Middle-upper arm circumference for nutritional surveillance in crisis-affected populations: Development of a method

Séverine Frison

Thesis submitted in accordance with requirements for the degree of Doctor of Philosophy. University of London. August 2016

London School of Hygiene and Tropical Medicine
Department of Population Health
Faculty of Epidemiology and Population Health

Supervisor: Dr. Marko Kerac
Associate Supervisor: Dr. Jennifer Nicholas

Funded by the Office of U.S. Foreign Disaster Assistance and the World Food Programme

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

DECLARATION OF OWN WORK

All students are required to complete the following declaration when submitting their thesis. A shortened version of the School's definition of Plagiarism and Cheating is as follows (*the full definition is given in the Research Degrees Handbook*):

"Plagiarism is the act of presenting the ideas or discoveries of another as one's own. To copy sentences, phrases or even striking expressions without acknowledgement in a manner which may deceive the reader as to the source is plagiarism. Where such copying or close paraphrase has occurred the mere mention of the source in a biography will not be deemed sufficient acknowledgement; in each instance, it must be referred specifically to its source. Verbatim quotations must be directly acknowledged, either in inverted commas or by indenting" (University of Kent).

Plagiarism may include collusion with another student, or the unacknowledged use of a fellow student's work with or without their knowledge and consent. Similarly, the direct copying by students of their own original writings qualifies as plagiarism if the fact that the work has been or is to be presented elsewhere is not clearly stated.

Cheating is similar to plagiarism, but more serious. Cheating means submitting another student's work, knowledge or ideas, while pretending that they are your own, for formal assessment or evaluation.

Supervisors should be consulted if there are any doubts about what is permissible.

DECLARATION BY CANDIDATE

I have read and understood the School's definition of plagiarism and cheating given in the Research Degrees Handbook. I declare that this thesis is my own work, and that I have acknowledged all results and quotations from the published or unpublished work of other people.

I have read and understood the School's definition and policy on the use of third parties (either paid or unpaid) who have contributed to the preparation of this thesis by providing copy editing and, or, proof reading services. I declare that no changes to the intellectual content or substance of this thesis were made as a result of this advice, and, that I have fully acknowledged all such contributions.

I have exercised reasonable care to ensure that the work is original and does not to the best of my knowledge break any UK law or infringe any third party's copyright or other intellectual property right.

To be completed by the candidate

NAME IN FULL (*Block Capitals*): ……SEVERINE GAELLE LISE FRISON……..

STUDENT ID NO: ……250557……..

SIGNED: ……SFrison…… DATE: ……12/08/16……

Registry
Last updated – 04/07/13

Abstract

Background: The assessment of prevalence of acute malnutrition in under-five children is widely used for the detection of nutritional emergencies, planning interventions, advocacy, and programme monitoring and evaluation. Current nutritional surveillance systems have important limitations. The aim of this thesis was to develop a new method for nutritional surveillance to assess acute malnutrition prevalence using PROBIT Methods based on Middle-Upper Arm Circumference (MUAC). Specific objectives were to: i) compare the appropriateness of MUAC versus other anthropometric measurements or indices to assess change in a population's nutritional status; ii) Examine assumptions behind the proposed PROBIT Methods; and iii) Assess outcomes of the proposed PROBIT Methods using estimation and classification approaches.

Methods: The first objective was achieved through a literature review. For the second objective, assumptions were tested on a database of 852 nutritional surveys including 668,975 children aged 6-59 months old. For the third objective, the Methods were assessed using data from 681,600 simulated surveys of eight different sizes.

Results: MUAC was identified as the most appropriate anthropometric measure to detect short-term changes in the nutritional status of a population; and the main assumptions behind the proposed Methods were verified. The PROBIT methods had better precision in the estimation of acute malnutrition than the Classic Method for all sample sizes tested and a better coverage for smaller sample sizes, while having relatively little bias. The classification approach performed well with a threshold of 5% acute malnutrition.

Conclusion: PROBIT Methods have a clear advantage in the assessment of acute malnutrition prevalence compared to the Classic Method. Their use would require much lower sample sizes and would enable great time- and resource-savings. There is great potential in their use in surveillance systems in order to produce timely and/or locally relevant prevalence estimates of acute malnutrition and to enable a swift and well-targeted response.

Acknowledgements

I would like to thank the following people and organisations for sharing the datasets used for this thesis: Grainne Moloney and Elijah Odundo from FSNAU, Mara Nyawo from UNICEF Khartoum, Dr. Sheila Isanaka from Epicentre/MSF Paris, Dr. Benjamin Guesdon and Cécile Salpeteur from Action Against Hunger-Paris, Dr. Anne-Marie Mayer and Gudrun Stallkamp from Concern Worldwide and Claudine Prudhon for sharing data from Goal.

We would also like to thank my advisory committee that helped me shape the project in the early stages of this thesis: Dr. Claudine Prudhon, Dr. Sheila Isanaka and Dr. Ulla Sovio.

I would like to thank Dr. Kathryn Ogden from WFP for facilitating the grants that enabled this project and Mary Marimotoo, Samantha Rowle, Simina Sacuiu as well as Jane Bruce from LSHTM for helping me navigate through the administration system of the LSHTM.

Sincere thanks to my friends and family for their encouragement and patience throughout this thesis and for their support that enabled me to finalise this project.

Finally, a huge thanks to the different supervisors I have had during this thesis: Dr. Marko Kerac and Dr. Jennifer Nicholas for their support and guidance during the last year and a half of my PhD; Jane Bruce for her great moral and technical support during the two years she kindly accepted to supervise me; Last but not least, I am extremely grateful to Dr. Francesco Checchi, my “original” supervisor. He is the reason why I decided to embark on this PhD adventure. His insight, encouragement and feedback helped me shape new ideas and improve my analysis.

Table of contents

Tables and figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

Abbreviations

ACF	Action Contre la Faim
CDC	Centres for Disease Control and Prevention
CI	Confidence Interval
CMAM	Community-based Management of Acute Malnutrition
Deff	Design Effect
DHS	Demographic and Health Surveys
ENA	Emergency Nutrition Assessment
ENN	Emergency Nutrition Network
FAO	Food and Agriculture Organisation
FSNAU	Food Security and Nutrition Analysis Unit
FSAU	Food Security Analysis Unit
FEWSNET	Famine Early Warning System Network
GAM	Global Acute Malnutrition
HFA	Height-For-Age or Length-For-Age
IASC	Inter-Agency Standing Committee
IDP	Internally Displaced Person
IQR	Inter-Quartile Range
LOESS	Locally Weighted Scatterplot Smoothing
MAM	Moderate Acute Malnutrition
MDG	Millennium Development Goals
MICS	Multiple Indicator Cluster Surveys
MSF	Médecins Sans Frontières
MUAC	Middle Upper Arm Circumference
NCHS	National Centre for Health Statistics
NGO	Non-governmental Organisation
OCHA	Office for the Coordination of Humanitarian Affairs
SAM	Severe Acute Malnutrition
SD	Standard Deviation
SE	Standard Error
SMART	Standardized Monitoring and Assessment of Relief and Transitions
SSF	Subscapular Skinfold
SSR	Standing-to-Sitting height Ratio
TSF	Triceps Skinfold
UNICEF	United Nations Children’s Fund
USAID	United States Agency for International Development
UNSCN	United Nation Standing Committee on Nutrition
WFA	Weight-For-Age
WFH	Weight-For-Height or Weight-For-Length
WHO	World Health Organisation

Key terminology

Malnutrition

"A broad term commonly used as an alternative to 'undernutrition', but which technically also refers to overnutrition. People are malnourished if their diet does not provide adequate nutrients for growth and maintenance or if they are unable to fully utilize the food they eat due to illness (undernutrition). They are also malnourished if they consume too many calories (overnutrition)" (Nutrition Glossary, UNICEF). In this thesis, malnutrition refers to undernutrition unless stated otherwise.

Anthropometry

"Anthropometry is the use of body measurements such as weight, height and mid-upper arm circumference (MUAC), in combination with age and sex, to gauge growth or failure to grow" (Nutrition Glossary, UNICEF).

Anthropometric status

"The growth status of an individual's body measurements in relation to population reference values" (Nutrition Glossary, UNICEF).

Public health surveillance

"Public health surveillance is the continuous, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice. Such surveillance can: serve as an early warning system for impending public health emergencies; document the impact of an intervention, or track progress towards specified goals; and monitor and clarify the epidemiology of health problems, to allow priorities to be set and to inform public health policy and strategies" (WHO website, health topics).

Early warning system

"An information system designed to monitor indicators that may predict or forewarn of impending food shortages, worsening of the nutritional situation or famine" (Nutrition Glossary, UNICEF).

Nutrition survey

"Nutrition surveys in emergencies assess the extent of undernutrition or estimate the numbers of children who might require supplementary and/or therapeutic feeding or other nutritional support" (Nutrition Glossary, UNICEF).

Rapid nutrition assessment

"An assessment which is carried out quickly to establish whether there is a major nutrition problem and to identify immediate needs of the population. Screening individuals for inclusion in selective feeding programmes is also a form of rapid nutrition assessment" (Nutrition Glossary, UNICEF).

Complex emergency

A complex emergency is "a humanitarian crisis in a country, region or society where there is total or considerable breakdown of authority resulting from internal or external conflict and which requires an international response that goes beyond the mandate or capacity of any single and/or ongoing UN country programme" (Inter-Agency Standing Committee, definition of complex emergencies).

Humanitarian emergency/crises

"A humanitarian emergency is an event or series of events that represents a critical threat to the health, safety, security or wellbeing of a community or other large group of people, usually over a wide area" (humanitarian coalition website).

Disaster

“A disaster is a sudden, calamitous event that seriously disrupts the functioning of a community or society and causes human, material, and economic or environmental losses that exceed the community’s or society’s ability to cope using its own resources. Though often caused by nature, disasters can have human origins” (IFRC website).

Normal distribution

The shape of the normal distribution (the characteristic "bell curve") is quantified by two parameters: the mean and the standard deviation, and follows important properties: (i) it is always symmetrical with equal areas on both sides of the curve; (ii) the highest point on the curve corresponds to the mean which equals the median and the mode; (iii) the spread of the curve is determined by the standard deviation; and (iv) as with all probability density functions the area under the curve must sum to the total probability of 1 (Essential Medical Statistics, B. Kirkwood).

Standard deviation

The standard deviation is a measure of dispersion in a frequency distribution, equal to the square root of the mean of the squares of the deviations from the arithmetic mean of the distribution (Collins English dictionary).

Probit function

The Probit function is the inverse of the cumulative distribution function of the standard normal distribution (Essential Medical Statistics, B. Kirkwood).

Preface

This PhD thesis includes a collection of research papers. These papers are related, though they have been published, or submitted, as independent research contributions.

As a result, some information has been repeated. Data collection and management is repeated in the method section of all research papers as well as some sections in the introduction.

There may also be different figures for the same condition as these may have changed over time and data was up-dated over the years.

PART I

INTRODUCTION

Chapter 1: Introduction

1.1 Humanitarian crises and actors

Humanitarian crises due to armed conflict, natural disasters, disease outbreaks and other hazards are a major and growing contributor to ill-health and vulnerability worldwide. According to the United Nations (UN), the number of people affected by humanitarian crises has almost doubled in the past decade[1]. In 2015, the Office for the Coordination of Humanitarian Affairs (OCHA) identified 125.3 million people in need of humanitarian assistance in around 37 countries in the world and 60 million displaced by conflicts or natural disasters [1].

The main actors responding to humanitarian crisis includes a wide range of agencies and organisations that can be categorised into the following groups: National and international non-governmental organisations (NGOs), the International Red Cross/Red Crescent Movement (IFRC), the UN agencies, local government institutions and donor agencies.

Need assessments of people affected by humanitarian crises is essential for strategic planning and timely and appropriate interventions. It also provides baseline information for monitoring and evaluation and helps determine the effectiveness of the humanitarian response. When carrying out an assessment, the following sectors should be prioritised: health, livelihood, water, sanitation and hygiene promotion, food and nutrition, safety, security and protection and shelter[2].

This thesis focuses on the assessment of nutritional status in crises affected populations.

1.2 Malnutrition

Malnutrition is defined as "any disorder resulting from a deficiency or excess of one or more essential nutrients"[3]. It includes:

i) undernutrition which encompasses acute malnutrition (AM) (wasting and/or oedematous malnutrition), chronic malnutrition (stunting), underweight, intra-uterine growth retardation, and micronutrient deficiencies and

ii) overnutrition (overweight and obesity).

Undernutrition is the underlying cause of child deaths associated with diarrhoea, pneumonia, malaria and measles. Overall, 45% of all death in young children is attributable to undernutrition[4, 5].

Measuring undernutrition involves the assessment of nutritional status of children (as proxy for the nutritional status of the general population) and can be done using anthropometric assessment. Weight, height or length and Middle-Upper-Arm Circumference (MUAC) are measured in order to determine anthropometric indices such as weight-for-age (WFA), weight-for-height (WFH), MUAC-for-age or height/length-for-age (HFA) and are used as indicators of underweight, wasting/AM and stunting/chronic malnutrition respectively. Low unadjusted MUAC is also used as an indicator of AM. Oedematous malnutrition is defined by the presence of bilateral pitting oedema (see paragraph below on acute malnutrition).

A child's anthropometric indices are usually compared to those of a reference population in order to classify his or her anthropometric status. The child's deviation from the central values of this distribution, as percentage of the reference median or standard deviations (SD or Z-scores) below or above the reference mean have been used to assess anthropometric status. Pioneer reference datasets for children's anthropometry were derived by Meredith from a small unrepresentative

sample of American children[6] in the 1940s. Tanner and Harvard Growth Curves were then compiled and used as reference in the 60s [7] and the Harvard growth curves were simplified and established by the World Health Organisation (WHO) as international growth reference in the late 1960s. From 1977 onwards, any child's anthropometric indices were compared to those of the National Center for Health Statistics (NCHS) reference of healthy, well-fed American children. In 2006, WHO developed new standards, developed on the basis of the results of the Multicentre Growth Reference Study (MGRS) including well-fed children from different backgrounds and cultural settings (Brazil, Ghana, India, Norway, Oman and the USA)[8]. Unlike the NCHS reference, these standards are prescriptive. They depict normal human growth under optimal environmental conditions and can be used to assess children everywhere, regardless of ethnicity, socio-economic status and type of feeding. The WHO standards were designed to be a gold standard rather than a reference and are now used worldwide.

1.3 Acute malnutrition

Acute malnutrition is particularly prominent in emergencies and is one of the basic indicators for assessing the severity of a humanitarian crisis [9]. The prevalence of acute malnutrition among children reflects the wider situation of crisis affected populations, including their food security, livelihoods, public health and social environment [10, 11].

Acute malnutrition is a major public health issue throughout the low and middle income countries. It is caused by a decrease in food consumption and/or illness and current definitions recognise two types — wasting (marasmus) and oedematous malnutrition (kwashiorkor). The United Nations Children's Fund's latest report on the State of the World's Children [12] estimates that out of the 898 million of children under 5 years old in least developed countries, approximately 81 million (9%) are wasted. Asia and West and Central Africa are the most affected regions with 15% and 11% of children wasted respectively. Out of the 6.9 million estimated deaths among children under 5 years old annually, 875 000 deaths (12.6%) [4] are attributed to wasting, which is a major determinant of child health and survival and can have devastating consequences [13-18]. Similar estimates are not available for oedematous malnutrition.

The characteristics of a marasmic child are prominent bones (ribs), skinny limbs, loose skin (on lifting) and loose skin around the buttocks (baggy pants), whereas the signs for a kwashiorkor child are presence of bilateral pitting oedema, hair changes (scanty, straight) and skin changes (dermatosis). These two types can sometimes occur simultaneously as marasmic kwashiorkor.

Acute malnutrition can be assessed and its severity classified through anthropometric measurements: low Weight-For-Height (WFH), low Middle-Upper Arm Circumference (MUAC) and the presence of bilateral pitting oedema (excessive fluids under the skin and in certain tissues, at a minimum on the dorsum of both feet) (Table 1).

The MGRS standards have systematically changed the ratio of Global Acute Malnutrition (GAM) to Severe Acute Malnutrition (SAM) prevalence. The proportion of SAM is increased; children previously classifies as Moderate Acute Malnutrition (MAM) (NCHS reference) are classified as SAM (WHO reference)[19]. Mid-Upper Arm Circumference (MUAC) was also adopted by WHO as a measure of acute malnutrition in 2005; SAM was defined by a MUAC < 110mm[20]. Discussions on the cut-off point defining SAM using MUAC led to a shift from MUAC<110mm to MUAC<115m in 2009[21]. The current classification of acute malnutrition is summarised in Table 1. Although the MUAC<125mm was not endorsed by WHO as a measure of GAM, it is widely used. Although acute malnutrition is a reliable measure to assess the nutritional situation, as well as a sensitive and objective indicator for nutritional emergency that can be assessed in almost all circumstances (i.e. even in emergency situations)[10, 11], its prevalence estimates need to be put into context, and other available data such as food security, morbidity and mortality, markets prices, access to food should be taken into account in order to interpret results. Furthermore, acute malnutrition may not always be sentive enough to detect a crisis e.g. in case of destruction of livelihoods in order to protect nutrition status.

Table 5: Acute Malnutrition definition and classification

Case definition	
Severe Acute Malnutrition (SAM)	WFH < -3 SD and/or oedema and/or MUAC<115 mm
Global Acute Malnutrition (GAM)	WFH < -2 SD and/or oedema and/or MUAC<125 mm

1.4 Nutritional surveillance: Current systems and tools for nutritional surveillance

Ongoing surveillance is an essential instrument for:

i) the detection of nutritional emergencies (early warning) and intervention;
ii) planning and advocacy;
iii) programme monitoring; and
iv) evaluation.

It provides information on trends and allows interpretation of malnutrition prevalence as compared to expected seasonal changes, i.e. what is normal for that population at that time of the year, and/or, in the absence of baseline data, arbitrary benchmarks for gravity of the nutritional situation [10, 22-26]. The use of threshold-based classifications to judge the severity of a situation was rejected by the Sphere Project[1] [27], however, WHO's classification (see table 2) of wasting prevalence based on the WFH index to assess the seriousness of a crisis is used by most humanitarian actors.

Table 6: The WHO classification of the severity of the prevalence of malnutrition[28]

Prevalence of WFH-wasting	Classification	Typical actions
< 5%	Acceptable	No action required
5% - 9%	Poor	Continue to monitor situation
10% - 14%	Serious	Intervene
≥ 15%	Critical	Immediate emergency intervention

Nutritional surveillance systems vary across settings, organisations and even within countries [29, 30]. Some of the first examples of the use of the prevalence of malnutrition to confirm humanitarian emergencies include the international

[1] Sphere Project consultative groups on Minimum Standards in Disaster Response

response to the Nigerian civil war in Biafra in the 1960s and the famine across the Sahel in the 1970s [10]. Systems like the United States Agency for International Development (USAID) Famine Early Warning System Network (FEWSNET [31]) were then developed in the 1980s. FEWSNET is now active in 25 countries and releases articles and reports on drought and food shortages. In 1994 the Food Security and Nutrition Analysis Unit (FSNAU [32]), a United Nations (UN) supported agency, was instituted in Somalia and has established a robust surveillance system which collects, analyses, and disseminates information on the food, nutrition (including nutritional anthropometric indicators) and livelihood security situation on a regular basis, based on surveys conducted throughout the country. Another important development is the Integrated Food Security Phase Classification (IPC) system, originally developed in Somalia under the FAO Food Security Analysis Unit (FSAU, now called FSNAU). It is a standardised tool that aims at providing information to classify food security using a common scale comparable across countries and is used in 42 countries. The IPC takes into account the various aspects of food security issues including health status, civil security, structural factors etc[33].

Table 3 describes the types of nutrition surveillance and the use of anthropometric measurements or indices for children under 5 years old.

Table 7: Anthropometric indices and nutrition surveillance

Type	Anthropometric indices (<5 years old)	Frequency	Cost[2]
Large scale national surveys (DHS/MICS)	WFH, WFA, HFA. Few MISC include Oedema and/or MUAC	Every 3 to 5 years	Very high
Repeated large scale surveys	WFH, WFA, HFA. Most include Oedema, MUAC	1 to 2 times/year to every 2 years	High
Repeated small scale surveys	WFH, WFA, HFA, Oedema, MUAC	2/3 times a year to every 2 years	Moderate
Sentinel site	WFH, WFA, HFA MUAC, oedema	2 to 4 times a year	Moderate/Low
Rapid assessments /screenings	MUAC/ oedema	N/A	Low
School height census	HFA	Monthly	Low
Health centre monitoring	WFA	Monthly	Low
Feeding programmes	WFH, Oedema, MUAC	Weekly or biweekly	N/A

Nutritional surveillance usually relies on cross-sectional anthropometric surveys to estimate the prevalence of malnutrition, often complemented by food security or mortality assessments [9-11]. National or large-scale surveys such as government-led Demographic and Health Surveys (DHS) and UNICEF's *Multiple Indicator Cluster Surveys* (*MICS)* are conducted in approximately 100 countries every three to ten years to follow trends on various indicators [34, 35]. Smaller-scale surveys are also conducted by various actors (INGOS, NGOs, IFRC). Since 2006, the Standardised Monitoring and Assessment of Relief and Transitions[3] (SMART) methodology has been increasingly adopted to conduct surveys at camp, district, regional or national level [9-11, 36]. It is "a standardised, simplified household-level survey methodology that provides representative and accurate nutrition and mortality data for effective decision making and resource allocation"[37]. Recently, the UNHCR developed a

[2] Very high: 0.8 million - 1.2million USD[2]; High: 80 000 USD / region; Moderate : 10,000 to 15,000 USD

[3] SMART is an inter-agency initiative, launched by a network of organizations and humanitarian practitioners.

methodology based on SMART with additional standardised questionnaires adapted to refugee settings called Standardised Expanded Nutrition Survey (SENS) [38]. Surveys are labour and resource intensive (in terms of time, logistics, and finance) especially in insecure settings, remote areas or when wide areas need to be covered (i.e. northern Kenya, Sudan)[11, 25, 26].

There are several alternatives to surveys. Sentinel sites are purposely selected sites that represent a particular population with specific livelihood systems or areas where the population is most at risk. Sentinel sites have several advantages. They can be monitored in a timely fashion, are likely to cost less and enable a more participatory approach. However, relying on these sentinel sites may result in unrepresentative findings and potential further bias due to continuous re-measurement of children[11, 39, 40]. Rapid nutrition assessments or mass screening are quick and although they are rarely representative of a population, they can help determine whether a more detailed assessment is required and are an important source of information in emergency settings. School height census has been used for nutrition surveillance but is not common practice. Although it is inexpensive and provides good population coverage, it is dependent on attendance rates. Health centre monitoring and statistics from feeding programmes can also inform on the nutritional status of a population.

1.5 The use of mean MUAC for nutritional surveillance

The case definition of AM has changed over time and while WFH was officially adopted by WHO as measure of AMdecades ago, MUAC was only adopted in 2007 for severe acute malnutrition[41]. The use of WFH versus MUAC is often debated but it is important to note that neither MUAC nor WFH can be considered a gold standard measure or index of nutritional status. MUAC-for-age is another index that assesses AM and could potentially be more sensitive to alterations in anthropometric status than MUAC alone. The use of MUAC-for-age is discussed in Annex 1.

MUAC versus WFH

The usefulness of MUAC is increasingly recognised [20] and there is accordingly a growing interest in MUAC-only programming [41-45]. There is a consensus that MUAC is a better predictor of mortality than WFH [23, 46-53] and it was reported that using MUAC alone is preferable for identifying high-risk malnourished children [54]. The MUAC tool – a tape measure – is advantageous since it is cheap, easy to transport and easy to use and interpret (e.g different colours for different status). It is the most field-appropriate anthropometric measure, with the addition of the presence of bipedal oedema, to screen and detect cases of malnutrition in communities [20, 55, 56].

Furthermore, using WFH to follow trends of AM may be problematic as it was shown that children often grow in height and in weight at different times of the year. During some seasons, children grow in length rapidly whereas their weight increases slowly, suggesting a deterioration of nutritional status in terms of WFH whereas it improves in terms of height-for-age (measure of stunting). This observation makes interpretation of these surveys difficult as for the long term outcome, reduction of stunting, is highly desirable. This difficulty is not applicable to MUAC. A reduction of the proportion of children with a low MUAC is always

positive and is easy to interpret[57]. MUAC has other advantages that are described in more detail in Chapter 2.

Advantages of assessing the mean of a nutritional measurement/ index

Nutritional indices are usually computed to estimate the prevalence of GAM or SAM. However, treating such indices as continuous variables can also give very useful information on trends and gravity levels; for example, a decrease in the mean WFH has long been recognised as a sign of a worsening nutritional situation [58-60]. The Arid Lands Resource Management Project in Kenya has included mean MUAC for many years among its indicators, and indeed MUAC trends in this setting do reflect bona fide changes in food security and burden of malnutrition [60].

Using mean MUAC as the primary anthropometric measure for nutritional surveillance would present two major operational advantages:

- Firstly, the sample size required could be smaller as the estimator would the mean of a continuous variable rather than prevalence based on a categorical variable [58]

- Secondly, using MUAC is more feasible in the community, as it does not entail transport or calibration of height and weight scales and allows for rapid screening. It does not appear to be more prone to measurement error than other indices; furthermore, focusing on one anthropometric measure would make the training as well as the supervision and analysis of the data collection easier and more feasible. An additional advantage of the use of MUAC for community surveillance is the better assessment of the number of children in need of treatment in MUAC based programmes where detection and treatment of children with a high risk of death is the priority[61]. The appropriateness of different measurement or index is discussed Chapter 2.

1.6 Rationale

Rationale for this thesis

Current nutritional surveillance systems have important limitations – they lack consistency in their frequency, and anthropometric surveys are intensive, expensive and usually detect a high prevalence of malnutrition only once it is already a serious problem. Some surveillance systems have been in place for many years but need to be improved in order to produce more comprehensive information (i.e. in Darfur or Kenya [29]). Other systems have been created but tend to disappear within a couple of years of their inception due to a lack of coordination, support and funding (i.e. Alert Site Surveillance Network created in 1997 in Somalia [56]). Broadly speaking, the current systems are not able to provide the frequency and geographic resolution of data that would assist in enabling a swift detection and response to crises before they are well established [11, 24-26, 30].

There is a consensus of opinion that ''the onset of a humanitarian disaster is often plagued by ambiguous and untimely information'' (Inter-Agency Standing Committee (IASC) Nutrition Cluster [4,62]). Furthermore, the need for a comprehensive nutrition surveillance system is clearly identified at the international level (United Nation Standing Committee on Nutrition UNSCN[5, 63]).

The rationale for this thesis was the need for evidence on how to best design nutritional surveillance systems as well as new field-practical approaches to facilitate effective data collection and analysis, and produce timely information [24-26].

[4] The IASC on nutrition is lead by UNICEF in collaboration with 33 UN, NGO, and Academic/research organizations.
[5] The UNSCN Working Group on Nutrition in Emergencies includes technical experts from the UN, INGOs, academics and independent nutritionists

The proposed Method

This thesis focused on the development of a new method to estimate acute malnutrition prevalence. The proposed approach was to estimate SAM/GAM prevalence (using MUAC based case definitions only) based on an estimate of the mean MUAC, and of its standard deviation (SD). I proposed that a point estimate and confidence interval (CI) for prevalence may be generated based on a small sample survey of mean MUAC and information about the distribution of MUAC SD values in the population, represented by the MUAC SDs observed in previous surveys within a given stratum. A database of surveys from different settings provided the required MUAC SD. The advantage of using MUAC SD from a database of previous surveys was to produce more robust estimates. The figure below summarises the proposed method.

