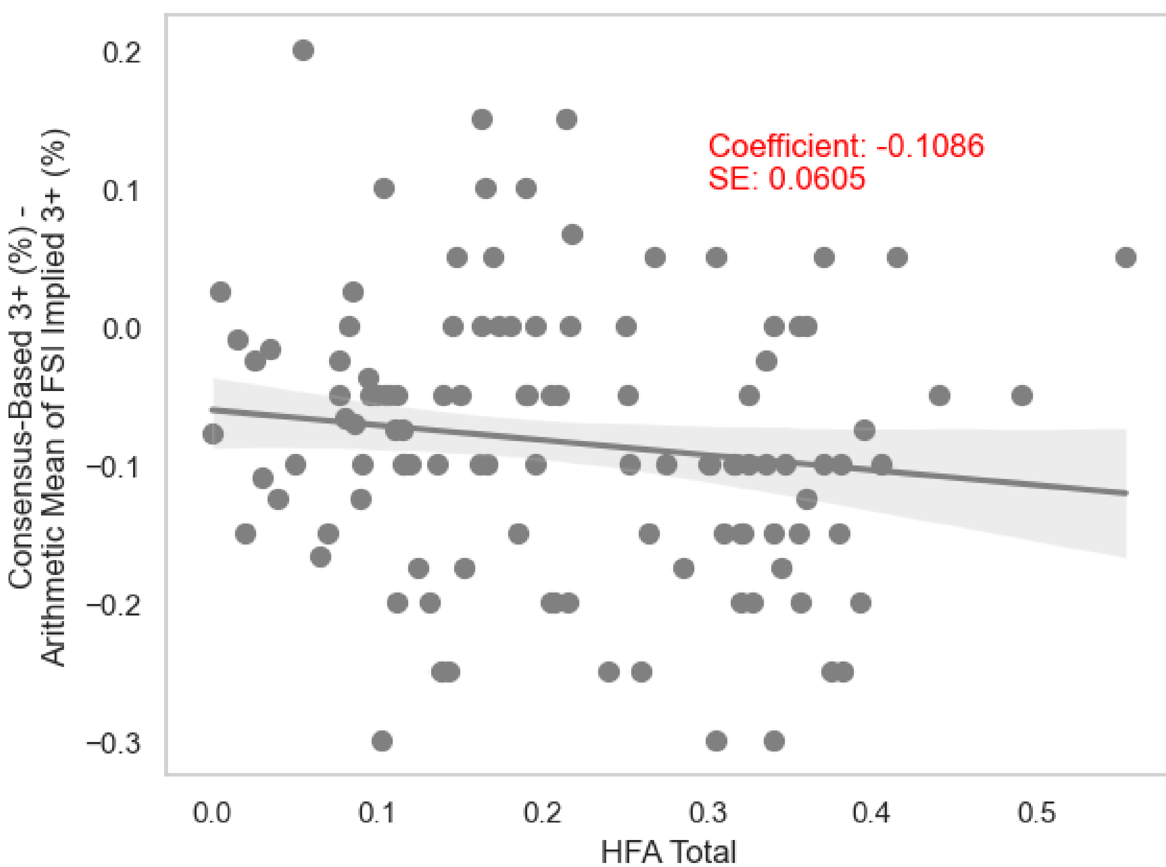
SI.4 Examining possible mechanisms

1) Case study: Humanitarian food assistance in Afghanistan

One possible driver of our results is that the TWG may take future influxes of aid into account, and adjust their assessments downwards in response. We explore this hypothesis for one country with robust data: Afghanistan. We conduct a regression analysis to examine the direction of the discrepancy between consensus-based population estimates and FSI-implied population estimates. We specifically estimate the effect of Humanitarian Food Assistance (HFA) on the value of the discrepancy. Information on HFA is from the Consolidated Data Tool (CDT), a repository where IPC compiles evidence supporting its consensus outcomes.

We plot the deviation between the consensus-based versus the average FSI-based estimate of the percent of population in phase 3+ on HFA (see Figure A7). We find that the presence of HFA is correlated with a slight decrease in the consensus-based estimates, but that this effect is not significantly different from zero at the ten percent level.

Figure A7. Relationship between the HFA and the Difference in Consensus-Based and FSI implied 3+ Population Estimates (%)



Caption: This scatter plot with a regression line illustrates the relationship between HFA (%) and the difference in consensus-based 3+ (%) and the arithmetic mean of FSI implied 3+ population estimates

(%). Each point represents the deviation of the FSI implied 3+ (%) from the consensus-based 3+ (%) for a given HFA Total, highlighting the trend with a red linear regression line (n=180).

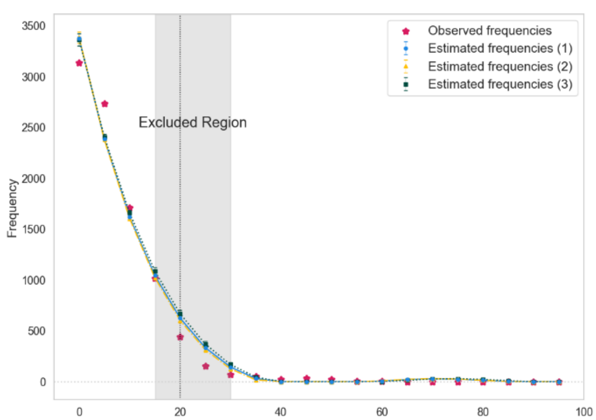
Note: HFA Total (%) is determined by multiplying the percent of the population that received humanitarian aid by the fraction of caloric needs that the humanitarian assistance, particularly food aid, fulfills for that population. For instance, if in certain areas 50 percent of households fulfill 50 percent of their caloric needs through humanitarian food assistance, we would assign a value of 25 percent as the HFA total (%) for those areas.

2) Examining bunching in population estimates of more severe food insecurity (IPC Phase 4+)

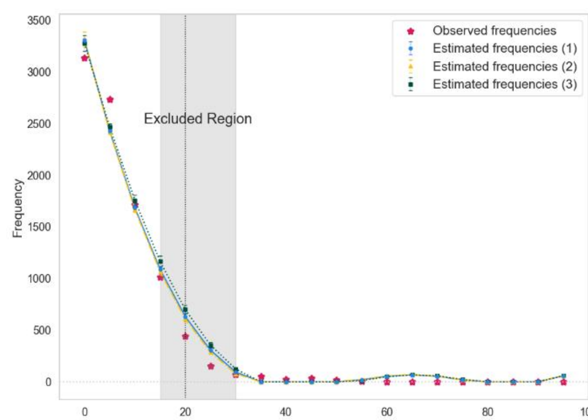
We also test for undercounting at the threshold of phase 4+, illustrated in Figure A8. We did not find indication of undercounting for the category of 4+. While all FSI have FSI-specific thresholds that differentiate between phases 2 and 3, fewer food security indicators have thresholds that can differentiate between phases 3 and 4 and between phases 4 and 5. For example, in the available data, the reduced Coping Strategies Index cannot differentiate between phases 3 and 4. Reducing the number of viable indicators to consider may contribute to decreased bunching. This is worth more investigation.

Figure A8. Observed and estimated distributions of IPC assessments defined by the proportion of population in phase 4+: 4, 5, 6, 7th polynomial counterfactual.

