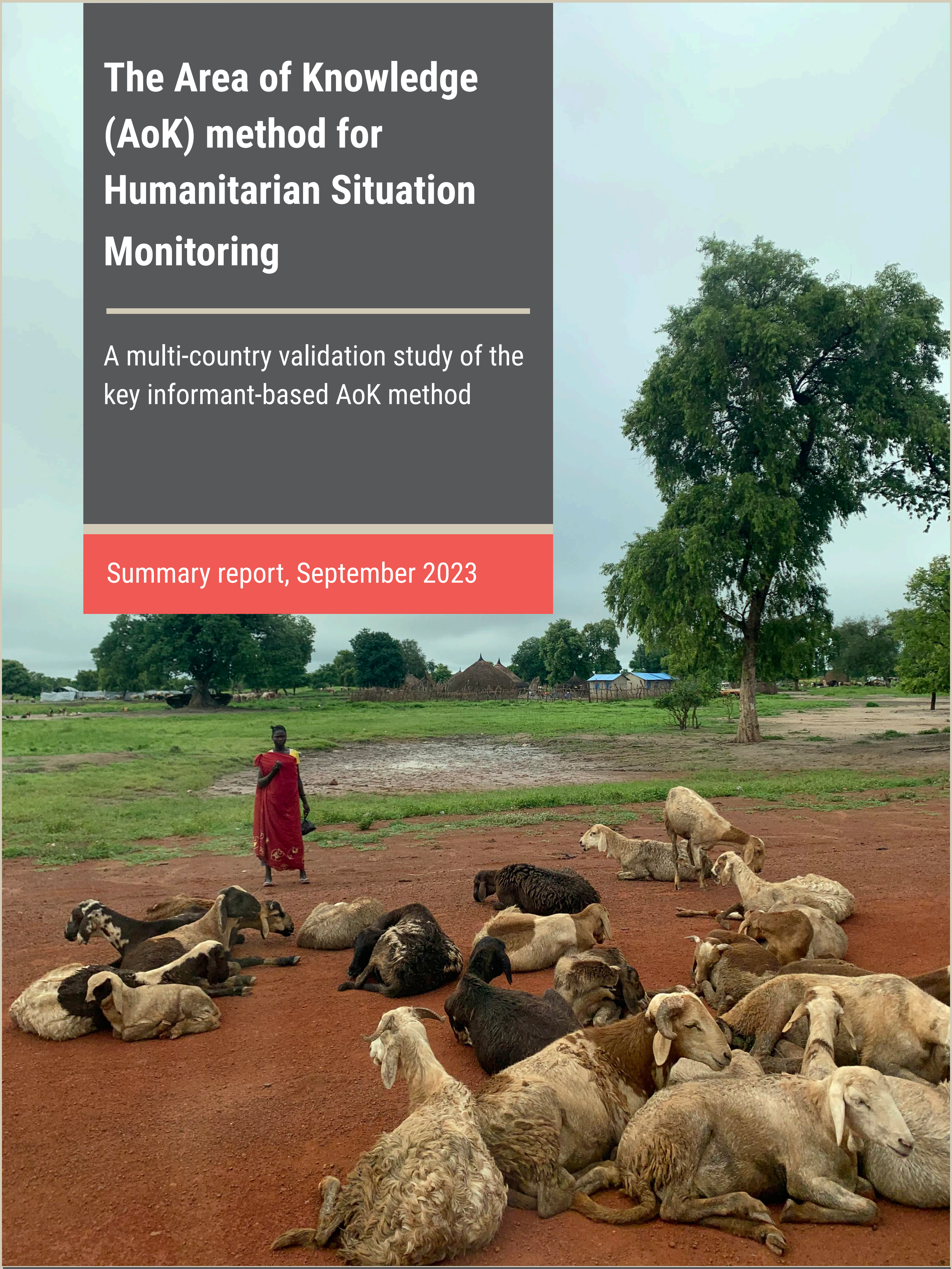
The Area of Knowledge (AoK) method for Humanitarian Situation Monitoring

A multi-country validation study of the key informant-based AoK method

Summary report, September 2023



USAID FROM THE AMERICAN PEOPLE

REACH Informing more effective humanitarian action

For more information about the study, please contact Jeremy Binz (jeremy.binz@impact-initiatives.org) and Joeri Smits (joeri.smits@impact-initiatives.org).

About REACH

REACH facilitates the development of information tools and products that enhance the capacity of aid actors to make evidence-based decisions in emergency, recovery and development contexts. The methodologies used by REACH include primary data collection and in-depth analysis, and all activities are conducted through inter-agency aid coordination mechanisms. REACH is a joint initiative of IMPACT Initiatives, ACTED and the United Nations Institute for Training and Research - Operational Satellite Applications Programme (UNITAR-UNOSAT). For more information please visit our website. You can contact us directly at: geneva@reach-initiative.org and follow us on Twitter @REACH_info.

Abstract

The scale and severity of humanitarian emergencies is increasing with no signs of abating, with the number of people in need at a record high. Further exacerbated by the climate crisis, in 2023, one in every 23 people now needs humanitarian assistance[1], yet funding commitments do not keep up with expanding humanitarian needs[2]. It is imperative therefore to ensure that dwindling humanitarian resources go to crisis-affected populations that are most in need and require life-saving assistance. As access constraints also increase globally, with attacks against aid workers becoming more commonplace[3], humanitarians are more limited in their ability to reach crisis-affected populations. Often, it is these areas and populations that are most in need that are the hardest-to-reach. The humanitarian situation is extremely dynamic, with conditions and needs changing quickly. Although household surveys collected through random sampling and with sufficient coverage to attain representativeness are the gold standard for needs assessments, the conditions on the ground evolve too quickly – especially in these hard-to-reach pockets – to allow for such large-scale, resource-intensive, and costly assessments. Moreover, representative surveys may be impossible in areas of fragile countries that are inaccessible due to conflict or natural disasters, and they don't produce answers quickly enough after a humanitarian disaster. In sum, there is an urgent need for alternative assessment methodologies to (i) capture conditions and humanitarian needs in harder-to-reach and inaccessible areas, and (ii) to monitor rapidly evolving needs where repeated representative surveys would be difficult or too costly to administer.

REACH developed the Area of Knowledge (AoK) approach to fill these information gaps. AoK entails interviewing (face to face or by phone) Key Informants (KIs) about humanitarian conditions in a community (typically a settlement) which they have declared to have recent knowledge of.[4] These KIs can include current residents of the areas of interest or people who have recently visited or have recently been displaced from these areas and are interviewed in their areas of displacement. They can include local leaders, basic service workers, traders, other private citizens, etc. A main limitation of AoK is that because it is not based on random sampling (but rather uses purposive sampling approaches which vary depending on the circumstances), it is considered to produce only "indicative", rather than representative, findings. However, the assumption that AoK is indeed indicative or predictive of the underlying humanitarian conditions it attempts to measure has never been empirically assessed.[5] This begs an important question: to what extent, if any, are AoK findings predictive of humanitarian conditions and needs on the ground? And how does that vary by sector, indicator, and design parameters (e.g., settlement coverage)?

In light of this and of the importance of the AoK methodology as a humanitarian assessment approach, and with the support of the United States Agency for International Development's (USAID) Bureau for

[1] OCHA's Global Humanitarian Overview 2023: https://reliefweb.int/report/world/global-humanitarian-overview-2023-enaresfr

[2] See OCHA's financial tracker: https://fts.unocha.org/home/2023/plans .

[3] https://reliefweb.int/report/world/aid-worker-security-report-2023-figures-glance

[4] Community KIIs form the basis of AoK. Displaced persons reporting on their area of origin (AoO, the original term for the research approach), was innovative when the approach was first deployed by REACH in Syria in 2014. Another distinguishing feature of AoK lies in its quantitative approach, that is, in the aggregation and reporting of KI responses. Quantitative researchers often deploy KIIs qualitatively, for hypothesis generation, and for the interpretation and contextualization of results. Likewise, qualitative researchers often report KII findings qualitatively or descriptively. In contrast, in AoK, KI responses are aggregated at the settlement level and then reported quantitatively at the area level (e.g., % of settlements where most KIs indicated most households in their settlement do not to have access to improved sanitation).

[5] In this report, we use the word predictive of, correlated with, and associated with, interchangeably, and take them to have the same meaning. The statistical predictions (nowcasting) are statistical correlations and do not imply anything regarding causality.