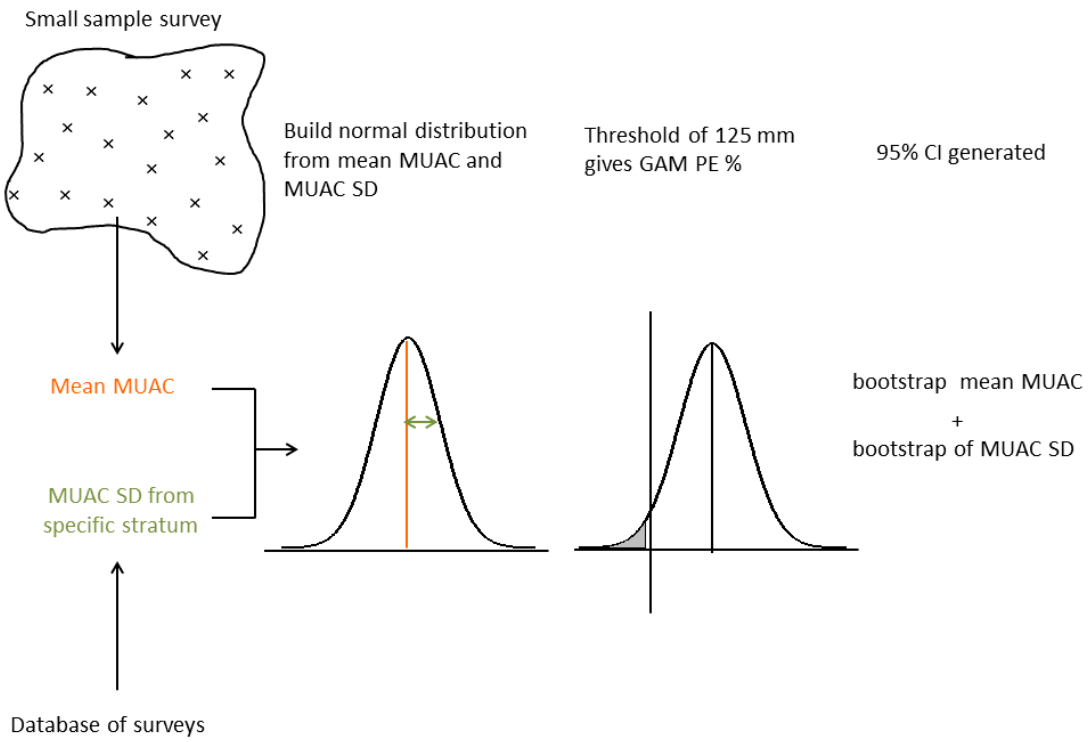


Figure 3: Overview of the proposed method

GAM : Global Acute Malnutrition ; PE: Point Estimate; SD: Standard deviation; CI : Confidence Interval

Assumptions of the above method

The above-mentioned method entails a number of critical assumptions that need to be thoroughly explored before it can be considered for field use. The following are the proposed method's key assumptions and potential limitations:

- MUAC is normally distributed in a large majority of populations and settings, or can be transformed mathematically so as to take a normal distribution; the same transformation needs to be used for *all* surveys within a given geographic stratum, meaning that at least one of the possible transformations needs to ensure normality in a large majority of surveys.

- In areas where kwashiorkor accounts for a non-negligible proportion of all SAM (e.g. parts of West, Central and Southern Africa), this method might yield considerable underestimates of SAM and GAM if MUAC cut-offs only capture a limited fraction of bilateral oedema cases: the overlap between low MUAC and oedema therefore needs to be analysed. If MUAC-based cut-offs are highly sensitive for oedema, as suggested by several studies [52, 64], the method would yield unbiased prevalence estimates, all else being equal; if on the other hand sensitivity of MUAC cut-offs for oedema were low, the method would result in under-estimation.

- The variability in MUAC SD from our database of nutritional surveys done in crisis settings is representative of the variability that we can expect in the future; furthermore, MUAC SD is itself not strongly associated with average nutritional status. If these assumptions on MUAC SD did not hold and/or could not be corrected for, prior estimates of SD for a given strata might not be applicable to the SD observed in future surveys, resulting in biased SAM or GAM prevalence.

- Within a given stratum, defined based on region, livelihood zones or residence status, there is little variability in the standard deviation (SD) of MUAC normal distribution or else estimates would be too imprecise to be useful.

An alternative approach

An alternative approach was explored in the case the proposed method above did not have the expected outcomes. The SD of the small sample survey could directly be used, along with the mean to build the desired normal distribution. This estimation of the prevalence by PROBIT computation could be done based on the small sample survey alone (i.e. not using information from previous surveys).

For the purpose of this thesis, the coverage of the method was defined as the proportion of SAM/GAM confidence intervals that contain the "true" proportion computed with the classical approach (from the full survey dataset). The precision was the range of values from the lower bound of the 95 % CI to the upper bound for each SAM/GAM estimates. Bias was defined as the estimated prevalence (from proposed method) minus "true" SAM/GAM prevalence (from full survey dataset).

1.7 Thesis aim

The overall aim of this thesis was to develop a new method for nutritional surveillance consisting of estimation of the SAM/GAM prevalence (MUAC based case definitions only) based on the mean MUAC and its standard deviation.

Thesis objectives

Specific thesis objectives are as follows:

I. Compare the appropriateness of MUAC versus other anthropometric measurements or indices

II. Examine the normality of MUAC distributions; and if necessary apply transformations to the data in order to achieve normality

III. Examine the association between MUAC and bilateral oedema

IV. Examine assumptions around MUAC SD: MUAC SD from our database is representative of the variability we can expect in the future; MUAC SD is not strongly associated with average nutritional status and; MUAC SD variability falls within a reasonably narrow range

V. Assess outcomes of the proposed PROBIT methods using estimation and classification approaches

1.8 Structure of the thesis

This thesis by publications is structured in three parts. Part I contains the introduction, literature review and the description of the data collection. Part II contains three research papers as well as an unpublished result chapter, and Part III discusses the research findings and the overall contribution of the thesis. Table 4 below summarises the structure of this thesis.

Part I

Chapter 1 provides a general introduction as well as the rational for this study, the aim and objectives.

Chapter 2 is a systematic literature review 'research paper' titled "Assessing change in the nutritional status of a population: a systematic literature review to identify the most appropriate anthropometric indicator". This paper summarises available evidence on the performance of the different anthropometric measurement and indices used in nutritional surveillance in order to appraise the appropriateness of mean MUAC for the proposed method (Objective I).

Chapter 3 describes the process of data collection and the creation of the databases used for this study.

Part II

Chapters 4 to 6 assess the different assumptions behind the proposed Method:

- Research paper 1: "Is Middle-Upper Arm Circumference "normally" distributed?" (Objective II)
- Research paper 2: "Omitting edema measurement: how much acute malnutrition are we missing?" (Objective III)

- Unpublished research results that examine the assumptions linked to the MUAC SD: Exploration of further method assumptions (Objective IV).

Chapter 7 presents the final research paper on the performance of the proposed method as well as the two alternative approaches. (I will add a title shortly) (Objective V).

Part III

Chapter 8 contains the discussion and conclusions. The main study findings are summarised and possible applications as well as the contribution of the study are discussed.

Table 4 below summarises the structure of my PhD thesis.

Table 8: Description of this thesis

Chapter	Objective	Type	Publication status
1	-	Introduction	-
2	I	Systematic literature review "research paper"	In peer review
3		Data collection & management	-
4	II	Research paper 1	Published
5	III	Research paper 2	Published
6	IV	Unpublished result	-
7	V	Research paper 3	Submitted
8	-	Discussion & conclusion	-

1.9 Publications from this thesis

- Frison S, Kerac M, Checchi Fi, Prudhon C. 2016. Anthropometric indicators to assess change in the nutritional status of a population: a systematic literature (in review; BioMed Central).

- Frison S, Checchi F, Kerac M, Nicholas J. Is Middle-Upper Arm Circumference "normally" distributed? Secondary data analysis of 852 nutrition surveys. Emerging Themes in Epidemiology. 2016 04 May;13:7. doi: 10.1186/s12982-016-0048-9. eCollection 2016. PMID: 27148390

- Frison S, Checchi F, Kerac M. Omitting edema measurement: How much acute malnutrition are we missing? American Journal of Clinical Nutrition. 2015 01 Nov;102(5):1176-81. doi: 10.3945/ajcn.115.108282Am J Clin Nutr ajcn108282

- Frison S, Checchi F, Kerac M, Nicholas J. 2016. A novel, efficient method for estimating the prevalence of acute malnutrition in resource-constrained and crisis-affected settings (submitted to PLOS One)

References

1. OCHA. *Global Humanitarian Overview*. 2016 [cited; Available from: https://docs.unocha.org/sites/dms/Documents/GHO-2016.pdf.
2. (ICRC), I.F.o.R.C.a.R.C.S. *Guidelines for assessment in emergencies*. 20087 [cited; Available from: http://www.ifrc.org/Global/Publications/disasters/guidelines/guidelines-emergency.pdf.
3. Wiley J, s., *International Dictionary of Medicine and Biology*, C. Livingstone, Editor. 1986: London.
4. Black, R., et al., *Maternal and child undernutrition and overweight in low-income and middle-income countries.* The Lancet, 2013. **382**(9890): p. 427-451.
5. Caulfield, L.E., et al., *Undernutrition as an underlying cause of child deaths associated with diarrhea, pneumonia, malaria, and measles.* The American journal of clinical nutrition, 2004. **80**(1): p. 193-8.
6. Meredith, H.V., *A physical growth record for use in elementary and high schools.* American Journal of Public Health & the Nation's Health, 1949. **39**(7): p. 878-85.
7. Stuart, H.C. and S.S. Stevenson, *Physical growth and development; in Nelson WE (ed): Text book of Pediatrics*. 5 ed. 1950, Philadelphia.
8. WHO, *Growth Standards; Length/Height-for-Age, Weight-for-Age, Weight-for-Length, Weight-for-Height and Body Mass Index-for-Age: Methods and Development*. 2006, World Health Organization.
9. Young, H., et al., *Public nutrition in complex emergencies.* Lancet, 2004. **364**(9448): p. 1899-909.
10. Young, H.J., S, *The meaning and measurement of acute malnutrition in emergencies: a primer for decision-makers.* Humanitarian Practices Network, 2006. **56**.
11. Shoham, J., F. Watson, and C. Dolan, *The use of Nutritional Indicators in Surveillance Systems.* International Public Nutrition Resource Group, 2001.
12. UNICEF, *The State of the World's Children*. 2015, UNICEF: New-York.
13. Black, R.E., et al., *Maternal and child undernutrition: global and regional exposures and health consequences.* Lancet, 2008. **371**(9608): p. 243-60.
14. Black, R.E., S.S. Morris, and J. Bryce, *Where and why are 10 million children dying every year?* Lancet, 2003. **361**(9376): p. 2226-34.
15. Collins, S., *Treating severe acute malnutrition seriously.* Arch Dis Child, 2007. **92**(5): p. 453-61.
16. Collins, S., et al., *Management of severe acute malnutrition in children.* Lancet, 2006. **368**(9551): p. 1992-2000.
17. Kapil, U. and H.P. Sachdev, *Management of children with severe acute malnutrition a national priority.* Indian Pediatr, 2010. **47**(8): p. 651-3.
18. Pelletier, D.L., et al., *The effects of malnutrition on child mortality in developing countries.* Bulletin of the World Health Organization, 1995. **73**(4): p. 443-8.

19. Seal, A. and M. Kerac, *Operational implications of using 2006 World Health Organization growth standards in nutrition programmes: secondary data analysis.* BMJ (Clinical research ed), 2007. **334**(7596): p. 733.
20. IASC, *Transitioning to the WHO Growth Standards: Implications for Emergency Nutrition Programmes*. 2008, Inter-Agency Standing Committee (IASC) Nutrition Cluster.
21. WHO and UNICEF, *WHO child growth standards and the identification of severe acute malnutrition in infants and children. A Joint Statement by the World Health Organization and the United Nations Children's Fund.* 2009.
22. Mason, J.B. and J.T. Mitchell, *Nutritional surveillance.* Bulletin of the World Health Organization, 1983. **61**(5): p. 745-55.
23. Young, H.J., S, *Review of Nutrition and Mortality Indicators for the Integrated Food Security Phase Classification (IPC): Reference Levels and Decision-Making*. 2009.
24. Tuffrey, V., *A perspective on the development and sustainability of nutrition surveillance in low-income countries.* BMC Nutrition, 2016. **2**(15).
25. Tuffrey, V., *A review of nutritional surveillance systems, their use and value.* Briefing paper, Save the Children UK and Transform Nutrition Research Consortium 2016.
26. Tuffrey, V. and A. Hall, *Methods of nutrition surveillance in low-income countries.* Emerging Themes in Epidemiology, 2016. **13**(4).
27. Sphere, *'Food Security, Nutrition and Food Aid', in Humanitarian Charter and Minimum Standards of Disaster Response*. 2004, Oxford, UK: The Sphere Project.
28. WHO, *Nutrition Landscape Information System (NLIS)*. 2010.
29. Busili, A., J. Frize, and J. Shoham, *A Review of Nutrition Information Systems in Kenya*. 2004, Save the Children UK.
30. Darcy, J.a.C.H., *According to need? Needs assessment and decision-making in the humanitarian sector*, in *Humanitarian Policy Group*. 2003.
31. *Famine Early Warning Systems Network (FEWSNET)*. [cited; Available from: www.fews.net.
32. *www.fsnau.org*. [cited.
33. *The Integrated Food Security Phase Classification (IPC)*
[cited 03/03/2016]; Available from: http://www.ipcinfo.org.
34. *The Demographic and Health Surveys Program (The DHS Program)*. [cited; Available from: http://www.dhsprogram.com/ (accessed 12 Novermber 2014).
35. UNICEF. *Multiple Indicator Cluster Survey (MICS)*. [cited.
36. SMART, *Measuring Mortality, Nutritional Status, and Food Security in Crisis Situations: SMART methodology*. 2006.
37. SMART. [cited; Available from: http://smartmethodology.org/.
38. UNHCR, *Standardised Expanded Nutrition Survey (SENS)*: http://sens.unhcr.org/.
39. Caleo, G.M., et al., *Sentinel site community surveillance of mortality and nutritional status in southwestern Central African Republic, 2010.* Popul Health Metr, 2012. **10**(1): p. 18.

40. Grellety, E., et al., *Observational Bias during Nutrition Surveillance: Results of a Mixed Longitudinal and Cross-Sectional Data Collection System in Northern Nigeria.* Plos One, 2013. **8**(5).
41. WHO, WFP, and UNICEF, *Community-based management of severe acute malnutrition. Joint statement*. 2007, WHO, WFP, UNICEF.
42. Collins, S., et al., *Key factors in the success of community-based management of severe malnutrition.* Food and nutrition bulletin, 2006: p. 49-79.
43. Dale N.M., et al., *Using Mid-Upper Arm Circumference to End Treatment of Severe Acute Malnutrition Leads to Higher Weight Gains in the Most Malnourished Children.* PLoS ONE, 2013. **8**(2): p. no pagination.
44. en-net, *Mid-Upper Arm Circumference and Weight-for-Height Z-Score as Indicators of Severe Acute Malnutrition: A Consultation of Operational Agencies and Academic Specialists to Understand the Evidence, Identify Knowledge Gaps and to Inform Operational Guidance. .* Emergency Nutrition Network, 2012.
45. Goossens, S., et al., *Mid-Upper Arm Circumference Based Nutrition Programming: Evidence for a New Approach in Regions with High Burden of Acute Malnutrition.* PLoS ONE, 2012. **7**(11): p. no pagination.
46. Briend, A., et al., *Usefulness of nutritional indices and classifications in predicting death of malnourished children.* British medical journal (Clinical research ed), 1986. **293**(6543): p. 373-5.
47. Briend, A., et al., *Nutritional status, age and survival: the muscle mass hypothesis.* European journal of clinical nutrition, 1989. **43**(10): p. 715-26.
48. Briend, A., B. Wojtyniak, and M.G. Rowland, *Arm circumference and other factors in children at high risk of death in rural Bangladesh.* Lancet, 1987. **2**(8561): p. 725-8.
49. de Onis, M., et al., *Comparison of the World Health Organization (WHO) Child Growth Standards and the National Center for Health Statistics/WHO international growth reference: implications for child health programmes.* Public Health Nutrition, 2006. **9**(7): p. 942-947.
50. Akinbami, F.O., et al., *Body mass composition: a predictor of admission outcomes among hospitalized Nigerian under 5 children.* Asia Pacific Journal of Clinical Nutrition, 2010. **19**(3): p. 295-300.
51. Alam, N., B. Wojtyniak, and M.M. Rahaman, *Anthropometric indicators and risk of death.* The American journal of clinical nutrition, 1989. **49**(5): p. 884-8.
52. Berkley, J., et al., *Assessment of severe malnutrition among hospitalized children in rural Kenya: comparison of weight for height and mid upper arm circumference.* JAMA, 2005. **294**(5): p. 591-7.
53. Vella, V., et al., *Anthropometry and childhood mortality in northwest and southwest Uganda.* American Journal of Public Health, 1993. **83**(11): p. 1616-8.
54. Briend, A., et al., *Mid-upper arm circumference and weight-for-height to identify high-risk malnourished under-five children.* Matern Child Nutr, 2012. **8**(1): p. 130-3.

55. Myatt, M., T. Khara, and S. Collins, *A review of methods to detect cases of severely malnourished children in the community for their admission into community-based therapeutic care programs.* Food and nutrition bulletin, 2006. **27**(3 Suppl): p. S7-23.
56. Prendiville, N., *Nutrition Surveillance in Somalia.* Field Exchange, 2001(14): p. 14.
57. Briend, A., T. Khara, and C. Dolan, *Wasting and stunting--similarities and differences: policy and programmatic implications.* Food and nutrition bulletin, 2015. **36**(1 Supplement): p. S15-S23.
58. Briend, A., et al., *Measuring change in nutritional status: a comparison of different anthropometric indices and the sample sizes required.* European journal of clinical nutrition, 1989. **43**(11): p. 769-78.
59. Myatt, M. and A. Duffield, *Weight-for-height and MUAC for estimating the prevalence of acute malnutrition*. 2007, IASC Global Nutrition Cluster: Global nutrition cluster meeting report.
60. Mude A, B.C., McPeak JG, Kaitho R, Kristjanson P, *Empirical Forecasting of Slow-Onset Disasters for Improved Emergency Response: an Application to Kenya's Arid North.* Food Policy, 2009. **34**(4): p. 329-339.
61. Briend, A., et al., *Low mid-upper arm circumference identifies children with a high risk of death who should be the priority target for treatment.* BMC Nutrition, 2016. **2**(63).
62. IASC, *Key Things to Know*, Inter-Agency Standing Commitee Nutrition Cluster.
63. UNSCN, *Standing Committee Nutrition statement*, in *United Nations Conference on Climate Change*: Copenhagen.
64. Sandiford, P. and F.H. Paulin, *Use of mid-upper-arm circumference for nutritional screening of refugees.* Lancet, 1995. **345**(8957): p. 1120.

Chapter 2 - Anthropometric indices and measures to assess change in the nutritional status of a population: a systematic literature review

This systematic literature review highlights the advantages of using mean MUAC for nutrition surveillance. It summarises available evidence on the performance of the different anthropometric measurement and indices used in nutritional surveillance in order to appraise the use of mean MUAC for the proposed method.

This chapter is supplemented by Annex I summarising the appropriateness of MUAC versus MUAC-for-Age. MUAC grows continuously with age, MUAC-for-Age could therefore potentially be more sensitive to alterations in anthropometric status than MUAC alone.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Severine Frison
Principal Supervisor	Marko Kerac
Thesis Title	Middle-upper arm circumference for nutritional surveillance in crisis-affected populations: Development of a method

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic peer review?	

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	BioMed Central Nutrition
Please list the paper's authors in the intended authorship order:	Severine Frison , Marko Kerac, Francesco Checchi, Claudine Prudhon
Stage of publication	Submitted and under peer-review

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am the main author and was involved in all stages from the conception and design, literature search, analysis and interpretation, to drafting the article and writing the final version to be published.

Student Signature: SFrison **Date:** 12/08/2016

Supervisor Signature: M. Kerac **Date:** 12/08/2016

The title of the article

Anthropometric indices and measures to assess change in the nutritional status of a population: a systematic literature review

Authors and affiliation

Severine Frison , Marko Kerac, Francesco Checchi, Claudine Prudhon

Severine Frison, Department of Population Health, London School of Hygiene and Tropical Medicine (LSHTM), Keppel Street, London, WC1E 7HT.
severine.frison@gmail.com
Claudine Prudhon, Save the Children, 207 *Old Street*, London EC1V 9NR.
C.Prudhon@savethechildren.org.uk
Marko Kerac, Department of Population Health, LSHTM, Keppel Street, London, WC1E 7HT. marko.kerac@lshtm.ac.uk
Francesco Checchi, Faculty of Public Health and Policy, LSHTM & Humanitarian Technical Unit, Save the Children, 207 *Old Street*, London EC1V 9NR.
f.checchi@savethechildren.org.uk

Corresponding author

Severine Frison

Abstract

Background: Undernutrition is a major public health issue highlighted by the 2015 Sustainable Development Goals, with target 2.2 aiming to 'end hunger' by 2030. On-going surveillance is an essential instrument for detecting nutritional stress in a population and is key to planning consequent interventions. Whilst methodologies of nutritional surveillance systems vary across different settings, organisations and even within the same country, the direct evidence-base underpinning these practices is limited. This paper aims therefore to: 1) compare the performance of different anthropometric indices/measurements for detecting change in the nutritional situation at population level; 2) discuss their properties and appropriateness for use in a surveillance system.

Methodology: This systematic literature review considered peer-reviewed and grey literature. Evidence was compiled from standard electronic databases, websites and snowballing. The search was performed in November 2015 by a single reviewer using the following terms to capture two concepts: 1) Undernutrition and 2) Nutrition surveillance. The search was limited to children under five and the period considered started in 1980. Languages included English and French. Articles had to assess whether the changes or trend observed at population level were statistically significant. All study designs were included.

Results: A total of 4563 articles were retrieved from the electronic database search. Most articles (3137, 89%) were not directly relevant based on title and abstract; 39 articles were reviewed in full. A total of 17 articles met the inclusion criteria and an additional 4 papers were added after snowballing. A number of measures and indices such as weight, weight-for-height/length, triceps skinfold and middle-upper arm circumference performed well in the detection of short term changes in the nutritional situation of a population. Height/Length-for-age responded the most to long term change. Applying a standard set of criteria (simplicity, acceptability, cost, independence of age, reliability and accuracy) to determine which is the most

appropriate measure or index identified middle-upper arm circumference as the one with the greatest net benefits.

Conclusion: Limited available evidence suggests that mid-upper arm circumference is the best measure to detect short term changes in the nutritional state of a population: this should receive higher priority in surveillance systems.

2.1 Introduction

Undernutrition is a major public health issue highlighted by the 2015 Sustainable Development Goals, target 2.2 aspiring to end hunger by 2030 [1]. United Nations Children's Emergency Fund's (UNICEF) latest report on the State of the World's Children [2] estimates that nearly half of all deaths in children under 5 are attributable to under-nutrition: this translates into the about 3 million young lives a year.

On-going surveillance is an essential instrument for the detection ofnutritional stress in a population, whether caused by natural or conflict related hazards. It is key to the planning of interventions. It provides information on trends and allows interpretation of malnutrition prevalence as compared to expected seasonal changes, i.e. what is normal for that population at that time of the year, and/or, in the absence of baseline data, to determine arbitrary benchmarks for gravity of the nutritional situation[3-5].

Methodologies used by nutritional surveillance systems vary across different settings, organisations and even within the same country [6, 7]. They usually rely on repeated cross-sectional anthropometric surveys [8-10]. They can also use clinic-based monitoring or sentinel sites selected to represent a particular population with specific livelihood systems or areas where the population is most at risk [10]. Common national surveys include government led Demographic and Health Surveys (DHS) and UNICEF's *Multiple Indicator Cluster Surveys* (*MICS*) that are conducted in approximately 100 countries every three to ten years [11, 12]. Many organizations also routinely use the Standardized Monitoring and Assessment of Relief and Transitions (SMART) methodology to conduct surveys at camp, district, regional or national level [13].

Children under five years are more at risk of malnutrition and more vulnerable to external shocks. It is therefore common practice to use the nutritional status of the

under-5-years population to draw conclusions about the situation of the whole population [14]. Commonly used anthropometric indices or measurements for nutrition surveillance are: weight-for-height/length (WFH/L) (wasting); Mid-Upper Arm Circumference (MUAC)acute malnutrition; Oedema (Oedematous malnutrition, also known as kwashiorkor); height/length-for-age (L/HFA) (stunting) and weight-for-age (WFA) (underweight) (table 1). Other less common indices or measurements include weight, height, birth weight, MUAC-for-Age, triceps skinfold thickness (TSF), TSF-for-Age, subscapular skinfold (SSF), head circumference and Muscular Circumference (MC) (MC=MUAC-π x TSF). To calculate nutritional indices, e.g. WFH/L, L/HFA and WFA, child's anthropometric measurements are compared to a well-fed, healthy reference population (main ones being the Harvard Growth curves in the 60s, the National Centre for Health Statistics (NCHS) reference distribution from 1978, the 2000 US Centers for Disease Control and Prevention (CDC) growth charts and the World Health Organisation (WHO) standards from 2006). The child's measurement deviation from the central values of this distribution, as percentage of the reference median or standard deviations (SD or Z-scores) below or above the reference mean have been used as estimates of anthropometric status. Measurements are then used directly or are compared to a specific threshold (e.g. MUAC< 115mm is used to diagnose severe acute malnutrition). Table one presents the most commonly used measurement and indices in children under five.

Table 1: Common anthropometric measurements and indices in children under five

Index	**Nutritional problem measured**	**Indicator**
Weight-for-Height/Length (WFH/L)	Severe wasting	WFH/L < -3 SD
	Moderate wasting	WFH /L< -2 SD and WFH /L≥ -3 SD
	Global wasting	WFH/L < -2 SD
Height/Length-for-age (H/LFA)	Severe stunting	H/LFA < -3 SD
	Moderate stunting	H/LFA < -2 SD and H/LFA ≥ -3 SD
	Global stunting	H/LFA< -2 SD
Weight-for-age (WFA)	Severe underweight	WFA < -3 SD
	Moderate underweight	WFA < -2 SD and WFA ≥ -3 SD
	Global underweight	WFA < -2 SD
Measurement	**Nutritional problem measured**	**Indicator**
MUAC	Severe Actue Malnutrition (SAM)	MUAC<115 mm
	Global Acute Malnutrition (GAM)	MUAC<125 mm
Oedema	Oedematous malnutrition	Bilateral oedema below the ankles: + Bilateral oedema up to knees: ++ Bilateral oedema up to arms and higher:+++

Nutritional surveillance generally measures point estimates/prevalence or incidence of malnutrition. However, treating nutritional indices as continuous variables can also give very useful information on trends and gravity levels; for example, a decrease in the mean and distribution of WFH, MUAC or weight has been recognised as a sign of a worsening nutritional situation [15-17].