4th

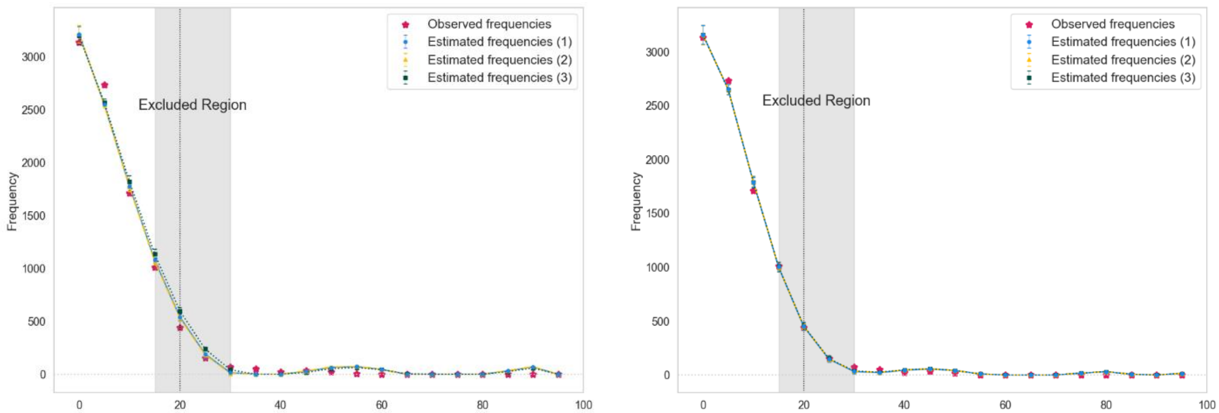


5th



6th

7th



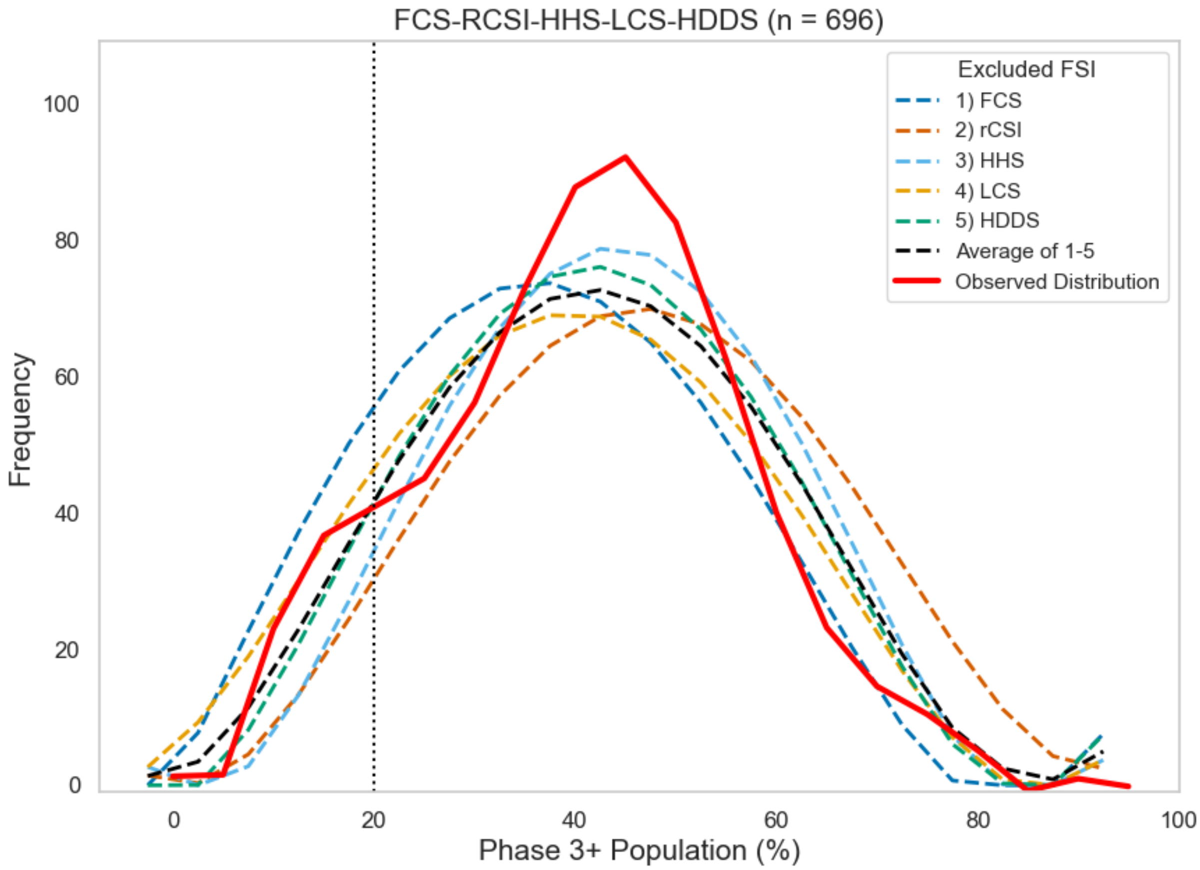
Caption: Above figures present distributional comparisons of 4+ population (%) based on different bin exclusion strategies around the threshold using bootstrapping. We apply alternative polynomial fits, specifically 4^{th}, 5^{th}, 6^{th}, and 7^{th} degree polynomials. We enforce non-negativity constraints to prevent frequency estimates from becoming negative for the entire analysis. For each subplot, the first set of estimated frequencies and their 95% confidence intervals, illustrated in circle markers and blue solid line, is derived by excluding bins in the range [-5%,15%] sequentially, aggregating coefficients across four different scenarios for each bin. The second set of estimated frequencies and 95% confidence intervals, illustrated in triangle markers and yellow dot-dashed line, is generated without excluding any bins. The third set of estimated frequencies and 95% confidence intervals, illustrated in square markers and green dotted line, excludes all bins in the range [-5%, 15%] simultaneously and extrapolates across the entire distribution. Red stars illustrate the observed frequencies of IPC assessments, and a vertical black dotted line indicates 20% population threshold, where a sub-national zone being assessed moves from being classified as phase 3 ("crisis") to being classified as phase 4 ("emergency") (n =9394).

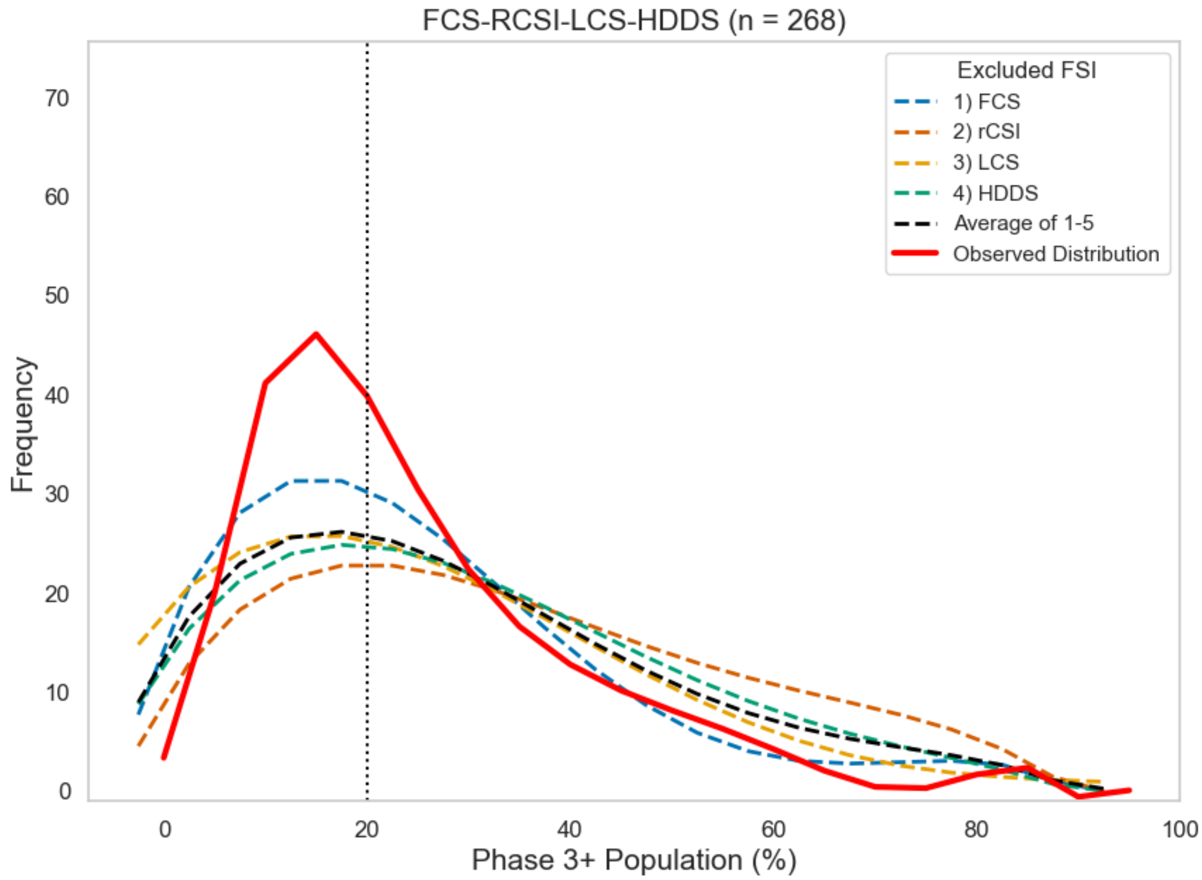
3) Results of sequentially removing one FSI by FSI grouping