Humanitarian Assistance (BHA), REACH conducted a validation study of AoK to assess whether findings obtained through AoK are actually predictive of conditions on the ground. To achieve this, AoK data was collected in Afghanistan, the Central African Republic (CAR) and South Sudan, and compared to other (benchmark) data sources, including concurrently collected representative household surveys (chiefly REACH's annual Multi-Sector Needs Assessments (MSNAs)[6]). Correlations between AoK and these other sources were measured across sectors and indicators, with the assumption that stronger correlations support AoK's ability to reflect conditions and needs on the ground. The role of settlement coverage (the proportion of sampled settlements in an area) for AoK reliability was also assessed.

Findings indicate that AoK has some merit for comparing humanitarian conditions and needs across areas, but the accuracy varies by sector and indicator, and much additional research & development remains to be done to test the method across different humanitarian settings and to iteratively improve and refine it. This study's validation approach of comparing AoK data against similar indicators in representative household surveys has several limitations, among which (i) differences in question phrasing and indicator construction between the two data sources, and (ii) the comparison source being survey data (rather than a census) introduces additional noise (error) which may have weakened some of the correlations. Notwithstanding these limitations, important insights transpire. **Taken together, the findings demonstrate that AoK can be a useful tool to assess certain conditions and needs for informing emergency prioritization and planning, especially seen in the context of a lack of existing alternatives (for many of the sectors and indicators) for measuring needs remotely and rapidly or at high enough frequency, cost effectively.**

While the results of this first AoK validation study were mixed, with some indicators not correlated, some indicators weakly correlated, and a few indicators moderately correlated with representative survey counterparts and remote sensing data, the findings provide some useful first insights. See below a summary of other key takeaways and implications from this study:

- The extent to which AoK indicators are reflective of humanitarian conditions (as evidenced by their correlations with representative survey data) varies by sector and indicator.
- Correlated indicators: AoK and representative survey indicators were most moderately correlated across study countries for proportions of households with access to improved water and sanitation. Other correlated indicators include security restrictions to movement and a minority of barriers to health and education, among which financial barriers, flooding and child marriage. This provides support to include these indicators in AoK assessments.
- Uncorrelated indicators include the proportion of children regularly attending school, most food security indicators, and the majority of barriers to education and health. REACH country teams are advised to reconsider the inclusion of these indicators, and to take care in their interpretation.
- Those findings are mostly in line with the existing literature on KI reliability, which suggests KI reports are more reliable for salient or publicly observable phenomena and matters that require little inference. This should inform the design of future AoK assessments.
- Most of the food security indicators are individually uncorrelated or only weakly correlated with their representative survey counterparts. However, combining food security and WASH AoK indicators may provide a more accurate picture, as they are jointly predictive of (correlate with) IPC Acute Food Insecurity (AFI) and IPC Acute Malnutrition (AMN) area classifications in South Sudan. But that needs replication out-of-sample, that is, testing the model on new data, such as a new round of IPC in South Sudan or other countries.
- The Household Hunger Score (HHS), the only indicator for which KIs were asked about their own household, is the single strongest predictor of IPC AMN and IPC AFI area classifications. This

[6] MSNA is a large-scale (in many countries nationwide), statistically representative household survey-based assessment of humanitarian needs and conditions, annually conducted by REACH and/or partners.

suggests including this question in AoK assessments that are not targeted to individuals in hierarchical positions (e.g., village leaders), as they may be more food secure than average in their settlement.

- Settlement coverage matters: increasing the proportion of settlements sampled in an area from a low of 5% to a high of 20% modestly increases the number of correlated indicators. The REACH guidelines have been updated accordingly to target at least 10% of settlements.
- This study relied on noisy comparison points (chiefly representative survey data) which may have weakened correlations with AoK, and more research is needed to validate and refine the method, test it against alternative (remote) methods, and assess important tradeoffs faced in the design of AoK-based assessments. Moreover, the current study was cross-sectional in nature. However, in humanitarian situation monitoring, AoK is used to assess changes in needs over time. Hence, longitudinal validation analyses is an avenue for future further research.

SUMMARY

The Area of Knowledge (AoK) methodology is a community key informant (KI) data collection and analysis approach developed by REACH to assess humanitarian conditions in cases where statistically representative data collection methods (such as randomly sampled household surveys) cannot be implemented. This can be the case either because the areas of interest are hard-to-reach due to security or logistical constraints, or because a lack of sufficient resources or time limits the feasibility or frequency of representative assessments.

The AoK methodology entails interviewing KIs about the humanitarian situation in an area (typically a settlement) which they have declared they have recent knowledge of, and aggregating results at community-level and then area-level (e.g. admin 2 area) to present findings. These KIs can include current residents of the areas of interest or people who have recently visited or have recently been displaced from these areas (including local leaders; basic service workers; aid workers; traders; other private citizens; etc.). The nature of AoK allows REACH to collect data in hard-to-reach areas where insecurity or logistical constraints preclude on-site data collection, as well as in accessible areas where resource constraints limit the possibility or frequency of representative assessments.

AoK is used by REACH for Humanitarian Situation Monitoring (HSM) which aims to provide timely and regularly updated information on humanitarian needs and conditions in crisis-affected areas, with the aim of informing emergency humanitarian planning, and at a higher frequency than large-scale annual representative assessments such as Multi-Sectoral Needs Assessments (MSNAs).[7]

AoK has also been used by REACH and subsequently other humanitarian actors for rapid assessments, for example following sudden-onset crises (such as earthquakes, extreme weather events, etc.). AoK is the methodology used by the United Nations Disaster Assessment and Coordination (UNDAC) team for conducting Multisectoral Initial Rapid Assessment (MIRA) following sudden-onset crises, based on a tool developed in collaboration with REACH. For instance, an AoK-based MIRA led by UNDAC with the support of REACH was jointly conducted with over 30 humanitarian organizations and UN agencies in response to the earthquakes in Türkiye in February 2023. As the first assessment of the crisis (beyond initial secondary data reviews), this MIRA was the main source of information for initial humanitarian planning in response to the earthquake.

AoK was initially developed by REACH for its HSM programming in Syria in 2013, where a highly dynamic context and accessibility issues due to the ongoing high-intensity conflict created information gaps which impeded humanitarian planning. It aimed to "inform aid planning and enhance the understanding of humanitarian context within Syria; improve humanitarian access to vulnerable groups; and indirectly monitor the direct outcomes and impacts of large-scale humanitarian interventions."[8] This original HSM RC focused on hard-to-reach areas exclusively, where direct data collection was not possible. Data was collected face-to-face in formal camps in Jordan and Iraq and in host community settings in Lebanon, exclusively through the purpose-developed Area of Knowledge (AoK) methodology. AoK consisted of Key Informant (KI) Interviews (KIIs) conducted with people having recently left these hard-to-reach areas of interest and arrived in the data collection area, and who reported indirectly on the sub-district (and later the community – i.e., village or neighborhood) they had left. Results were reported at sub-district (admin 2) level.

[7] By 2023, HSM research cycles have become core components of REACH's work in 13 countries: Afghanistan; Burkina Faso; the Central African Republic (CAR); Colombia (Venezuela); the Democratic Republic of Congo (DRC); Ethiopia; Mali; Niger; Nigeria; Somalia; South Sudan; Syria; and Ukraine.

[8] REACH, Assessment of needs and humanitarian situation in Syria, Terms of Reference, Version 5, May 2015

Since its origins, AoK has evolved to fit a wide variety of contexts and objectives, as a tool to plug information gaps in dynamic contexts where access and/or resource constraints limit the possibility or frequency of more granular assessments (such as representative household surveys), in hard-to-reach but also fully or partially accessible areas.

Validating key assumptions in AoK research design.

Since its inception, AoK has been built and designed around standard practices for KI data collection in humanitarian response, which undergird a key set of assumptions:

1. Assumption 1: AoK data represents the humanitarian situation and needs in settlements of interest with sufficient accuracy (i.e., KIs are reliable sources of information to understand the humanitarian situation and needs at the settlement level).

2. Assumption 2: Aggregated AoK data represents the humanitarian situation and needs in areas of interest with sufficient accuracy. Thus, comparing data across geographic regions, administrative zones, or otherwise between spatial areas produces a valid comparison of underlying need and represents real differences in the humanitarian situation. Comparing HSM data aggregated at the area-level across time also produces a valid comparison of needs and represents real evolutions.