Though they are commonly used, there is a very limited direct evidence-based exploring the usefulness of the different indices at detecting change in nutritional status of a population. This review aims to 1) compare the performance of the different anthropometric indices/measurements in the detection of change in the nutritional situation at population level (long term i.e. over a year and short term i.e. few months/season) and 2) discuss their properties and appropriateness for use in a surveillance system.

2.2 Methodology

This systematic literature review considered peer-reviewed and grey literature. Evidence was compiled from standard electronic databases, websites and snowballing (reference list from relevant primary studies and review articles).

Exclusion criteria

The search excluded paper on adults and adolescents and was limited to children under five. Articles prior to 1980 were not considered. Languages included English and French. Articles had to assess whether the changes or trend observed at population level were statistically significant. All study designs were included.

Search strategy

The peer-reviewed literature search was conducted using Embase, Global Health and Medline. The search was performed in November 2015 by a single reviewer using the following terms to capture two concepts: 1) Undernutrition: ((arm or midarm or mid-arm or mid-upper arm) and circumference) or MUAC or weight-for-height or weight-for-length or WHZ or WHM or weight- for- age or WAZ or height-for-age or length-for-age or HAZ or kwashiorkor or oedema or WAM or HAM or weight or height or anthropometry or anthropometric indices or anthropometric indicators or stunting or wasting or acute malnutrition or marasmus or underweight AND 2) Nutrition surveillance: Nutrition$ assessment or nutrition$ survey or nutrition$ surveillance or nutrition$ situation or malnutrition prevalence or nutrition$ monitoring or nutrition$ screening or nutrition$ evaluation or nutrition$ early warning system or nutrition$ change or nutrition$ variation or nutrition$ impact or season$ change or season$ variation.

Grey literature undertaken by searching the following websites: Emergency Nutrition Network (ENN), The United Nations System Standing Committee on Nutrition and the Community-based Management of Acute Malnutrition (CMAM) forum [18-20].

Data extraction, analysis and reporting

Returned citations were downloaded to Endnote software and a five-stage screening process applied (see Figure 1). Articles that met the inclusion criteria were selected and data abstracted in an excel sheet. The following data were extracted from each paper: i) study authors, ii) year; iii) study country and collection period, iv) setting, (v) type of study, vi) sample size, vii) age group, viii) independent variables, ix) dependent variables, x) reference and unit, xi) outcome of the study. The outcome of the study included prevalence, means and Odd Ratios (OR) with associated p-values. Descriptive analysis was used and the systematic review methodology adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement[21]. We did not register the review protocol and this review does not include a bias analysis. Papers included were too different to be able to do a synthesis and very few of them had as objective to assess performance of different anthropometric measurements or indices in the detection of change in the nutritional status of the study population.

2.3 Results and discussion

Performance of nutritional measurement/indices to detect changes in nutritional situations

A total of 4563 articles were retrieved from the electronic database search (1837 articles from Embase, 1102 from articles Global Health and 1624 articles from Medline) out of which 1033 duplicates were excluded leaving 3530 articles to review. A large majority of articles (3137, 89%) were found out of topic and 39 articles were left for full review. A total of 17 articles met the inclusion criteria. All potential articles found in the grey literature had been published and therefore included in the above search. An additional 4 papers were added after snowballing. A total of 21 articles were included in this review. Figure 1 flow diagram summarises the search.

Over half of the articles included were published in the 2000s (12, 57%), a fifth (4, 19%) in the 90s and a quarter (5, 24%) in the 80s which translates recent interest in the topic. Although the African continent is overrepresented (52% of studies), we believe this does not affect the generalisability of the findings as we are interested in the capacity to detect change within the same population. Most studies were conducted in rural areas (15, 71%) while few were implemented in urban (3, 14%) or both urban and rural (3, 14%) settings. Different types of design were used to conduct the studies included which made it difficult to compare outcomes. Longitudinal (9, 43%) and repeated cross sectional studies (7, 33%) were the predominant types. Most studies examined the effect of seasonality on malnutrition (17, 80%). Different sets and numbers of dependent variable as well as different references and types of analysis were used which made comparison and generalisations difficult. The main dependent variable analysed was weight-for-height/length (18, 86%) followed by weight-for-age (13, 62%), height/length-for-Age (13, 62%) and mid-upper arm circumference (6, 29%). Most studies included three or more dependent variables. Half (10, 50%) of the analysis treated

dependent variables as continuous (mean) and binary (prevalence). Just over half of the analyses (11, 55%) used the NCHS reference and Z-score was the most common unit (14, 67%) (Table 2). The detailed characteristics of each study can be found in the extraction sheet table 3(at the end of the chapter).

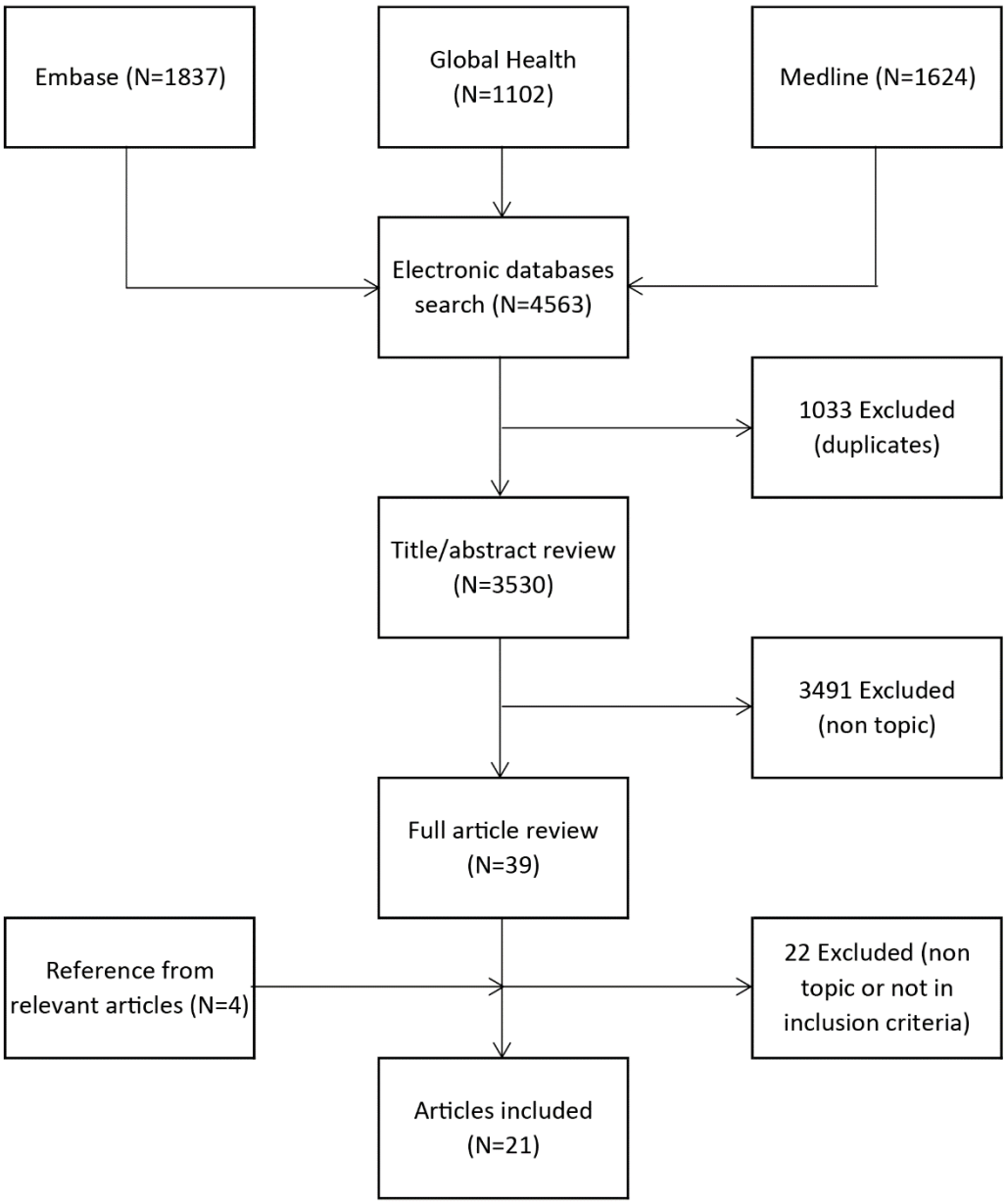


Figure 1: Search flow diagram

Table 2: Characteristics of the studies included (N=21)

Continent[6]	**N (%)**
Africa	11 (52)
Asia	9 (43)
Latin America	1(5)
Setting	**N (%)**
Rural	15 (71)
Urban	3 (14)
Both	3 (14)
Type of study	**N (%)**
Longitudinal study	9 (43)
Repeated cross-sectional studies	7 (33)
Cohort	3 (14)
Secondary data analysis	1(5)
Growth monitoring data (health centre)	1(5)
Age group	**N (%)**
0 to 59 months	4 (19)
6 to 59 months	4 (19)
12 to 59 months	3 (14)
6 to 36 months	2 (10)
Other[7]	8 (38)
Independent variable	**N (%)**
Seasonal change	17 (81)
Devaluation of CFA franc	1(5)
Drought and financial crisis 1997/1998	1(5)
Herd dynamic, food, biophysical and	1(5)
Seasonal change and change over the years	1(5)
Dependent variable	**N (%)**
WFHL	18 (86)
WFA	13 (62)
H/LFA	12 (57)
MUAC	7 (33)
Weight	7 (33)
Height	7 (33)
MUAC-for-Age	3 (14)
Other[8]	6 (29)
Number of dependent variables	**N (%)**
One	4 (19)
Two	1(5)

[6] Africa: 2 in Ethiopia, 2 in Kenya and 1 in each of the following: Chad, Congo, Malawi, Niger, Senegal, Zimbabwe; Asia: 4 in Bangladesh and 1 in each of the following: India, Indonesia, Vietnam, West Timor, Nepal; South America: Peru

[7] 0 to 36 months, 6 to 36 months, 6 to 72 months, 0 to 50 months, 6 to 24 months, 12 to 36 months, 24 to 59 months, 6 months to 10 years

[8] Body Mass Index (BMI), Head Circumference (HC), Triceps SkinFold (TSF), TSF-for-Age, Subscapular SkinFold (SSF), Muscle Circumference (MC), birth weight

Three	8 (38)
Four	1(5)
Five	4 (19)
Six +	3 (14)
Variable treatment	**N (%)**
Change in mean and prevalence of indice(s)	10 (48)
Change in mean indice(s)	7 (33)
Change in prevalence of indice(s)	6 (29)
Standard	**N (%)**
NCHS	11 (52)
WHO	7 (33)
Harvard	2 (10)
CDC-2000	1(5)
Unit	**N (%)**
Z-score	14 (67)
% median	5 (24)
% of median & Z-score	2 (10)

HH, Household; WFH, Weight-For-Height; WFL, Weight-For-Lenght; HFA, Height-for-Age; LFA, Lenght-for-Age WFA, Weight-for-Age; MUAC, Middle-Upper Arm Circumference; NCHS, National Center for Health Statistics; WHO, World Health Organisation; CDC, Centers for Disease Control and Prevention

Table 4 (at the end of the chapter) presents the summary of the study outcome for each measurement/index. Means and/or prevalence of the measurement and/or indices examined in the 21 papers generally varied significantly between seasons or before/after external hazards. Few studies showed no or few differences. Egata et al[22] showed no difference in mean WFH and mean MUAC. They argued that good food security was common regardless of the seasonal variation. Huong et al[23] found no change in weight, height, WFA, H/LFA and WFH/L but the small sample sizes (around 200 children 24 to 59 months) involved as well as the design of the study (repeated cross sectional studies) were not ideal to detect differences. Loutan et al[24] showed no differences in WFH and MUAC but had a very small sample size (around 30 children under five years).

Out of the 21 studies included, 4 (19%) compared the change in mean and/or prevalence of several measurements/indices. Benefice et al[25] presented variations of mean MUAC, WFH, TSF and MC in a longitudinal study in rural Senegal. Mean

WFH/L was the only index that was not changing significantly and TSF showed the largest differences. Briend et al[15] examined the effect of seasonal change as well as the change between the first two years of the study and the last two on weight, WFA, H/LFA, WFH/L, and MUAC. This study revealed that Weight, MUAC and WFH were the nutritional indices that changed the most between seasons. Mean H/LFA, WFA and MUAC were significantly higher during last 2 years. This was more pronounced for HFA. In a study assessing the fluctuations of the mean weight, height increment, WFA, HFA, WFH, MUAC-for-Age and TSF-for-Age, Brown et al[26] found that TSF-for-Age had the greatest seasonal change. WFA and MUAC-for-Age followed the same patterns and magnitude while WFH had greater range but similar coefficient of variation as MUAC and smaller than WFA. Finally, Garenne et al[27] study looked at seasonal changes of mean WFA, WFH, MUAC, TSF, weight, height, BMI, MC, SSF and HC. The highest contrast value was observed for mean MUAC which made it the best measurement for the detection of short term changes. Mean weight, height and head circumference had the highest responsiveness. Responsiveness was defined as a measure of the change over a semester compared with the variation of the indicator in the population (change divided by the standard deviation of the same indicator). These indices were the most appropriate to monitor growth velocity of children in a stable situation (Table 4).

Unsurprisingly, H/LFA was mainly out of phase compared to other measures of undernutrition and was a good measure of long term change (Briend et al[15], Brown et al[26], Huong et al[23], Marin et al[28], Martin-Prevel et al[29], Miller et al[30], Panter-Brick et al[31]) (Table 4).

The capacity to detect change in the nutritional status of the population did not seem to differ whether the anthropometric measurements/indices were treated as continuous or binary. However, the sample size requirement differs whether assessing the mean of a continuous variable or looking at the prevalence of a binary

variable. Using means allows for smaller sample size which has important implications in terms of logistics, costs and timeliness. This was confirmed in Briend et al paper.

Appropriateness for use in surveillance systems

In addition to its responsiveness to nutritional stress, a number of important criteria need to be taken into account to identify the most appropriate and relevant measure or index to be used to detect changes: simplicity, acceptability, cost, independence of age, reliability and accuracy, sensitivity and specificity [32].

Simplicity: Any index that includes an age component requires that age be ascertained accurately and it is widely acknowledged that determining age correctly is problematic in many developing countries [33-36]. The use of multi-component indices (i.e. WFA, HFA, WFH, MUAC-for-Age, TSF-for-Age) is usually more complex [33, 37]. Moreover, transporting and carrying weight scales as well as height/length board is more logistically challenging than for MUAC tapes.

Acceptability: The measurement of weight, height and MUAC is widely accepted and commonly performed in nutrition surveillance and interventions. A study reported that younger children tended to become upset and agitated during both weight and height measurement but not during MUAC measurement [37]. TSF index is not currently used for surveillance or programming and would probably not be as acceptable as the measures above as it requires the measurement of the width of a fold of skin taken over the triceps muscle using a skinfold caliper.

Cost: The measurement of height and weight requires fairly costly equipment [37-39] while the MUAC tool – a tape measure – is cheap and easy to transport. A caliper is also relatively costly and may be harder to procure.

Independence of age: One way to ascertain age-independence is to adjust indices for age (i.e. WFA, HFA, MUAC-for-Age) but the issue of the accuracy of age remains. MUAC is relatively age and sex independent among 1-5 year olds[32, 37, 40-43] as well as

WFH [39, 44]. It was also shown that MUAC alone, without correction for age, was a better predictor of death than indices based on height, weight and age [41-43, 45].

Reliability and accuracy: Although weight and height alone were shown to be more precisely measured [46-49], it was reported that MUAC has a better reliability than WFH and shows better performance in screening programmes [50]. It was also shown that in field conditions, minimally trained workers make fewer and smaller errors in screening children with MUAC than with WFH [37]. Indices usually require finding values in tally sheets or calculations that can lead to further errors. A recent paper shows that MUAC is more reliable than WFH [51] and another that MUAC outperforms weight-based measures of nutritional status in children with diarrhoea [52]. It was also shown to be less affected by dehydration than WFH [53]. As mentioned above, any index requiring the age (i.e. WFA or HFA) of the child is likely to be less accurate.

Sensitivity and specificity (to mortality): MUAC is increasingly recognised as a very useful index of nutritional status [50]. There is a consensus that MUAC is a better predictor of mortality than WFH [40, 45, 54-60] and it was recently reported that using MUAC alone is preferable for identifying high-risk malnourished children [61].

Table 5 summarises the characteristics of all relevant measures and indices reviewed. We focus on measures and indices that are currently in use in nutrition programming and nutrition surveillance (i.e. we did not discuss TSF, TSF-for-Age, MUAC-for-Age, MC, birth weight). Table 5 highlights the advantages of using MUAC over other measures or indices detecting short term changes.

These findings are consistent with the increasing interest in MUAC-only nutrition programming and use for admission and discharge to feeding programmes [62-66]. This concordance makes the findings of this review applicable and of interest to international policy makers and programme managers.

Table 3: Characteristics of measures and indices

Measure or index	Detect short term change	Detect long term change	Simplicity	Acceptability	Cost	Independence of age	Reliability and accuracy
WFH/L	+++	++	+	+++	++	++	+
WFA	++	++	+	+++	++	+++	+
H/LFA	++	+++	+	+++	++	+++	+
MUAC	+++	++	+++	+++	+++	++	++

WFA, Weight-For-Age; H/LFA, Height/Length-For-Age; WFH/L, Weight-For-Height/length; MUAC, Middle Upper Arm Circumference;

+++ = Good; ++ = Fair ; +=Poor

Limitations

We acknowledge the limitations to our review, the most important being:

1) Great heterogeneity (population; setting; study design; methods; time periods; primary research question) between the studies found: this makes it difficult to carry out any quantitative analysis / meta-analysis to compare the performance of different measures and indices
2) A single reviewer performed the search which may have lead to errors or omissions
3) Publication bias: studies that were unable to assess changes or trends at population level are less likely to be published
4) The observational nature of the studies: it is not possible to directly 'test' the performance of one indicator against another in an interventional study
5) There is no gold standard measure of population nutritional status. Where no change is observed, we cannot know whether there really was no change in the population or whether a real change was simply not detected by indices used (i.e. not sensitive enough)
6) We did not look at over-nutrition. MUAC might not be the best index when measuring obesity, an increasing problem even in resource poor settings[67].

Strengths

Balancing these limitations, a major strength of our review is that we explore a highly policy/practice-relevant question using a systematic approach. By highlighting the overall limited evidence base we hope to stimulate both more and better-quality future research in this area. We also provide a framework whereby policy makers and managers can think about the different aspects of indicator performance: different indicators may suit different questions and in choosing which is 'best' it is vital to consider context. Different aspects of malnutrition that may be better monitored by different sets of indictors such as in DHS or MICS. The measurement or index to use also depends on the nature and intensity of the crises. In some crises where diets might still be sufficient to maintain weight but have lost adequacy in micronutrient, the change in stunting might be significant but not in wasting. This has been the case in recent conflicts [68]. Finally a strength is that we highlight an indicator – MUAC - that is still missing from many major surveys such as DHS. This is an important gap given MUAC's good performance for detecting short term changes in population nutritional status. This has major implications for early warning systems or other assessments systems which only allow for limited field data collection because of time or budget constraints.

Future research should look at cost-effectiveness and logistics issues of different systems as this is critical to successful and sustained large-scale rollout of any system. Especially with the large number of SDGs, there is increasing pressure to make efficient use of resources.

2.4 Conclusion

A number of measures and indices such as weight, WFH, TSF and MUAC perform well in the detection of short term changes in the nutritional situation of a population. However, after applying a set of criteria which are critical to successful large-scale rollout (simplicity; acceptability; cost; independence of age; reliability; and accuracy) MUAC stands out strongly as the best measure to use in nutritional surveillance systems to detect short term changes in the nutritional status of a population.

References

1. *Sustainable Development Goals*. [cited 08/08/2016]; Available from: https://sustainabledevelopment.un.org/.
2. UNICEF, *The State of the World's Children*. 2015, UNICEF: New-York.
3. *The Integrated Food Security Phase Classification (IPC)* [cited 03/03/2016]; Available from: http://www.ipcinfo.org.
4. WHO, *Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee*, in *World Health Organization Technical Report Series*. 1995. p. 1-452.
5. Tuffrey, V. and A. Hall, *Methods of nutrition surveillance in low-income countries.* Emerging Themes in Epidemiology, 2016. **13**(4).
6. Busili, A., J. Frize, and J. Shoham, *A Review of Nutrition Information Systems in Kenya*. 2004, Save the Children UK.
7. Darcy, J.a.C.H., *According to need? Needs assessment and decision-making in the humanitarian sector*, in *Humanitarian Policy Group*. 2003.
8. Young, H., et al., *Public nutrition in complex emergencies.* Lancet, 2004. **364**(9448): p. 1899-909.
9. Young, H.J., S, *The meaning and measurement of acute malnutrition in emergencies: a primer for decision-makers.* Humanitarian Practices Network, 2006. **56**.
10. Shoham, J., F. Watson, and C. Dolan, *The use of Nutritional Indicators in Surveillance Systems.* International Public Nutrition Resource Group, 2001.
11. *The Demographic and Health Surveys Program (The DHS Program)*. [cited; Available from: http://www.dhsprogram.com/ (accessed 12 Novermber 2014).
12. UNICEF. *Multiple Indicator Cluster Survey (MICS)*. [cited.
13. SMART, *Measuring Mortality, Nutritional Status, and Food Security in Crisis Situations: SMART methodology*. 2006.
14. WFP and CDC, *A Manual: Measuring and Interpreting Malnutrition and Mortality*. 2005.
15. Briend, A., et al., *Measuring change in nutritional status: a comparison of different anthropometric indices and the sample sizes required.* European journal of clinical nutrition, 1989. **43**(11): p. 769-78.
16. Myatt, M. and A. Duffield, *Weight-for-height and MUAC for estimating the prevalence of acute malnutrition*. 2007, IASC Global Nutrition Cluster: Global nutrition cluster meeting report.
17. Mude A, B.C., McPeak JG, Kaitho R, Kristjanson P, *Empirical Forecasting of Slow-Onset Disasters for Improved Emergency Response: an Application to Kenya's Arid North.* Food Policy, 2009. **34**(4): p. 329-339.
18. *Emergency Nutrition Network (ENN)*. [cited; Available from: http://www.ennonline.net/.
19. *United Nation Systems Standing Committee on Nutrition (UNSCN)*. [cited; Available from: http://www.unscn.org/.
20. *Community-based Management of Acute Malnutrition (CMAM) forum*. [cited; Available from: http://www.cmamforum.org/.

21. Moher, D., et al., *Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement.* BMJ (Online), 2009. **339**(7716): p. 332-336.
22. Egata, G., Y. Berhane, and A. Worku, *Seasonal variation in the prevalence of acute undernutrition among children under five years of age in east rural Ethiopia: a longitudinal study.* BMC public health, 2013. **13**: p. 864.
23. Huong le, T., et al., *Diet and nutritional status among children 24-59 months by seasons in a mountainous area of Northern Vietnam in 2012.* Glob Health Action, 2014. **7**: p. 23121.
24. Loutan, L. and J.M. Lamotte, *Seasonal variations in nutrition among a group of nomadic pastoralists in Niger.* Lancet, 1984. **1**(8383): p. 945-947.
25. Benefice, E., S. Chevassus-Agnes, and H. Barral, *Nutritional situation and seasonal variations for pastoralist populations of the Sahel.* Ecology of Food and Nutrition, 1984. **14**: p. 229-247.
26. Brown, K.H., R.E. Black, and S. Becker, *Seasonal changes in nutritional status and the prevalence of malnutrition in a longitudinal study of young children in rural Bangladesh.* American Journal of Clinical Nutrition, 1982. **36**(2): p. 303-313.
27. Garenne, M., et al., *Adequacy of child anthropometric indicators for measuring nutritional stress at population level: a study from Niakhar, Senegal.* Public Health Nutrition, 2012: p. 1-7.
28. Marin, C.M., et al., *Seasonal change in nutritional status among young children in an urban shanty town in Peru.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 1996. **90**(4): p. 442-445.
29. Martin-Prevel, Y., et al., *Deterioration in the nutritional status of young children and their mothers in Brazzaville, Congo, following the 1994 devaluation of the CFA franc.* Bulletin of the World Health Organization, 2000. **78**(1): p. 108-118.
30. Miller, J., et al., *Seasonal variation in the nutritional status of children aged 6 to 60 months in a resettlement village in West Timor.* Asia Pacific Journal of Clinical Nutrition, 2013. **22**(3): p. 449-456.
31. Panter-Brick, C., *Seasonal growth patterns in rural Nepali children.* Annals of Human Biology, 1997. **24**(1): p. 1-18.
32. Myatt, M., T. Khara, and S. Collins, *A review of methods to detect cases of severely malnourished children in the community for their admission into community-based therapeutic care programs.* Food and nutrition bulletin, 2006. **27**(3 Suppl): p. S7-23.
33. Hamer, C., et al., *Detection of severe protein-energy malnutrition by nurses in The Gambia.* Arch Dis Child, 2004. **89**(2): p. 181-4.
34. Jelliffe EFP, J.D., *The arm circumference as a public health index of protein-calorie malnutrition of early childhood.* Journal of tropical pediatrics, 1969. **15**(4): p. 177-260.
35. Bairagi, R. and R.I. Ahsan, *Inconsistencies in the findings of child nutrition surveys in Bangladesh.* American Journal of Clinical Nutrition, 1998. **68**(6): p. 1267-71.

36. Bairagi, R., B. Edmonston, and A.D. Khan, *Effects of age misstatement on the utility of age-dependent anthropometric indicators of nutritional status in rural Bangladesh.* American Journal of Public Health, 1987. **77**(3): p. 280-2.
37. Velzeboer, M.I., et al., *The use of arm circumference in simplified screening for acute malnutrition by minimally trained health workers.* Journal of tropical pediatrics, 1983. **29**(3): p. 159-66.
38. Davis, L.E., *Epidemiology of famine in the Nigerian crisis: rapid evaluation of malnutrition by height and arm circumference in large populations.* American Journal of Clinical Nutrition, 1971. **24**(3): p. 358-64.
39. Ross, D.A., et al., *Measuring malnutrition in famines: are weight-for-height and arm circumference interchangeable?* Int J Epidemiol, 1990. **19**(3): p. 636-45.
40. Alam, N., B. Wojtyniak, and M.M. Rahaman, *Anthropometric indicators and risk of death.* The American journal of clinical nutrition, 1989. **49**(5): p. 884-8.
41. Bairagi, R., *On validity of some anthropometric indicators as predictors of mortality.* American Journal of Clinical Nutrition, 1981. **34**(11): p. 2592-4.
42. Briend, A. and S. Zimicki, *Validation of arm circumference as an indicator of risk of death in one to four year old children.* Nutrition Research 1986(6): p. 249-261.
43. Rasmussen, J., et al., *Mid-upper-arm-circumference and mid-upper-arm circumference z-score: The best predictor of mortality.* European Journal of Clinical Nutrition, 2012. **66**(9): p. 998-1003.
44. Waterlow, J.C., *Classification and definition of protein-calorie malnutrition.* British Medical Journal, 1972. **3**(5826): p. 566-9.
45. Briend, A., et al., *Usefulness of nutritional indices and classifications in predicting death of malnourished children.* British medical journal (Clinical research ed), 1986. **293**(6543): p. 373-5.
46. Sicotte, M., et al., *Reliability of anthropometric measures in a longitudinal cohort of patients initiating ART in West Africa.* BMC Medical Research Methodology, 2010. **10**: p. 102.
47. Ulijaszek SJ, K.D., *Anthropometric measurement error and the assessment of nutritional status.* Br J Nutr 1999. **82**(3): p. 165-77.
48. Mwangome, M.K. and J.A. Berkley, *The reliability of weight-for-length/height Z scores in children.* Matern Child Nutr, 2014. **10**(4): p. 474-80.
49. Ayele, B., et al., *Reliability of Measurements Performed by Community-Drawn Anthropometrists from Rural Ethiopia.* PLoS ONE [Electronic Resource], 2012. **7**(1).
50. IASC, *Transitioning to the WHO Growth Standards: Implications for Emergency Nutrition Programmes*. 2008, Inter-Agency Standing Committee (IASC) Nutrition Cluster.
51. Mwangome, M.K., et al., *Reliability and accuracy of anthropometry performed by community health workers among infants under 6 months in rural Kenya.* Tropical Medicine & International Health, 2012. **17**(5): p. 622-629.

