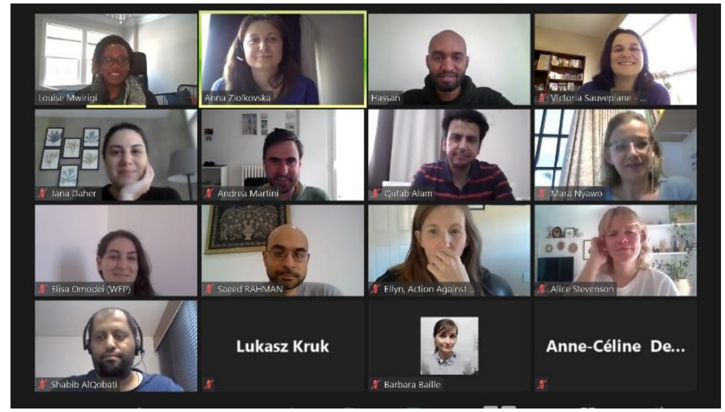
Global Nutrition Cluster Predictive Analytics workshop - Summary

22 February & 3 March 2021 (1-5 pm GMT I 2 – 6 pm GMT+1 I 8 am – 12 pm ET)



Objectives of the workshop: To define the GNC approach on Predictive Analytics for the forecasting of Nutrition outcomes. The GNC, under the auspicious of the GNC Technical Alliance (GNC-TA), is embarking on a new initiative on Nutrition Predictions. When exploring the existing work in this area, there are several initiatives and projects looking at projections in general and nutrition predictions in particular. To avoid duplication of efforts, maximize learning, and decide on the best way forward with nutrition projections, two half-day workshops were organized to learn about the pros and cons of different approaches to projections that have been tried so far, and what would be needed to develop a forecasting model that is useful for all countries and decision-makers for Nutrition. Participants are detailed in *Annex 1.*

Summary of the main agreements and points for follow-up:

- Mobilize resources to develop the nutrition forecasting model based on the presentation by the OIC GNC Coordinator;
- Develop a Terms of Reference for a Taskforce on Nutrition Predictive Analytics to lead the GNC AI work (general interest from participants to take part in it) that would be part of the Nutrition Information Systems Technical Working Group of the GNC-TA and invite the participants of this workshop to be its members;
- Consult with colleagues to gain from experiences with food security and nowcasting, namely World Bank, Cornell University and another university in Texas to expand learning out;
- Organize a follow-up call with MERIAM in March (as this is the end of their project) to further learn from their experiences.

Agenda and notes: Facilitators included Hassan Ali Ahmed, Louise Mwigiri & Anna Ziolkovska, notetaker Victoria Sauveplane Stirling

Session title	Presenter (s)	Remarks	Main Notes
1) IPC Acute Malnutrition (AMN) forecasting	Douglas Jayasekaran, FAO	This is methodology currently used by most of the countries for nutrition projections (does not involve AI or any mathematical modeling; mostly trend analysis with assumptions)	• Involves both current and projection situation analysis for Acute Malnutrition based on seasonality; • Looks at most likely changes in light of assumptions to determine projection based on *Most likely scenario* which doesn't necessarily represent the worst or base case scenario; • Does not use mathematical modeling, rather the likelihood of areas moving into different IPC AMN phase based on current levels and historical trends and consensus reached based on identified assumptions; • No point estimate is generated for the projection period.
2) Geospatial data solutions for humanitarian action	Margareet Barkhof, UNICEF	The purpose of this session was to introduce the work developing global level resources for humanitarian action (i.e., the humanitarian elements of the UNICEF GeoHub)	• Aim: to build global level resources for geospatial data solutions to support humanitarian action • Inputs to GeoHub building blocks: definitions, geospatial data catalogue and data repository, geospatial data analysis tools • Implement use cases works with country offices, implement global estimates, children's climate and vulnerability analsysi – scoping of next steps. Geospatial data 'use case' contains country profiles in terms of policies, standards, capacity for geospatial data readiness, and a data catalog of what is available and accessible.
3) IPC Advanced Technology and Artificial Intelligence (ATARI)	Nicholas Haan, IPC secretariat	The IPC Secretariat has established the ATARI working group to improve forecasting for the IPC FS analysis. This presentation was to share experiences that can be potentially transferred for enhancing AM preditions.	• Food Security nowcasting and forecasting: focus on data management (gatheting, processing, and sharing) and analysis (Importance of starting with a problem-based approach when looking at data management (gathering, processing and sharing) and analysis issues (improving analysis and classification process in terms of consensus building, efficiency, time and human resources required, quality, scalability, frequency etc) • Use of artificial Swarm intelligence that leverages human knowledge to make a conclusion to a question being asked with predefined and pre-set options for the group of individuals to decide. This focuses on consensus-building in an non-biased way as it maintains unanimity. • Nowcasting vs. forecasting - how can machine learning support real-time analysis of the situation, before looking at projections; • 3 pilots: Haiti, Malawi and South Africa. SwarmpProcess took around 12 hours. Importance of open data as an essential input for this API.
4) Hunger map	Elisa Omodei<,WFP	The purpose of this session was to present the predictive model	• The model consolidates all data sources in one central systems, with daily updates from WFP's near real-time monitoring systems – ability to nowcast the food security