The main limitation of AoK is that because it is not based on random sampling (but rather uses purposive sampling approaches, which vary depending on the circumstances), it is considered to produce only indicative – rather than representative – data. However, the assumption that AoK is indeed indicative of the underlying humanitarian conditions it attempts to measure has never been validated. Anecdotal evidence gathered from REACH field teams suggests that AoK has at times been supported by triangulation with secondary data sources, but this has not been systematically investigated to produce more robust evidence of the validity of AoK. This begs an important question: to what extent, if any, are AoK findings predictive of humanitarian conditions and needs on the ground? And how does that vary by sector, by indicator, and by design parameters (e.g., settlement coverage)?

In light of this and of the importance of the AoK methodology as a humanitarian assessment approach, and with the support of the United States Agency for International Development's (USAID) Bureau for Humanitarian Assistance (BHA), REACH conducted a validation study of AoK to assess whether findings obtained through AoK can predict humanitarian conditions on the ground.

For the purposes of validating the AoK methodology, AoK data was collected in Afghanistan, the Central African Republic (CAR) and South Sudan, and compared to other (benchmark) data sources, including concurrently collected representative household surveys (chiefly MSNAs) and remote sensing indicators. Correlations between AoK and these other sources were measured across sectors and indicators, with the assumption that stronger correlations indicate a stronger ability of AoK to point to the same conclusions as these other sources in assessing humanitarian conditions on the ground. Further, this study sought to test (aspects of) the sampling approach, by assessing how the predictive ability of AoK indicators may vary depending on the coverage of sampled settlements within assessed areas.

Due to the larger sample size in terms of administrative areas (see Table 1), more results were obtained from South Sudan. In most sampled counties of South Sudan, and all of Afghanistan and CAR sample, Key Informant Interviews (KIIs) were conducted face-to-face. In Afghanistan, settlements for AoK were sampled randomly (population proportional to size); in South Sudan, individuals were interviewed at central towns and markets in each county; and in CAR, the MSNA and AoK data collection was conducted by the same survey team and an alternating strategy was applied: the first KI was to be sampled from the settlement itself, the second from another settlement in the area, the third from the settlement itself, etc.

Table 1: Sample coverage across study countries, and mode of KI interview for the AoK survey.

	AoK data			Representative household survey			# of admin areas that overlap between AoK and HH survey
	Coverage: # of admin areas	# of KIIs	Mode of interview	Survey	Coverage: # of admin areas	Sample size (# HHs)	
Afghanistan	34 provinces x urbanity = 68	10,250	In-person	MSNA	34 provinces x urbanity = 68*	20,696	68
Central African Republic	21 admin2s	683	In-person	MSNA	66 admin2s	12,328	21
South Sudan	74 admin2s	5822	In-person / phone	ISNA, FSNMS	74 (ISNA) / 77 (FSNMS) admin2s	ISNA: 13,961; FSNMS**: 8,532	71 (ISNA) / 72 (FSNMS)

* Afghanistan is the only country where the MSNA is representative at the province/admin 1 level by urbanity (urban/rural) only (and not at the admin 2/district level).
** Rough (not exact) sample size, with 108 households sampled from 79 counties.

This study sought to address the following research questions and hypotheses:

<u>Overall predictiveness of AoK indicators with respect to representative household survey data</u>. If the joint (or global) null hypothesis that across sectors and indicators, each representative survey indicator is uncorrelated with its corresponding AoK proxy, cannot be rejected, then that would mean AoK does no better than randomly guessing the representative survey data. This global null hypothesis is thus the first bar to pass in validating the AoK methodology.

<u>Reliability of individual AoK indicators with respect to the corresponding household survey indicators</u>. KI reliability will vary across concepts and indicators, as does the extent to which the condition or need varies within an area. Hence, correlations between AoK indicators and household survey counterparts are analyzed for each (pair) of indicator(s).

<u>Correlations between AoK indicators and IPC Acute Food Insecurity (AFI) and IPC Acute Malnutrition (AMN) area classifications</u>. AoK indicators are often used jointly to analyze the severity of needs, and the IPC AFI and IPC AMN area classifications likewise use a convergence of evidence approach. This begs the question whether AoK proxies can jointly predict IPC AFI and IPC AMN area classifications.

<u>Differences in the reliability of AoK indicators across sectors</u>. Our literature review indicated that Key Informants tend to be more reliable when reporting on salient (rather than routine) events and on publicly observable phenomena. This may suggest, for instance, that KIs may report more accurately on, say, the prevailing sanitation type in a settlement or on child marriage than on indicators such as the consumption frequency of food categories or certain barriers to education or healthcare.

<u>Settlement coverage as a moderating factor of the quality of AoK-derived results at the admin2 level.</u> Intuitively, the higher the proportion of settlements in an area from which KIs are sampled, the more reliable the aggregated findings will tend to be, but this adds to survey cost.

<u>AoK's predictive ability across geographies (e.g., urban vs. rural)</u>. There is a lack of evidence on where AoK may perform better.

Predictiveness of Household Hunger Score (HHS) administered on the KI's own household as proxy indicator for area-level HHS. This one question was included in the AoK questionnaire that asks about the KI's own household. How does its predictive power compare to traditional KI questions about the settlement?

AoK indicators' correlation with administrative data (e.g., travel time to nearest health facility) and remote sensing-based environmental indicators at the settlement level. These comparison data sources are available at the settlement level, thus providing a larger sample for validation.

Several limitations to the study can be noted. Whereas the study was mostly conducted at the level at which AoK data is typically represented (namely, the area level), this also means that it is hard to identify what explains a lack of correlation between a representative survey indicator and its corresponding AoK proxy. First, the data sources used as benchmarks are themselves imperfect (and noisy) measures of real conditions on the ground. These sources are nevertheless used because of the costs of producing a "ground truth" (i.e., through a census) as a comparator were prohibitive in the framework of this study. The sampling error in the MSNA indicators dilutes (and hence weakens) the correlations with their corresponding AoK indicators. Second, phrasing of questions (sometimes inevitably) differs between AoK and MSNA, which again, may partly explain a lack of (stronger) correlation for some indicators. Third, a lack of correlation may also be due to a combination of heterogeneity of needs within an area and a settlement coverage that is too low (although in South Sudan, a higher than usual settlement coverage was attained) and/or sample selection bias in AoK (given the non-random sampling of settlements for AoK), and/or issues in the aggregation of KI responses both within settlements, and across settlements to the area level.

What follows is a synopsis of the research findings by research question.

Overall predictiveness of AoK indicators with respect to representative household survey data

- Findings. The global null hypothesis that none of the AoK indicators are predictive of the corresponding representative household survey indicators at the area level, could be confidently rejected in all three countries analyzed (AFG, CAR, SSD), as indicated by joint tests across indicators and sectors.
- Implications. There was robust evidence across three countries studied that AoK has the potential to be indicative of comparative needs across areas in a country, and that it can be a useful decision tool for humanitarian prioritization.
- Caveats. The accuracy of AoK in terms of indicating needs depends on several factors (e.g., sampling, indicators), as outlined in other parts of this report, and continued research is needed to keep improving the AoK methodology. Furthermore, the current analysis was cross-sectional in nature, that is, it analyzed correlations between AoK other data sources at a given point in time, and the ability of AoK to pick up changes over time in humanitarian conditions and needs is a topic for further research.

Reliability of individual AoK indicators with respect to the corresponding household survey indicators

- Findings. Some, but not all, AoK proxy indicators are predictive of corresponding representative HH survey-based indicators at the area-level. This provides partial but robust evidence for the validity of the AoK methodology in providing indicative measures of how some humanitarian conditions and needs vary across areas within a country. For indicators where AoK and the representative survey data are correlated, AoK indicators seem more able to rank areas in terms of need than reflect the absolute level of need (Fig. 2).

Indicators that correlate with representative survey data at the area level include livelihood coping strategies involving cattle (Fig. 1), proportions of HHs having access to improved water, sanitation or shelter, several of the indicators on barriers to accessing health care (including financial barriers), barriers to education (including flooding, child marriage), and safety restrictions on movement (Fig. 3).

Indicators lacking area-level correlation with survey data include the share of children regularly attending school, the Food Consumption Score, and some of the barriers to education and health (Fig. 1 and 3). KIs were also not able to estimate the population size of their settlement (about half the KIIs failed to even produce an estimate).