52. Modi, P., et al., *Midupper Arm Circumference Outperforms Weight-Based Measures of Nutritional Status in Children with Diarrhea.* J Nutr, 2015. **145**(7): p. 1582-7.
53. Mwangome, M.K., et al., *Are diagnostic criteria for acute malnutrition affected by hydration status in hospitalized children? A repeated measures study.* Nutrition Journal, 2011. **10**.
54. Briend, A., et al., *Nutritional status, age and survival: the muscle mass hypothesis.* European journal of clinical nutrition, 1989. **43**(10): p. 715-26.
55. Briend, A., B. Wojtyniak, and M.G. Rowland, *Arm circumference and other factors in children at high risk of death in rural Bangladesh.* Lancet, 1987. **2**(8561): p. 725-8.
56. de Onis, M., et al., *Comparison of the World Health Organization (WHO) Child Growth Standards and the National Center for Health Statistics/WHO international growth reference: implications for child health programmes.* Public Health Nutrition, 2006. **9**(7): p. 942-947.
57. Young, H.J., S, *Review of Nutrition and Mortality Indicators for the Integrated Food Security Phase Classification (IPC): Reference Levels and Decision-Making*. 2009.
58. Akinbami, F.O., et al., *Body mass composition: a predictor of admission outcomes among hospitalized Nigerian under 5 children.* Asia Pacific Journal of Clinical Nutrition, 2010. **19**(3): p. 295-300.
59. Berkley, J., et al., *Assessment of severe malnutrition among hospitalized children in rural Kenya: comparison of weight for height and mid upper arm circumference.* JAMA, 2005. **294**(5): p. 591-7.
60. Vella, V., et al., *Anthropometry and childhood mortality in northwest and southwest Uganda.* American Journal of Public Health, 1993. **83**(11): p. 1616-8.
61. Briend, A., et al., *Mid-upper arm circumference and weight-for-height to identify high-risk malnourished under-five children.* Matern Child Nutr, 2012. **8**(1): p. 130-3.
62. Collins, S., et al., *Key factors in the success of community-based management of severe malnutrition.* Food and nutrition bulletin, 2006: p. 49-79.
63. Dale N.M., et al., *Using Mid-Upper Arm Circumference to End Treatment of Severe Acute Malnutrition Leads to Higher Weight Gains in the Most Malnourished Children.* PLoS ONE, 2013. **8**(2): p. no pagination.
64. en-net, *Mid-Upper Arm Circumference and Weight-for-Height Z-Score as Indicators of Severe Acute Malnutrition: A Consultation of Operational Agencies and Academic Specialists to Understand the Evidence, Identify Knowledge Gaps and to Inform Operational Guidance. .* Emergency Nutrition Network, 2012.
65. Goossens, S., et al., *Mid-Upper Arm Circumference Based Nutrition Programming: Evidence for a New Approach in Regions with High Burden of Acute Malnutrition.* PLoS ONE, 2012. **7**(11): p. no pagination.
66. WHO, WFP, and UNICEF, *Community-based management of severe acute malnutrition. Joint statement*. 2007, WHO, WFP, UNICEF.

67. Grijalva-Eternod, C.S., *The Double Burden of Obesity and Malnutrition in a Protracted Emergency Setting: A Cross-Sectional Study of Western Sahara Refugees.* PLoS Medicine, 2012. **9**(10): p. no pagination.
68. Rahim H.F.A., et al., *Maternal and child health in the occupied Palestinian territory.* The Lancet, 2009. **373**(9667): p. 967-977.

Table 4: Extraction sheet (N=21)

Author	Year	Country	Setting	Type of study	Data collection period	Sample size	Age	Independent variables	Dependent variable	Standard	Unit
Baigari, R.	1980	Bangladesh	Rural	Longitudinal study	Aug 1974 - Nov 1975	376, 326 and 356	12-36 m	Seasonal change	WFA	Harvard	% median
Bechir, M. et al	2010	Chad	Rural	Repeated cross-sectional studies	May/June 2007 and October 2007	653, 644 and 579, 539	0-59 m	Seasonal change	WFH	WHO	Z-score
Benefice, E. et al	1984	Senegal	Rural	Longitudinal study	March 1980, Aug/Sept 1980, Jan/Feb 1981, June 1981	114, 106, 88 and 90	12-59 m	Seasonal change	MUAC, MC, TSF, WFH	NCHS	% median
Block, S. A. et al	2003	Indonisia	Both	Repeated cross-sectional studies	1997/98. Repeated every 3 months approx	From 5450 to 10553	0-59 m	Drought and financial crisis 1997/1998	WFA, WFH	NCHS	Z-score
Branca, F. et al	1993	Ethipia	Rural	Longitudinal study	May 1987 - June 1988	60 for WFH, 40 for HFA in 0-59	0-59 m	Seasonal change	Height increment, WFH/L, H/LFA	NCHS	Z-score

Briend, A. et al	1989	Bangladesh	Rural	Longitudinal study	Dec 1984-Dec 1987	413 average (351 - 514)	6-36 m	Seasonal change & change over the years	Weight, WFA, H/LFA, WFH/L, MUAC	NCHS	% median & Z-score
Brown, K. et al	1982	Bangladesh	Rural	Longitudinal study	April 1978 - June 1979	174	6-59 m	Seasonal change	% expected Weight & Height/leng th gain, WFA, H/LFA, WFH/L, MUAC-for-age and TSF-for-age	NCHS	% median
Chikhungu, L. C. et al	2014	Malawi	Both	Repeated cross-sectional studies	March 2004 - Feb 2005	4012 and 2675	6-59 m	Seasonal change	WFH/L, WFA, HFA	WHO	Z-score
Egata, G. et al	2013	Ethiopia	Rural	Longitudinal study	July 2010 - Feb 2012	2132	6-36 m	Seasonal change	Weight, WFH/L, MUAC	WHO	Z-score
Garenne et al	2012	Senegal	Rural	Cohort	May 1983 - Nov 1983 - May 1984 - Nov 1984	775, 988 and 1040	6-23 m	Seasonal change	Weight, length, HC, MUAC, TSF, SSF, MC, BMI, WFA, WFH/L	CDC - 2000	Z-score
Hillbruner, C. et al	2008	Bangladesh	Urban	Repeated cross-sectional studies	Aug 2002, Feb 2003 and Aug/Sept 2003	185	6-72 m	Seasonal change	% expected growth, WFH/L, H/LFA	WHO	Z-score

Huong, L. T. et al	2014	Vietnam	Rural	Repeated cross-sectional studies	March, June, Sept, Dec 2012	195, 237, 196 and 225	24-59 m	Seasonal change	Height, weight, WFH/L, WFA,H/LFA	WHO	Z-score
Loutan, L. et al	1984	Niger	Rural	Cohort	Aug 1980 - Sept 1981	29 and 32	12-59 m and 0-59 m	Seasonal change	WFH/L, MUAC, TSF	Harvard	% median
Marin, C. M. et al	1996	Peru	Urban	Longitudinal study	Jan 1987 - Oct 1993	Min 100 per month, 4023 to 7946 per year	0-35 m	Seasonal change	WFH/L, WFA, H/LFA	NCHS	Z-score
Martin-Prevel, Y. et al	2000	Congo	Urban	Repeated cross-sectional studies	1993 then 1996	2581 and 1576	4-23 m	Devaluation of CFA franc	Birth weight, WFH/L, H/LFA	NCHS	Z-score
Meshram, I. I. et al	2014	India	Rural	Repeated cross-sectional studies	June/Sept 2007, Oct/Jan and Feb/May 2008	833, 527 and 555 children	12-59 m	Seasonal change	WFH/L, WFA, H/LFA	WHO	Z-score
Miller, J. et al	2013	West Timor	Rural	Nested cohort in cross-sectional survey	March 2010 - Nov 2010	80	6 to 59 months	Seasonal change	WFH/L, H/LFA, WFA, MUAC-for-age , TSF-for-age	WHO	Z-score

Mude, A. G. et al	2006	Kenya	Rural	Secondary data analysis	Feb 2000 - May 2005	between 17 and 58	6-59 m	herd dynamic, food, biophysical and seasonality	MUAC	NCHS	Z-score
Panter-Brick, C.	1997	Nepal	Rural	Longitudinal study	1982 (8 rounds: Feb/march to Sept/Oct)	53 to 71	0-50 m	Seasonal change	Weight, Height, WFH/L, WFA, L/HFA	NCHS	Z-score
Shell-Duncan, B.	1995	Kenya	Rural	Longitudinal study	From feb 1990 to Jan 1991	54	6 m-10 y	Seasonal change	Weight, Height, MUAC, BMI, WFH/L, MUAC-for-Age, WFA, H/LFA	NCHS	% median & Z-score
Wright, J. et al	2001	Zimbabwe	Both	Growth monitoring	Jan 1988-March 1993 and Jan 1994-Dec 1995	50 districts	0-59 m	Seasonal change	WFA	NCHS	% median

WFA, Weight-For-Age; HFA, Height-For-Age; WFH, Weight-For-Height; MUAC, Middle-Upper Arm Circumference BMI, Body Mass Index; HC, head circumference; TSF, Triceps skinfold; SSF, subscapular skinfold; MC, muscle circumference; NCHS, National Centre for Health Statistics; CDC, Centers for Disease Control and Prevention; WHO, World Health Organisation

Table 5: Outcome of the studies included (N=21)

	WFA		H/LFA		WFH/L		MUAC		MUAC/A		TSF		TSF/A	W	H	BMI	MC	SSF	HC	BW
Author	**Mean**	**%**	**Mean**	**%**	**Mean**	**%**	**Mean**	**%**	**Mean**	**%**	**Mean**	**%**	**Mean**	**Mean**	**Mean**	**Mean**	**Mean**	**Mean**	**Mean**	**Mean**
Baigari, R.	+	+																		
Bechir, M. et al						+														
Benefice, E. et al					-		+				+						+			
Block, S. A. et al	+				+															
Branca, F. et al			+		-										+					
Briend, A. et al	-/+	-/+	+	+	+	+	+	+						+						
Brown, K. et al	-		+		+				+				+	+	+					
Chikhungu, L. C. et al		+		+		-														
Egata, G. et al					-	+	-							+						
Garenne et al	+				+		+				+			+	+	+	+	-	-	
Hillbruner, C. et al			+	+	+	+									+					
Huong, L. T. et al	-	-	-	-	-	-								-	-					
Loutan, L. et al						-		-				+								
Marin, C. M. et al	+		-		+															
Martin-Prevel, Y. et al			+	+	+	+														+
Meshram, I. I. et al	+	+	+	+	+	+														
Miller, J. et al	+		-	+	+	+			-				+							
Mude, A. G. et al							+													
Panter-Brick, C.	+	+	+	+	+	+								+	+					
Shell-Duncan, B.		-		+		+	+			+				+	+	+				
Wright, J. et al		-/+																		

+, Statistically significant change;- No statistically significant change; WFA, Weight-For-Age; H/LFA, Height/Length-For-Age; WFH/L, Weight-For-Height/Length; MUAC, Middle-Upper Arm Circumference; ; TSF, Triceps skinfold; W, Weight; H, Height BMI, Body Mass Index; MC, muscle circumference; SSF, subscapular skinfold; HC, head circumference; BW, Birth Weight

Declarations

List of abbreviations

CDC: Centers for Disease Control and Prevention

CMAM: Community-based Management of Acute Malnutrition

DHS: Demographic and Health Surveys

ENN: Emergency Nutrition Network

FSNAU: Food Security and Nutrition Analysis Unit

H/LFA: Height/Length-For-Age

MUAC: Middle-Upper Arm Circumference

MICS: Multiple Indicator Cluster Surveys

NCHS: National Centre for Health Statistics

SMART: Standardized Monitoring and Assessment of Relief and Transitions

SSF: Subscapular Skinfold

TSF: Triceps SkinFold

UNICEF: United Nations Children's Fund's

WFA: Weight-For-Age

WFH/L: Weight-For-Height/Length

WHO: World Health Organization

Ethics approval and consent to participate

Not applicable.

Availability of data and materials

Not applicable.

Competing interests

No financial or non-financial competing interests.

Author contribution

Severine Frison is the main author and was involved in all stages from the conception and design, literature search, analysis and interpretation to drafting the article and writing the final version to be published.

Claudine Prudhon, Marko Kerac and Francesco Checchi were involved in the conception and design as well as in critically revising different draft versions and approving the version to be published.

Funding

This work was supported by the Office of U.S. Foreign Disaster Assistance (OFDA) and the World Food Programme (WFP) (grant number ITDCZD07). OFDA and WFP had no role in the design, analysis or writing of this article.

Acknowledgements

We would like to acknowledge Jane Falconer (LSHTM) for assisting with the peer-reviewed search. No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Chapter 3 – Description of source data for the thesis and data management

This chapter describes the different steps undertaken to fulfil objective 1 of this thesis, namely to assemble a large database of nutritional surveys from different settings in order to enable analyses, including testing the assumptions for the proposed method and the validation of the method through statistical simulation approaches.

This chapter describes data collection, data management, the creation of the database used for this thesis as well as the additional variables generated for the analysis.

Description of source data for the thesis and data management

3.1 Assembling a large survey database

Data sources

Several organisations were contacted to obtain the datasets required. The organisation approached were the main actors in the field of nutrition. There was no specific number of surveys to obtain but the goal was to acquire as many as possible and to include as many countries and regions as possible. Memorandums of Understanding were signed with six organisations that agreed to share data: UNICEF, Food Security and Nutrition Analysis Unit (FSNAU), Medecin Sans Frontiere (MSF), Action Contre la Faim (ACF), Concern Worldwide (Concern) and Goal. A total of 1068 cross-sectional survey datasets were collected from October 2011 to July 2012.

Data management

Eligible datasets had to have one row per individual child-observation, and include: location, date, cluster, sex, age, Middle-Upper Arm Circumference (MUAC), oedema, weight and height. Meta-data required for each survey included region, country, livelihood zone and residence status. The formats of the datasets shared included Excel, Emergency Nutrition Assessment (ENA) formats, SPSS, STATA, and text files. Part of the meta-data information was only included in the survey reports and the variables for livelihood and residence were mostly added by consulting survey reports or, where relevant, maps of livelihood zones produced by famine early warning systems. All the files were transferred into R and STATA 13 software. The World Health Organisation's (WHO) "Child Growth Standards" package[1] was used to re-calculate all Weight-For-Height (WFH), Weight-For-Age (WFA) and Height-For-Age (HFA) indices that were added to the main database.

Of the 1068 cross-sectional surveys collected, 852 surveys were included (figure 1). Datasets were excluded if any of the required variables was missing. Other exclusion criteria included corrupted files, duplicates, number of cluster under 25 for cluster sample surveys (it was shown that cluster sample surveys should include a minimum of 25 clusters to be statistically representative [2, 3]) and a quality score less than 0 (see section 3 below for details on how the quality score variable was created). Low quality score above 0 were kept in order to represent as much as possible data collected in the field. Surveys with a low quality score above 0 were kept in order to include as many datasets as possible and to reflect the reality of data quality. A large majority of the surveys included were cluster sample surveys (797, 93.5%) while a small proportion were exhaustive surveys (55, 6.5%).

The 852 surveys contained 694 108 child observations of which 25 134 presented highly improbable values or missing values and were excluded from the analysis leaving 668 975 children eligible for analysis. Highly improbable values for WFH, WFA and HFA were defined using WHO standard flags: WFA<-6.0 SD or WFA>+5.0 SD; HFA<-6.0 SD or HFA>+6.0 SD and WFH<-5.0 SD or WFH>+5.0 SD. There is no standard for MUAC “extreme” values and a minimum of 85 mm and maximum of 200 mm was used after consultation with expert in the field (see Figure 1).

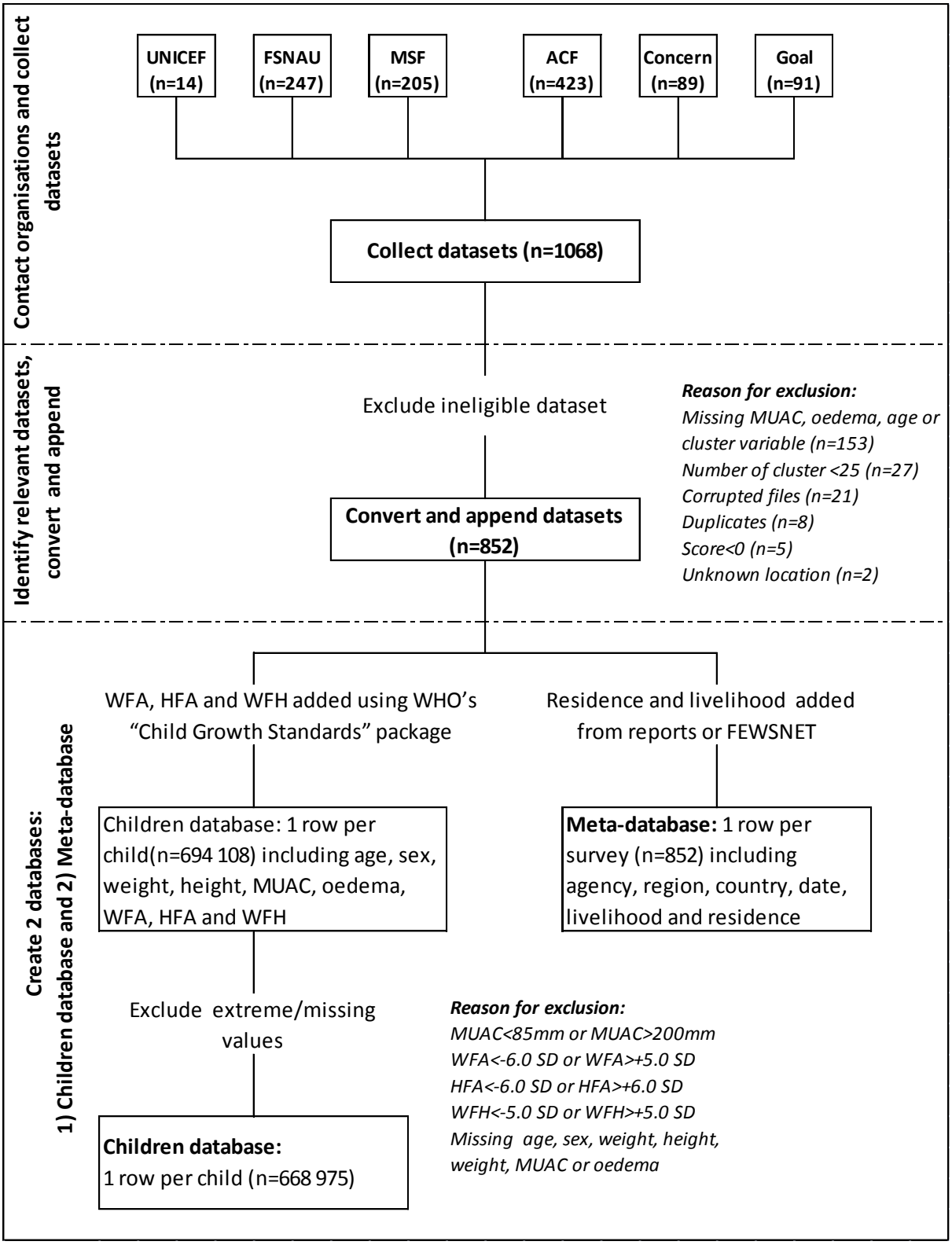


Figure 1: Data management

3.2 Description of the data

Table 1 and 2 below describe the variables included in the meta-database as well as the children database.

Table 1: Description of the meta-database

Variable	Format and coding
Cluster	Integer
Date	Integer, month/year
Region	Character: East Africa, West Africa, Central & Southern Africa Caribbean, Asia
Country	Character: 38 countries see annex II
Livelihood	Character: Agriculture, Agro-pastoral, Pastoral, Other
Residence	Character: Rural, Urban, Displaced, Other

Table 2: Description of the children database

Variable	Format and coding
Gender	1=Male, 2=Female
Age	Integer; month
Oedema	0=No, 1=Yes
Height	Fixed decimal (1 decimal place); cm
Weight	Fixed decimal (1 decimal place); kg
MUAC	Integer; mm
Weight-for-Age	Numeric; Z-scores
Height-for-Age	Numeric; Z-scores
Weight-for-Height	Numeric; Z-scores

Surveys were conducted in 38 different countries (annex 1 presents the number of surveys per countries) from 1992 to 2011, with 95% of them from 2000 (Figure 2). The sample size of the surveys varied from 122 to 3491 children. The mean sample size was 785 and the median was 815. Before the development of the Standardized Monitoring and Assessment of Relief and Transitions (SMART) methodology for anthropometric and mortality surveys[4], most nutritional surveys were conducted

using a 30 by 30 cluster survey approach which translates into a large number of surveys with a sample size close to 900 children (Figure 3).

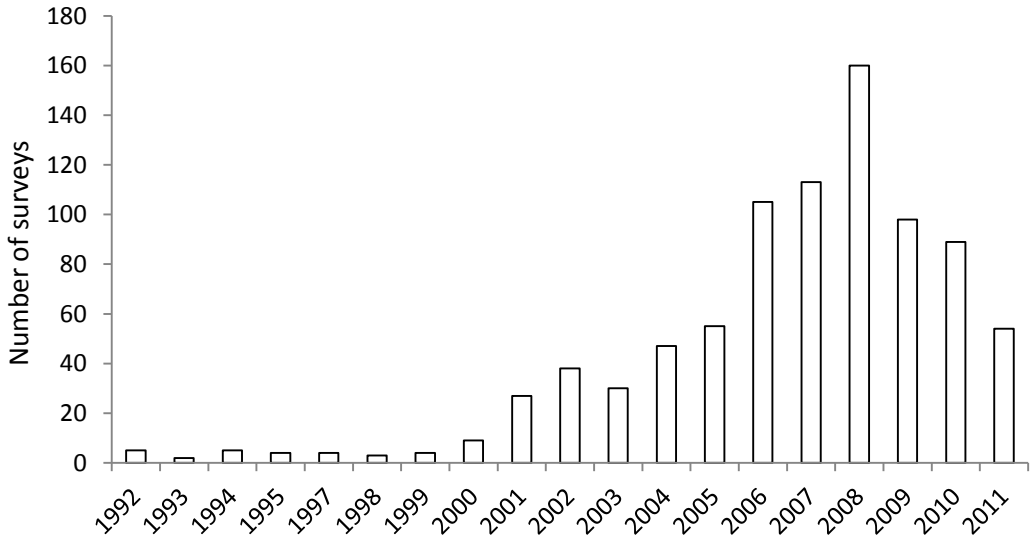


Figure 2: Number of surveys over time

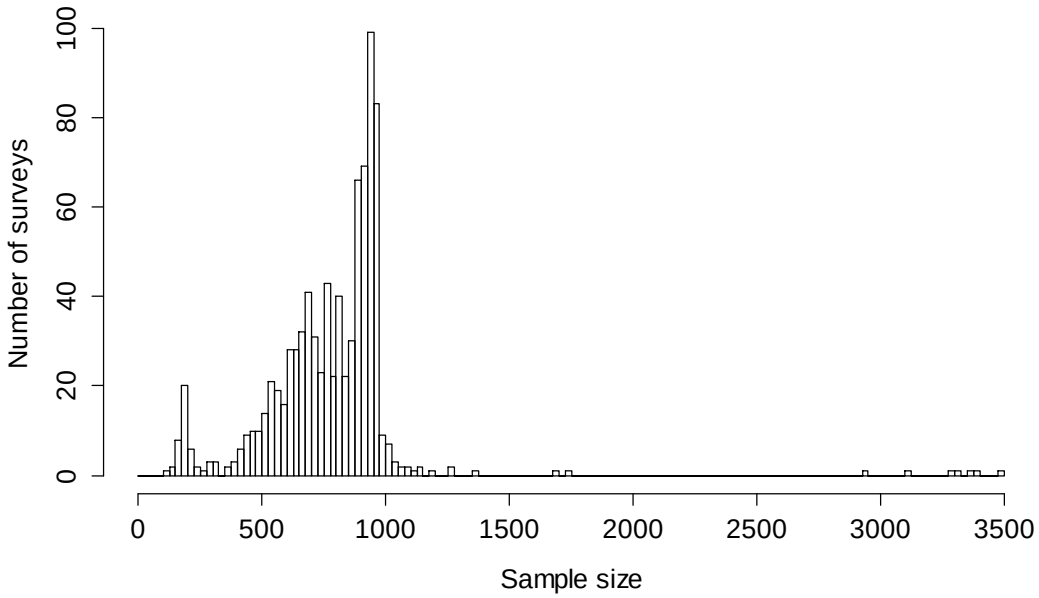


Figure 3: Distribution of sample sizes

Table 3 describes the surveys included in the analysis. A large proportion of the surveys (554) were conducted in East Africa including 187 Surveys from FSNAU Somalia were included in the database. The majority of the surveys were agriculture or agro-pastoral livelihood zones (41.2 % and 28.9 % respectively) and a smaller proportion in a pastoral zone (14.4 %). The "other" category includes fishing, riverine and mixed livelihood zones. Most of the surveys were conducted in rural areas (64.6 %).

The proportion of males and females was roughly the same. The different age categories are unevenly represented especially for the 24-29, 36-41 and 42-47 months old age categories (Table 4).

The prevalence of Global Acute Malnutrition (GAM) observed overall, according to MUAC varies a lot across surveys. GAM varied from 1 to 47.7% and Severe Acute Malnutrition (SAM) from 0% to 14.7%. Figure 4 shows the distribution of SAM and GAM prevalence measured with MUAC measure alone.

