LINK NCA TESTS THE USE OF SECONDARY DATASETS FOR RESULT SYNTHESIS

GRACE HEYMSFIELD, LINK NCA ANALYST

EXECUTIVE SUMMARY

Recent secondary datasets that include anthropometric measurements of children can provide key pointers to the multi-causality of undernutrition in a set context. When re-assessed per Link NCA Guidelines, correlations between indicators in key sectors and child nutrition outcomes can guide hypotheses development, substitute for a Risk Factor Survey, or buttress the findings of a Risk Factor Survey. The Analyst must take into consideration the relevance of the secondary dataset to the zone of the study and time required for analyses. Based on experiences in Myanmar, Uganda, and Somalia, it is recommended that all Link NCA Analysts review availability of secondary datasets during the preparatory stage of the study.

INTRODUCTION

Link NCA is a mixed-methods approach, combining both qualitative and quantitative research methods, drawing conclusions from a synthesis of results. While quantitative components of the study answer questions of "how many," "which," and "what;" qualitative components of the study answer questions of "how" and "why."

The Link NCA methodology includes 5 stages, ranging from Preparatory Phase to Response Planning (Figure 1). The third stage, or primary data collection, includes a qualitative inquiry and a quantitative Risk Factor Survey (RFS). In this stage, qualitative inquiry can precede the quantitative Risk Factor Survey, vice versa, or the two can run in parallel. When completed before the RFS, the qualitative inquiry often is missing quantitative "pointers" and thus must be generally wide sweeping, covering a variety of risk factors in semi-structured interviews, focus group discussions, and observations. When completed before qualitative inquiry, the RFS often misses qualitative "insights" and must also be very broad, which often results in a time-consuming questionnaire that burdens participating households. If completed in parallel, the qualitative inquiry and RFS have limited opportunity to influence each other but can save time in overall study timeline.

Recent Link NCA's have explored the USE OF SECONDARY DATASETS to substantiate result synthesis prior to stage two and/or as part of stage 3. This approach, with variant timing, has recently been tested in Uganda[1], Somalia[2], and Myanmar[3].

1 Mid-North, Karamoja, and West Nile regions

2 Four livelihood zones- Agro-pastoral livelihood zone SO 15-16, Riverine livelihood zone SO 13, Pastoral livelihood zone SO 05, Settlement for Internally Displaced Population SO 19