Global NUTRITION CLUSTER

GLOBAL NUTRITION CLUSTER Technical Alliance

		that WFP developed to nowcast the prevalence of people with insufficient food consumption (https://hungermap.wfp.org/).	situations in the countries WFP works, as well as most lower and lower-middle income countries, and converts data into user-friendly visualisations. • Done through continuous data collection conducted remotely through live calls, using the some IPC indicators (FCS, rCSI, LhCSI) = provides representative data on the food security situation in a country with data anaysed automatically on a daily basis. • Information is consolidated into a unified data lake; • Predictive model is trained using historial food security data spanning 63 countries across 14 years (2006-2019) =algorithm used is a non-linear regression method: gradient boosted tree ensembles (xgboost); model is trained on a large subset of the historical data and then validated on the remaining subset to ensure the goodness of the model; ; • the explainability of the model is key to consider with visual explanations of each predictor's contribution.
5) MERIAM	Alice Stevenson and Ellyn Yakowenko, ACF	Modelling Early Risk Indicators to Anticipate Malnutrition (MERIAM)	• A four-year research project funded by the UK government, which seeks to identify, test, and scale up cost-effective means to improve the prediction and monitoring of acute malnutrition, through the use of open access secondary data. Focus on climate- and conflict-affected regions using 2 separate but complementary modelling approaches: computational models work at more a granular-level (household) vs. econometric is more broad. More specifically, computational is an evidence-driven model vs. econometric looks at subnational regional analysis (focusing on GAM, aggregating regional-level factors, covering 29 countries) and multilevel analysis (focusing on risk of individual child with separate models for Kenya, Uganda, Mali and Nigeria). A scenario-based interactive tool may be useful to consider. • Limitations include data availability, coverage, resolution and accuracy; restricted to countries in sub-Saharan Africa, and not all findings link to specific practical interventions.
6) South Sudan forecasting	Saeed Rahman Qutab Alam	Used predictive modeling techniques to estimate the burden of acute malnutrition and severity mapping in the absence of recent anthropometric data	• Due to the COVID 19 pandemic, relevant global guidance suspended the collection of this anthropometric data in the most recent round of Food Security and Nutrition Monitoring System (FSNMS) due to the concern of spreading the virus with equipment such as height boards and weighting scales. • The South Sudan Nutrition Cluster, with support from REACH South Sudan, UNICEF and WFP through the Nutrition Information Working Group led this work to estimate the burden of acute malnutrition and severity mapping in absence of anthropometric data. • Used training data from 3 rounds of FSNMS data and applicd predive algorithms including regression and Random Forest that may be used for proxy GAM and SAM.