Table 3: Description of the surveys

Region	**Number of surveys**	**(%)**
East Africa	554	65.0
West Africa	97	11.4
Central & South Africa	128	15.0
Caribbean	13	1.5
Asia	60	7.0
Livelihood	**Number of surveys**	**(%)**
Agriculture	351	41.2
Agro-pastoral	246	28.9
Pastoral	123	14.4
Other	132	15.5
Residence status	**Number of surveys**	**(%)**
Rural	550	64.6
Urban	66	7.8
Displaced	145	17.0
Other	91	10.7

Table 4: Characteristics of the children

Age in months	**N**	**%**
6-11	68 385	10.2
12-17	78 747	11.8
18-23	71 440	10.7
24-29	92 321	13.8
30-35	64 901	9.7
36-41	92 588	13.8
42-47	86 746	13.0
48-53	57 492	8.6
54-59	56 355	8.4
Total	668 975	100
Sex	**N**	**%**
Female	331 932	49.5
Male	337 043	50.5
Total	668 975	100

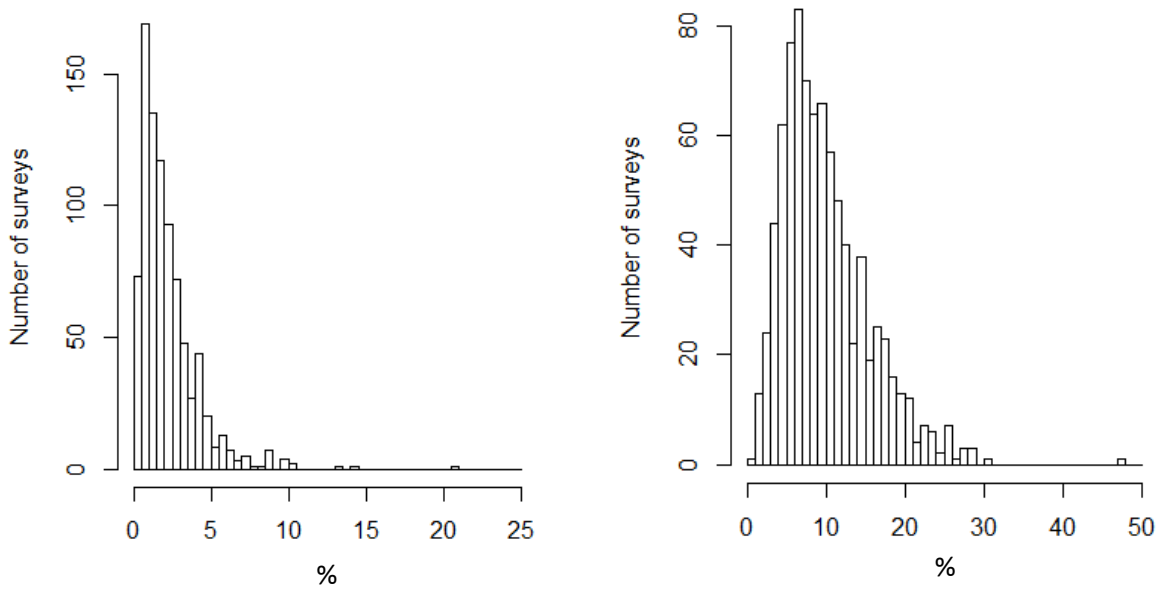


Figure 4: Distribution of SAM (MUAC<115mm) (left) and GAM (MUAC<125mm) (right) prevalence

3.3 Variables created for data analysis

Several variables were created for the analysis (Table 5):

- **Quality score** based on the digit preference of MUAC. The digit preference (score) variable was calculated as follow:

$$Score=1-\Sigma\ Abs|0.1-p_{\alpha}|$$

Where p_{α} is the proportion p of each digit α from 0 to 9.

Assuming that the proportion of measurements ending with 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 should equal 10% and therefore that the highest score was 1, the lesser the digit preference, the higher the score. A score equal or over 0.75 corresponded to a low digit preference, and under 0.75 to a high digit preference

- **Sampling weight**: in analysis across surveys, the sample size could not be used for weighting because of varying design, such as exhaustive versus cluster suveys and differing numbers of clusters and children per cluster. Instead, the effective sample size (sample size/design effect) was used in weighting[5].

Quality scores (based on digit preference) were also considered. There is no clear guidance on how to combine variables (effective sampling weight and quality scores) into one weight variable. When using quality weight, the impact on results was minimal and the analysis was conducted using effective sampling weight only.

- The normality of the distribution of MUAC using Shapiro-Wilk test[6]: "**normal**". Binary variable yes/no presenting whether the distribution showed significant deviation from normal distribution at p-value<0.05.

- The **skewness** of MUAC as binary (yes/no) using D'Agostino test[7] and as continuous. Skewness is a measure of the asymmetry of a distribution around its mean. The binary variable measured whether the distribution was significantly skewed at p-value<0.05.

- The **Kurtosis** of MUAC as binary (yes/no) using Anscombe-Glynn test[8] and as continuous. Kurtosis indicates heavy tails and "peakedness" relative to a normal distribution. The binary variable measured whether the distribution was peaked significantly at p-value<0.05.

- **Design effect (Deff)** of surveys. In a cluster sample surveys, deff quantifies the extent to which the expected sampling error in a survey departs from the sampling error that can be expected under simple random sampling.

- MUAC standard deviation (**MUAC SD**). SD is a measure that is used to quantify the amount of variation or dispersion of a set of data values around the mean.

- **Large sample size**: Survey sample size category: large (>900) or small (<900).

- **GAM category**: GAM categories based on MUAC (<5%, 5-9%, 10-14%, ≥15%).

- **Date category**: date classified before/equal to 2006 or after 2006.

- **Coverage**: the proportion of SAM /GAM confidence intervals that contain the "true" proportion computed with the classical approach (from the full survey dataset.

- **Precision**:
Absolute precison: Half the widths of the CIs of the SAM /GAM estimates.
Relative precison: (Absolute precision * 100)/Prevalence estimate

-**Bias**: the estimated prevalence (from proposed method) minus "true" SAM /GAM prevalence (from full survey dataset).

- **Probability**: the probability of classifying GAM correctly for different thresholds (5%, 10% or 15%)

The table below describes the variable created for different part of the analysis.

Table 5: Variables created for the analysis

Variable	Format and coding	Objective
Quality score	Numeric from 0 to 1 (1 decimal place)	1
Sampling Weight	Integer	2 - 5
Deff	Numeric (2 decimal places)	2 – 5
MUAC SD	Numeric (1 decimal places)	2 – 5
Normal	1=Yes; 0=No	2
Skewness (2 variables)	1=Yes; 0=No and Numeric	2
Kurtosis (2 variables)	1=Yes; 0=No and Numeric	2
Large sample size (n>900)	1=Yes; 0=No	2
GAM levels (MUAC)	1=<5%; 2=5-9%; 3=10-14; 4=≥ 15%	4-5
Date category	1=before 2006; 2=after 2006	4-5
Coverage	Numeric (1 decimal place)	5
Precision	Numeric (1 decimal place)	5
Bias	Numeric (1 decimal places)	5
Probability	Numeric (1 decimal place)	5

References

1. WHO. *Child growth standards package*. 2011 [cited; Available from: http://www.who.int/entity/childgrowth/software/en/.
2. Binkin, N., et al., *Rapid nutrition surveys: how many clusters are enough?* Disasters, 1995. **16**: p. 99-103.
3. Spiegel, P.B., et al., *Quality of malnutrition assessment surveys conducted during famine in Ethiopia.* Jama-Journal of the American Medical Association, 2004. **292**(5): p. 613-618.
4. SMART. *Internet: http://smartmethodology.org/ (accessed 12 November 2014)*. [cited.
5. Furlow-Parmley C. , et al., *Combining estimates from two surveys: An example from monitoring 2009 influenza A (H1N1) pandemic vaccination.* Statistics in Medicine, 2012. **31**(27): p. 3285-3294.
6. G, A.D. and B.J. Martin, *Statistics notes: The normal distribution.* BMJ, 1995. **310**(298).
7. D'Agostino, R.B., *Transformation to Normality of the Null Distribution of G1.* Biometrika, 1970. **57**(3): p. 679-681.
8. Anscombe, F.J. and W.J. Glynn, *Distribution of kurtosis statistic for normal statistics.* Biometrika, 1983. **70**(1): p. 227-234.

PART II

RESULTS

Chapter 4 – Is Middle-Upper Arm Circumference “normally” distributed?

One of the main assumptions behind the proposed method is that Middle-Upper Arm Circumference (MUAC) is normally distributed in a large majority of populations and settings, or can be transformed mathematically so as to take a normal distribution.

This research paper describes the different steps undertaken to fulfil objective II “Examine the normality of MUAC distributions; and if necessary apply transformations to the data in order to achieve normality”. It assesses the normality of the MUAC distribution graphically and statistically and explores different smoothing techniques and transformations in order to reach “normality”.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Severine Frison
Principal Supervisor	Marko Kerac
Thesis Title	Middle-upper arm circumference for nutritional surveillance in crisis-affected populations: Development of a method

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Emerging Themes in Epidemiology. 2016, 13:7		
When was the work published?	04/05/2016		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic peer review?	Yes

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am the main author and was involved in all stages from the conception and design, literature search, analysis and interpretation to drafting the article and writing the final version.

Student Signature: SFrison **Date:** 12/08/16

Supervisor Signature: M. Kerac **Date:** 12/08/16

Frison *et al. Emerg Themes Epidemiol (2016) 13:7*
DOI 10.1186/s12982-016-0048-9

Emerging Themes in Epidemiology

RESEARCH ARTICLE Open Access

Is Middle-Upper Arm Circumference "normally" distributed? Secondary data analysis of 852 nutrition surveys

Severine Frison[1*], Francesco Checchi[3], Marko Kerac[1] and Jennifer Nicholas[2]

Abstract

Background: Wasting is a major public health issue throughout the developing world. Out of the 6.9 million estimated deaths among children under five annually, over 800,000 deaths (11.6 %) are attributed to wasting. Wasting is quantified as low Weight-For-Height (WFH) and/or low Mid-Upper Arm Circumference (MUAC) (since 2005). Many statistical procedures are based on the assumption that the data used are normally distributed. Analyses have been conducted on the distribution of WFH but there are no equivalent studies on the distribution of MUAC.

Methods: This secondary data analysis assesses the normality of the MUAC distributions of 852 nutrition cross-sectional survey datasets of children from 6 to 59 months old and examines different approaches to normalise "non-normal" distributions.

Results: The distribution of MUAC showed no departure from a normal distribution in 319 (37.7 %) distributions using the Shapiro–Wilk test. Out of the 533 surveys showing departure from a normal distribution, 183 (34.3 %) were skewed (D'Agostino test) and 196 (36.8 %) had a kurtosis different to the one observed in the normal distribution (Anscombe–Glynn test). Testing for normality can be sensitive to data quality, design effect and sample size. Out of the 533 surveys showing departure from a normal distribution, 294 (55.2 %) showed high digit preference, 164 (30.8 %) had a large design effect, and 204 (38.3 %) a large sample size. Spline and LOESS smoothing techniques were explored and both techniques work well. After Spline smoothing, 56.7 % of the MUAC distributions showing departure from normality were "normalised" and 59.7 % after LOESS. Box-Cox power transformation had similar results on distributions showing departure from normality with 57 % of distributions approximating "normal" after transformation. Applying Box-Cox transformation after Spline or Loess smoothing techniques increased that proportion to 82.4 and 82.7 % respectively.

Conclusion: This suggests that statistical approaches relying on the normal distribution assumption can be successfully applied to MUAC. In light of this promising finding, further research is ongoing to evaluate the performance of a normal distribution based approach to estimating the prevalence of wasting using MUAC.

Keywords: Normal distribution, Middle-Upper Arm Circumference, Child malnutrition, Wasting, Probit

Background

Wasting is a major public health issue throughout the developing world. The United Nations Children's Fund's (UNICEF) latest report on the State of the World's Children [1] estimates that 10 % of children under 5 years old in least developed countries are wasted. Out of the 6.9 million estimated deaths among children under five annually, over 800,000 deaths (12.6 %) are attributed to wasting [2]. Wasting is quantified as Weight-For-Height (WFH) < −2 standard deviations (SD) from the World Health Organization (WHO) reference median and/or Mid-Upper Arm Circumference (MUAC) < 125 mm. MUAC has been adopted by the World Health Organisation (WHO) as a measure of wasting and is increasingly

*Correspondence: severine.frison@gmail.com
[1] Department of Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
Full list of author information is available at the end of the article

recognised as a very useful measure of anthropometric status [3, 4].

Many statistical procedures are based on the assumption that the data follow a normal distribution. The shape of the normal distribution (the characteristic "bell curve") is quantified by two parameters: the mean and the standard deviation, and follows important properties: (1) it is always symmetrical with equal areas on both sides of the curve; (2) the highest point on the curve corresponds to the mean which equals the median and the mode; (3) the spread of the curve is determined by the standard deviation; and (4) as with all probability density functions the area under the curve must sum to the total probability of 1 [5]. The distribution of many characteristics in nature is normal or follows some form that can be derived from the normal distribution and specific statistical approaches are based on the properties of a normal distribution. For example, the probit approach [5, 6] estimates the prevalence of wasting as the cumulative probability of lying below the relevant MUAC cut-point based on the mean and standard deviation (SD) of the observed data [5, 6].

There are graphical and statistical methods for evaluating normality. Graphical methods include histograms and normality plots. Statistical methods include diagnostic hypothesis tests for normality, and a normal distribution has a skewness of 0 and kurtosis of 3 [7, 8]. Skewness is a measure of the asymmetry of a distribution around its mean while Kurtosis indicates heavy tails and "peakedness" relative to a normal distribution [9, 10]. The ability to detect departure from a normal distribution can be sensitive to local peaks and troughs in the distribution. A way to deal successfully with this issue is to apply smoothing techniques (fit a smooth curve to a set of noisy observations) using different methods such Spline function or Locally Weighted Scatterplot Smoothing (LOESS) [11–13]. For distribution originating from cluster surveys, it may be expected that high clustering in observations (large design effect) lead to asymmetric distributions, e.g. featuring a long tail of low MUAC observations. When a variable is not normally distributed for a reason other than the ones above, it can often be transformed and tested for normality using power transformations such as the Box-Cox transformation [14, 15].

Although the violation of the normal distribution assumption often increases chances of committing either a type I or II error, very few researchers test whether the assumption does indeed hold before carrying out statistical analyses [16, 17]. Previous studies have assessed the distribution of WFH [18–20] but there are no equivalent studies on the distribution of MUAC. This paper assesses the normality of the MUAC distribution graphically and statistically, and explores different transformations and smoothing techniques in order to reach normality. Findings presented pertain to a broader project to develop a more efficient method for estimating the prevalence of wasting using MUAC as the primary index, which relies heavily on MUAC distributions meeting normality criteria.

Methods

Study design and inclusion criteria

A total of 1068 cross-sectional survey datasets from various settings were shared by six organisations (UNICEF, Food Security and Nutrition Analysis Unit, Epicentre/Médecins Sans Frontières, Action Against Hunger, Concern Worldwide and Goal). The study size depended on availability of surveys and on specific inclusion criteria. Eligible datasets had to: (1) include MUAC, oedema, age, weight and height as well as meta-data on country, livelihood, residence, cluster (if cluster surveys) and date; (2) have a minimum of 25 clusters if cluster surveys [21, 22]. The last criteria aimed to minimise selection bias, as surveys with a small number of clusters may not be representative of the population. The surveys were exhaustive or clustered surveys. The datasets were cleaned and records with extreme or missing values were excluded: Children were excluded if any of the following data were missing: age; sex; height; weight; MUAC; oedema. Those with highly improbable extreme values ('flags') were also excluded from analysis: MUAC < 85 mm or MUAC > 200 mm, age < 6 months or age > 59 months, Weight-For-Age (WFA) < −6.0 SD or WFA > +5.0 SD, Height-For-Age (HFA) < −6.0 SD or HFA > +6.0 SD, WFH < −5.0 SD or WFH > +5.0 SD (WHO "flags" were applied on SD for WFH, WFA and HFA [23]).

Database

Out of the 1068 surveys collected, 852 surveys were included in the secondary data analysis (55 exhaustive surveys and 797 clustered surveys). The 852 surveys contained 668,975 children of which 25,134 (3.76 %) presented highly improbable values and were excluded from the analysis. The database included six variables for anthropometry (sex, MUAC, oedema, age, weight and height), six meta-data variables (organisation, country, livelihood, residence, cluster (when cluster surveys) and date). Other variables were computed for the purpose of this analysis: (1) the normality of the distribution (binary: 1 = yes/0 = no using Shapiro–Wilk test), (2) the skewness and Kurtosis of MUAC as continuous and binary (binary: 1 = yes/0 = no whether the data was skewed or peaked using D'Agostino and Anscombe–Glynn tests respectively), (3) the design effect of surveys (large over 3) (4) digit preference of MUAC. The digit preference variable was equal to 1—absolute (0.1-proportion of

each digit preference). Assuming that the proportion of measurements ending with 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 should equal 10 % and therefore that the highest score was 1, the lesser the digit preference, the higher the score. A score equal or over 0.75 corresponded to a low digit preference, and under 0.75 to a high digit preference, and (5) survey size category (large size over 900).

Data analysis

The normality of the MUAC distributions was assessed graphically looking at histograms of MUAC distributions and Q–Q plots (probability plot, "Q" stands for quantile). Q–Q plots show sorted values from the data set against the expected values of the corresponding quantiles from the standard normal distribution. The measure of departure from normality was also investigated statistically through Shapiro–Wilk test as well as the D'Agostino test to assess the skewedness and Anscombe–Glynn test to assess the peakedness of MUAC distributions. For each statistical test, a p value less than 0.05 indicates evidence for departure from a normal distribution.

Different methods were explored to transform non-normal distributions into normal: (1) Spline smoothing (using a spline function) and LOESS (locally weighted scatterplot smoothing using local polynomial regression fitting) techniques were applied to all distribution showing departure from a normal distribution (Shapiro–Wilk test). While smoothing the data, three criteria were applied: the mean MUAC and MUAC SD, after back-transformation of the smoothed data must be almost unchanged from the non-smoothed mean and SD (properties of a normal distribution is defined by the mean and the SD), and the Shapiro–Wilk test p value has to exceed 0.05. (2) Box-Cox power transformation was applied to all survey showing departure from a normal distribution, and (3) Box-Cox power transformation was applied on surveys showing departure from "normality" after smoothing techniques had been applied.

Spline smoothing fits a spline with knots at every data point (x) by estimating its parameters minimizing the usual sum of squares plus a roughness penalty (λ). If $\lambda \rightarrow 0$ imposes no penalty (very close fit), but the resulting curve could be very noisy as it follows every detail in the data. As $\lambda \rightarrow \infty$ the penalty dominates and the solution converges to the ordinary least square line. LOESS is a fairly direct generalization of traditional least-squares methods for data analysis. It fits a polynomial surface determined by one or more numerical predictors, using local fitting. That is, for the fit at point x, the fit is made using points in a neighbourhood of x, weighted by their distance from x (with differences in 'parametric' variables being ignored when computing the distance). The size of the neighbourhood is controlled by α (set by span).

The Box-Cox method transforms data into a "normal" shape using parameter λ corresponding to different transformations (i.e. $\lambda = 1.00$: no transformation needed; $\lambda = 0.50$: square root transformation $\lambda = 0.29$: for a transforming power between cube and fourth root $\lambda = 0.33$: cube root transformation $\lambda = 0.25$: fourth root transformation $\lambda = 0.00$: natural log transformation $\lambda = -0.50$: reciprocal square root transformation $\lambda = -1.00$: reciprocal (inverse) transformation and so forth). The most appropriate value of λ was identified as that which minimised the departure from a normal distribution on the Shapiro–Wilk test.

R studio and STATA 13 were used for all analyses [24, 25].

Ethics approval for the project was sought and obtained from the Ethics Committee of the London School of Hygiene and Tropical Medicine (LSHTM Ethics reference 6158).

Results

The distribution of MUAC showed no departure from a normal distribution in 37.4 % (319 out of 852) of the MUAC distributions using the Shapiro–Wilk test. Out of the 533 surveys showing departure from a normal distribution, 183 (34.3 %) were skewed (D'Agostino test), 196 (36.8 %) had a kurtosis different to the one observed in the normal distribution (Anscombe–Glynn test) and 70 (13.1 %) showed both features. The sensitivity of the Shapiro–Wilk test to departure from normality is influenced by the presence of local peaks and troughs in the distribution such as those caused by digit preference (poor data quality), the design effect (high design effect may lead to asymmetric distributions), and sample size (large sample size results in greater power to detect small departures from a normal distribution). Out of the 533 surveys showing departure from normal distribution, 294 (55.2 %) showed high digit preference (score < 0.75) 164 (30.8 %) had a large design effect (over 3), and 204 (38.3 %) a large sample size (>900) (Table 1). The skewness and kurtosis of surveys showing departure from normality included values above and below the value for a normal distribution (0 for skewedness and 3 for kurtosis) indicating surveys skewed to right as well as to the left and survey with a distribution flatter or more peaked than the normal distribution (Table 2).

Figure 1 shows examples of distributions of MUAC and their respective Q–Q plots for two surveys with very "non-normal" distribution (very low p-value Shapiro–Wilk test) but skewness and kurtosis close to those observed in a normal distribution. Visually neither of the distribution seemed skewed or peaked but digit preferences were visible in both cases which suggest this might be the reason behind the low p-value (Shapiro–Wilk test).

Table 1 Characteristics of surveys showing departure from a normal distribution (Shapiro–Wilk test, p < 0.05) and effect of transformation and smoothing on specific characteristics (N = 533)

Surveys failing Shapiro–Wilk test (p < 0.05) N = 533 (62.6 %)	N (%)	N (%) with "normal" distribution after transformation or smoothing		
		Box-Cox	Spline	Loess
All surveys[a]	533 (100)	301 (56.5)	318 (59.7)	304 (57.0)
By key survey characteristics				
Skewed[b]	183 (34.3)	113 (61.7)	81 (49.4)	89 (48.6)
Non-normal kurtosis[c]	196 (36.8)	62 (31.6)	137 (69.9)	139 (70.9)
Skewed and non-normal kurtosis[b, c]	70 (13.1)	23(32.9)	41 (58.6)	43 (61.4)
Large design effect (>3)	164 (30.8)	86 (52.4)	81 (49.4)	92 (56.1)
High digit preference (score < 0.75)	294 (55.2)	143 (48.6)	170 (57.8)	178 (60.6)
Large sample size (n > 900)	204 (38.3)	122 (59.8)	95 (46.6)	101 (49.5)

[a] Shapiro–Wilk test (p < 0.05); [b] D'Agostino test (p < 0.05); [c] Anscombe–Glynn test (p < 0.05)

Table 2 Skewness and kurtosis of survey showing departure from a normal distribution (n = 533)

	Minimum	Lower quartile	Median	Mean	Upper quartile	Maximum
Skewness	−0.61	−0.15	−0.01	−0.01	0.11	0.91
Kurtosis	2.26	2.3	3.2	3.24	3.45	5.27

Figure 2 shows examples of distributions of MUAC and their respective Q–Q plots for two surveys with "non-normal" distribution (low p-value Shapiro–Wilk test) and also skew or kurtosis different to that observed in the normal distribution. The distribution and Q–Q-Plot for the survey shown in panel A has skewed distribution (D'Agostino test) and that shown in panel B has peaked distribution (Anscombe–Glynn test). The distribution in panel A was slightly skewed to the left and there were very visible digit preferences. Peaks are visible in the distribution in panel B as well as digit preferences.

Table 1 as well as Figs. 1 and 2 suggest the main reason for departure from a normal distribution is due to local peaks and troughs.

Smoothing techniques

Spline and LOESS smoothing techniques were explored and both techniques work well. After applying Spline smoothing to the distributions showing departure from normality, 301 (56.5 %) of the MUAC distributions showed no departure from normality and 318 (59.7 %) after LOESS (Table 3).

The average mean MUAC change after Spline smoothing was 0.1 and the mean SD MUAC change was 0.8. All surveys had an average mean MUAC change under 10 and 90 % had a SD change under 10 %. After LOESS smoothing, the average mean MUAC change was 0.2 and the average SD MUAC change was 0.9. All surveys had a mean MUAC change under 10 and 84 % had a SD change under 10 %.

The effect of Spline and Loess smoothing on "non-normal" distributions with large design effect, high digit preference, large sample size as well as on skewed distributions and distributions with a kurtosis different from a normal distribution (flat or peaked) was considerable. Approximately half of surveys with large design effect were normalised after Spline and LOESS (49.4 and 56.1 % respectively), about two-third of surveys with high digit preference had a distribution approximating normal after Spline and LOESS (60.6 and 57.8 % respectively), half of surveys with large sample size (46.6 and 49.54 % respectively) as well as half of skewed distributions (49.4 and 48.6 % respectively) and over two-third of surveys with kurtosis different from normal were approximating a normal distribution after Spline and Loess smoothing (69.9 and 70.9 % respectively) (Table 3).

Box-Cox power transformation

Power transformations are typically used to "normalise" skewed distributions. Common power transformations include log, reciprocal, square and square root transformations. After applying the Box-Cox transformation to the 533 distributions showing departure from normality, 304 (57 %) of the distribution were converted to "normal" (Table 3).

The summary statistics of the Box-Cox transformation coefficient [Lambda (λ)] suggest that a variety of different power transformations were required for different surveys and few Lambda values corresponded to common power transformations (Table 4).

The effect of Box-Cox transformation on skewed distributions was sizable with almost two-third of skewed distribution approximation a normal distribution after Bo-Cox

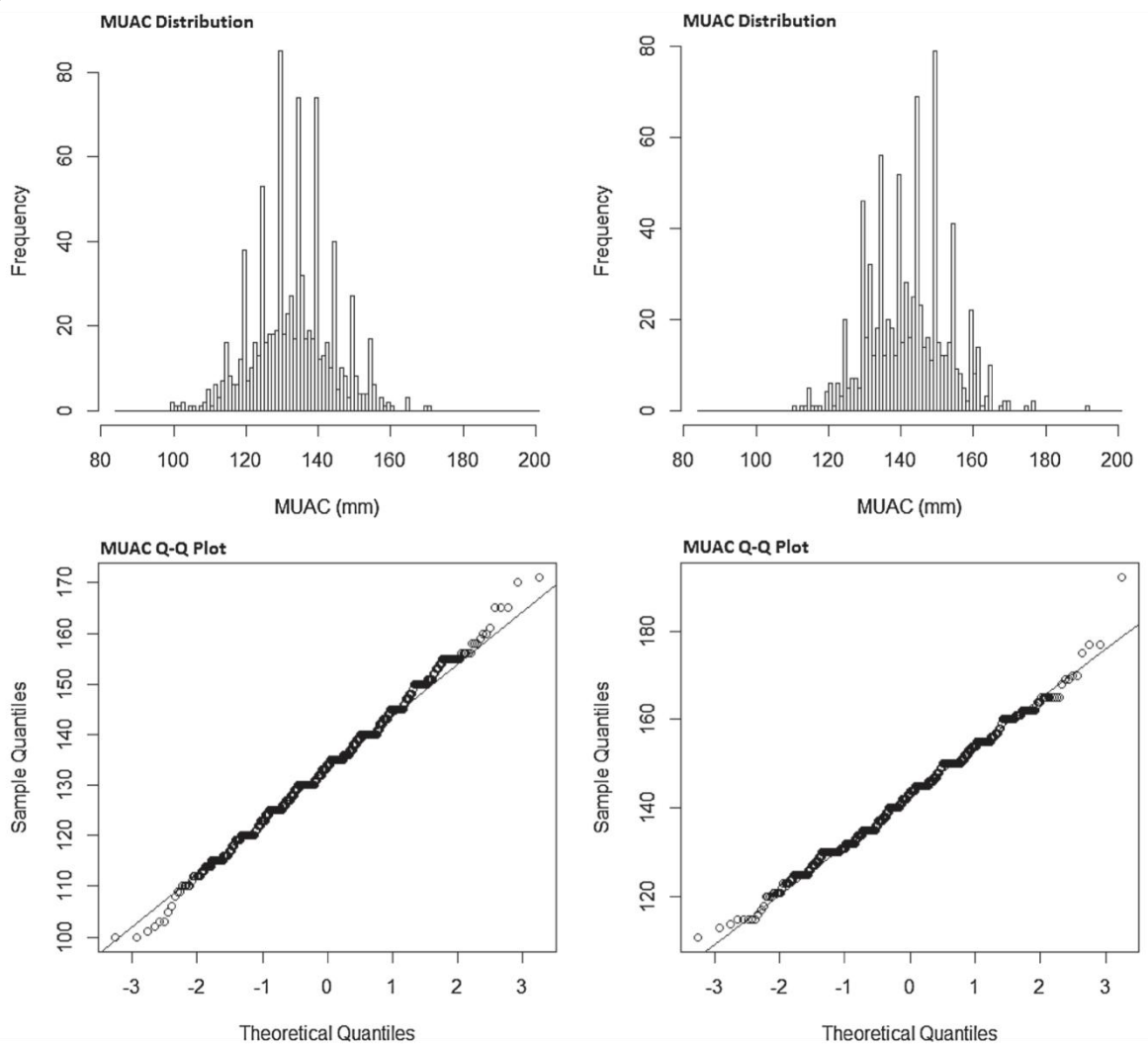


Fig. 1 Examples of non-normal (Shapiro–Wilk test) MUAC distributions and their respective Q–Q plot

transformation (61.7 %). About half of surveys with large design effect, high digit preference and large sample size distribution were approximating a normal distribution after Box-Cox (52.4, 48.6 and 59.8 % respectively). The effect on distributions with a kurtosis different from normal was less marked with a third (31.6 %) approximation a normal distribution after Box-Cox transformation (Table 3).