3 Chin State

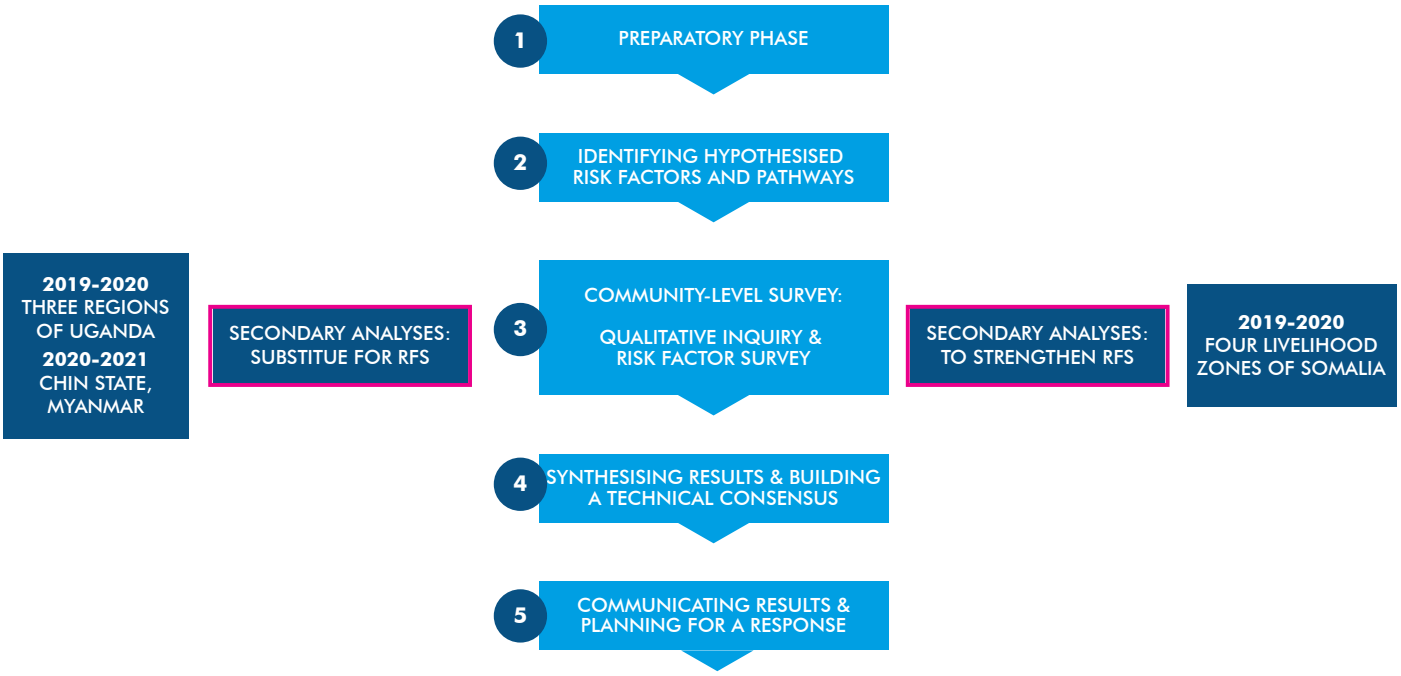

FIGURE 1: Stages of a Link NCA, with complementary secondary analyses tested in Somalia, Myanmar, and Uganda

METHODOLOGY

TO STRENGTHEN FINDINGS FROM THE RFS: FOUR LIVELIHOOD ZONES OF SOMALIA The 2019-2020 Link NCA in the Agro-pastoral livelihood zone of Baidoa District, Riverine livelihood zone of Beltweyne District, Pastoral livelihood zone of Goldogob District, and Settlement for Internally Displaced Population of Kahda District of Somalia utilized secondary analyses in Stages One and Two. The main objective of a preparatory phase of a Link NCA is to define key parameters of the study, including its objectives, geographical coverage and feasibility. Due to the extensive geographic scope of this Link NCA, as well as the variance in risk factors expected across livelihood zones, preliminary secondary data analyses were conducted in addition to the literature review, to define the structure of the study and aid pre-selection of hypotheses, ultimately strengthening findings from the Risk Factor Survey. As risk factors were perceived as highly subjective to the rainy seasons (Dey, the short rainy season, and Gu, the long rainy season), use of secondary datasets also enabled exploration of risk factors both historically and seasonally.

Thus, prior to primary data collection, a thorough complementary analysis of post Deyr and Gu seasonal datasets, 2014-2018, from the Food Security and Nutrition Analysis Unit (FSNAU) was conducted[4]. This data is collected semi-annually by the FSNAU in Somalia to generate the malnutrition indicators specific to the Gu and Deyr seasons. It is a population-representative household survey. Data on demographics, gender, food intake, food security and livelihoods, health environment, and morbidity were present in all FSNAU surveys that contained both the nutrition and food security components. Associations with wasting and stunting were identified with binary logistic (WHZ <-2, HAZ <-2) and linear regression (WHZ, HAZ) analyses, with significance reported, as appropriate. All statistical computing was performed in R.

4 2018 Bay Region, 2015 Deyr and 2017 Hiraan/ Beletweyne datasets not available for analyses.

BEFORE A QUALITATIVE INQUIRY: CHIN STATE MYANMAR In 2017, the Maternal And Child Cash Transfer (MCCT) Programme Baseline Survey[5] was conducted in the geographic area of interest for the Link NCA, Chin State to establish baseline values for pregnant and lactating women enrolled in the program. While the survey was not a population-representative household level survey, it included numerous indicators at household and maternal level relevant to probable risk factors for the study. Due to the extensive scope of the relevant indicators, as well as the onset of the COVID-19 pandemic which would impede implementation of an RFS, MCCT secondary analyses were selected for triangulation with a pending qualitative inquiry, in place of an RFS. Associations with undernutrition were identified with binary logistic and linear regressions for wasting (MUAC, WHZ, MUAC and/or WHZ), stunting (HAZ), and WaST (WHZ <-2 and HAZ <-2). All statistical computing was preformed in STATA 16.