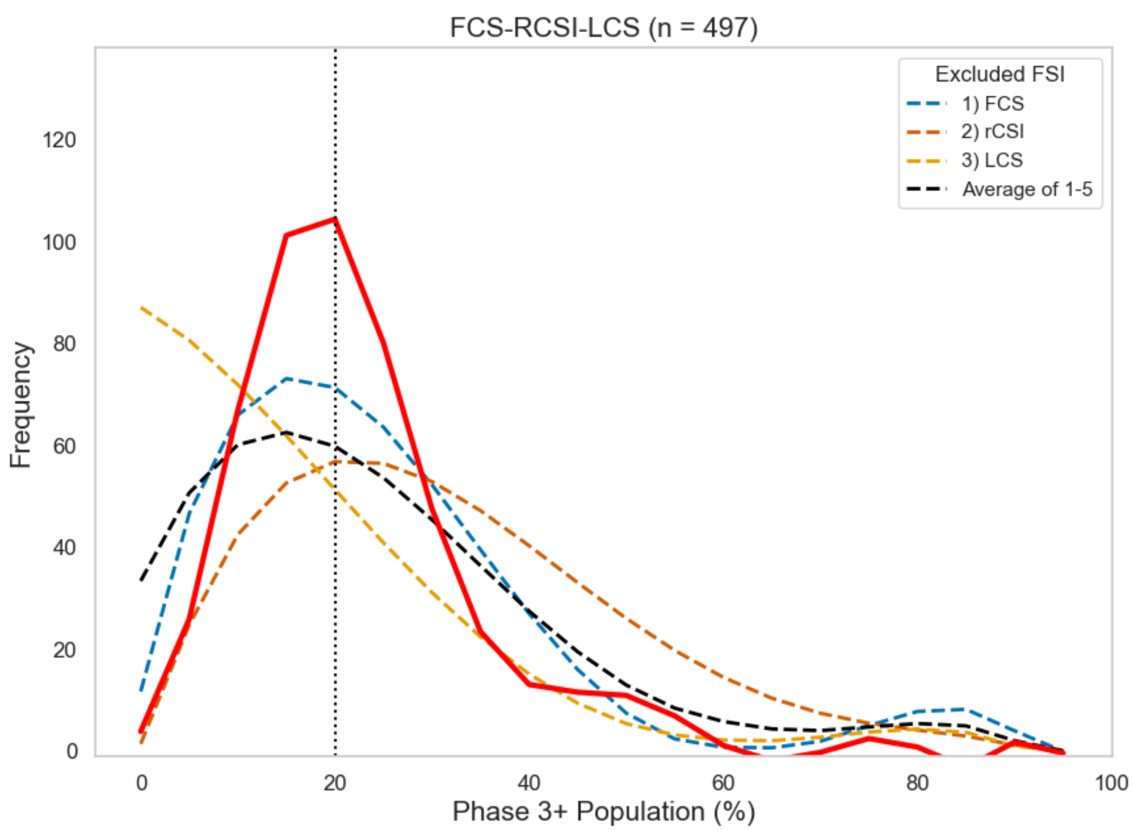
To assess whether any single FSI has an outsize influence on the results, we ran Monte Carlo simulations for four separate groups of indicators using the available FSI dataset. We omit one FSI from the average calculation of FSI-implied population estimates and determine whether doing so would significantly affect our findings. The results presented in Figure A9 suggest that removing any single FSI does not affect our conclusions.

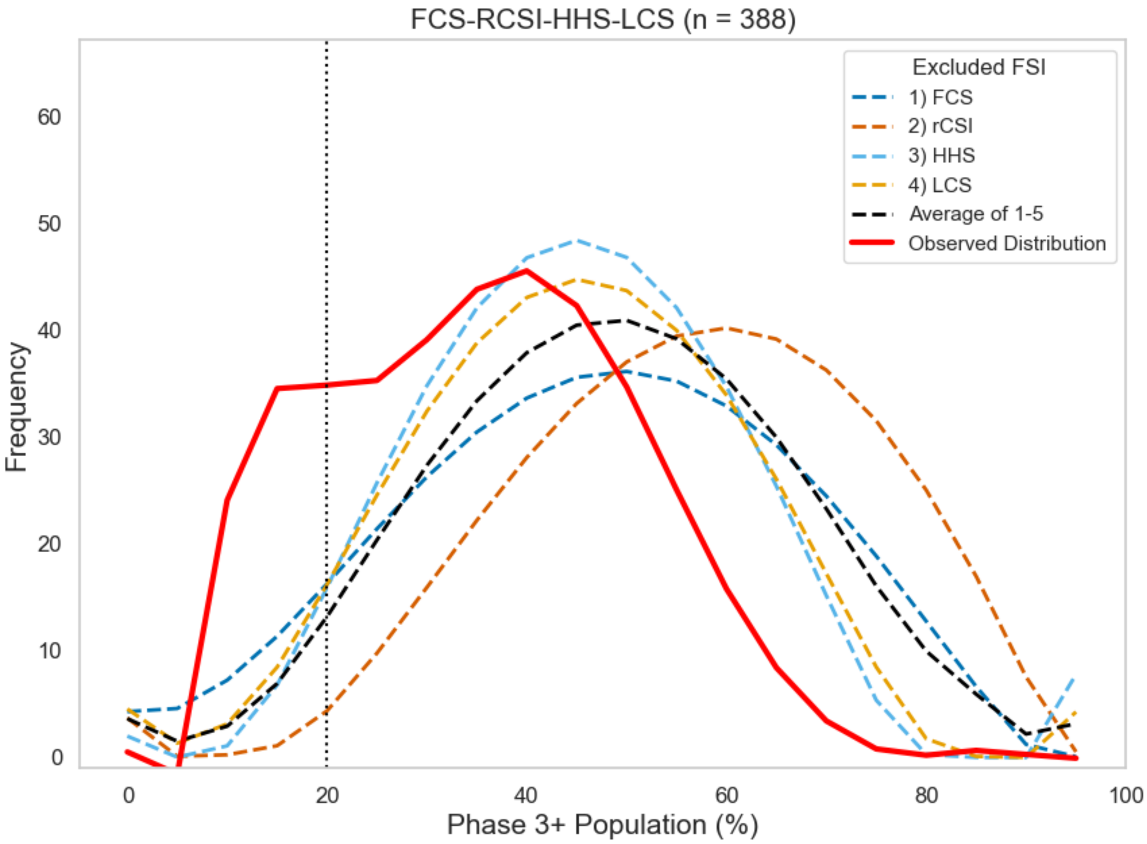
In fact, we observe that removing some FSIs results in a large shift of the distribution further to the right, suggesting much higher percentages of population in phase 3+ than reported. Using the average of the Monte Carlo simulated FSI-implied population (minus one FSI) estimates as counterfactual distributions remains more right-shifted compared to the observed distribution. This outcome is consistent across various FSI exclusion scenarios and among different groups. The one possible exception is removing LCS for the group of countries with only 3 FSIs.

Figure A9. Monte-Carlo simulation results: analyzing the impact of excluding any one FSI on the estimated distribution of IPC assessments defined by the proportion of population in phase 3+ by FSI group




FCS-RCSI-LCS-HDDS (n = 268)
Excluded FSI
1) FCS
2) rCSI
3) LCS
4) HDDS
Average of 1-5
Observed Distribution
Frequency
Phase 3+ Population (%)
0
10
20
30
40
50
60
70
0
20
40
60
80
100


FCS-RCSI-LCS (n = 497)
Excluded FSI
1) FCS
2) rCSI
3) LCS
Average of 1-5
Frequency
Phase 3+ Population (%)
0
20
40
60
80
100
120
0
20
40
60
80
100



Caption: These figures illustrate a series of simulations for four distinct groupings of available indicators in the FSI dataset, accompanied by a 10th-degree polynomial curve that represents the observed distribution of the IPC assessments defined by the proportion of population in phase 3+, marked as a red bold line. The simulations are run by sequentially excluding one FSI at a time, with each exclusion represented by differently colored dashed lines in the legend: blue for FCS, orange for rCSI, sky blue for HHS, vermilion for LCS, yellow for HDDS, and black representing the average of all simulations. These simulations use 4th-degree polynomial fitting to analyze the data at 5% bin intervals. A black vertical dotted line at the 20% mark is a threshold where classifications shift from phase 2 to 3. This visualization allows readers to observe the impact of excluding certain FSIs on the overall distribution of the population in phase 3+. In our simulations, we enforce a non-negativity constraint to prevent frequency estimates from becoming negative (n = 1849).