			• Results still needs to be further tested and validated with SMART/household survey data.
7) Overview on Obesity/overweight surveillance systems	Alex Hutchison, the data lab	Analyses of the "Growing up in Scotland (GUS)" child cohort to inform the design of obesity/overweight surveillance systems internationally". More information about the [Data for Children Collaborative with UNICEF](https://www.dataforchildrencollaborative.com)	• Partnership between UNICEF, Scottish Government, University of Edinburgh • Looking at the use of data science to improve outcomes for children. Analysis was done based on a cohort started in 2005 looking at infancy to teenagers and the predictos of obesity, i.e. if you are obese by age 12, then what are the patterns or risk factors? 2800 children acorss a representative population of Scotland was used, all data was anonymized. • Results are not yet published but a blog on some key conceptual points was released on March 3rd: https://www.dataforchildrencollaborative.com/news-from-the-unicef-data-for-children-collaborative/obesity-gus-output
8) Prediction of acute malnutrition prevalence in three countries in Eastern Africa	Mara Nyawo, UNICEF	UNICEF Eastern and Southern Africa Regional Office working with LSHTM to develop a model that will predict GAM and SAM for small geographical areas.	• Concept of the research: to reduce the need for nutrition surveys while enhancing the availability of information, we are aiming to develop and validate a statistical approach that can predict GAM and SAM in (near) real-time, at sub-national level for small geographical units. • Uses survey reports, raw datasets from a wide range of sources, data on plausible predictors, and a causal framework to fit a model to estimate acute malnutrition prevalence, then to test accuracy of model for future forecast of AM prevalence before it can ultimately predict acute malnutrition prevalence. • 71 eligible SMART surveys for Somalia, while 175 in South Sudan – this data was used to developed using logistic regression, linear regression and Random Forest prediction algorithms. • Analyses are ongoing – to revise the statistical approach, in particular adding models with MUAC – with MUAC, better results in our work on mortality prediction. AM measured with MUAC is more strongly associated with mortality than WFH so the PI (Principal Investigator) hopes that what predicts mortality well will predict AM based on MUAC well. The PI also believe we should move towards MUAC only programming so it's good to test with MUAC. • Not sure what time-periods will the model be able to predict for – it may just be real-time.
9) Initial GNC project approach	Anna Ziolkovska, GNC	Preliminary GNC NIS WG thoughts on the forecasting (prepared before the workshop)	• The goal of the GNC project is to leverage artifical intelligence to monitor in real time the level of malnutrition in communities at risk in order to provide a timelier response. • First focus on GAM of children U5, and then expand to different forms of micronutrient malnutrition, obesity/overweight, U5 mortality, IYCF practices

			• Model analyzing data on the causes of malnutrition and their correlation with malnutrition using structured data (such as provided by the UN agencies, clusters and on-site surveys), and unstructured data such as live feeds on social media and new outlets, the radio, Google trends, etc. by text extraction and translation, along with image processing from real-time satellite data and imagery to ultimately produce outputs estimating the risk of malnutrition, in near real-time at the community level.
10) WCARO Experiences	Anne-Céline Délinger, UNICEF	Oral sharing of experiences given COVID-19 mobility restrictions	• Used national nutrition surveys from 2019 across the Sahel, along with national and regional programming data from last 5 years. • Used 23 nutrition and food security indicators and aggravating factors to model an adjusted incidence factor per quarter for planning and response strategies during COVID-19.

Facilitated discussion on how GNC should approach forecasting for Nutrition outcomes:

Proposed approach: To leverage artifical intelligence to monitor in real time the level of malnutrition in communities at risk in order to provide a timelier response. First focus on GAM of children U5, and then expand to different forms of micronutrient malnutrition, obesity/overweight, U5 mortality, IYCF practices. The model would analyze data on the causes of malnutrition and their correlation with malnutrition using structured data (such as provided by the UN agencies, clusters and on-site surveys), and unstructured data such a live feeds on social media and new outlets, the radio, Google trends, etc. by text extraction and translation, along with image processing from real-time satellite data and imagery to ultimately produce outputs estimating the risk of malnutrition, in near real-time at the community level.

- Alignment with ESAR ongoing work, predictions available on a more regular basis and for small areas that are more emergency or conflict affected – the model can also reduce the need to do so many surveys.
    - Quite comprehensive with data sources and if that can tie into well and can information on different causal pathways;
- Building on MERIAM work – so many similarities with intentions and what was tried and tested – very good progress, and how to work with validated prospective forecasts at individual level, looking at MUAC and GAM, with various lead times (3/6/6+ months)
    - To identify the specific lead time needed for improved decision-making concretely with the intended users as it may be much longer than 3-6 months, for example to better target interventions;
    - To consider how to build on existing systems at national-level – may be next phase for MERIAM
    - The causal framework – not as relevant as long as the data that had the streams
- To define how and for whom (very specifically) at the design phase: 1) cluster or sector coordination teams; 2) to recognize the donors and the funders, having the confidence in the predictions

Infrastructure of the data – often underestimated how much work and resources, solid infrastructure from the beginning to collect and structure the data, and very importantly how to use the model in real-time with a data flow system and how to update it more easily and how to integrate data that would be available later on and how they would interact

- For SMART surveys, the team is currently building an aggregator in the new tool and this would facilitate the exchange and dissemination of data
- Phased-approach is recommended, to build on existing approaches in V1 of the model with more traditional structured data and then V2 expand scope with other data to ensure that the basics are covered
    - If no structured data, then to bring in unstructured data – scrapping can be important
    - To build on South Sudan and Somalia's data sources, as contexts will be different in terms of data sources
- 2 types: outcome is AM but also to predict caseload admission estimation. Perhaps not to predict GAM outcome itself (GAM and SAM is more tricky) – but perhaps focus on admissions? OTP admissions on a regular basis, may be easier to predict if real-time monitoring is the main goal. Admissions trends – very easy if strong model is in place, may not need so much time to get information