- Implications. The findings suggest that among the indicators suitable for inclusion in AoK surveys are those capturing publicly visible phenomena, e.g., questions related to water source, and types of sanitation and shelter, as those AoK proxies have been shown to correlate with the corresponding measures based on representative survey data. In the food security and livelihoods (FSL) domain, the household hunger scale (HHS) administered to the KIs own household is one of the more predictive AoK indicators. Given the varying AoK-survey data correlations across indicators in the Food Security and Livelihoods (FSL) domain (e.g., across Livelihood Coping Strategy Index (LCSI) indicators), it is suggested to analyze AoK indicators jointly, rather than treating indicators in isolation. Asking KIs about settlement population size is not recommended, at least in countries with similarly low levels of numeracy as CAR and South Sudan. Finally, seeking opportunities to align AoK with household surveys and to identify comparable secondary data sources to compare AoK data with (and hence "validate" AoK) wherever possible to build the evidence base, is suggested.
- Caveats. Due to infeasability of direct observation of the ground truth (e.g., a census), household surveys were relied on for AoK validation. These household surveys contain sampling error and measurement error. The noise (errors) in the comparator (the household survey indicators AoK indicators are compared against) may have inflated (weakened) correlations between AoK indicators and the corresponding household survey indicators. More broadly, a lack of correlation for some indicators could be explained by several factors, including the nature of the question that may not lend itself well to the AoK methodology, differences in phrasing of questions between AoK and representative HH survey, and heterogeneity within administrative areas coupled with non-random sampling of settlements in AoK. A key outstanding question is how to interpret the degree of correlation between corresponding AoK and HH indicators, given that many of the indicators are not directly comparable, and (often unavoidably) measured in different ways and/or on different scales. Finally, whereas the current study was cross-sectional in nature, longitudinal analyses are needed to reveal whether AoK is reflective of changes in conditions and humanitarian needs over time.

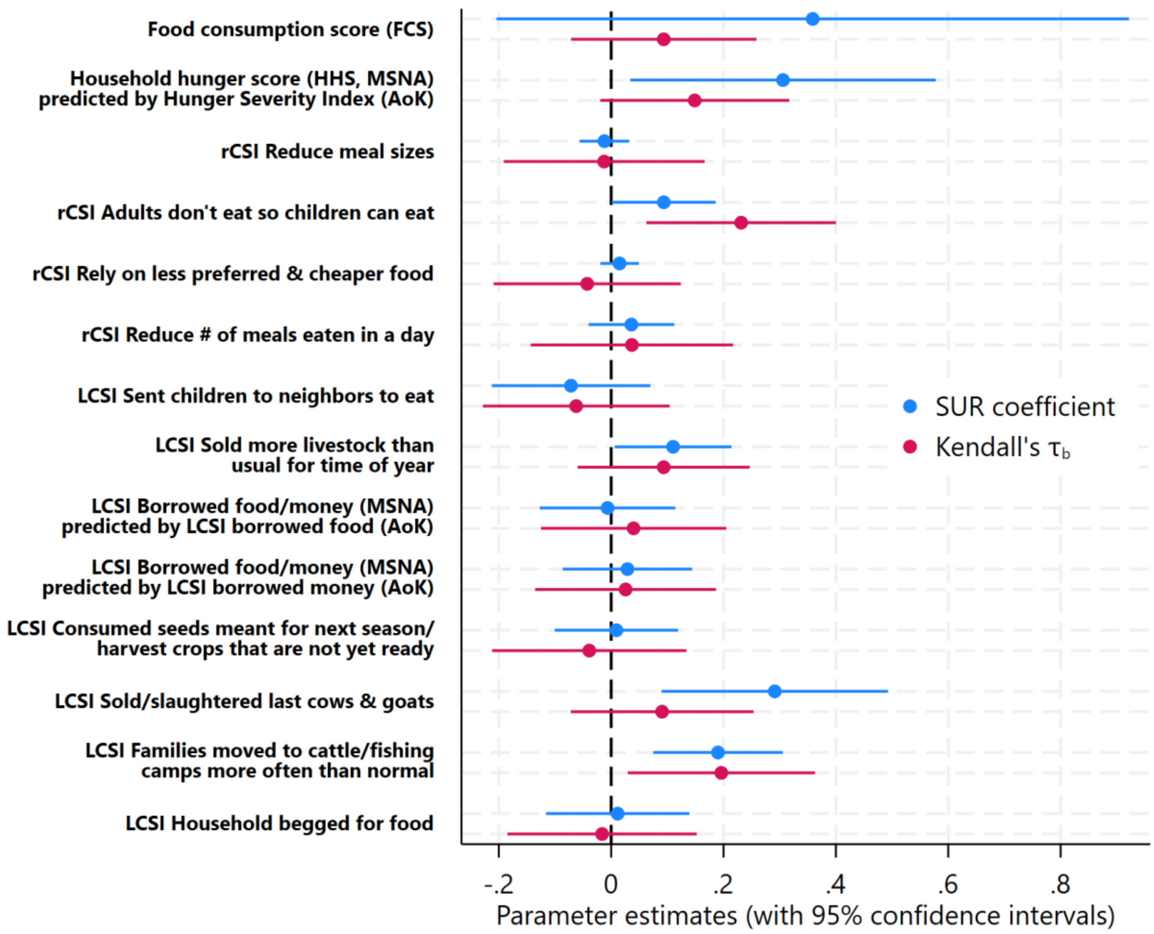


Fig. 1: Estimated slope coefficients from Seemingly Unrelated Regressions (SUR) of MSNA indicators on AoK proxies, and estimated Kendal's Tau-b rank correlations between MSNA indicators and corresponding AoK proxies, in the Food Security and Livelihoods (FSL) domain. In both models, a parameter of 1 is the theoretical optimum (indicating a perfect correlation) that one would not expect to be attained here due to e.g., differences between MSNA and AoK in terms of question phrasing and aggregation. A parameter of 0 would imply the absence of (rank) correlation for the MSNA-AoK indicator pair, that is, AoK would be as good as randomly guessing the (rank of) the MSNA indicator. The AoK "Hunger severity scale" is constructed from the KI question "In the last 30 DAYS, how was hunger for MOST HHs because they were not able to access enough food in this settlement ?", with answer options: "No or almost no hunger - the majority of households had access to food everyday over the last 30 days"; "hunger is minor - most households have only RARELY no access to food (during the last 30 days, most households had no access to food during a maximum of 2 days in total)"; "hunger is moderate - most households have SOMETIMES no access to food (during the last 30 days, most households had no access to food during 3 to 10 days in total)"; "hunger is severe - most households have OFTEN no access to food (during the last 30 days, most households had no access to food during more than 10 days in total)". Confidence intervals for the SUR coefficients are based on heteroskedasticity-robust standard errors with small-sample degrees of freedom adjustment; the confidence intervals for the Kendall's tau-b rank correlation coefficients are based on bias-corrected bootstrap (with 1000 replications). In the SUR estimation, the equation for "LCSI borrowed food/money (MSNA)" has both the "LCSI borrowed food" and "LCSI borrowed money" AoK proxies as predictors, and it is those coefficients that are displayed. Hence, the number of SUR equations is one less (13) than the number of Kendall's tau-b rank correlations estimated (14). The joint F-test that all 14 SUR slope coefficients are zero is rejected at the 5% significance level.