SI.5 Limitation:

1) Representativeness of FSI available sample (Sample B in Table A1)

Because we only observe FSI data for a subset of our assessments, we investigate the generalizability of our results from the FSI analyses by evaluating whether we observe systematic differences between the samples with and without FSI data (sample B versus sample A from Table A1 in the main text). We employ a logistic regression to explore if and how various country-level characteristics —including GDP, measured in purchasing power parity (7), ln(conflict-related fatalities) (8), democratic measures (9), ln(average population) (10), and the average phase classification at the country level— predict the likelihood of a country having underlying FSI information (1 if a country has FSI information, 0 otherwise). We find that locations are generally similar across these dimensions, with the notable distinction being higher conflict-related fatalities observed in areas with FSI data.

Table A5. Determinants of FSI availability: A logistic regression across 33 countries

	Coef.	Std. Err.	z	$P>\lvert z\rvert$
Constant	-3.48	8.68	-0.40	0.69
ln(Fatalities)	0.77	0.35	2.22	0.03
Democracy index	-0.17	0.36	-0.46	0.64
ln(GDP)	0.04	0.64	0.07	0.95
ln(Population)	-0.25	0.45	-0.55	0.58
Overall phase	1.23	1.11	1.11	0.27

Caption: This presents logistic regression analysis results that examine the influence of various factors on the availability of food security indicator data (1 if FSI available and 0 otherwise). IPC analyses periodically report the 'Tri-National Region' (El Salvador, Guatemala, and Honduras) collectively, which we treat as a 33rd “country.”

SI.6 Robustness Tests: Bunching Analysis

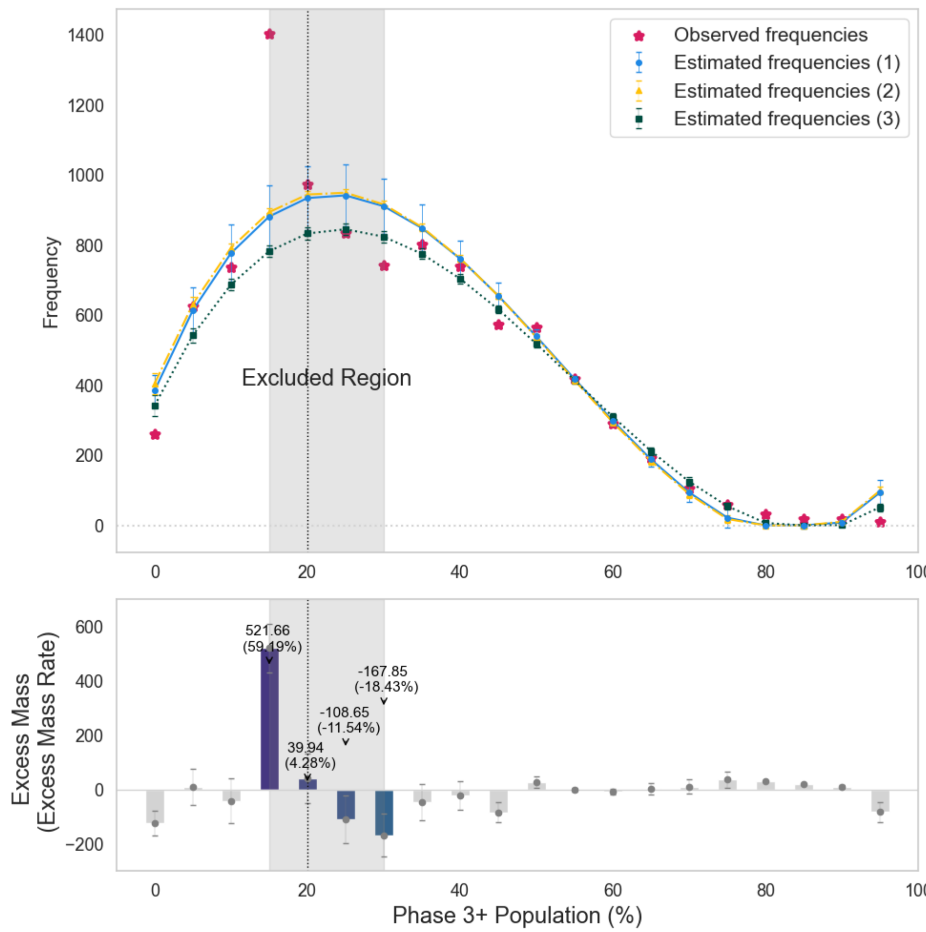
We examine five robustness tests for our bunching analysis, including examining (1) polynomial choice, (2) alternate selection of bins around the threshold, (3) sensitivity to the excluded bin, (4) sample with unrounded population estimates, and (5) subsample analysis.

1) Polynomial choice

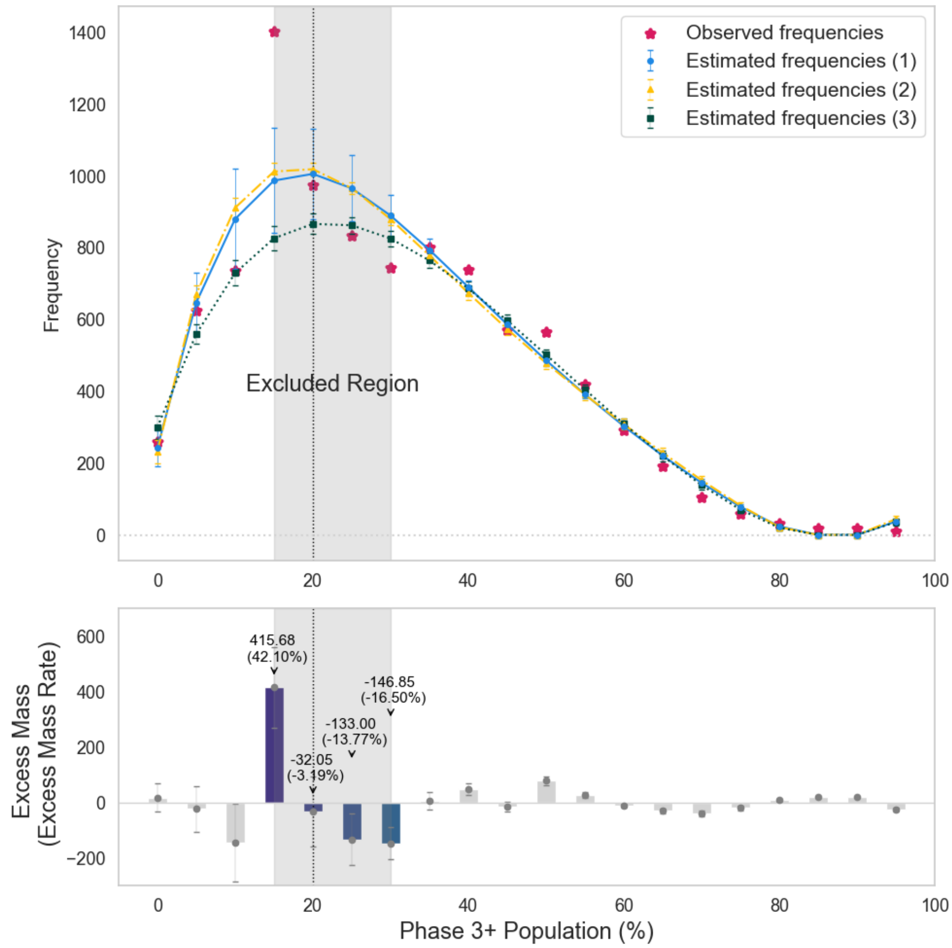
For bunching, we test specifications using multiple polynomial levels to construct alternative counterfactual distributions (Figure A10).