Model-related discussion (define the outcome as the regression problem, accuracy, lead time to realistically achieve, what admin level can we predict the outcome)

- Multi-linear regression – and how they relate to malnutrition, since it is so complex and a lot of interactions with different predictors = cannot capture as well
- Machine learning like Random Forrest may be better suited (DHS data from Bangladesh) for acute malnutrition
- Framework of logistic regression would be correct, beyond just linear regression as this won't be enough given all the indicators – Random Forrest (algorithm) to be applied to obtain classification = two separate points
- Discussions with data scientists would be required to flesh out these technicalities, unclear in a swarm analysis would be relevant as primarily we would be analyzing historical nutrition data and using other contextual data, vs. trying to reach a consensus
- Neural network can be tested in terms of accuracy
    - How this is going to be used would dictate the type of method – huge point of learning from MERIAM
    - MERIAM models are validated through retrospective forecasting - currently working to finalize the prospective forecasting elements.
        1. Big difference for a model for anticipative action – good precision and accuracy, to better pre-position supplies to make better decision-making in both types of modelling - econometric (more insightful in what is going on) = more phase 2 and 3
        2. Phases 4 and 5 prediction = more severe crises, not yet been cracked and to think of out-of-the-box methodologies
- Need to have both: real-time monitoring (MERIAM) prediction 3-6 months, but also nowcasting – different applications to try to fill in data gaps, in contexts without access

Annex 1 – Participants:

Name	Email	Representation
Alice Stevenson	astevenson@actionagainsthunger.org	ACF, MERIAM
Ellyn Yakowenko	eyakowenko@actionagainsthunger.org	ACF, MERIAM
Hassan Ali Ahmed	hahmed@actioncontrelafaim.ca	Co-chair for the NIS WG, ACF
Louise Mwirigi	lmwirigi@unicef.org	Co-chair for the NIS WG, UNICEF
Cristina Majorano	cristina.majorano@wfp.org	Global Cluster Representative - Food Security
Lukasz Kruk	lukasz.kruk@reach-initiative.org	Global Cluster Representative - WASH
Anna Ziolkovska	aziolkovska@unicef.org	GNC-CT
Magnat Kavuna	mkavuna@unicef.org	GNC-CT
Shabib AlQobati	salqobati@unicef.org	GNC-CT
Victoria Sauveplane	vsauveplane@unicef.org	GNC-CT
Douglas Jayasekaran	Douglas.Jayasekaran@fao.org	IPC Acute Malnutrition
Nicholas Haan	nickjhaan@gmail.com	IPC, ATARI
Manuel Garcia-Herranz	mgarciaherranz@unicef.org	Magic Box, UNICEF
Viviana Rocio Canon	vcanon@unicef.org	Magic Box, UNICEF
Tinka Valentijn	tinka.valentijn@un.org	OCHA, Predictive analytics unit
Jana Daher	jdaher@actioncontrelafaim.ca	SMART +
Saeed Rahman	saeed.rahman@reach-initiative.org	SS Nutrition Cluster, REACH initiative
Qutab Alam	qalam@unicef.org	SS Nutrition Cluster, UNICEF
Mara Nyawo	mnyawo@unicef.org	UNICEF RO, ESARO
Marjorie Volege	mvolege@unicef.org	UNICEF RO, ESARO
Anne Celine Delinger	adelinger@unicef.org	UNICEF RO, WCARO
Barbara Baille	bbaille@unicef.org	UNICEF RO, WCARO
Chika Hayashi	chayashi@unicef.org	UNICEF, Division of Data, Analytics, Planning and Monitoring
Margareet Barkhof	mbarkhof@unicef.org	UNICEF, Geospatial database
Elisa Omodei	elisa.omodei@wfp.org	WFP RAM – Hunger Monitoring Unit
Hermann Ouedraogo	houedraogo@unicef.org	SS Nutrition Cluster, UNICEF
Andrea Martini	amartini@unicef.org	Magic Box, UNICEF
Giulia Martini		WFP RAM – Hunger Monitoring Unit
Sejal Jaiswal		WFP RAM – Hunger Monitoring Unit

Alex Hutchison	alex.hutchison@thedatalab.com	Data for Children Collaborative with UNICEF