Smoothing and Box-Cox transformation

Applying Box-Cox transformation on surveys showing departure from a normal distribution after Loess or smoothing techniques increased further the number of "normal" distributions with 401 distributions (82.7 %) after Loess and Box-Cox and 439 (82.4 %) after Spline and Box-Cox (Table 3).

Discussion

Over a third of MUAC distributions showed no departure from normality without any transformation and three quarters showed no departure once the data were smoothed or after Box-cox transformation. Applying Box-Cox transformation on surveys showing departure from normality after smoothing resulted in over 80 % of surveys approximating a normal distribution.

Loess smoothing had slightly better outcome then Spline smoothing or Box-Cox transformation alone

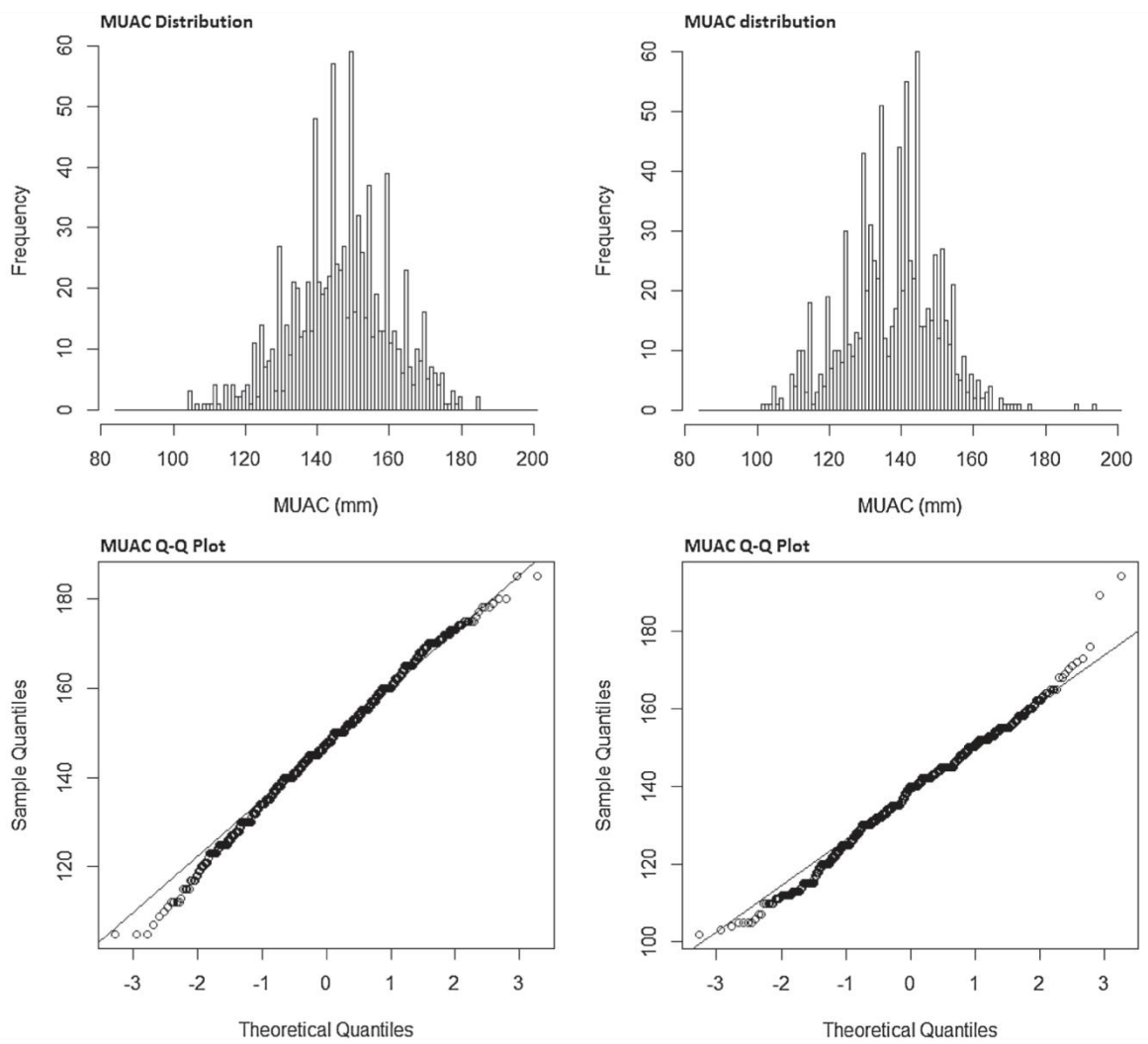


Fig. 2 Examples of a skewed and a peaked distribution and their respective Q–Q plots (D'Agostino and Anscombe–Glynn tests respectively)

Table 3 Smoothing and transformation of surveys showing departure from a normal distribution (n = 533)

Type of transformation or smoothing technique applied to "non-normal" distributions (n = 533)		N (%) "normal" distributions
Smoothing	Spline	301 (56.5)
	Loess	318 (59.7)
Box-Cox transformation	Power transformations	304 (57.0)
Smoothing and Box-Cox transformation	Box-Cox after Spline	439 (82.4)
	Box-Cox after Loess	441 (82.7)

in terms of number of distributions approximating a normal distribution but had a change in mean and SD slightly higher (but acceptable) than Spline smoothing. Although Box-Cox transformation performed well, data transformations change the nature of the variable, and any Lambda (λ) less than 0.00 has the effect of reversing the order of the data. Even though back transformation restores the data, care should be taken when applying this function [15].

The normality of MUAC distributions is affected by sample size, high digit preference, kurtosis different than

Table 4 Summary statistics of the Box-Cox transformation coefficient (Lamdba) for surveys showing departure from normality (n = 533)

	Minimum	Lower quartile	Median	Mean	Upper quartile	Maximum
Lambda (λ)	−1.2	0.61	1.08	1.03	1.51	2.73

a normal distribution and skewness. Datasets with larger sample size increase the power of the test to detect small differences when applying normality tests. Digit preference reflects the quality of the data. Training measurers to increase accuracy and precision would decrease digit preference. Both effects were lessened (two-third for digit preference and half for sample size) applying smoothing techniques to the distributions as well as applying Box-Cox transformation (half for both high digit preference and large sample size surveys). Although a third of surveys showing departure from a normal distribution were skewed or had a kurtosis different from a normal distribution, half and over two-thirds (respectively) of these were "normalised" after smoothing. Box-Cox transformation was effective on skewed distributions (almost two-third of skewed distribution "normalised") but didn't perform as well on distributions with a kurtosis different from normal (a third of distributions approximated a normal distribution after Box-Cox).

Few studies have assessed the distribution of WFH. Two looked at the standard deviations of the WFH distributions. In 1977, Waterlow et al. [19]. showed that the WFH distributions were skewed at the upper centiles. Their analysis was performed on data from surveillance or surveys involving nutrition and anthropometry in young children up to the age of 10 years. In 2006, Mei et al. [18] analysed data from 51 DHS surveys representing 34 developing Countries. They found a mean WFH and SD WFH (z-scores) of 0.06 and 1.40 respectively. The mean ranged from −0.91 to 0.83 and the SD range from 1.03 to 1.55. They concluded that their analysis confirms the WHO assertion that the SD remains in a relatively small range (i.e. close to SD from a standard normal distribution), no matter the Z-score mean although the observed range of SD for was consistently wider. Finally, in 2013, Blanton and Bilukha showed that based on the Shapiro–Wilk test for normality, 6 surveys out of the 10 surveys included in their analysis were "non-normal". All of the surveys had a small amount of skewness ranging from −0.17 and 0.31 as well as a relatively small amount of kurtosis ranging from 0.15 to 0.75.

Regarding the assessment of MUAC distributions, no equivalent studies were conducted. In 2013, data analysis from 560 cross sectional surveys conducted by Dale et al. [26]. mention the use of Box-Cox transformation to normalise MUAC and WFH data but do not give further details.

There is one main limitation to this study. The database was built based on available small scale surveys that were mainly conducted in areas where there was suspicion of a problem (i.e. high wasting prevalence) compared to national DHS and MICS surveys that are conducted every 3–5 years and show long term trends. However, we do not believe this affects the generalisability of the study. Future research might explore similar analysis on different datasets.

Conclusions

Over a third of the MUAC distributions of our database were normally distributed. MUAC distributions can easily be normalised applying simple smoothing techniques if the distribution is noisy or displays digit preference and then Box-Cox transformation if indicated (i.e. if data is skewed). This suggests that statistical approaches relying on the normal distribution assumption can be successfully applied to MUAC. In light of this promising finding, further research is ongoing to evaluate the performance of a normal distribution based approach to estimating the prevalence of wasting using MUAC.

Abbreviations

LOESS: Locally Weighted Scatterplot Smoothing; MUAC: Middle-Upper Arm Circumference; UNICEF: United Nations Children's Fund's; SD: standard deviation; WFH: Weight-For-Height; WHO: World Health Organization.

Authors' contribution

SF wrote the first draft of the article and had the primary responsibility for the final content. SF was involved in all stages from the conception and design, data acquisition, analysis and interpretation. FC was involved in the conception, design and data acquisition as well as in data analysis and in critically revising different draft versions. JN was involved in data analysis and interpretation as well as in critically revising different draft versions. MK was involved in data interpretation and in critically revising different draft versions. All authors read and approved the final manuscript.

Author details

[1] Department of Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. [2] Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. [3] Faculty of Public Health and Policy, LSHTM and Humanitarian Technical Unit, Save the Children, 207 Old Street, London EC1V 9NR, UK.

Acknowledgements

We would like to thank the following people and organisations for sharing the datasets used for this research: Grainne Moloney and Elijah Odundo from FSNAU, Mara Nyawo from UNICEF Khartoum, Dr. Sheila Isanaka (Nutritional epidemiologist) from Epicentre/MSF Paris, Dr. Benjamin Guesdon and Cécile Salpeteur from Action Against Hunger-Paris, Dr. Anne-Marie Mayer and Gudrun Stallkamp from Concern Worldwide and Claudine Prudhon for sharing data from Goal. We would also like to thank Jane Bruce for supporting the Ph.D. from which this study arose.

Competing interests
The authors declare that they have no competing interests.

Funding
This work was supported by the Office of U.S. Foreign Disaster Assistance (OFDA) and the World Food Programme (WFP) Grant Number ITDCZD07. OFDA and WFP had no role in the design, analysis or writing of this article.

Received: 31 July 2015 Accepted: 19 April 2016
Published online: 04 May 2016

References

1. UNICEF. The State of the World's Children. New York: UNICEF; 2014.
2. Black R, Victora C, Walker S, Bhutta Z, Christian P, De Onis M, Ezzati M, Grantham-Mcgregor S, Katz J, Martorell R, Uauy R. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet. 2013;382:427–51.
3. Cluster I-ASCN: Transitioning to the WHO growth standards: implications for emergency nutrition programmes. 2008.
4. Myatt M, Khara T, Collins S. A review of methods to detect cases of severely malnourished children in the community for their admission into community-based therapeutic care programs. Food Nutr Bull. 2006;27:S7–23.
5. Kirkwood BR, Sterne JAC: Essential medical statistics. Blackwell Publishing; 2003.
6. Finney DJ. Probit analysis. Cambridge: Cambridge University Press; 1971.
7. Bulmer MG. Principles of statistics. New York: Dover Publications Inc; 1979.
8. DeCarlo LT. On the meaning and use of kurtosis. Psychol Methods. 1997;2:292–307.
9. Thode HC. Testing for normality. New York: Marcel Dekker; 2002.
10. D'Agostino RB, Stephens MA. Goodness-of-fit techniques. New York: Marcel Dekker; 1986.
11. De Boor C. A practical guide to splines (revised edition); 2001.
12. Cleveland WS. Lowess—a program for smoothing scatterplots by robust locally weighted regression. Am Stat. 1981;35:54.
13. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc. 1979;74:829–36.
14. Carroll RJ, Ruppert D. On prediction and the power transformation family. Biometrika. 1981;68:609–15.
15. Osborne JW. Improving your data transformations: Applying the Box-Cox transformation. In: Improving your data transformations: applying the Box-Cox transformation (editor), vol 15, City;2010.
16. Osborne JW. Notes on the use of data transformations. In: Notes on the use of data transformations. Practical assessment, research and evaluation (editor). City;2002.
17. Altman DG, Martin BJ. Statistics notes: the normal distribution. BMJ. 1995;310:298.
18. Mei Z, Grummer-Strawn LM. Standard deviation of anthropometric Z-scores as a data quality assessment tool using the 2006 WHO growth standards: a cross country analysis. Bull World Health Organ. 2007;85:441–8.
19. Waterlow JC, Buzina R, Keller W, Lane JM, Nichaman MZ, Tanner JM. The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years. Bull World Health Organ. 1977;55:489–98.
20. Blanton C, Bilukha O. The PROBIT approach in estimating the prevalence of wasting: revisiting bias and precision. Emerg Themes Epidemiol. 2013;10:8.
21. Binkin N, Sullivan K, Staehling N, Nieburg P. Rapid nutrition surveys: How many clusters are enough? Disasters. 1995;16:99–103.
22. Spiegel PB, Salama P, Maloney S, van der Veen A. Quality of malnutrition assessment surveys conducted during famine in Ethiopia. JAMA. 2004;292:613–8.
23. WHO. 2006. World Health Organization (WHO) Child Growth Standards. http://www.who.int/childgrowth/en/. Accessed 23 June 2015.
24. Software: STATA 13. In: STATA 13 (editor). pp. StataCorp. 2013. Stata Statistical Software: Release 2013. College Station, TX: StataCorp LP.
25. RStudio Team (2015). RStudio: integrated development for R. RStudio, Inc., Boston, MA. http://www.rstudio.com/.
26. Dale NM, Myatt M, Prudhon C, Briend A. Assessment of the PROBIT approach for estimating the prevalence of global, moderate and severe acute malnutrition from population surveys. Public Health Nutr. 2013;16:858–63.

Chapter 5 – Omitting oedema measurement: how much acute malnutrition are we missing?

In areas where kwashiorkor accounts for a non-negligible proportion of all SAM (e.g. parts of West, Central and Southern Africa), the proposed PROBIT Methods based on MUAC alone might yield considerable underestimates of SAM and GAM if MUAC cut-offs only capture a limited fraction of bilateral oedema cases: the overlap between low MUAC and oedema needs to be analysed. If MUAC-based cut-offs are highly sensitive for oedema, as suggested by one study, the method would yield unbiased prevalence estimates, all else being equal; if on the other hand sensitivity of MUAC cut-offs for oedema were low, the method would result in under-estimation.

This research paper describes the different steps undertaken to fulfil objective III to "examine the association between MUAC and bilateral oedema". It describes prevalence estimates from surveys collected; assesses the overlap between oedematous malnutrition and wasting overall and per region; and evaluates the overall and regional contribution of oedematous malnutrition to prevalence estimates.

Supplemental online figures and tables can be found Annex III and IV respectively. The paper is also complemented by previous work I had done that looked at the sensitivity and specificity of different cut-off points of MUAC in the detection of oedema cases by building receiver operating characteristic (ROC) curves (Annex V). That work supports finding from this paper: the sensitivity of MUAC cut-offs for oedema differs regionally.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Severine Frison
Principal Supervisor	Marko Kerac
Thesis Title	Middle-upper arm circumference for nutritional surveillance in crisis-affected populations: Development of a method

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	American Journal of Clinical Nutrition. 2015 102: 1176-1181		
When was the work published?	16/09/2015		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic peer review?	Yes

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am the main author and was involved in all stages from the conception and design, literature search, analysis and interpretation to drafting the article and writing the final version.

Student Signature: SFrison **Date:** 12/08/16

Supervisor Signature: M. Kerac **Date:** 12/08/16

AJCN. First published ahead of print September 16, 2015 as doi: 10.3945/ajcn.115.108282.

Omitting edema measurement: how much acute malnutrition are we missing?[1,2]

Severine Frison,[4] Francesco Checchi,[5] and Marko Kerac[4]*

[4]Department of Population Health, London School of Hygiene and Tropical Medicine (LSHTM), London, United Kingdom, and [5]Faculty of Public Health and Policy, LSHTM & Humanitarian Technical Unit, Save the Children, London, United Kingdom

ABSTRACT

Background: Acute malnutrition is a major public health issue in low-income countries. It includes both wasting and edematous malnutrition, but the terms *wasting* and *acute malnutrition* are often used interchangeably. Little is known about the burden of edematous malnutrition, and few large-scale surveys measure it.

Objective: Most acute malnutrition might be captured by the measurement of wasting alone, but this is unknown. This article aims to fill this gap.

Design: This article presents a secondary data analysis of 852 nutrition cross-sectional survey data sets of children aged 6–59 mo. The data sets assembled included surveys from East, West, South, and Central Africa; the Caribbean; and Asia. The overlap between edematous malnutrition and wasting was assessed, and the impact of including/excluding edema on acute malnutrition prevalence estimates was evaluated.

Results: The prevalence of edematous malnutrition varied from 0% to 32.9%, and children were more likely to have bilateral edema in Central and South Africa (OR: 4; 95% CI: 2.8, 5.6). A large proportion of children with edematous malnutrition were not wasted [62% and 66% based on midupper arm circumference (MUAC) and weight-for-height (WFH), respectively], and most were not severely wasted (83% and 86% based on MUAC and WFH, respectively). When wasting and global acute malnutrition prevalence estimates as well as severe wasting and severe acute malnutrition prevalence estimates overall were compared, the differences between estimates were small (median of 0.0% and mean of 0.3% based on WFH and MUAC for global estimates and slightly higher median of 0.1% and mean of 0.4% based on MUAC and WFH, respectively, for the severe forms), but the picture was different at the regional level.

Conclusions: The terms *acute malnutrition* and *wasting* should not be used interchangeably. The omission of the measurement of edema can have important repercussions, especially at the nutrition program level. *Am J Clin Nutr* doi: 10.3945/ajcn.115.108282.

Keywords: acute malnutrition, nutrition surveillance, edematous malnutrition, wasting, kwashiorkor, mid-upper arm circumference, weight-for-height

INTRODUCTION

Acute malnutrition is a major public health issue throughout the developing world. Current definitions recognize 2 types—wasting (marasmus) and edematous malnutrition (kwashiorkor). UNICEF's latest report on the State of the World's Children (1) estimates that 10% of children aged <5 y in least developed countries are wasted. Of the 6.9 million estimated deaths among children aged <5 y annually, >800,000 deaths (12.6%) (2) are attributed to wasting (3–8). Similar estimates are not available for edematous malnutrition.

Edematous malnutrition is characterized by the presence of bilateral pitting edema, at a minimum on the dorsum of both feet, and is an independent criterion for identifying severe acute malnutrition (SAM).[6] Wasting is defined as weight-for-height (WFH) <−2 SD from the WHO mean reference value and/or midupper arm circumference (MUAC) <125 mm. Severe wasting is defined as WFH <−3 SD and/or MUAC <115 mm. Global acute malnutrition (GAM) is characterized as WFH <−2 SD and/or MUAC <125 mm and/or edema. SAM is defined as WFH <−3 SD and/or MUAC <115 mm and/or edema (9).

Ongoing surveillance of acute malnutrition is an essential instrument for the detection of nutritional emergencies and for planning interventions. Although there are debates about its use (10), the WHO classification of acute malnutrition prevalence is used by most organizations to assess the severity of a crisis. Although it is meant to be applied to GAM estimates, it is based on the prevalence of wasting by using WFH alone: the "reference

[1] Supported by the Office of US Foreign Disaster Assistance (OFDA) and the World Food Programme (WFP) (grant ITDCZD07).

[2] OFDA and WFP had no role in the design, analysis, or writing of this article.

[3] Supplemental Tables 1 and 2 and Supplemental Figures 1 and 2 are available from the "Supplemental data" link in the online posting of the article and from the same link in the online table of contents at http://ajcn.nutrition.org.

*To whom correspondence should be addressed. E-mail: severine.frison@gmail.com.

[6] Abbreviations used: GAM, global acute malnutrition; MUAC, midupper arm circumference; SAM, severe acute malnutrition; WFH, weight-for-height.

Received February 5, 2015. Accepted for publication August 20, 2015.
doi: 10.3945/ajcn.115.108282.

estimates" used for policy and program planning are mainly based on Demographic and Health Surveys and Multiple Indicator Cluster Surveys, which rarely take edema cases into account (11–13).

Although the importance of bilateral edema measurement has been discussed for several decades (14) and the case fatality rate among children with edematous malnutrition ranges from 1.5% to 27% (15–19), it is not part of the World Health Assembly Indicators (20). It was not discussed in the latest *Lancet* series on maternal and child nutrition (21) and is mentioned only once in the notes of the latest Global Nutrition Report (22). Furthermore, the Generation Nutrition campaign, a network of major nutrition civil society organizations (23), has been using *wasting* and *acute malnutrition* interchangeably. This is largely because little is known about the burden of edematous malnutrition.

Because of the common pathways between wasting and edematous malnutrition (24), most edemas cases could be expected to be included in the measurement of wasting, but this is unknown. Experts highlighted this gap in 2013 (25, 26), and this article aims to fill it by *1*) describing prevalence estimates from surveys collected, *2*) assessing the overlap between edematous malnutrition and wasting overall and per region, and *3*) evaluating the overall and regional contribution of edematous malnutrition to prevalence estimates.

METHODS

Study design and inclusion criteria

A total of 1068 cross-sectional survey data sets from various settings were shared by 6 organizations (UNICEF, Food Security and Nutrition Analysis Unit, Epicentre/Médecins Sans Frontières, Action Against Hunger, Concern Worldwide, and Goal). No formal sample size calculation was used. The study size depended on the availability of surveys and on specific inclusion criteria. Eligible data sets had to *1*) include anthropometric data, including MUAC, edema, age, weight, and height, as well as meta-data on country, livelihood, residence, cluster (if cluster surveys), and date, and *2*) have a minimum of 25 clusters if cluster surveys (27, 28). The last criterion aimed to minimize selection bias. The surveys were exhaustive or clustered surveys. The data sets were cleaned and records with extreme or missing values were excluded. Children were excluded if any of the following data were missing: age, sex, height, weight, MUAC, and edema. Those with highly improbable extreme values ("flags") were also excluded from analysis: MUAC <85 mm or >200 mm, age <6 mo or >59 mo, weight-for-age <−6.0 SD or >+5.0 SD, height-for-age <−6.0 SD or >+6.0 SD, and WFH <−5.0 SD or >+5.0 SD (WHO "flags" were applied on SD for WFH, weight-for-age, and height-for-age).

Database

Of the 1068 surveys collected, 852 were included in the secondary data analysis (55 exhaustive surveys and 797 clustered surveys). The 852 surveys contained 694,108 children, of whom 25,134 had highly improbable values and were excluded from the analysis. The database included 6 variables for anthropometric measures (sex, MUAC, edema, age, weight, and height), 6 metadata variables [organization, country, livelihood, residence, cluster (when cluster surveys), and date], and 3 indexes based on WHO standards (WFH, weight-for-age, and height-for-age) added by using the WHO's "Child Growth Standards" package (29).

Data analysis

The prevalence of severe and global wasting and acute malnutrition (based on MUAC and WFH) and the prevalence of edema cases were computed for each survey and described (mean, median, range, variance) overall. A mixed-effect multivariable logistic regression with edema as a binary dependent variable was computed with variables showing univariable associations with edema. The overlaps between the number of wasted and severely wasted children (MUAC or WFH) and edema cases were examined by building Venn diagrams overall and by region. The impact of edematous malnutrition on global estimates was assessed by plotting wasting/severe wasting vs. GAM/SAM estimates overall and per region. The summary statistics of the differences between wasting and GAM and severe wasting and SAM were computed for MUAC and WFH. RStudio (RStudio Inc.) (30) and STATA 13 (StataCorp) (31) were used for all analyses.

Ethical standards disclosure

Because the project involved secondary data analysis, ethics approval for the project was sought and obtained from the Ethics Committee of the London School of Hygiene and Tropical Medicine.

RESULTS

In total, 852 surveys from 38 countries were included in the analysis, and the final database comprised a total of 668,975 children aged 6–59 mo. Surveys were conducted from 1992 to 2011, with 95% from 2000. The most represented region was East Africa, with 65.0% of the surveys conducted in the region (**Table 1**). The sample size of the surveys varied from 122 to 3491 children. The mean sample size was 785, and the median was 815. Before the development of the Standardized Monitoring and Assessment of Relief and Transitions method for anthropometric and mortality surveys (32), most nutritional surveys were conducted by using a 30-by-30 cluster survey approach, which translates into a large number of surveys with a sample size close to 900 children.

Summary of prevalence results from the 852 surveys

The database included 3230 edema cases, and 65,680 children were classified as wasted according to MUAC and 93,406 according to WFH. The prevalence of wasting and severe wasting

TABLE 1
Surveys per region (n = 852)

Region	Surveys, n (%)
East Africa	554 (65.0)
West Africa	97 (11.4)
Central and South Africa	128 (15.0)
Caribbean	13 (1.5)
Asia	60 (7.0)

as well as the prevalence of SAM and GAM observed overall varied across the surveys. Wasting varied from 1% to 47.7% based on MUAC and from 0.4% to 42.8% based on WFH. Similarly, GAM ranged between 1% and 48.7% and 1.4% and 42.9% based on MUAC and WFH, respectively. The prevalence of bilateral edema across surveys ranged from 0% to 32.9%, and 325 of 852 (38.1%) surveys had no cases (**Table 2**). The median prevalence of edema cases was very low (≤0.2%) in all regions expect Central and South Africa (0.6%). The mean was also fairly low in all regions (≤0.6% or less) except in Central and South Africa (1.2%) (**Supplemental Table 1**).