DURING A QUALITATIVE INQUIRY: THREE REGIONS OF UGANDA At the onset of the Link NCA in Mid-North, Karamoja, and West Nile regions of Uganda, preliminary secondary data and literature reviews were conducted to define the scope of the study. Considering likely availability of quantitative data from the Karamoja Food Security and Nutrition Assessment (FSNA), and that UNICEF was conducting an FSNA in Mid-north and West Nile regions, it was decided that only qualitative data would be collected and triangulated with results of the FSNAs. The overall aim of the FSNA is to determine the extent and severity of malnutrition in different age groups and food security of households as well as monitor selected food security, nutrition, health, water and sanitation indicators. It is a population-representative household level survey. Thus, secondary quantitative data for Karamoja region was drawn from FSNAs conducted by WFP in June 2017 and UNICEF in January 2018[6], allowing for investigation of seasonality[7]; while that for Mid-North and West Nile Regions was provided by UNICEF in November 2019[8]. Associations with undernutrition were identified with binary logistic and linear regressions for wasting (WHZ, MUAC, WHZ and/or MUAC) and stunting (HAZ). All statistical computing was preformed in in STATA 13 and SPSS 23 software.

ADVANTAGES

- **LARGE SAMPLE SIZE** National, regional, and sub-regional datasets typically have larger sample sizes than is feasible for a Risk Factor Survey to achieve. For example, in the case of Karamoja FSNA, over 6,000 children were measured and eligible for analyses. As Link NCA analyses detect variation in risk factors between malnourished and non-malnourished children, increasing the sample size may increase the total number of malnourished children in the sample. This theoretically reduces the risk of insufficient variation in the sampled population for analyses.
- **INDICATORS INFEASIBLE FOR LINK NCA** Some secondary datasets include anthropometric indices for analyses which are infeasible, or deemed too time costly, for a Link NCA RFS; namely, anemia and maternal height/weight for calculation of maternal BMI.
- **NATIONALLY RECOGNIZED DATASETS** Internal validation mechanisms exist by context, i.e. Technical Working Groups. Secondary datasets may already have more 'buy-in' from key stakeholders, having undergone quality review, compared to new, primary data.
- **POTENTIAL TO INFLUENCE HYPOTHESES** When conducted to inform hypothesized risk factors, secondary dataset analyses can yield key pointers to vulnerable groups- perhaps even informing pockets of malnutrition to study in the qualitative inquiry.
- **COST EFFECTIVENESS** Risk Factor Surveys are expensive. In light of limited budget, use of secondary quantitative data, when suitable data is available, saves costs associated with primary data collection.

5 LIFT. https://www.lift-fund.org/mcct-chin-baseline-survey

6 Though hoped that WFP would conduct an FSNA in Karamoja, this turned out not to be the case and thus data from an older FSNA of 2017 was used. OPM, UNICEF, WFP, FAO (2017). Food Security and Nutrition Assessment in Karamoja Sub-Region Karamoja. OPM, UNICEF, WFP, FAO (2018). Food Security and Nutrition Assessment in Karamoja Sub-Region Karamoja.

7 The majority of communities in Karamoja have one agriculture season per year, leading to a lean season hunger gap of March to August and post-harvest season of September to February.

8 Uganda Bureau of Statistics (UBOS), Makerere University School of Public Health (MakSPH) and UNICEF (2020). Situation of Food Security and Nutrition Assessment in Northern Uganda 2019. Kampala, Uganda