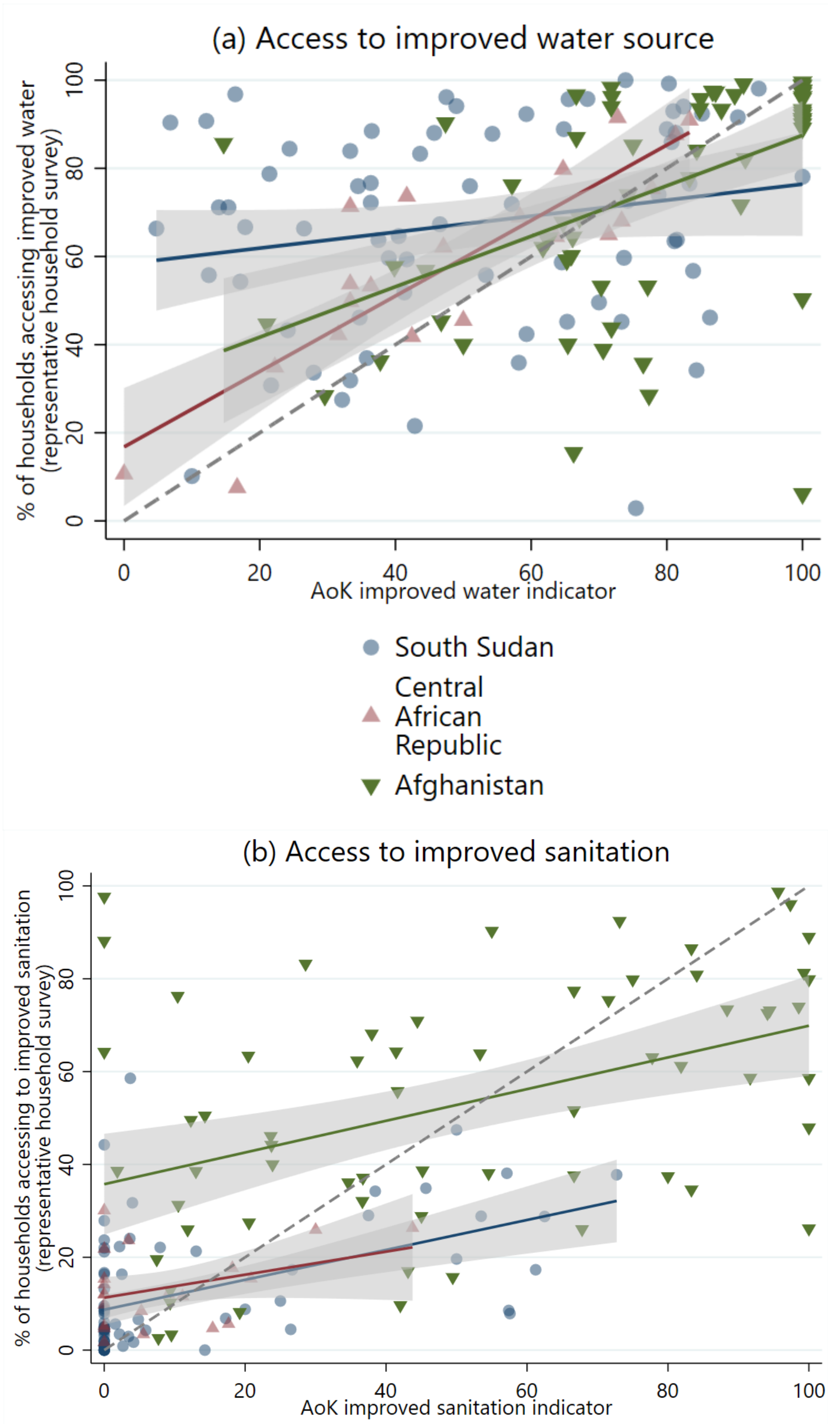


Fig. 2: Estimated lines of best linear fit based on a linear regression of the representative survey indicator on the corresponding AoK indicator (with 95% confidence bands), for access to improved water (panel (a)) and access to improved sanitation (panel (b)). The dashed line is the 45-degree line.

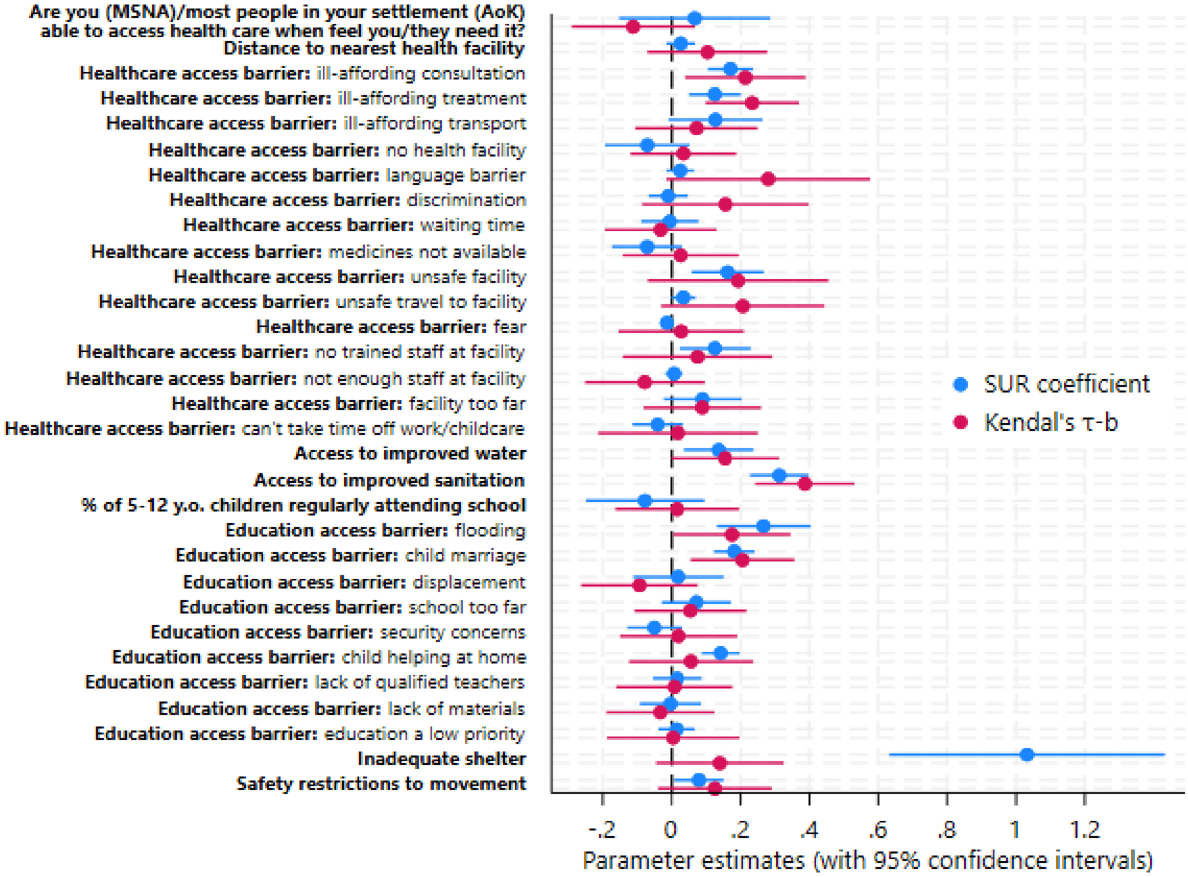
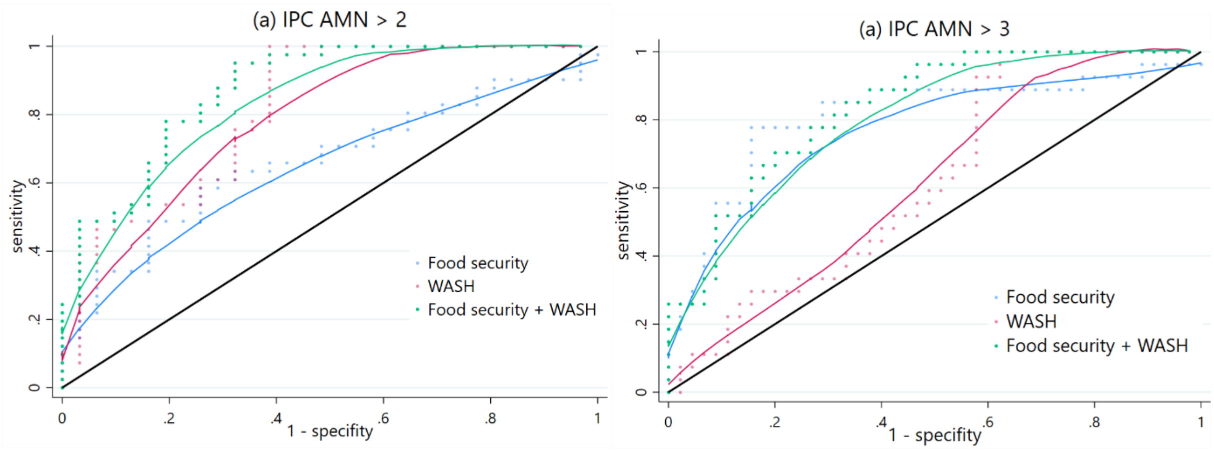
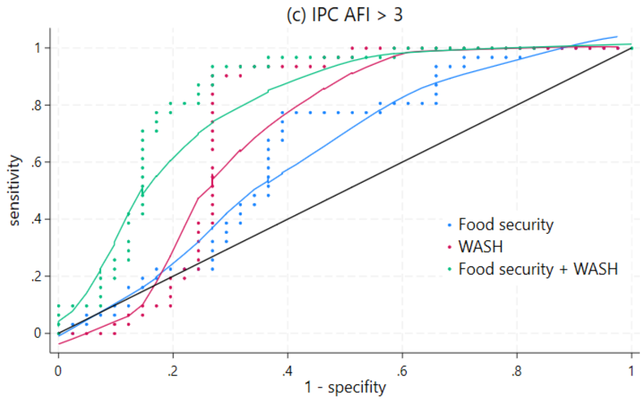


Fig. 3: Estimated slope coefficients from Seemingly Unrelated Regressions (SUR) of MSNA indicators on AoK proxies, and estimated Kendal's Tau-b rank correlations between MSNA indicators and corresponding AoK proxies in South Sudan, in sectors other than FSL. Confidence intervals for SUR are based on heteroskedasticity-robust standard errors with small-sample degrees of freedom adjustment; the confidence intervals for the Kendall's tau-b rank correlation coefficients are based on bias-corrected bootstrap (with 1000 replications). The joint F-test that all 31 SUR slope coefficients are zero is rejected at the 5% significance level, as are the joint F-tests that all healthcare access barrier SUR slope coefficients are zero, and the joint F-test that all 11 education access barrier SUR slope coefficients are zero.