Figure A10. Observed and estimated distributions of IPC assessments defined by proportion of population in phase 3+: 3, 5, 6, 7th polynomial counterfactual distributions

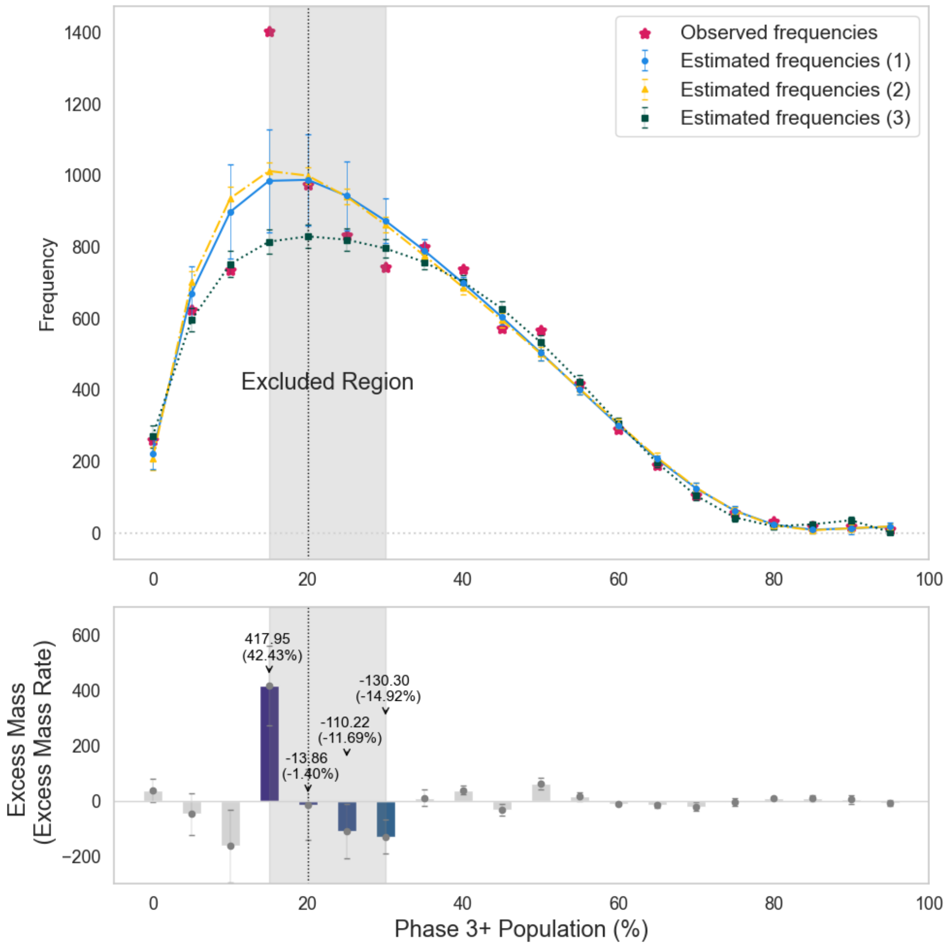
3rd Polynomial



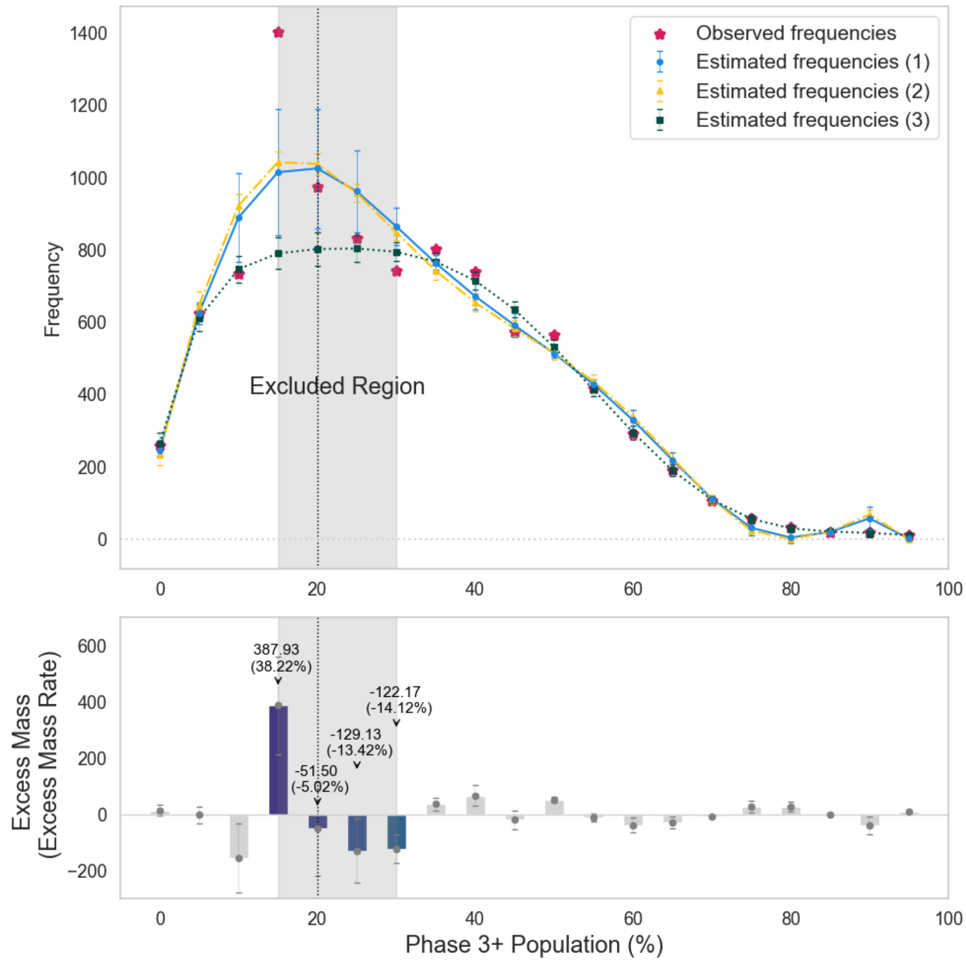
5^{tth} Polynomial



6^{th} Polynomial



7th Polynomial



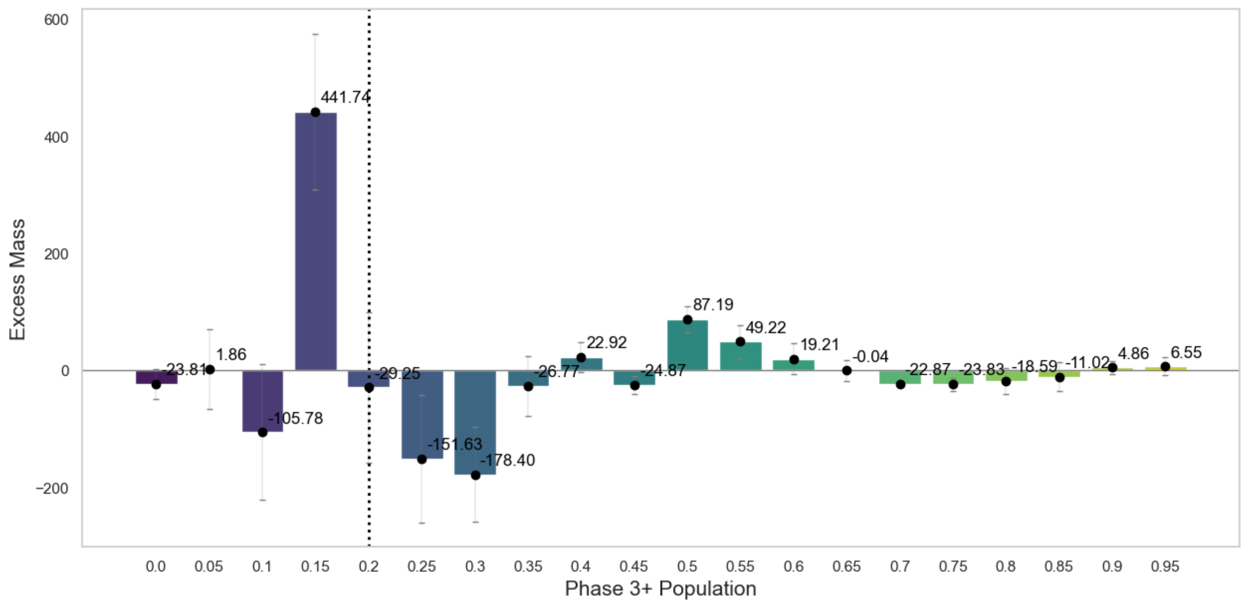
Caption: Above figures present distributional comparisons of 3+ population (%) based on different bin exclusion strategies around the threshold using bootstrapping. Unlike Figure 2, where we used a 4th degree polynomial fit, here we apply alternative polynomial fits, specifically 3rd, 5th, 6th, and 7th degree polynomials. We enforce non-negativity constraints to prevent frequency estimates from becoming negative for the entire analysis. For each subplot, the first set of estimated frequencies and their 95% confidence intervals, illustrated in circle markers and blue solid line, is derived by excluding bins in the range [15%, 30%] sequentially, aggregating coefficients across four different scenarios for each bin. The second set of estimated frequencies and 95% confidence intervals, illustrated in triangle markers and yellow dot-dashed line, is generated without excluding any bins. The third set of estimated frequencies and 95% confidence intervals, illustrated in square markers and green dotted line, excludes all bins in the range [15%, 30%] simultaneously and extrapolates across the entire distribution. Red stars illustrate the observed frequencies of IPC assessments, and a vertical black dotted line indicates 20% population threshold, where a sub-national zone being assessed moves from being classified as phase 2 ("stressed") to being classified as phase 3 ("crisis") (n =9394).