DISADVANTAGES

- QUALITY CHECKS NOT BASED ON RAW DATA Unlike primary data collection, there is no opportunity for the Link NCA Analyst to revisit original questionnaires or verify responses. In case a concern about quality arises for an indicator or subset of indicators, the Analyst must either rely on contacting the Survey Manager, or dropping the indicator with justification.
- NON-CUSTOMIZED SCOPE The secondary datasets included in these analyses were part of large, regionally representative surveys, including more clusters than could feasibly be achieved in a Risk Factor Survey. However, surveys used for secondary analyses are not always specific to the geographic area of the Link NCA, and the sample size may not have been powered to that geographic area, meaning the generalizability of findings may be limited.
- LIMITED INSIGHT REGARDING CHALLENGES AND CONTEXT The Link NCA Analyst depends on the final report for any challenges that may have influenced data collection. Furthermore, if contact with the survey lead and/or initial data analyst cannot be established, it can be very challenging to understand confounders for counterintuitive findings.
- EXCLUSION OF ANTHROPOMETRICS Some surveys do not include edema as an anthropometric outcome, which is typically used in Link NCA analyses for calculation of cGAM, so a composite measure of GAM with WHZ and/or MUAC was used. Other surveys do not collect MUAC measurements, which limits assessment of wasting to WHZ, though this was not the case in the secondary datasets discussed in this article.
- PERMISSIONS ASSOCIATED WITH DATA USE Permission for use of secondary data can require time. It is important to submit requests for secondary data use during the preparatory phase, to ensure adequate time for analyses.
- INABILITY TO INFLUENCE INDICATOR SELECTION In the case of secondary datasets, the Link NCA Analyst is unable to mold the questionnaire to the hypotheses for the study. Due to the multisectoral nature of undernutrition, Technical Experts are usually encouraged not to disproportionately rank more risk factors by a particular sector, often resulting in a roughly even spread of validated risk factors across the five sectors of the study. While this varies by context, secondary datasets tend to yield fewer sectoral indicators than a Risk Factor Survey, especially for MHCP and Gender indicators. From operational experience, most secondary datasets include more Health/ NUT or FSL indicators. For example, in Somalia, the Ratio of Indicators Assessed to Hypothesized Risk Factors, by sector, ranged from 0 for MHCP to 2.5 for HEALTH/ NUT (10 Indicators: 4 Hypothesized Risk Factors). In comparison, the Liberia RFS ratio ranged from 1.8 for FSL (9 Indicators: 5 Hypothesized Risk Factors) to 4.7 for HEALTH/ NUT (14 Indicators: 3 Hypothesized Risk Factors).

Comparing an example standard Link NCA, to secondary analyses in Somalia, Uganda, and Myanmar, the overall sectoral distribution is inherently more subjective to the intent of the original survey than can be influenced in an RFS. For example, while the distribution of indicators is roughly even in the Liberia RFS, the Somalia FSNAU data was heavily skewed to HEALTH/NUT indicators, while the MCCT dataset was slightly more evenly distributed. It's important to note that no WASH or MHCP indicators were analyzed in the Mogadishu IDP FSNAU analyses, which is something that the Analyst cannot retrospectively adjust for.