Correlations between AoK indicators and IPC Acute Food Insecurity and IPC Acute Malnutrition area classifications

- Findings. The AoK hunger severity indicator and some of the AoK LCSI questions correlated with the IPC AMN County classifications, and those indicators as well as the AoK food consumption score (FCS) correlated with the IPC AFI county (admin 2) classifications. Combining these FSL indicators and WASH (water, sanitation, and hygiene) indicators increased their predictive ability with respect to the IPC AMN phase classifications (Fig. 4).





	% (of admin 2 areas) correctly classified		
AoK predictors:	FSL	WASH	FSL & WASH
Outcome:			
IPC AMN >2	66.7%	83.3%	86.1%
IPC AMN >3	81.9%	59.7%	84.7%
IPC AFI >3	62.5%	80.6%	87.5%
	Area under the Curve (AUC) [95% CI]		
AoK predictors:	FSL	WASH	FSL & WASH
Outcome:			
IPC AMN >2	0.728 [0.517 ; 0.801]	0.813 [0.694 ; 0.916]	0.895 [0.765 ; 0.948]
IPC AMN >3	0.875 [0.699 ; 0.926]	0.661 [0.474 : 0.761]	0.875 [0.726 : 0.922]
IPC AFI >3	0.643 [0.486 : 0.767]	0.787 [0.615 : 0.876]	0.882 [0.726 : 0.931]

Fig. 4: Receiver Operating Characteristic (ROC) curves of logistic regression-based predictions of the July - October 2022 IPC Acute Malnutrition (AMN) area classifications in South Sudan exceeding 2 (panel (a)) and the area classifications exceeding 3 (panel (b)) and the October – November 2022 IPC Acute Food Insecurity (AFI) area classifications exceeding 3, using AoK food security and/or AoK WASH indicators. The solid curves are locally weighted running line smoothers (LOESS) of the ROCs. The horizontal axis shows 1 – specificity, which is the rate of false positives (e.g., in panel (a), the proportion of counties in South Sudan predicted to have IPC AMN > 2 where the IPC AMN is actually 2 or lower), and the vertical axis shows the sensitivity, which is the true positive fraction (e.g., in panel (a), the proportion of counties predicted to have IPC AMN > 2 where the IPC AMN is actually 3 or higher). The sample size is 72 admin 2 areas (counties). The FSL predictors include the AoK FCS proxy indicator, the AoK Hunger scale, an AoK rCSI index (with empirical weights for the rCSI indicators based on Principal Component Analysis (PCA)), an AoK LSCI index (constructed from the PCA-weighted LCSI indicators), and the AoK indicator of whether "most households have access to food"; the WASH predictors are the AoK improved water and AoK improved sanitation indicators. The table on the right bottom shows the percentage correctly classified (i.e., the percentage of counties where the actual IPC area classification being above or below the treshold aligns with the prediction), as well as the Area Under the ROC Curve (AUC) with 95% confidence intervals (Cis) based on 10-fold cross-validation. For instance, predictions from a logistic regression of whether the IPC AMN exceeds 2 using only the AoK WASH indicators (on access to improved water and sanitation) results in 83.3% of the counties being correctly classified in terms of whether IPC AMN exceeds 2, and the AUC is 0.813, with a 95% CI of (0.694 ; 0.916).

- Implications. The results showed that some of the AoK FSL and WASH indicators are correlated with areas' IPC AFI and IPC AMN phase classifications in South Sudan (where this analysis was done), respectively. These first findings on the use of AoK results in predicting IPC area classifications are still tentative and call for replication. Given that AoK rCSI indicators often negatively predicted IPC AMN and IPC AFI phase classifications, their inclusion is questionable or at the very least, those AoK indicators should be interpreted with care. It should also be noted that correlation does not imply causation, and this exercise is merely one of predicting (nowcasting) IPC classifications; so, the model estimates do not imply anything regarding causative factors for acute food insecurity and malnutrition.

- Caveats. Given the limited sample size of 72 counties in South Sudan (where this exercise was run), the same dataset was used for model training and testing, hence overestimating predictive ability for new data. Though cross-validation was used to mitigate this overoptimism, out-of-sample prediction (to new countries and/or assessment periods) of IPC phase classifications using the trained model remains an important avenue for further research. Given the relatively large number of AoK indicators in the full model compared to the sample size, there is also a risk of overfitting. To (at least partially) address this, created rCSI and LCSI indices were created with empirical weights assigned to each of the indicators making up those indices (obtained from Principal Component Analysis (PCA)). While this use of country-specific empirical weights in rCSI and LCSI index construction reduces the number of AoK predictors and increase predictive power, it may have further limited generalizability to other countries and assessment periods, as the weights are likely different in other countries and assessment periods. Another caveat for the IPC AFI classification models is that KII data can be used as indirect evidence in the IPC AFI classification process, so there is some risk of leakage from the predictor into the predictand (a form of circular logic); however, the IPC process never relies solely on KII or AoK data to classify an area. Finally, future iterations for IPC AMN models could include AoK health indicators.

Differences in the reliability of AoK indicators across sectors

- Findings. AoK appears more predictive of the humanitarian situation for some indicators than for others, and for some indicators no statistically significant correlation or concordance between AoK and the superior comparator was found. For visible infrastructure like WASH, the AoK-household survey correlations are highest, whereas for some of the LCSI and education indicators, correlations are absent. Indeed, AoK questions in the WASH realm are most predictive of the representative household survey data: whether most households have access to improved water, and sanitation, respectively. Shelter, another publicly observable indicator, also correlates with its AoK proxy across countries. For food security, findings are more mixed, but in South Sudan and CAR, where more aligned food security indicators were available than in Afghanistan, it can be conclusively rejected that overall, AoK food security proxies are not indicative of the food security conditions according to representative household survey data. As for education, AoK seems not a viable strategy to estimate the proportion of children regularly attending school. For barrier-type questions, AoK's predictive ability depends on the barrier (and the country). For barriers to accessing healthcare, ill-affording cost of consultation or treatment seems predictive of the corresponding household survey-based indicators; for education the barriers where AoK seems to work relate to flooding and child marriage, the latter also being a protection indicator.
- Implications. AoK can be used to indicate needs pertaining to publicly visible phenomena (WASH, shelter), financial and physical (e.g., flooding) barriers to accessing health and education, and salient events (e.g., child marriage). There is also some evidence that AoK can be indicative of the food security and livelihood situation, but here the picture is more mixed, and AoK's correlation with household survey indicators in this sector varies by indicator and requires further confirmatory research. Finally, AoK does not seem to be useful for estimating granular phenomena like the percentage of children regularly attending school.
- Caveats. Our analysis was cross-sectional, i.e., concurrent correlations between AoK and representative household survey data. Whereas this is relevant to AoK use cases, e.g., where a household survey is impossible due to inaccessibility, AoK is most often used to assess changing

needs over time. A longitudinal analysis would thus be beneficial, which could represent an avenue for future further research.