The second panel of each subplot presents the difference between the observed and estimated frequencies at different phase 3+ population bins [15%, 30%], using the approach by Chetty et al. (11) and Allen et al. (12). These differences are generated from estimated frequencies from the upper panel of Figure 2. Values on each bar are the excess mass (i.e., observed - estimated frequency). Values in parentheses represent the excess mass as a percent of the estimated frequency (i.e., (observed - estimated frequency) / estimated frequency).

2) Alternate selection of the bins around the threshold

Second, we consider a wider range of bins around the threshold. The method we use to estimate our counterfactual smooths the estimates around any particular bin close to the threshold, thus it includes increasing the expected number of assessments in the bins on both sides of the 15 % bin. We observe suggestive undercounting just below the 15% bin as well as above the threshold of 20%. This effect can be seen in Figure A11 which illustrates the excess mass across all 5% level bins ranging from 0% to 100%. This undercounting may be an artifact of the mass point at 15% or might stem from the TWG being indifferent between a classification of either 10% or 15%, given that both of these bins result in a classification of Phase 2 (i.e., 'stressed' and not in Phase 3 'crisis').

Figure A11. Differences between observed and estimated frequencies of IPC assessments defined by the proportion of population in phase 3+ by 5% bin

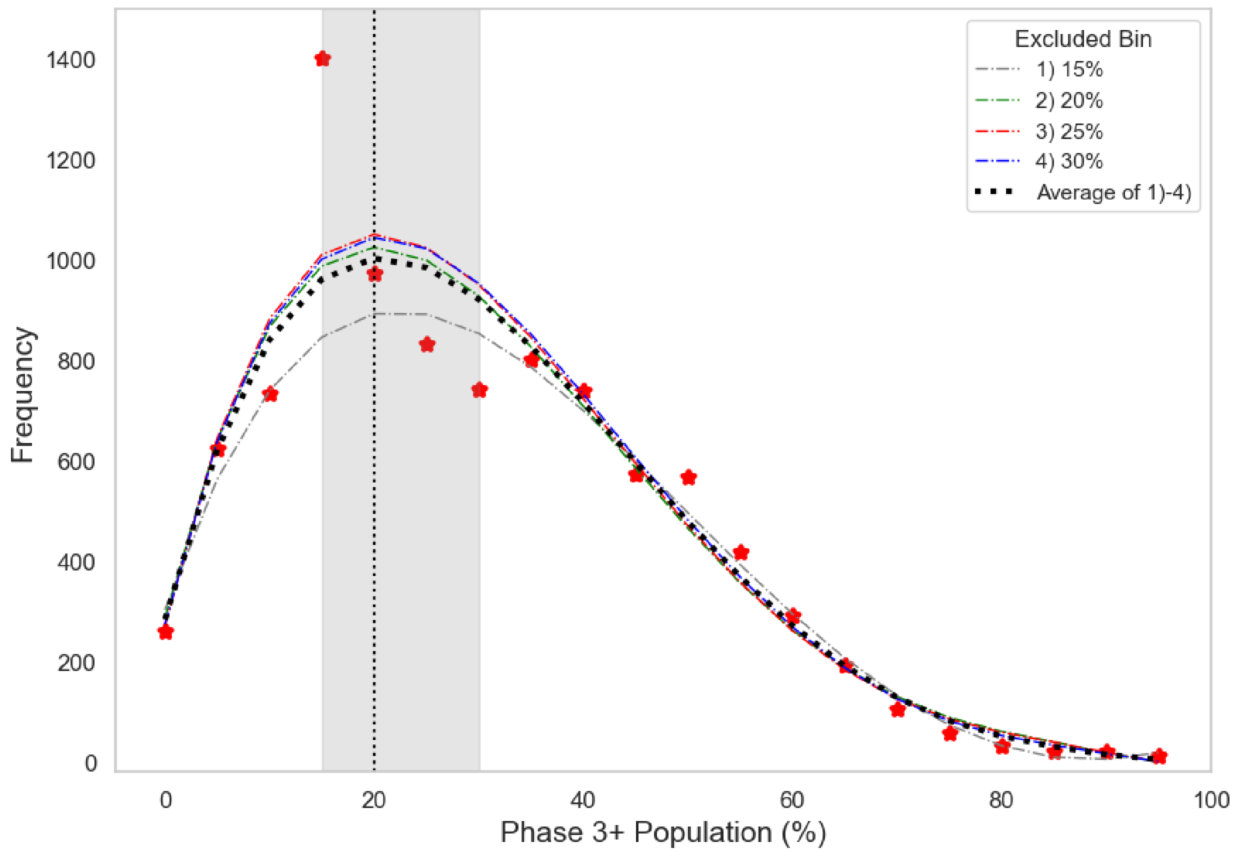


Caption: We present the difference between the observed and estimated number of assessments across all 5% bins. These differences are generated from estimated frequencies (1) from the top panel of Figure 2. Values on each bar are the excess mass (i.e., observed – estimated frequency). Values in parentheses represent the excess mass as a percent of the estimated frequency (i.e., (observed – estimated frequency) / estimated frequency). The data come from our full sample of assessments (n = 9394).

3) Sensitivity to the excluded bin

We conducted multiple simulations to test the robustness of our findings by excluding one bin at a time. Each simulation excluded a specific bin within the range of [15%, 30%] to generate a distinct set of coefficients using a 4th-degree polynomial fit. Our results show consistency: the variance across these simulations is minimal when different bins are excluded. This implies that the bunching is not sensitive to the exclusion of any particular bin in this range except for when 15% bin is excluded.

Figure A12. Bunching analysis results: analyzing the impact of excluded bins



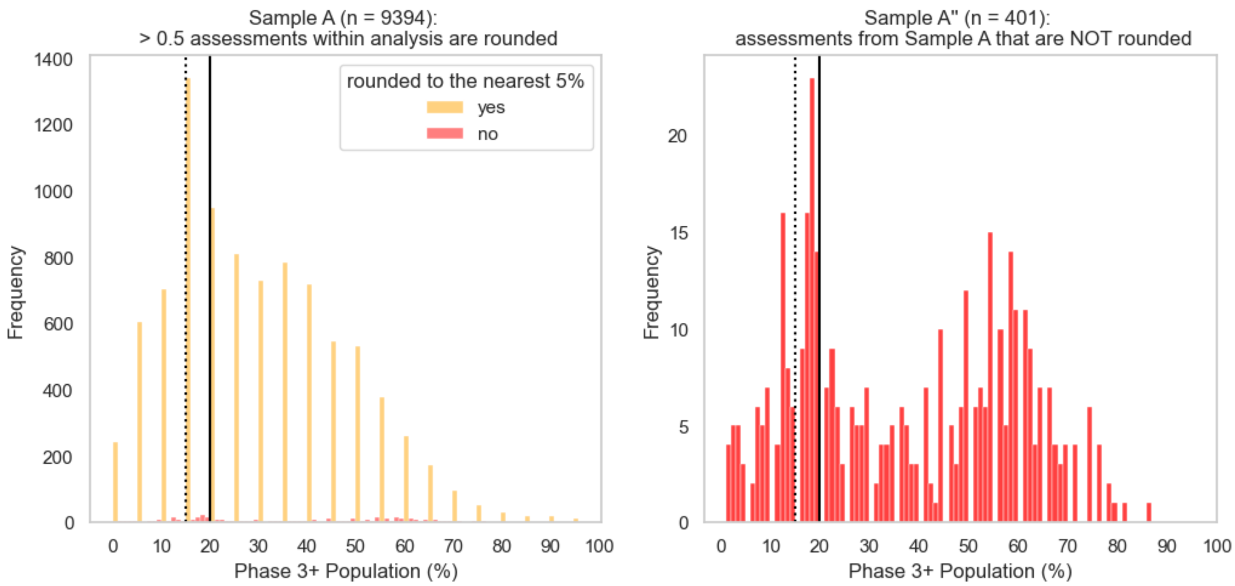
Caption: This figure illustrates the results of a bunching analysis using the proportion of population in phase 3+. Each simulation excludes a specific midpoint percentage (15%, 20%, 25%, 30%) from the analysis. For each estimation, a 4^{th}-degree polynomial fitting is applied to the 5% binned data, and the mean coefficients are calculated. These polynomial lines are represented with colors and patterns to distinguish each excluded midpoint: 15% is gray and solid, 20% is green and dashed, 25% is red and dotted, and 30% is blue and dash-dotted. A black bold dashed line shows the averaged effect across all simulations, and a black dotted vertical line at the 20% mark serves as a reference point.