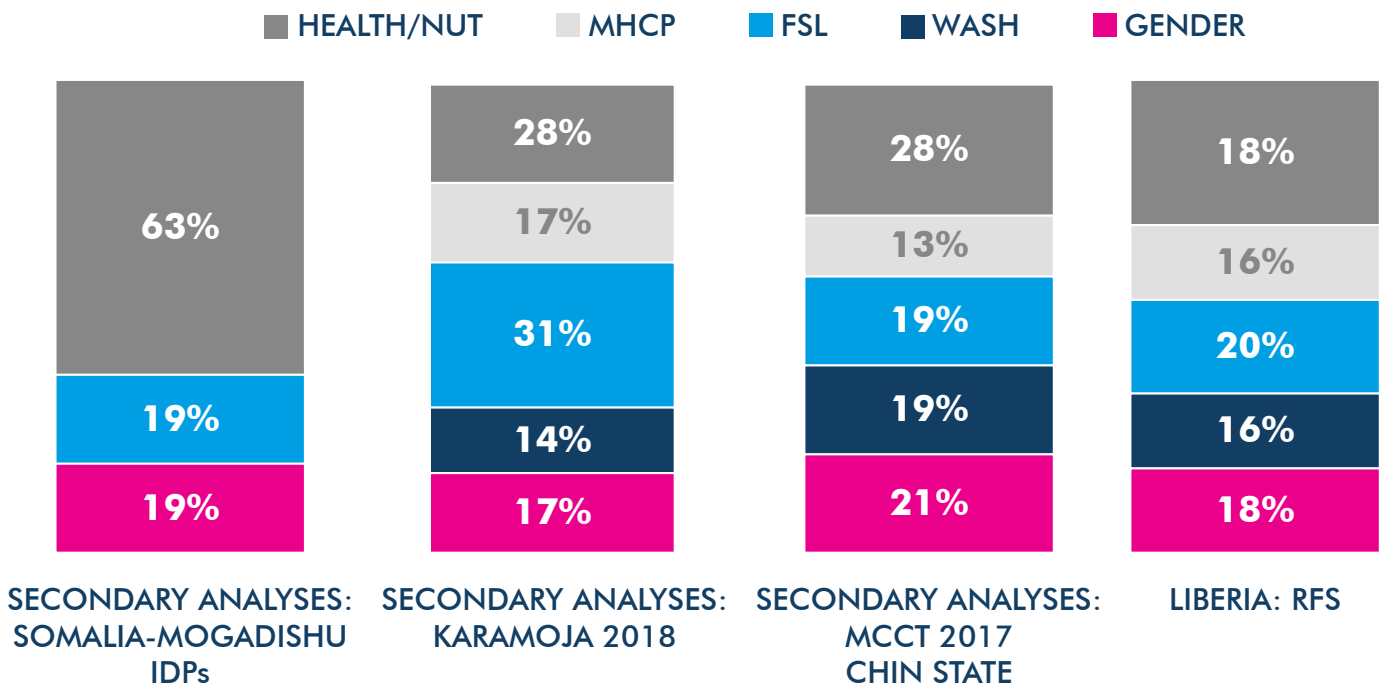

FIGURE 2: Indicators by Sector

Furthermore, it is rare that secondary datasets include semi-structured observations which have been tested in other Link NCA contexts, such as child cleanliness, kitchen hygiene, caregiver/ child interactions, etc.

RECOMMENDATIONS

Based on the experiences above, it is recommended that Link NCA Analysts probe the possibility of using secondary analyses to support the primary findings of the qualitative inquiry and Risk Factor Survey, at minimum. The following steps are recommended:

1. PROBE AVAILABILITY OF DATASETS Explore datasets during the literature phase- this includes searching nutrition/mortality surveillance studies in the geographic area of interest- i.e. DHS, MICS, SMART Surveys, FSNA, etc. The Link NCA focal point can recommend others that may be available, as well. For purposes of triangulation, a survey must include anthropometric measurements and an identifying variable to merge the child to his/her household, and his/her mother. If a potentially suitable dataset is identified:
    - Was it completed recently enough (within a two year margin) to be relevant to the current context? If a major crisis occurred between the time of the stud and the time of data collection, what are the implications affecting risk factors for child undernutrition[9]?
    - What is the process for requesting the dataset? Is it publicly available, as is typically the case with DHS, or will you require special permission to receive the dataset?
    - Is there a codebook, which provides the Analyst with provides information on the structure, contents, and layout of a data file, including coding of variables?
    - Is the questionnaire available, which will provide key insights into the structure of the household interview and also will be crucial if a codebook is not available?

9 In the case of Liberia, for example, the most recent DHS data preceded the 2016 Ebola crisis, and sizable differences were anticipated in the studied population.

2. REVIEW THE WORKPLAN For Analysts' skilled in statistical analyses, secondary dataset analyses can range from 5-14 days of work, depending on the state of the dataset. Does the Analyst have time to complete these analyses, and to complete them well?
3. MAKE CONTACT WITH THE SURVEY MANAGER AND/OR COMMISSIONING SURVEY BODY In case clarifying questions arise during data analyses, and to glean recommendations for handling the data/ suggested analyses.
4. CONDUCT A QUALITY CHECK The same quality checks conducted for Risk Factor Surveys should be applied to secondary datasets, including anthropometric plausibility checks in ENA per the SMART guidelines, identifying outlier values, calculating percentage of incomplete data, etc.
5. RE-ANALYZE RELEVANT DATA Anthropometric z-scores should be calculated from the raw anthropometric data, even if available as an indicator in the dataset. Scaled indices, such as dietary diversity scores, MAHFP, and others, should be analyzed per the Link NCA guidelines, even if a composite indicator is present in the dataset.
6. APPLY NECESSARY FLAGS WHO or SMART flags should be applied. In addition to following the Link NCA protocols for data management, justification should be documented for all deleted entries and suspected miscodes. Thresholds for erroneous data should be discussed with technical advisors and well-documented. For example, in a context where presence of more than 10 adult females in one household is considered unlikely, a threshold of ten female household members should be justified based on previous survey precedents, technical expert consensus, and then documented throughout analyses.
7. APPROPRIATELY FRAME FINDINGS In Link NCA analyses, the denominator usually represents children, even for questions presented to the entire household or mother. Thus, prevalence's will be different than those reported in most population-representative survey reports. It is important to provide the necessary disclaimers, as well as explain the data management and analysis process, to avoid concerns about discrepancies with initial results published from the secondary dataset.