The logistic regressions of edema as a dependent variable showed statistically significant univariable associations with the following independent variables: livelihoods, residence, region, and age (categorical variable). Adding region, residence, livelihood, or age in a multivariable logistic regression resulted in a statistically significant improvement in model fit (likelihood ratio test). The model is presented in **Table 3**. Edema cases were 4 times more likely to be found in Central and South Africa (OR: 4.0; 95% CI: 2.8, 5.6; $P < 0.001$) and 0.7 times less likely to occur in children aged >36 mo (OR: 0.7; 95% CI: 0.6, 0.7; $P < 0.001$).

Overlap between edematous malnutrition and wasting overall and per region

A large proportion of children with edematous malnutrition were not wasted. About two-thirds of children with bilateral edema were not wasted, based on MUAC (62%) or WFH (66%). Most children with edema were not severely wasted based on MUAC (83%) or WFH (86%) (**Figure 1**). The proportions of overlap were similar in all regions (**Supplemental Figure 1**). Central and South Africa had the largest proportion of cases (1394 of 3230).

Although more children were classified as wasted by using WFH compared with MUAC overall (92,302 using WFH vs. 64,447 using MUAC), the overlap between MUAC and edema was larger than between WFH and edema (MUAC included 1233 edema cases, whereas WFH included 1104) (see Figure 1).

TABLE 2
Summary estimates of the surveys included in the analysis (n = 852)[1]

	Minimum/maximum, %	Lower/upper quartile, %	Median/mean, %
MUAC			
<125 mm	1.0/47.7	5.9/12.7	8.8/9.9
<115 mm	0.0/20.6	0.9/2.9	1.7/2.2
MUAC			
GAM	1.0/48.7	6.0/13.1	9.0/10.2
SAM	0.0/35.3	1.1/3.3	1.9/2.6
WFH			
<−2 SD	0.4/42.8	8.7/19.2	13.4/14.2
<−3SD	0.0/19.1	1.4/4.4	2.7/3.2
WFH			
GAM	1.4/42.9	9.0/19.6	13.7/14.5
SAM	0.0/35.2	1.7/4.8	3.0/3.7
Bilateral edema	0.0/32.9	0.0/0.5	0.1/0.5

[1]GAM, global acute malnutrition; MUAC, midupper arm circumference; SAM, severe acute malnutrition; WFH, weight-for-height.

TABLE 3
Logistic regression of bilateral edema (852 surveys)

Independent variable	OR[1] (95% CI)	SE	χ^2	P value[2]
Region				
East Africa	—	—	—	—
Asia	0.679 (0.405, 1.136)	0.179	−1.47	0.141
Caribbean	2.359 (0.979, 5.681)	1.058	1.91	0.056
Central and South Africa	3.976 (2.837, 5.572)	0.685	8.02	<0.001
West Africa	0.743 (0.497, 1.110)	0.152	−1.45	0.147
Residence				
Rural	—	—	—	—
Displaced population	1.496 (1.080, 2.074)	0.249	2.42	0.016
Other	0.840 (0.567, 1.245)	0.169	−0.87	0.385
Urban	0.475 (0.297, 0.761)	0.114	−3.09	0.002
Livelihood				
Agriculture	—	—	—	—
Agropastoral	0.569 (0.418, 0.775)	0.090	−3.58	<0.001
Other	1.102 (0.775, 1.567)	0.198	0.54	0.589
Pastoral	0.980 (0.666, 1.441)	0.193	−0.1	0.917
Age group, mo				
6–35	—	—	—	—
36–59	0.659 (0.610, 0.711)	0.026	−10.67	<0.001

[1]Robust OR (mixed-effect logistic regression cluster by survey).
[2]Wald test. The reference group was the largest

Overall and regional impact of edematous malnutrition on prevalence estimates

The scatterplots of wasting vs. GAM suggested that only few outlier surveys had substantially different estimates based on MUAC or WFH. The inclusion of edema cases had a stronger impact when looking at severe wasting vs. SAM (**Figure 2**). A similar pattern was observed in each region, especially in East Africa, West Africa, and Asia (see **Supplemental Figure 2**).

The differences between wasting and GAM and severe wasting and SAM overall are summarized in **Table 4**. Differences between global estimates were small overall (median = 0.0% and mean = 0.3%) and slightly higher between severe wasting and SAM (median = 0.1% and mean = 0.4%) for MUAC and WFH. Although the impact of edema inclusion was greater in the Caribbean and Central and South Africa, where the prevalence of edema were higher, the differences between global estimates were not substantial. The differences were more important in Central and South Africa, especially when looking at the differences between severe wasting and SAM (median = 0.4% and 0.5% for MUAC and WFH, respectively; mean = 1.0% and 1.1% for MUAC and WFH, respectively) (Supplemental Figure 2 and **Supplemental Table 2**).

DISCUSSION

Although *wasting* and *acute malnutrition* are often used interchangeably, our results emphasize that they are not equivalent. This has different implications depending on the purpose of the nutritional assessment. Two-thirds of edema cases are missed by measuring wasting only, and >80% missed by assessing severe wasting whether using MUAC or WFH. The difference in prevalence between wasting and GAM and between severe wasting and SAM is rarely statistically significant overall, but the picture is different at the regional level.

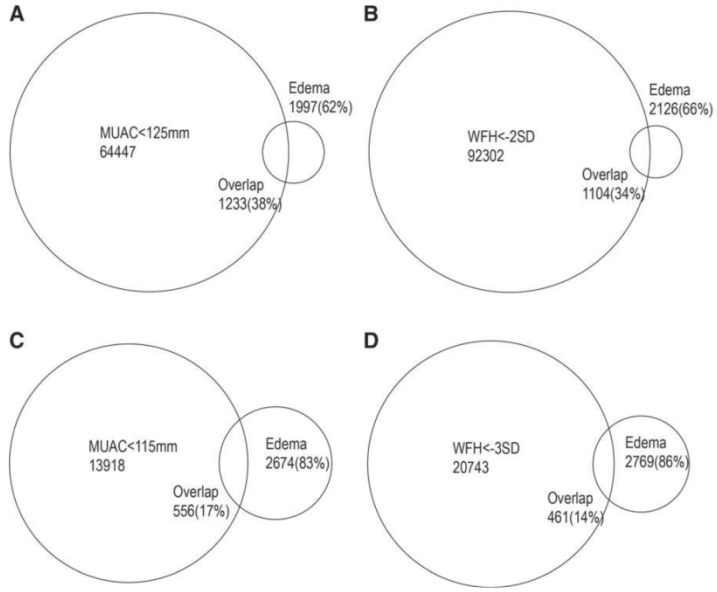


FIGURE 1 Overall overlap between bilateral edema and wasting based on MUAC (A) and WFH (B) and between edema bilateral and severe wasting based on MUAC (C) and WFH (D) (n = 852). MUAC, midupper arm circumference; WFH, weight-for-height.

At the nutrition program level, the implications are considerable. Using wasting indexes only in any context of screening for potential admission to a treatment program will capture a minority of edema cases, thereby excluding a numerically rare but highly lethal form of malnutrition from the possibility of treatment. Furthermore, nutrition programs treating SAM rely on caseload estimates for program planning, supply chain management, and human resource requirements. The annual burden of SAM is calculated as follows: burden = population 6–59 mo × prevalence × 2.6, where the population refers to the population of children aged 6–59 mo in the program area, prevalence is the prevalence of SAM for children aged 6–59 mo, and 2.6 is a factor to convert prevalence into incidence based on the expected duration of SAM episodes (33). Whether the prevalence of SAM used includes edematous malnutrition has an important impact on the burden calculation in countries with a higher proportion of edema cases (i.e., in Central and South Africa). Currently, many national burden estimates are based on Multiple Indicator Cluster Surveys and Demographic and Health Surveys that do not include edematous malnutrition and therefore underestimate the burden of SAM in certain countries, which has a detrimental impact on programs.

At the population level, the implications are less marked. When looking at trends of global prevalence over time to predict whether a situation is getting worse or is worse than the previous year at the same time, including or not including edema makes relatively little difference. The differences between wasting and GAM are very small overall. Using MUAC for the estimation of wasting included more edema cases than WFH overall. Surveys used as an early warning system or to confirm the severity of a situation could help classify a situation based on wasting alone, particularly if using MUAC. In Central and South Africa, where the largest number and prevalence of edema cases were observed, the differences between estimates were higher, which indicates this may not be applicable to all countries.

There are 2 main limitations to this study. One is the selection of surveys. The database was built based on available small-scale surveys that were mainly conducted in areas where there was suspicion of a problem compared with national Demographic and Health Surveys and Multiple Indicator Cluster Surveys that are conducted every 3–5 y and show long-term trends. This explains the greater number of African surveys in the sample compared with other parts of the world: our data are thus not representative of global epidemiology. This may skew the burden of edematous malnutrition observed in our data set. The second is the fact that edematous malnutrition should be assessed by using incidence rather than prevalence measures. This study uses prevalence, which is not an optimal basis to measure the impact of the inclusion of edema cases on the measure of acute malnutrition. The use of incidence is likely to show bigger differences between the estimates of acute malnutrition and wasting because the duration of kwashiorkor episodes is very short, and kwashiorkor cases are less likely to be picked up by a cross-sectional survey (26, 33, 34).

Other limitations include the small number of surveys (13) from the Caribbean and the fact that they were all from Haiti. Also important to take account of in future work is the seasonality of SAM: the timing of the surveys might affect the prevalence of edema cases. Measurement errors might also influence survey results. There is no reason to believe that it affects our findings where hundreds of surveys are pooled, but in a single survey, different final prevalence results might arise if observers are not well trained or supervised and there is different intraobserver/interobserver variability assessing different anthropometric measures. Finally, the extent to which to results presented here

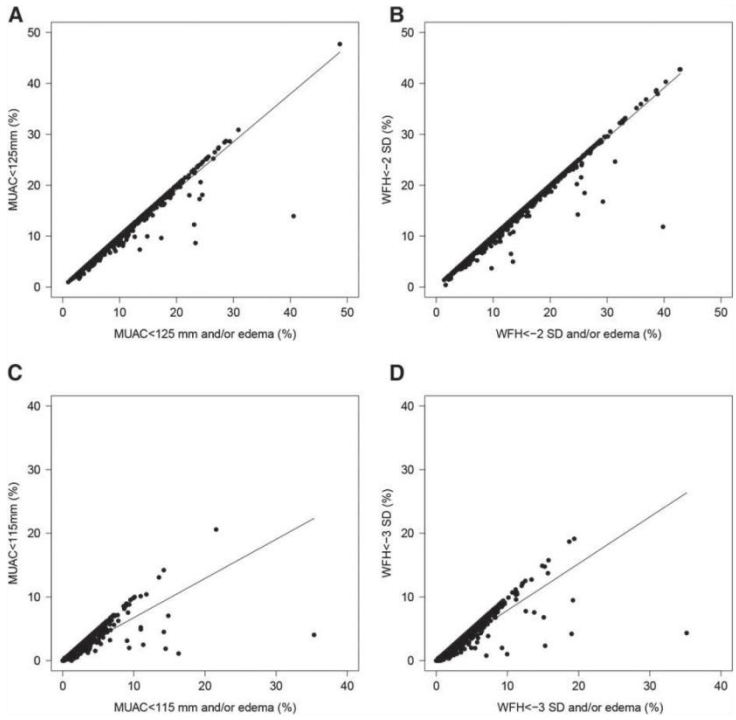


FIGURE 2 Overall wasting vs. acute malnutrition on MUAC (A) and WFH (B) and severe wasting vs. severe acute malnutrition with MUAC (C) and WFH (D) (n = 852). MUAC, midupper arm circumference; WFH, weight-for-height.

reflect the national program is unclear, and it would thus be interesting to look at the proportion of admissions based on bilateral edema per country.

In conclusion, although *wasting* is often used as a substitute for *acute malnutrition*, the 2 terms are not interchangeable. Using wasting alone instead of acute malnutrition can have important repercussions, particularly at the nutrition program level. Although wasting alone can be used to classify the severity of a situation based on global estimates in countries with a low burden of edema cases, edematous malnutrition should be included in nutrition surveys. The measurement of bilateral edema does not imply heavy time or cost implications. Furthermore, wasting and edematous malnutrition have common causes (24), are both detected at the community level and in health centers, and are managed in the same programs with the same treatments. We thus recommend that both should be systematically included in nutrition surveillance.

We thank the following people and organizations for sharing the data sets used for this research: Grainne Moloney and Elijah Odundo from the Food Security and Nutrition Analysis Unit, Mara Nyawo from UNICEF, Sheila Isanaka (nutritional epidemiologist) from Epicentre/Médecins Sans Frontières, Benjamin Guesdon and Cécile Salpeteur from Action Against Hunger, Anne-Marie Mayer and Gudrun Stallkamp from Concern Worldwide, Goal, and Claudine Prudhon. We also thank Jane Bruce for supporting the dissertation (by SF) from which this study arose.

The authors' responsibilities were as follows—SF: wrote the first draft of the article and had the primary responsibility for the final content; SF, FC, and MK: were involved in conception and design, data acquisition, analysis, and interpretation; FC and MK: critically revised different draft versions; and all authors: approved the final version of the article. The authors reported no conflicts of interest related to this study.

TABLE 4
Summary statistics of the differences between estimates of wasting and GAM and severe wasting and SAM using MUAC or WFH overall (n = 852)[1]

	Minimum	Lower quartile	Median	Mean	Upper quartile	Maximum
GAM–wasting (MUAC)	0.0	0.0	0.0	0.3	0.3	26.7
GAM–wasting (WFH)	0.0	0.0	0.0	0.3	0.3	28.0
SAM–severe wasting (MUAC)	0.0	0.0	0.1	0.4	0.4	31.3
SAM–severe wasting (WFH)	0.0	0.0	0.1	0.4	0.4	30.8

[1]Values are percentages. GAM, global acute malnutrition; MUAC, midupper arm circumference; SAM, severe acute malnutrition; WFH, weight-for-height.

REFERENCES

1. UNICEF. The state of the world's children. New York: UNICEF; 2014.
2. Black RE, Victora C, Walker S, Bhutta Z, Christian P, De Onis M, Ezzati M, Grantham-Mcgregor S, Katz J, Martorell R, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet 2013;382:427–51.
3. Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, Mathers C, Rivera J; Maternal Child Undernutrition Study Group. Maternal and child undernutrition: global and regional exposures and health consequences. Lancet 2008;371:243–60.
4. Black RE, Morris SS, Bryce J. Where and why are 10 million children dying every year? Lancet 2003;361:2226–34.
5. Collins S. Treating severe acute malnutrition seriously. Arch Dis Child 2007;92:453–61.
6. Collins S, Dent N, Binns P, Bahwere P, Sadler K, Hallam A. Management of severe acute malnutrition in children. Lancet 2006;368:1992–2000.
7. Kapil U, Sachdev HP. Management of children with severe acute malnutrition a national priority. Indian Pediatr 2010;47:651–3.
8. Pelletier DL, Frongillo EA Jr., Schroeder DG, Habicht JP. The effects of malnutrition on child mortality in developing countries. Bull World Health Organ 1995;73:443–8.
9. WHO, UNICEF. WHO child growth standards and the identification of severe acute malnutrition in infants and children: a joint statement by the World Health Organization and the United Nations Children's Fund. Geneva (Switzerland): World Health Organization; 2009.
10. SPHERE. Food security, nutrition and food aid. In: Humanitarian charter and minimum standards of disaster response. Oxford (United Kingdom): The Sphere Project; 2004.
11. The DHS Program. The Demographic and Health Surveys Program [Internet]. Rockville (Md): DHS; 2014 [cited 2014 Nov 12]. Available from: http://www.dhsprogram.com/.
12. UNICEF. Multiple Indicator Cluster Survey (MICS) [Internet]. New York: UNICEF; 2014 [cited 2014 Nov 12]. Available from: http://www.unicef.org/statistics/index_24302.html.
13. Shoham J, Watson F, Dolan C. The use of nutritional indicators in surveillance systems. London: International Public Nutrition Resource Group; 2001.
14. Franklin RR, Dikassa LN, Bertrand WE. The impact of oedema on anthropometric measurements in nutritional surveys: a case from Zaire. Bull World Health Organ 1984;62:145–50.
15. Bachou H, Tumwine J, Mwadime R, Tylleskar T. Risk factors in hospital deaths in severely malnourished children in Kampala, Uganda. BMC Pediatr 2006;6:7.
16. Prudhon C, de Radigues X, Dale N, Checchi F. An algorithm to assess methodological quality of nutrition and mortality cross-sectional surveys: development and application to surveys conducted in Darfur, Sudan. Popul Health Metr 2011;9:57.
17. Schofield C, Ashworth A. Why have mortality rates for severe malnutrition remained so high? Bull World Health Organ 1996;74:223–9.
18. Ciliberto M, Manary M, Ndekha M, Briend A, Ashorn P. Home-based therapy for oedematous malnutrition with ready-to-use therapeutic food. Acta Paediatrica. 2006;95:1012–5.
19. Sadler K, Kerac M, Collins S, Khengere H, Nesbitt A. Improving the management of severe acute malnutrition in an area of high HIV prevalence. J Trop Pediatr 2008;54:364–9.
20. WHO. Sixty-fifth World Health Assembly. Geneva (Switzerland): World Health Organization; 2012.
21. Horton R, Lo S, Black RE, Alderman H, Bhutta ZA, Gillespie S, Haddad L, Horton S, Lartey A, Mannar V, et al. Series on maternal and child nutrition. The Lancet 2013.
22. Global Nutrition Report Study Group, Independent Expert Group. Global Nutrition Report 2014 Nov 13.
23. Generation Nutrition. Closing the gap: towards a 2030 wasting target. London: Generation Nutrition; 2014.
24. UNICEF. Conceptual framework of the causes of malnutrition. New York: UNICEF; 1991.
25. Briend A, Myatt M, Dent N, Brown R. Putting kwashiorkor on the map. CMAM Forum 2013.
26. Briend A, Collins S, Golden M, Manary M, Myatt M. The burden of severe acute malnutrition estimated in the Lancet series is a minimum rather than an actual estimate. Lancet 2013;382:1549.
27. Binkin N, Sullivan K, Staehling N, Nieburg P. Rapid nutrition surveys: how many clusters are enough? Disasters 1992;16:97–103.
28. Spiegel PB, Salama P, Maloney S, van der Veen A. Quality of malnutrition assessment surveys conducted during famine in Ethiopia. JAMA 2004;292:613–8.
29. WHO. Child growth standards [Internet]. [cited 2014 Nov 12]. Available from: http://www.who.int/childgrowth/software/en/.
30. RStudio [Internet]. [cited 2014 Nov 12]. Available from: http://www.rstudio.org/.
31. STATA 13 [Internet]. [cited 2014 Nov 12]. Available from: http://www.stata.com/products/.
32. SMART [Internet]. Measuring mortality, nutritional status, and food security in crisis situations: SMART methodology. 2006. [cited 2014 Nov 12]. Available from: http://smartmethodology.org/.
33. Garenne M, Willie D, Maire B, Fontaine O, Eeckels R, Briend A, Van den Broeck J. Incidence and duration of severe wasting in two African populations. Public Health Nutr 2009;12:1974–82.
34. Isanaka S, Grais R, Briend A, Checchi F. Estimates of the duration of untreated acute malnutrition in children from niger. Am J Epidemiol 2011;173:932–40.

Chapter 6 – Exploration of further method assumptions

Chapter 4 and 5 examined the first two assumptions behind the proposed method: the normality of MUAC distributions (Objective 2) and the overlap between MUAC thresholds and oedematous malnutrition (Objective 3).

This chapter assesses the assumptions linked to MUAC SD (Objective 4):

- the variability in MUAC SD from our database of nutritional surveys done in crisis settings is representative of the variability that we can expect in the future
- MUAC SD is itself not strongly associated with average nutritional status
- within given strata, defined based on region, livelihood status or residence status, the standard deviation (SD) of the MUAC normal distribution falls within a reasonably narrow range

Exploration of further method assumptions

Introduction

An important assumption behind the proposed method is that the variability in the standard deviation (SD) of the Middle-Upper Arm Circumference (MUAC) from our database of nutritional surveys done in crisis settings is representative of the variability that we can expect in the future and that MUAC SD is itself not strongly associated with average nutritional status. Furthermore, within given strata of the world, defined based on region, livelihood status and/or residence status, the SD of the MUAC distribution falls within a reasonably narrow range.

The aim of this chapter was to assess the three assumptions above:

- Describe the MUAC SD over time and discuss its representativeness

- Assess the association of MUAC with average nutritional status

- Identify strata that maximise differences in the SD of MUAC across strata and minimise differences within strata.

Method

Please refer to chapter 3 for details on data collection and on the description of the database used.

Weighted and un-weighted summary statistics and box-plots of MUAC SD were computed overall. The variability of MUAC SD was assessed by plotting box-plots of MUAC SD over time and by computing a linear regression with MUAC SD as dependent variable and date (categorical) as independent variable. The year 2001 was used as base as the number of surveys in previous years was very low (2 to 9 per year).

The association between MUAC SD and Global Acute Malnutrition (GAM) prevalence based on MUAC was examined: i) MUAC SD was plotted against GAM ; ii) a weighted univariable linear regression was computed with MUAC SD as dependent variable and GAM as continuous independent variable and iii) MUAC SD was described for different GAM categories (<5%; 5-9%; 10-14%; ≥ 15%) using summary statistics and box-plots.

Different approaches were used to stratify the database: i) regression trees: regression trees are a nonparametric technique that can identify the combination of variables and cut-off values for these variables that optimally partition observations (in our case, surveys) into the most similar groups possible (i.e. in this case into strata with relatively homogeneous SD values). Region, livelihood and residence were imputed in the regression. Every possible binary split on every field was assessed and the algorithm selected the split that minimized the sum of the squared deviations from the mean in the two separate partitions. The package “rpart” was used in R to compute the regression trees. The minimum number of observations per node in order for a split to be attempted was set to 20 and the complexity parameter (any split that does not decrease the overall lack of fit by the complexity

factor is not attempted) was set to 0.01; ii) MUAC SD was described using box-plots and summarized by categories for each of the following variables: region, livelihood and residence; and iii) weighted univariable linear regressions were computed with MUAC SD as dependent variable and region, livelihood or residence as independent variables. R-squared of the univariable regressions were compared. R-squared or coefficient of determination reflects the proportion of variance in the dependent variable that is explained by the independent variable.

The analysis was conducted using STATA and R and was weighted using effective sampling weight (see Chapter 3).

Results

Description of MUAC SD overall

The MUAC SD was computed for each survey. The minimum was 9.5 mm and the maximum 19.3 mm with a mean of 12.4mm and a median of 12.4 or 12.5 mm if weighted (see Table 1 and Figure 1). This suggests reasonable homogeneity in SD across the entire dataset.

Table 1: Summary of MUAC SD – Unweighted and weighted

MUAC SD (n=852)	Unweighted	Weighted
Minimum - Maximum (mm)	9.5 – 19.3	9.5 – 19.3
Lower - Upper quartile (mm)	11.5 – 13.2	11.6 – 13.2
Median , Mean (mm)	12.4 , 12.4	12.4 , 12.5

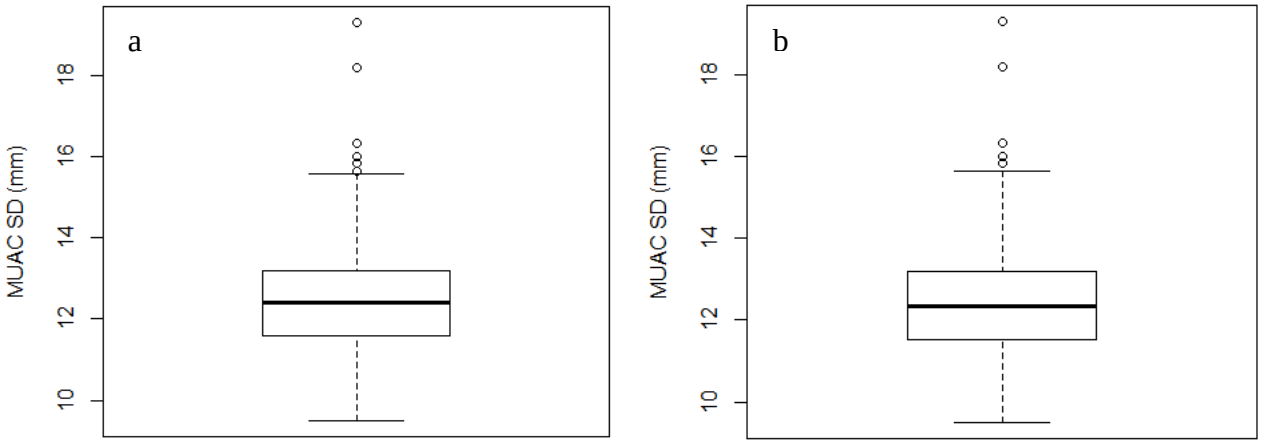


Figure 1: Boxplot of MUAC SD across all surveys (unweighted (a), weighted (b))

The database comprised of surveys from very different settings, regions and livelihood zones which suggest a good representativeness.

Variability of MUAC SD over time

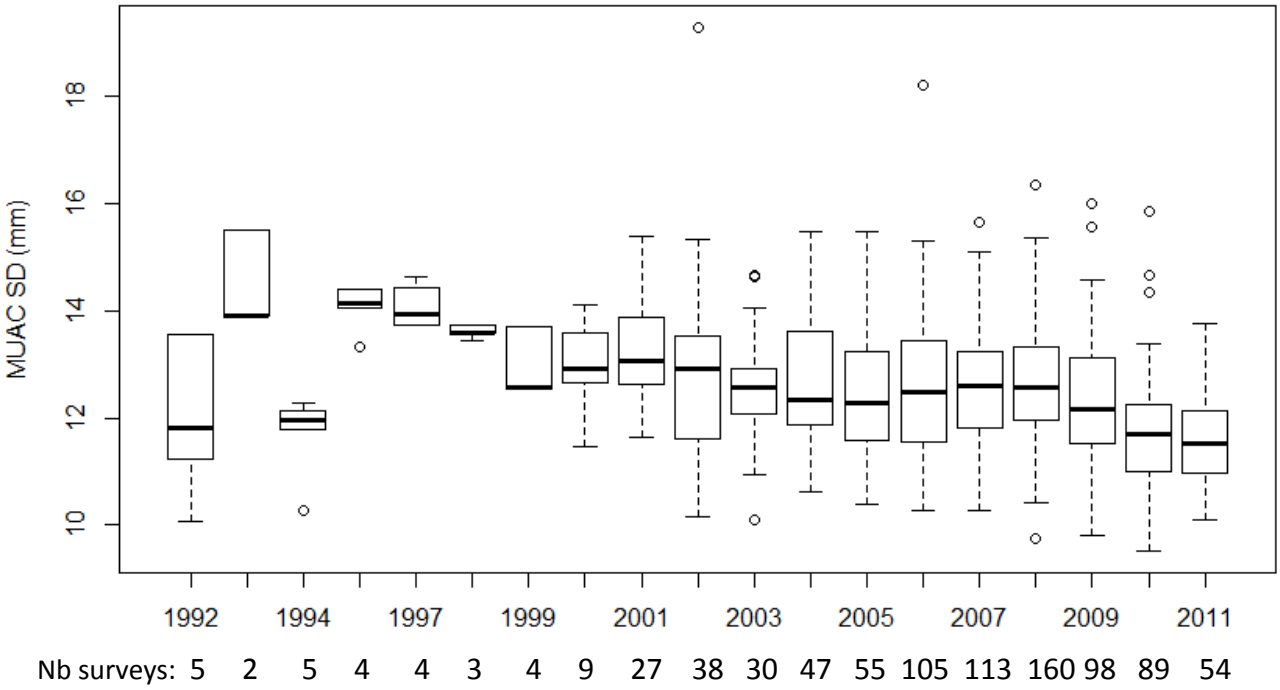


Figure 2: Box plots of MUAC SD over time

The surveys included in the database were collected over a long period of time; the earliest dataset was from 1992 and the latest from 2011. The number of surveys per year varied a lot with very few surveys before 2001 which explained partly the variability observed within years between early 90s and the 2000s (see Table 2). MUAC SD varied little overall and over time (see Figure 1 and 2) but it did seem to decrease slightly from mid/late 90s to 2011 (see Table 2). The minimum mean SD was 11.4 in 1992 and the maximum was 14.6 in 1993.