4) Sample with unrounded population estimates

In the context of bunching analysis, one might be concerned that the presence of rounding in data reporting practices could inadvertently lead to artificial bunching. According to the IPC manual, population estimates (%) can be rounded to the nearest 5%. Using data from the IPC population tracker (accessed on 12/02/2023) comprising 10,890 observations, we refined our sample based on the criterion that a majority (>50%) of each subset is rounded to the nearest 5%. For instance, in Afghanistan's August 2018 data, all 45 area assessments showed 3+ population estimates (%) rounded in the raw data, qualifying them to be in our primary sample (Sample A in Table A1). In contrast, the November 2017 data from Afghanistan, with only 11 out of 23 area assessments meeting this criterion, was excluded from our main sample. This criterion reduced our sample size from 10,890 to 9,394 observations.

The left panel of Figure A13 presents the distribution for the main sample (n=9,394). Notably, there is a substantial mass at the 15% bin, the category immediately below the critical 20% threshold, illustrated by the vertical black dashed line. We analyzed the distribution of the unrounded data and observed significant masses in the 17-19% range, just below the 20% threshold (right panel of Figure A13). This pattern suggests that the conservative reporting behavior around the threshold is not solely a consequence of the rounding process.

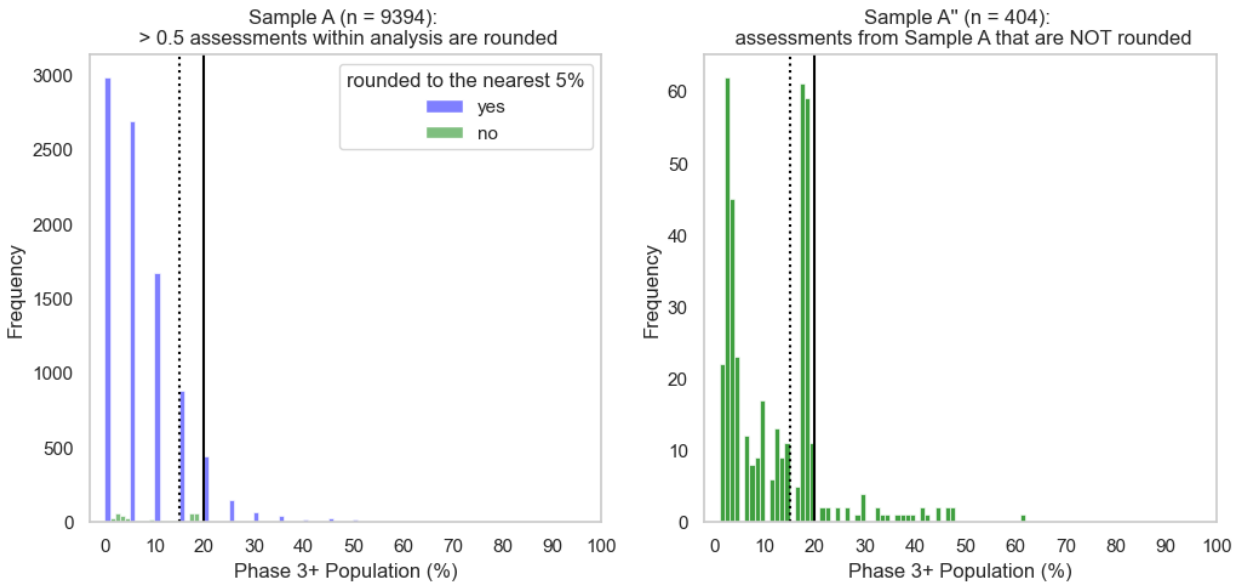
Figure A13. Histograms of rounded and unrounded IPC assessments defined by the proportion of population in phase 3+ (%)



Caption: This figure presents a comparison of two histograms depicting the distribution of the proportion of population in phase 3+. The left panel shows the full dataset (Sample A in Table A1), with data points color-coded to indicate whether they have been rounded to the nearest 5% (yellow for 'yes' and red for 'no'). Both 15% and 20% thresholds are marked by vertical dashed black line and solid black line respectively. The right panel focuses specifically on the subset of Sample A that has not been rounded to the nearest 5%, showing the frequency of these unrounded assessment areas in red.

Notably, when we consider the unrounded data in phase 4+, we do see some evidence of bunching just below the 20% threshold (see the left panel in Figure A13), which might suggest that bunching in the case of phase 4 is being masked by a combination of a smaller sample of observations combined with rounding. However, the unrounded data comprise a much smaller sample than the rounded data, so this evidence is suggestive only.

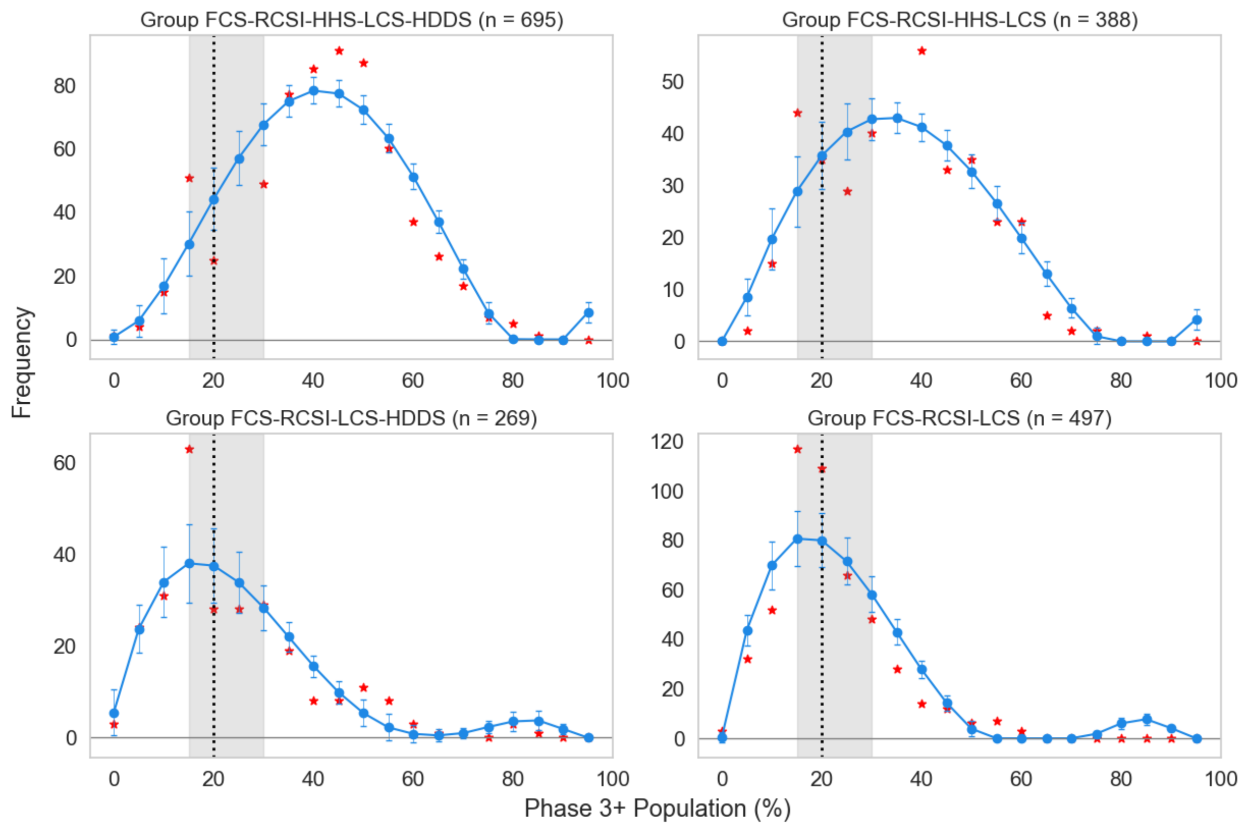
Figure A14. Histograms of rounded and unrounded IPC assessments defined by the proportion of population in phase 4+ (%)



Caption: This figure presents a side-by-side comparison of two histograms depicting the distribution of the observed consensus-based 4+ population estimates (%). The left panel shows the full dataset (Sample A in Table A1), with data points color-coded to indicate whether they have been rounded to the nearest 5% (blue for 'yes' and green for 'no'). Both 15% and 20% thresholds are marked by vertical dashed black line and solid black line respectively. The right panel focuses specifically on the subset of data that are not been rounded to the nearest 5%, showing the frequency distribution of these unrounded assessment areas in green (n = 9394).