CONCLUSION

Use of secondary datasets yielded meaningful supportive evidence in the recent and ongoing Uganda, Somalia, and Myanmar Link NCA studies. As is true of Risk Factor Survey analyses, secondary data analyses can explore risk factors associated with undernutrition in the studied area. Choosing secondary datasets to substitute for the primary source of quantitative data often is accompanied with a trade-off, however: large sample size, saved costs, but limited number of relevant risk factors. The stages of a Link NCA's primary data collection remain flexible to the needs of any particular study; however, based on these experiences, secondary data is best analyzed prior to or in coordination with qualitative inquiry, so the Analyst has quantitative pointers that will be further refined during the RFS. Secondary analyses can be particularly helpful in larger study zones, where a geographically comprehensive survey is infeasible and/or expensive, or to create pointers for formulating hypotheses.

ACKNOWLEDGEMENTS

We would like to thank all participating governmental institutions, implementing organizations and donors, which made these Link NCA studies possible, namely: The Federal Government of Somalia, Building Resilient Communities for Somalia (BRCiS) Consortium, Food Security and Nutrition Analysis Unit (FSNAU) Somalia and UK's Department for International Development (DFID), who conducted the Link NCA study across four livelihood zones of Somalia; the Office of the Prime Minister (OPM), UNICEF Uganda, Action Against Hunger (AAH) Uganda, Development Initiative for Northern Uganda (DINU) and EU, who conducted the Link NCA study across three regions of Northern Uganda; The Government of Myanmar, Catholic Relief Services and Save The Children, who conduct the Link NCA in Chin State of Myanmar.

ANNEX 1: Undernutrition Outcomes Assessed, by study

Context	Sample size, children under five years	Infants < 6 months assessed	Undernutrition outcomes assessed							
			Underweight (WAZ)	Stunting (HAZ)	Wasting- (MUAC)	Wasting- (WHZ)	Wasting- WHZ and/or MUAC	cGAM- WHZ, MUAC. and/or edema	WaST WHZ and HAZ	CHBB
2019-2020 Somalia										
Bay Agropastoral 2014-2017	3725[9] Gu 4009[10] Deyr 7734 **Total**	No	No	Yes	No	Yes	No	No	No	No
Hiraan / Belet Weyne Riverine, 2014-2016, 2018	2606[11] Gu 2668[12] Deyr 5274 **Total**	No	No	Yes	No	Yes	No	No	No	No
Mogadishu IDPs, 2014-2018 Gu	3891[13] Gu 4422[14] Deyr 8313 **Total**	No	No	Yes	No	Yes	No	No	No	No
Mudug Hawd Pastoral, 2014-2018 Gu	3852[15] Gu 3554[16] Deyr 7406 **Total**	No	No	Yes	No	Yes	No	No	No	No
2019 Uganda										
Karamoja, June 2017	6,525*	Yes	No	Yes	No	No	Yes	No	No	Yes
Karamoja, January 2018	6,389*	Yes	No	Yes	No	No	Yes	No	No	Yes
Mid-North 2019 FSNA	1,959	No	No	Yes	No	No	Yes	No	No	Yes
West Nile, 2019 FSNA	3,508	No	No	Yes	No	No	Yes	No	No	Yes
2020 Myanmar										
Chin State, 2017	3,445	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No

9 2014: 946, 2015: 933, 2016: 703, 2017: 1143
10 2014: 1110, 2015: 1132, 2016: 677, 2017: 1100
11 2014: 766, 2015: 825, 2016: 522, 2018: 493
12 2014: 748, 2015: 727, 2016: 544, 2018: 649
13 2014: 663, 2015: 865, 2016: 695, 2017: 844, 2018: 824
14 2014: 925, 2015: 843, 2016: 680, 2017: 800, 2018: 1174
15 2014: 563, 2015: 867, 2016: 601, 2017: 958, 2018: 863
16 2014: 508, 2015: 796, 2016: 643, 2017: 850, 2018: 757