Settlement coverage as a moderating factor of the quality of AoK-derived results at the admin2 level

- Findings. Settlement coverage has a modest influence on the correlation between AoK indicators and corresponding indicators from representative household surveys. In South Sudan, field teams were encouraged to increase settlement coverage from the arbitrary informal guideline REACH has in place of a minimum of 5% of settlements at the area level. The median coverage attained in the full sample was 19% of settlements across administrative areas. What would the AoK-survey data correlations have been, had settlement coverage not been pushed as high? In an analysis across indicators, the number of statistically significant negative correlations (where the AoK and representative household survey data are negatively correlated) decreases from 3-6 indicators out of a total of 46 indicators with a median settlement coverage of 5.0% to only 2 indicators in the full sample (19.4% median coverage) according to one method (seemingly unrelated regressions), whereas the number of indicators that are positive and statistically significant is somewhat stable. According to another method, rank correlation, increasing the median settlement coverage from 5.0% to 19.4% increases the number of positive and statistically significant AoK indicators from 2-3 to 4 at the 1% significance level, and from 5-6 to 8-10 indicators at the 5% level.
- Implications. Based on the patterns of results obtained when restricting our AoK indicators to only the first X% of sampled settlements (while varying X), the REACH guidance will be updated to at least 10% settlement coverage, along with the note that 15% will lead to results that are more indicative of the humanitarian situation, for more indicators. Based on the evidence obtained so far, the recommendation is to focus on increasing settlement coverage rather than on aiming to sample multiple KIs per settlement – at least when aiming to report results at the aggregate (e.g. admin 2) level.
- Caveats. The master settlement list used for the denominator in settlement coverage is based on the Common Operational Datasets used by all humanitarian actors and developed over time, with some addition of settlements mapped by REACH. This settlement list is itself unlikely perfect or exhaustive. The pattern of results by indicator when reducing settlement coverage is not consistent, that is, AoK-comparator correlations do not monotonically strengthen for all indicators when increasing settlement coverage, due to sampling error. An avenue for further research is the importance of random sampling of settlements – whereas random sampling is preferred, it may not always be feasible or involve significant tradeoffs (e.g., survey cost).

AoK's predictive ability across geographies (e.g., urban vs. rural)

- Findings. The main finding that AoK is most reliable for WASH indicators, holds across the study countries (South Sudan, CAR, Afghanistan). However, there are some differences across the study countries in terms of the correlation between AoK and the comparator representative household survey across indicators. For instance, LCSI indicators were jointly statistically significant predictors of the corresponding indicators from representative household survey data, but not in Afghanistan. The household hunger scale was also not correlated with MSNA in Afghanistan. This may in part be due to differences across countries in the sampling of KIs. In CAR, most KIs were sampled from the same settlements as the comparator survey, whereas in South Sudan, KIs were in general sampled from different settlements. As a result, there were some AoK indicators that were only correlated with the household survey in CAR, but not in

South Sudan, the percentage of girls regularly attending school being one of them. Because of differences in sampling and because of the larger sample size in South Sudan, more emphasis is placed on our results from South Sudan. In Afghanistan, despite differences in sampling of KIs, the predictive ability of AoK does not differ statistically significantly between urban and rural areas (the representative of MSNA by urban/rural was unique to Afghanistan, so that analysis was only conducted in that country; for details see the Methodology Section).

- Implications. AoK proxies for WASH infrastructure and the hunger severity score appeared to be indicative of representative household survey data across study countries, as are indicators capturing child marriage (as a barrier to education in the case of South Sudan, as an LCSI indicator in the case of Afghanistan). Hence, those indicators seem promising to include also in other countries not covered in this study. Most of our sample in the countries under study was rural, but our findings from Afghanistan indicate that AoK may work both in urban and rural areas. While it is not a direct implication from our empirical findings, contextualization may be important for some indicators (e.g., livelihood strategies for LCSI).
- Caveats. It is hard to interpret differences across countries, given differences in sampling of KIs, sample size, mode of interview, and settlement coverage. Further, most of our sample in the countries under study was rural, so the evidence generated for urban areas is a bit thinner than for rural areas. Sampling is necessarily different across urban and rural areas, and so may sampling guidelines. The AoK questionnaires in CAR and South Sudan were similar, but different from the one in Afghanistan, making the reasons for different correlations between Afghanistan on the one hand and CAR and South Sudan on the other hand, harder to interpret.

Predictiveness of Household Hunger Score (HHS) administered on the KI's own household as proxy indicator for area-level HHS

- Findings. The HHS administered on the KIs own household (among KIs who either live in the settlement or were recently displaced from it at the time of the interview) statistically significantly predicts the area-level HHS based on representative household survey data in both CAR and South Sudan (HHS was not administered on the KI's household in Afghanistan). Moreover, HHS administered on the KIs own household is also among the strongest predictors of the IPC AFI classification in South Sudan.
- Implications. HHS administered to KI's household can be a useful proxy for hunger severity where a household survey is impossible, and it should be considered for incorporation into AoK questionnaires wherever the KI resides or has very recently resided in the settlement (s)he reports about. These findings lend support to the "Famine Monitoring System" approach used by REACH for example in Nigeria in collaboration with UNICEF and others, whereby household outcome indicators are measured in purposively sampled new IDP arrivals.
- Caveats. In South Sudan, KIs were sampled at markets and transit points, and are thus in most cases private individuals. If instead, village chiefs are selected as KIs, they might be relatively well-off in their settlement, and their HHS may not be as reflective of the settlements hunger level. Further research is needed to evaluate how KI profile affects the predictive ability of HHS on KI's own household with respect to area level food security situation. Exploring further which household outcome indicators beyond HHS might be integrated in AoK-type tools in a "hybrid" approach is also a promising avenue for further research. Another avenue this result suggests is worth analyzing the validity of, is AoK-N(eighbor), whereby the KI is asked about her/his direct neighbors rather than the settlement as a whole. AoK-N has been applied by REACH in several contexts.

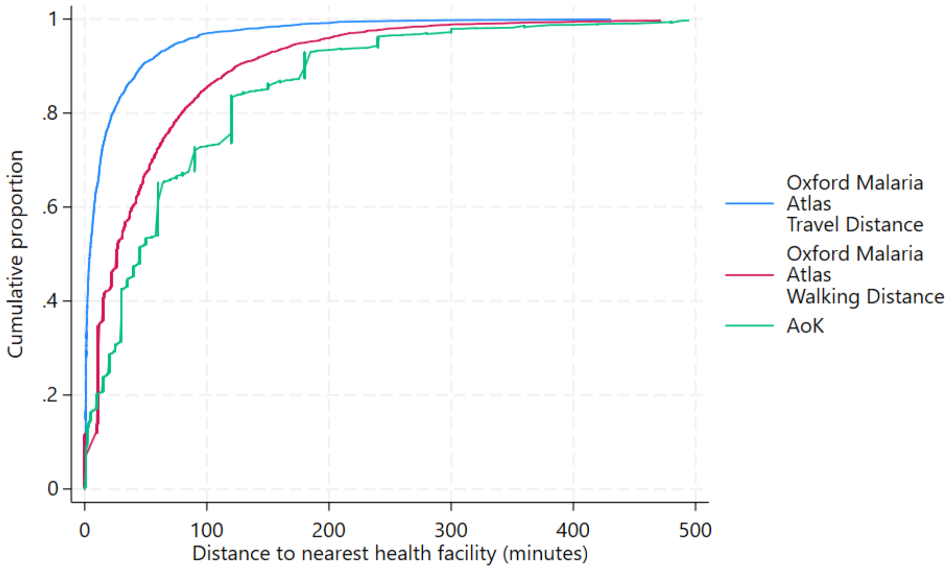


Fig. 5: Empirical cumulative distribution functions (CDFs) of distance to the nearest health facility according to AoK and according to the Oxford Malaria Atlas at the settlement level in South Sudan. Whereas the two variables are statistically significantly (at the 5% level) correlated in a linear regression and in terms of their Kendall's tau-b rank correlation, KIs systematically over-report travel distance to the nearest health facility.