5) Subsample analysis: data groups and year specific distributions

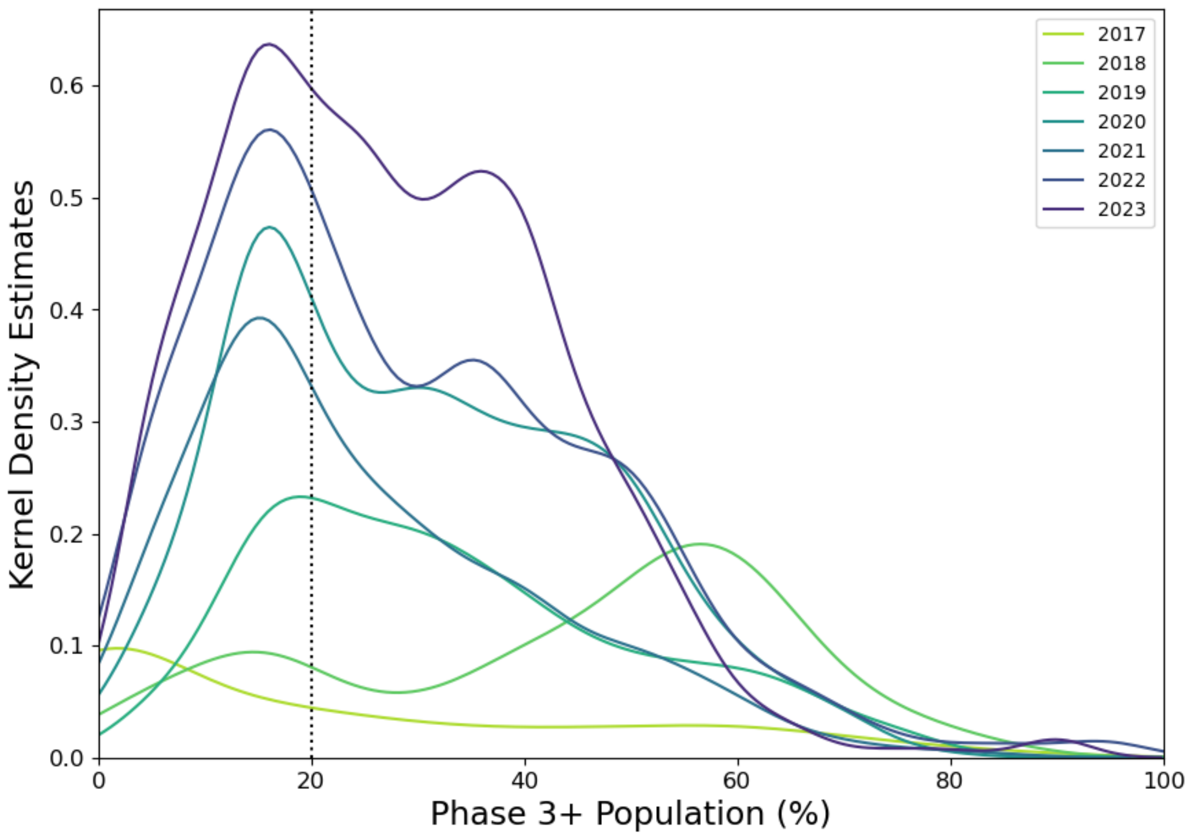
Figure A15. Observed and estimated distributions of IPC assessments defined by the proportion of population in phase 3+by available group of FSIs



Caption: The figure provides a comparative distribution analysis of the population in phase 3+ across four groups with underlying FSI data. Each subplot combines observed frequencies of 3+ population (red stars) and estimated frequencies (4th polynomial fitting by 5% bin) with 95% confidence intervals. We employ the same bin exclusion strategy used to calculate estimated frequencies (1) from Figure 2 in the main text. The data comes from our Sample B in Table A1. Each panel uses a sample based on different combination of FSIs, reflecting the fact that available FSIs tend to vary across assessment areas. Lastly, we enforce a non-negativity constraint to prevent frequency estimates from becoming negative. (n = 1849).

Furthermore, our kernel density estimates of the 3+ population estimates (%) by year, reveal that the largest concentrations of observations are consistently found below the threshold, with the exception of 2017 when our sample is smallest, specifically in the range of [15%, 20%). This observation further substantiates our findings, indicating a widespread tendency for reports to cluster just below the 20% threshold across different time periods and geographical locations. These distributions are illustrated in Figure A16.

Figure A16. Year-specific distribution of IPC assessments defined by the proportion of population in phase 3+ (%)



Caption: The figure illustrates gaussian kernel density estimates for the percent of population assessed by the IPC technical working groups to be in phase 3+ over seven consecutive years, from 2017 to 2023. Each year's distribution is represented by a distinct curve, which provides a visual comparison between the years. A dotted black line represents the 20% threshold that moves an assessment area to be in phase 2 (“stressed”) to phase 3 (“crisis”) (n = 9394).

7. References

1. IPC. n.d. “Population Tracking Tool” Accessed December 2, 2023. https://www.ipcinfo.org/ipc-country-analysis/population-tracking-tool/en/

2. Maxwell, Daniel, Bapu Vaitla, and Jennifer Coates. 2014. “How Do Indicators of Household Food Insecurity Measure up? An Empirical Comparison from Ethiopia.” *Food Policy* 47 (August): 107–16. https://doi.org/10.1016/j.foodpol.2014.04.003.

3. Vaitla, Bapu, Jennifer Coates, Laura Glaeser, Christopher Hillbruner, Preetish Biswal, and Daniel Maxwell. 2017. “The Measurement of Household Food Security: Correlation and Latent Variable Analysis of Alternative Indicators in a Large Multi-Country Dataset.” *Food Policy* 68 (April): 193–205. https://doi.org/10.1016/j.foodpol.2017.02.006.

4. IPC. 2021. “Technical Manual Version 3.1.” Accessed April 13, 2024. https://www.ipcinfo.org/fileadmin/user_upload/ipcinfo/manual/IPC_Technical_Manual_3_Final.pdf.

5. Barrett, Garry F., and Stephen G. Donald. 2003. “Consistent Tests for Stochastic Dominance.” *Econometrica* 71 (1): 71–104. https://doi.org/10.1111/1468-0262.00390

6. Lee, Kyungho, and Yoon-Jae Whang. 2023. “PySDTest: A Python Package for Stochastic Dominance Tests.” https://doi.org/10.48550/ARXIV.2307.10694

7. IMF. 2024. “GDP Based on PPP.” https://www.imf.org/external/datamapper/PPPSH@WEO/OEMDC/ADVEC/WEOWORLD

8. Raleigh, Clionadh, Rew Linke, Håvard Hegre, and Joakim Karlsen. 2010. “Introducing ACLED: An Armed Conflict Location and Event Dataset.” *Journal of Peace Research* 47 (5): 651–60. https://doi.org/10.1177/0022343310378914

9. EIU. 2022. “Economist Intelligence Unit (2023), Economist Intelligence Unit (2021) via Gapminder – Processed by Our World in Data.” https://ourworldindata.org/grapher/democracy-index-eiu?tab=table

10. The World Bank. 2024. “Population.” https://data.worldbank.org/indicator/SP.POP.TOTL

11. Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri. 2011. “Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records.” *The Quarterly Journal of Economics* 126 (2): 749–804. https://doi.org/10.1093/qje/qjr013

12. Allen, Eric J., Patricia M. Dechow, Devin G. Pope, and George Wu. 2017. “Reference-Dependent Preferences: Evidence from Marathon Runners.” *Management Science* 63 (6): 1657–72. https://doi.org/10.1287/mnsc.2015.2417