AoK indicators' correlation with administrative data (e.g., travel time to nearest health facility) and remote sensing-based environmental indicators at the settlement level

- Findings. The AoK and Oxford Malaria Atlas Project[9] measures of health facility distance are correlated, albeit weakly. KIs systematically report a longer distance (in minutes) to nearest health facilities compared to the Oxford Malaria Atlas' measures (Fig. 5). More importantly, the KIs reporting the amount of rain to be higher (lower) than normal is (weakly) correlated in the expected direction with respect to the remote sensing-based Standard Precipitation Index (SPI)[10]. AoK reports about how last growing season's yield compares to previous seasons is negatively (so in the opposite direction as expected) correlated with the Normalized Difference Vegetation Index (NDVI)[11].
- Implications. The results showed that some of the AoK FSL and WASH indicators show weak correlations with areas' IPC AFI and IPC AMN phase classifications in South Sudan (where this analysis was done), respectively. These first findings on the AoK results correlating with IPC area classifications are tentative and call for replication with stronger representation of the range of conceptually important elements, greater sample sizes and other contexts. AoK proxies corresponding to only 4 out of 5 household rCSI indicators were collected. Given that some of the individual AoK rCSI proxies negatively correlate with IPC AMN and IPC AFI phase classifications, those AoK indicators should for now be interpreted with care; and they need further testing and validation in other contexts to assess their value. It should also be noted that correlation does not imply causation, and this exercise is merely academic and should not be

[9] https://developers.google.com/earth-engine/datasets/catalog/Oxford_MAP_accessibility_to_healthcare_2019
[10] See https://library.wmo.int/doc_num.php?explnum_id=7768. The 6-months and 12-months SPI were used in the analysis.
[11] Landsat Normalized Difference Vegetation Index (NDVI) is used to quantify vegetation greenness and is useful in understanding vegetation density and assessing changes in plant health. See e.g., https://www.usgs.gov/landsat-missions/landsat-normalized-difference-vegetation-index.

used to actually predict IPC AFI or AMN classifications; so, the model estimates do not imply anything regarding causative factors for acute food insecurity and malnutrition.

- Caveats. Given the limited sample size of 72 counties in South Sudan (where this exercise was run), their was insufficient data to develop separate training and testing datasets, hence overestimating predictive ability for new data. Though cross-validation was used to mitigate this overoptimism, out-of-sample prediction (to new countries and/or assessment periods) of IPC phase classifications using trained models remains an important avenue for further research. Given the relatively large number of AoK indicators in the full model compared to the sample size, there is also a risk of overfitting. To (at least partially) address this, a composite AoK rCSI index (based on the four AoK rCSI proxies that were collected corresponding to 4 out of the 5 standard rCSI indicators) and a composite AoK LCSI index were created with empirical weights assigned to each of the indicators making up those indices (obtained from Principal Component Analysis (PCA)). While this (non-standard) use of country-specific empirical weights in rCSI and LCSI index construction reduces the number of AoK predictors and increase their correlation with IPC AFI classifications, it may have limited generalizability to other countries and assessment periods. Another caveat for the IPC AFI classification models is that KII data can be used as indirect evidence in the IPC AFI classification process, so there is some risk of leakage from the predictor into the predictand (a form of circular logic); however, the IPC process never relies solely on KII or AoK data to classify an area. Additionally, models for IPC AMN should in theory be underpinned by conceptual frameworks, such as the UNICEF conceptual framework for malnutrition being the most common for acute malnutrition. However, the model does not account for all the different types of determinants such as health or morbidity data, and therefore has certain conceptual gaps.

Conclusion

This study found that the Area of Knowledge (AoK) approach holds promise for humanitarian situation monitoring: the null hypothesis that none of the AoK indicators are correlated with corresponding indicators from representative household surveys, is confidently rejected. The extent to which AoK indicators and household survey indicators are correlated depends on sectors, and within sectors, on the indicator. AoK shows the strongest correlation with publicly visible phenomena, like access to improved water or sanitation, or adequate shelter. It also correlates with household survey findings on salient events like child marriage, observable shocks such as flooding as a barrier to education, as well as safety and security restrictions on movement. AoK indicators are also predictive of financial barriers to accessing health care.

The picture is mixed in the food security domain, with most AoK indicators lacking correlation with the corresponding MSNA indicators, and the rCSI indicators in some models negatively predicting IPC AMN and IPC AFI phase classifications, casting doubt around this indicator or at least implying that care should be taken in its interpretation. The HHS administered to the KIs own household is the single strongest predictor of IPC AMN and IPC AFI area classifications, suggesting that asking KIs about their own households (or their neighboring households, as in the "neighborhood method" or AoK-N) can augment or complement the traditional AoK approach. The AoK approach does not appear to work well for granular indicators like the proportion of children regularly attending school. In addition, key informants in our study areas are unable to estimate and report the population size of the settlements about which they report. A lack of quality settlement population estimates prevented us from weighing settlement-level KI data by settlement population size, which might have increased the correlation between AoK and its comparators. Micro-censuses coupled with state-of-the art population modelling is thus called for, both to improve measurement of the ground truth, and to explore AoK weighing of settlement values by settlement size.

The results of this study show that AoK is indicative: as described above, some of the AoK indicators are correlated with corresponding household survey indicators, but the correlation is weak to moderate. Some correlations may have been diluted (and hence weakened) by noisy predictands, that is, the representative survey measures themselves are but noisy representations of underlying phenomena. For example, the AoK indicator capturing distance to the nearest health facility correlates with the healthcare accessibility indicator from the Oxford Malaria Atlas but not with the representative household survey-derived measure. Nevertheless, as it stands, the limited evidence generated by this initial study does not allow us to conclusively state that AoK can be used to measure absolute levels of need in a way that sufficiently reflects the absolute levels of need as measured by representative household surveys.

Increasing settlement coverage modestly increases the number of AoK indicators that are correlated with MSNA, and REACH will update its internal guidance to reflect that. Among the many outstanding questions is the importance of random sampling of settlements. In Afghanistan, settlements were sampled randomly but in South Sudan they were not. Theoretically, random sampling is preferred but it would involve tradeoffs in some countries (e.g., survey cost).

The current study generated evidence that AoK can be useful to assist in ranking areas' needs within a country. As the study was cross-sectional in nature, comparing AoK with other data sources at a given point in time, an important outstanding research question concerns the extent to which AoK indicators capture trends in humanitarian conditions and needs over time. Ancillary questions are whether one should follow the same KIs over time (risking respondent fatigue) or whether one would need to recruit KIs from the same settlement.

There are several open questions that merit further study to improve the measurement of humanitarian needs across space and time where collecting representative surveys is impossible:

(1) The current study was conducted using community-level KI responses aggregated at the area (i.e. admin 2) level. A lack of correlations between AoK indicators and corresponding indicators from representative survey data or remote sensing data may be due to a failure of assumption 1 (AoK giving an accurate depiction at the settlement level) and/or a failure of assumption 2 (comparing AoK indicators across areas produces a valid comparison) stated above. Even if and where assumption 1 holds, assumption 2 may fail due to a multitude of reasons, including: non-random sampling of settlements or KIs, low settlement coverage, or the aggregation of KI responses to the settlement level and from the settlement level to the area level. Conducting a household survey with a larger sample size per settlement (so that it is representative at the settlement level), would allow for a clearer attribution of (a lack of) correlation between AoK and representative survey data to (a lack of) KI reliability, by allowing the use of KI responses aggregated at community-level rather than area-level.

(2) The current study was cross-sectional in nature, analyzing correlations between AoK and other data sources at one point in time. However, the AoK approach is often deployed to capture changes over time in the humanitarian situation, and further research is needed to assess the extent to which AoK is able to reliably capture changes over time.

(3) The design of AoK assessments involves choices and tradeoffs, which merit careful examination in a thereto designed study that changes one parameter at a time to disentangle and attribute differences in predictive accuracy. Should settlements from which KIs are interviewed, be sampled randomly? This should increase representativeness of findings reported aggregated at the area level, but in certain settings it would only be possible by phone (rather than face-to-face) due to access or resource constraints – and likely only partially. Or should KIs be selected through purposive or quota sampling (e.g., by intercepting individuals travelling through central towns or at bus stops), allowing for face-to-face interviews and for attaining higher settlement coverage at lower cost? The former approach would imply interviewing individuals in a hierarchical position within a settlement such as village chiefs, who may be more connected and more informed, whereas the latter approach would capture KI profiles closer to the average person whose responses to questions about her or his own household may be informative (as shown in this study's HHS question example). Should data collection resources be spent on achieving a maximum settlement coverage or on triangulating KI responses within a settlement (by interviewing several KIs per settlement)? Yet another tradeoff exists between accuracy (achieved through random sampling and increasing coverage) and assessment frequency.

(4) How does AoK's predictive accuracy compare to alternative remote data collection approaches? Does asking KIs their own households or about their direct neighbors rather than an entire community, as in the Area of Knowledge Neighborhods (AoK-N) method, result in more accurate findings for some indicators, perhaps meriting the combination of the two approaches in assessments?

(5) The prediction of food crises – both nowcasting (as was explored in this study) and forecasting, needs more work, including out-of-sample testing of trained predictive models, and combining AoK proxies with other data (e.g., food prices, remote sensing indicators, news sources[12]) to increase predictive accuracy.

[12] Balashankar, A., Subramanian, L., & Fraiberger, S. P. (2023). Predicting food crises using news streams. Science Advances, 9(9), eabm3449.