Table 2: Number of surveys per year

Date	1992	1993	1994	1995	1997	1998	1999	2000	2001	2002
Nb surveys	5	2	5	4	4	3	4	9	27	38
Mean SD	11.4	14.7	11.7	14.0	14.2	13.6	13.0	13.0	13.3	12.9
Date	2003	2004	2005	2006	2007	2008	2009	2010	2011	
Nb surveys	30	47	55	105	113	160	98	89	54	
Mean SD	12.5	12.7	11.7	12.5	12.6	12.6	12.2	11.7	11.7	

The linear regression with MUAC SD as dependent variable and date (categorical) as independent variable showed a statistically significant association ($p<0.001$) (table 3).

Table 3: Overall outcome of the linear regression with MUAC SD as dependent variable and date (categorical) as independent variable

	F- statistic	**R-squared**	**P-value**
Date (categorical)	28.73	0.13	<0.001

The change in MUAC SD could be due to different factors. The quality of surveys might have improved with years. The methodologies and training have change and improved over the last decade. Another hypothesis is the difference in setting and crises over time that could affect the MUAC SD in a different way.

Although the variety of setting and time period covered by the database suggested that the MUAC SD included in the database was representative of the variability that we can expected in the future, the proposed method was examined in simulated surveys from different time periods in order to verify performance did not differ markedly for more recent surveys compared to earlier ones.

Association between MUAC SD and mean levels of GAM

The plot of MUAC SD against GAM prevalence suggests a slight increase in the MUAC SD when the prevalence of GAM based on MUAC increases (see Figure 3). The outliers observed on the figure below were investigated and kept in the linear regression as they were considered as "real" values. The regression was run without outliers and the coefficient for each regression were very similar.

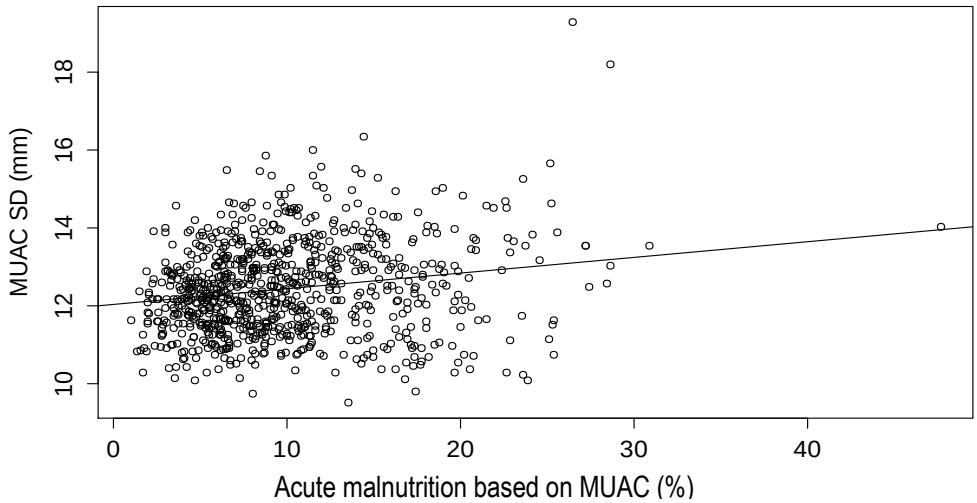


Figure 3: Plot of MUAC SD against GAM

The results of the univariable regression with MUAC SD as dependent variable and GAM as independent variable are presented in the table below.

Table 4: Univariable association between MUAC SD (mm) and GAM prevalence (%) based on MUAC

	Coefficient	95% CI	t-statistic	P-value

GAM (MUAC)	0.037	0.022 – 0.050	4.92	<0.001

The change in MUAC SD is reasonably small but statistically significant (p<0.001) (Table 4). The mean and median change in MUAC SD in different GAM categories varies from 12.1mm to 12.8 mm. The larger MUAC SD was found in the 10-14 % GAM category (see Table 5 and Figure 4). Although significant, the slight difference in MUAC SD found for different GAM prevalence would be expected to have little impact on the performance of the method. This is examined further chapter 7.

Table 5: Summary of MUAC SD per level of GAM

MUAC SD per malnutrition category	GAM <5% (n=143)	GAM : 5-9% (n=361)	GAM : 10-14% (n=205)	GAM ≥ 15% (n=143)
Minimum - Maximum (mm)	10.1 – 14.6	9.7 – 15.8	9.5 – 16.3	9.8 – 19.3
Lower - Upper quartile (mm)	11.5 – 12.7	11.6 – 13.1	11.8 – 13.6	11.5 – 13.6
Median - Mean (mm)	12.1 , 12.1	12.3, 12.4	12.8, 12.8	12.6, 12.5

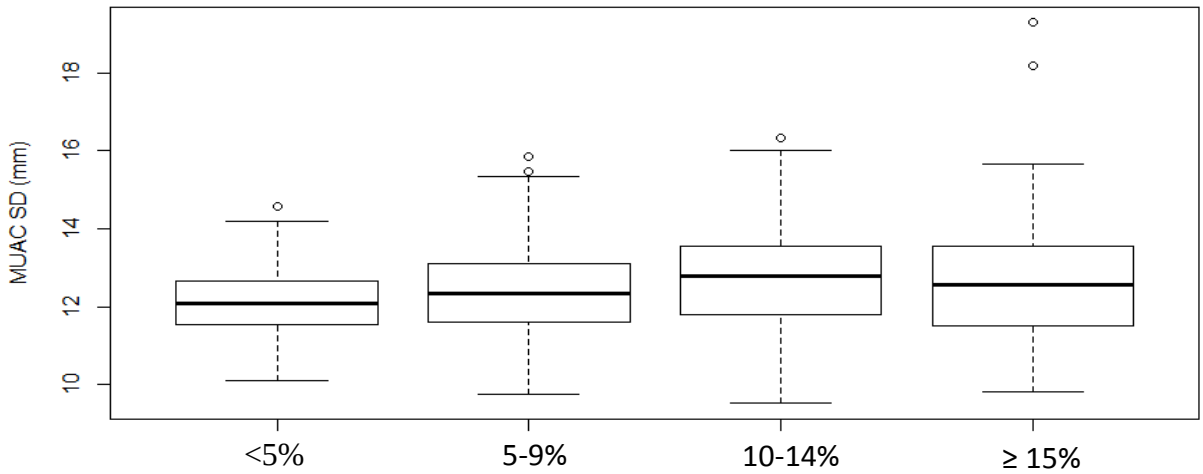


Figure 4: MUAC SD in each GAM category

Stratification of the datasets in order to minimise the variability of MUAC SD

The regression tree below grouped surveys in order minimise the variability of MUAC SD (Figure 5). The number below each grouping corresponds to the mean MUAC SD in each group. Mean MUAC SD varied from 11.8 to 14.1. Although the regression tree below minimised the variability of MUAC SD, it grouped surveys in a fairly unpractical manner. It would be difficult to develop guidelines for a method with such categories (i.e. use 14.1 as MUAC SD if in Asia or East Africa, in Urban area or Displaced and in an agriculture livelihood zone).

The first spilt devided regions into two groups which suggests that deviding the suveys by region might be the best and simplest way to stratify surveys (Figure 5).

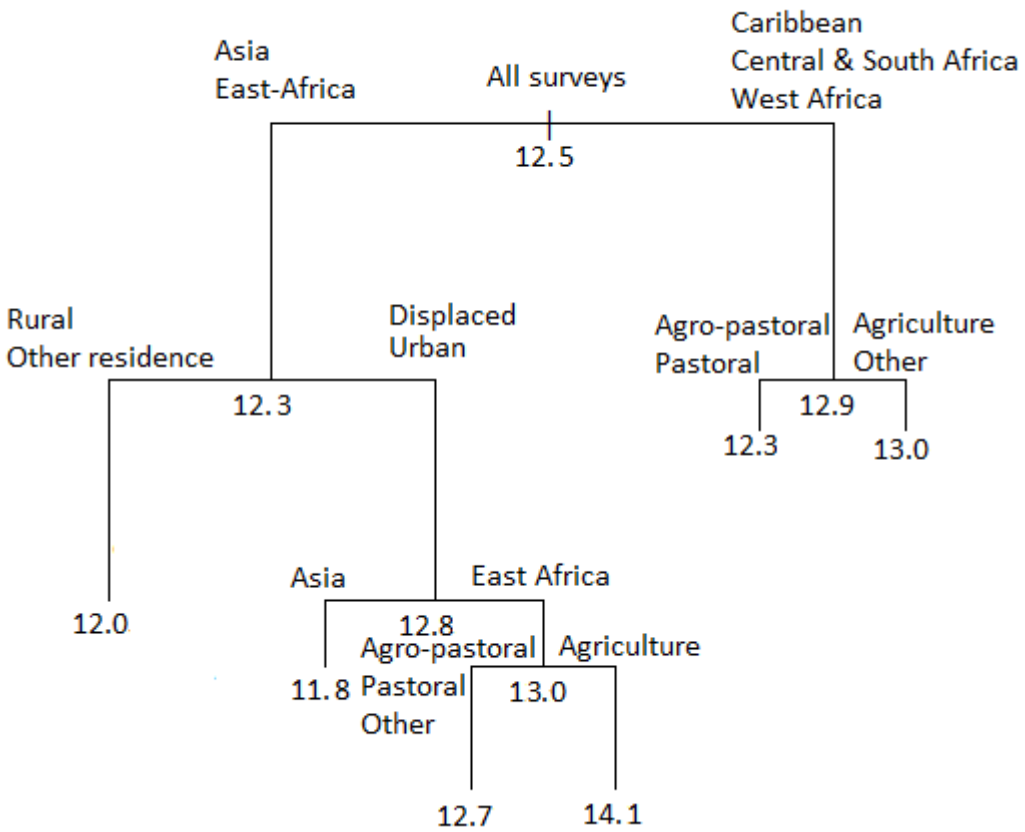


Figure 5: Tree regression by region, residence and livelihood (n=852)

Table 6-8 and Figure 6-8 describe MUAC SD stratified by region, livelihood zone and residence status.

MUAC SD per region

Table 6: Summary of MUAC SD per Region

MUAC SD per Region	Asia (n=60)	Caribbean (n=13)	Central & South Africa (n=128)	East Africa (n=554)	West Africa (n=97)
Minimum - Maximum (mm)	10.1 – 14.9	12.2 – 13.5	10.2 – 16.3	9.5 – 19.3	10.6 – 16.0
Lower - Upper quartile (mm)	11.0 – 12.9	12.5 – 13.1	12.3 – 13.7	11.5 – 13.1	11.8 – 13.9
Median - Mean (mm)	11.8, 12.0	12.7, 12.8	12.9,13.0	12.2, 12.3	12.7, 12.9

West Africa, Central and South Africa and the Caribbean have higher SD compared to Asia and East Africa. Mean MUAC SD varies from 11.7 in Asia to 12.9 in Central and South Africa (Table 6 and Figure 6). Stratifying by region reduces the variability of SD.

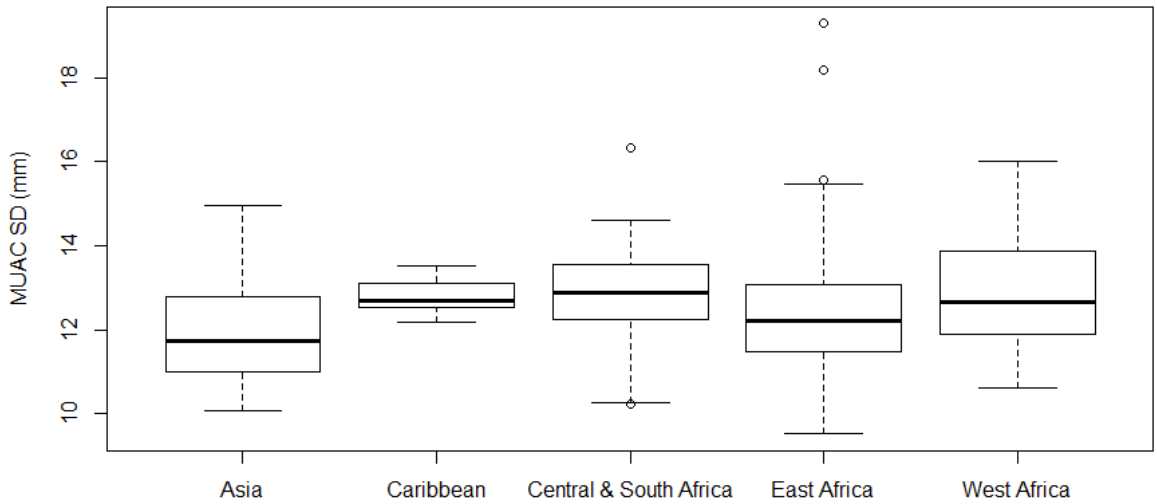


Figure 6: Box-plot of MUAC SD in each region

MUAC SD per livelihood zone

The mean MUAC SD varies from 12.1 in agro-pastoral livelihood zone to 12.8 in the "other" category. The "other" livelihood category includes fishing, riverine and mixed livelihood zones. This suggests that heterogeneous groups have a larger MUAC SD. There is some minimal reduction in the variability of SD when stratifying by livelihood zones (Table 7 and Figure 7).

Table 7: Summary of MUAC SD per livelihood zone

MUAC SD per Livelihood	Agriculture (n=351)	Agro-Pastoral (n=246)	Other[9] (n=123)	Pastoral (n=133)
Minimum - Maximum (mm)	9.5 – 16.3	9.8 – 18.2	10.7 – 19.3	10.3 – 15.5
Lower - Upper quartile (mm)	11.5 – 13.5	11.5 – 12.8	11.9 – 13.6	11.7 – 13.1
Median - Mean (mm)	12.6 , 12.6	12.1, 12.2	12.8, 12.8	12.4, 12.4

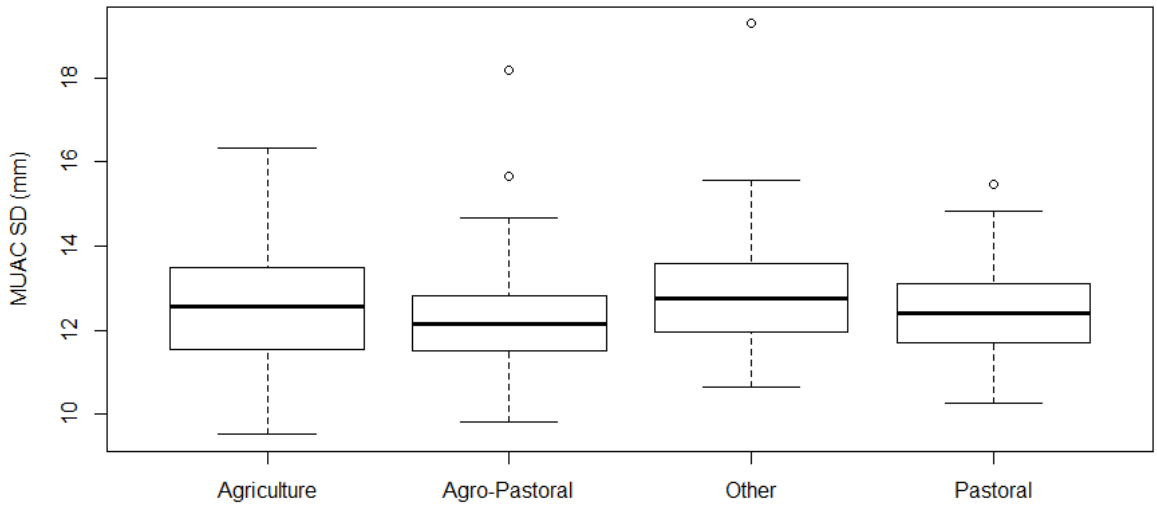


Figure 7: Box-plot of MUAC SD per livelihood zone

[9] fishing, riverine and mixed livelihood zones

MUAC SD per residence status

Mean MUAC SD varies from 12.2 in the Rural and other category to 12.9 in the displaced population. The MUAC SD from displaced population (refugees and internally displaces) as well as urban population are larger (Table 8 and Figure 8). The "other" includes mainly mixed rural/urban populations with higher proportions of rural. Displaced populations as well as urban population tend to have a more mixed background which confirms the fact that heterogeneous groups have a larger MUAC SD.

Table 8: Summary of MUAC SD per residence status

MUAC SD per residence status	Displaced (n=145)	Other[10] (n=91)	Rural (n=551)	Urban (n=66)
Minimum - Maximum (mm)	10.1 – 15.6	10.3 – 14.6	9.5 – 19.3	11.0 – 14.9
Lower - Upper quartile (mm)	12.1 – 13.5	11.6 – 13.2	11.5 – 13.0	12.1 – 13.2
Median - Mean (mm)	12.9 ,12.9	12.2, 12.4	12.2, 12.3	12.6, 12.7

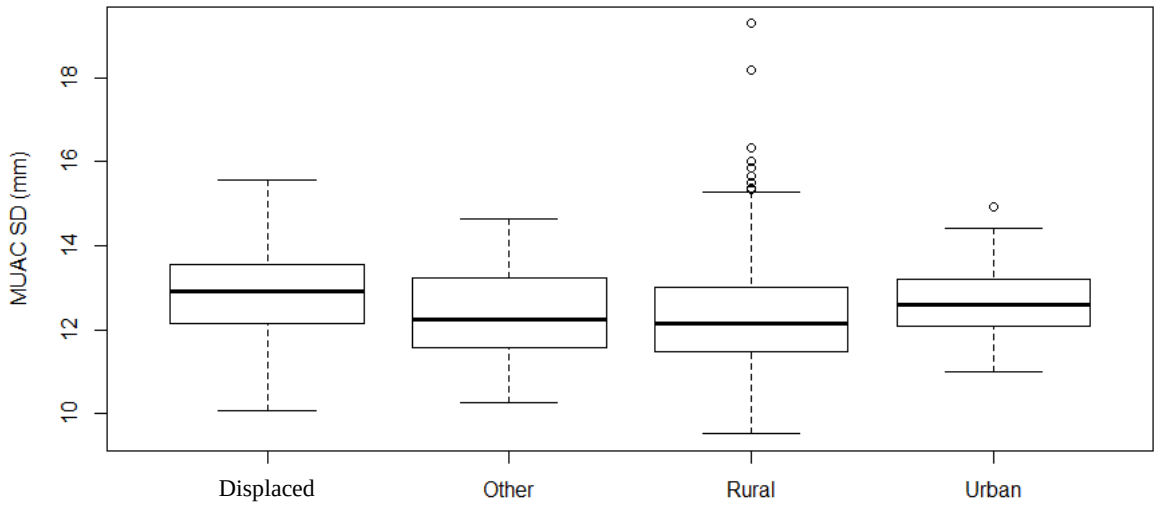


Figure 8: Box-plot of MUAC SD by residence status

[10] Mixed residence status

Based on the above descriptive statistics, categorising by region appeared to be the most promising strategy to minimise the variability MUAC SD but further evidence was needed to determine which stratification would be the most appropriate. Univariable linear regressions with MUAC SD as dependent variable and region, livelihood zone or residence status were therefore computed.

Univariable linear regressions

The linear regressions of MUAC SD as dependent variable showed significant univariable associations with the following independent variables: livelihood, residence, region (see Table 9).

Table 9: Univariable association between MUAC SD (mm) and regions, livelihood zones or residence status

Region	Coefficient	95% CI	t-statistic	P-value
East Africa	-	-	-	-
Asia	-0.430	-0.712; -0.148	-2.99	0.003
Caribbean	0.469	-0.178; 1.115	1.42	0.155
Central & South Africa	0.603	0.393; 0.812	5.65	< 0.001
West Africa	0.495	0.275; 0.714	4.41	< 0.001
Livelihood	Coefficient	95% CI	t-statistic	P-value
Agriculture	-	-	-	-
Agro-Pastoral	-0.368	-0.548; -0.188	-4.01	< 0.001
Pastoral	-0.114	-0.360; 0.132	-0.91	0.364
Other	0.256	0.017; 0.496	2.1	0.036
Residence	Coefficient	95% CI	t-statistic	P-value
Rural	-	-	-	-
IDP	0.573	0.376; 0.770	5.71	< 0.001
Urban	0.371	0.109; 0.634	2.78	0.006
Other	0.085	-0.159; 0.329	0.68	0.494

The table below present the R-squared corresponding to the different independent variables used for the univariable regressions with MUAC SD as dependent variables. The three R-squared are very low and region is the variable that explains the largest amount of the variance in MUAC SD, although most of the variability in SD remains unexplained.

Table 10: R-squared values for linear regression with MUAC SD as dependent variable and region, livelihood or residence

Independent variable	R-squared
Region	0.06
Livelihood	0.03
Residence	0.03

Stratifying the database on livelihood zones and Regions or other combinations could potentially reduce the variability of the MUAC SD further but it would be detrimental to the simplicity of the proposed method. A mulitivariable regression could have been used to further investigate the best way to stratify the surveys. Regression trees were unhelpful and it is essential to keep the method as simple as possible. A multivariable regression was therefore not investigated.

In light of the analysis above, stratification of the datasets by region was found to be the best way to stratify the database and minimise the variability in MUAC SD.

Conclusion

Three assumptions were assessed in this chapter:

- The variability of MUAC SD varied slightly over time. The proposed method was examined in simulated surveys from different time periods in order to verify performance did not differ markedly for more recent surveys compared to earlier ones (see Chapter 7).

- The assumption regarding the association between levels of GAM and MUAC did not hold but the MUAC SD variation across GAM categories are minor and are not expected to cause serious bias in the estimates. The performance of the proposed method was examined in different categories of GAM (see Chapter 7).

- Once stratified, MUAC SD fell within a reasonably narrow range. It differed most significantly from one region to another. The proposed method was developed by stratifying the MUAC SD by region (see Chapter 7).

Chapter 7: A novel, efficient method for estimating the prevalence of acute malnutrition in resource-constrained and crisis-affected settings

This aim of this thesis is to develop new methods using Middle-upper arm circumference for nutritional surveillance in crisis-affected populations. The previsous chapters introduced the methods and assessed all assumptions behind them.

This Chapters presents the outcomes of the developed methods themselves:

(i) PROBIT Method I, which takes the mean MUAC from the survey sample data and the MUAC Standard Deviation (SD) from a database of previous surveys; and

(ii) PROBIT Method II, which applies both the mean and SD of MUAC as observed in the survey sample.

I examined the performance of both methods for estimation and classification purposes. Supporting information can be found Annex VI.

An additional analysis on the relative presicison of both methos and the classic method can be found Annex VII.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Severine frison
Principal Supervisor	Marko Kerac
Thesis Title	Middle-upper arm circumference for nutritional surveillance in crisis-affected populations: Development of a method

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic peer review?	

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	PLoS One
Please list the paper's authors in the intended authorship order:	Severine Frison , Marko Kerac, Francesco Checchi, Jennifer Nicholas
Stage of publication	Submitted

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am the main author and was involved in all stages from the conception and design, literature search, analysis and interpretation, to drafting the article and writing the final version to be published.

Student Signature: SFrison **Date:** 12/08/16

Supervisor Signature: M. Kerac **Date:** 12/08/16

Title page

1. The title of the article

A novel, efficient method for estimating the prevalence of acute malnutrition in resource-constrained and crisis-affected settings

Short title: A novel, efficient method for estimating the prevalence of acute malnutrition

2. Name and address of department(s) and institution(s)

Severine Frison[1], Marko Kerac[1], Francesco Checchi[2], Jennifer Nicholas[1]

[1] Department of Population Health, London School of Hygiene and Tropical Medicine (LSHTM), Keppel Street, London, WC1E 7HT

[2]Francesco Checchi, Faculty of Public Health and Policy, LSHTM, Keppel Street, London, WC1E 7HT

3. Corresponding Author

Severine Frison, severine.frison@gmail.com

4. Author's contribution

SF wrote the first draft of the article and had the primary responsibility for the final content. SF was involved in all stages from the conception and design, data acquisition, analysis and interpretation. JN was involved in data analysis and interpretation as well as in critically revising different draft versions. MK was involved in data interpretation and in critically revising different draft versions. FC was involved in the conception, design and data acquisition as well as in data analysis and in critically revising different draft versions. All authors approved the final version of the article.

Abstract

Background: The assessment of the prevalence of acute malnutrition in children under five is widely used for the detection of nutritional emergencies, planning interventions, advocacy and programme monitoring and evaluation. This study examined the use of PROBIT Methods which convert parameters (mean and standard variation (SD)) of a normally distributed variable to a cumulative probability below any cut-off to estimate acute malnutrition in children under five using Middle-Upper Arm Circumference (MUAC).

Methods: We assessed the performance of PROBIT Method I, which takes the mean MUAC from the survey sample data and the MUAC Standard Deviation (SD) from a database of previous surveys; and PROBIT Method II, which applies both the mean and SD of MUAC as observed in the survey sample. We assessed the performance of both methods. Specifically, we generated sub-samples from 852 survey datasets, simulating 100 surveys for eight different sample sizes (25, 50, 75, 100, 125, 150, 175, 200). Overall the methods were tested on 681 600 simulated surveys.

Findings: This study suggests that PROBIT methods relying on sample sizes as small as 50 had better performance than the classic method for estimating and classifying the prevalence of acute malnutrition. The PROBIT methods had better precision in the estimation of acute malnutrition than the classic approach for all sample sizes and a better coverage for smaller sample sizes, while having relatively little bias. They classified situations accurately for a threshold of 5% acute malnutrition.

Conclusions: PROBIT Methods have a clear advantage in the assessment of acute malnutrition prevalence based on MUAC, compared to the classic method. Their use would require much lower sample sizes, and would thus enable great time and resource savings. There is great potential in their use in surveillance systems in order to produce timely and/or locally relevant prevalence estimates of acute malnutrition and enable a swift and well-targeted response/intervention.

Background

Acute malnutrition (AM) is a major public health issue throughout low-middle income countries. Indices of AM include low Weight-for-Height/Length (WFH), low Middle-Upper-Arm Circumference (MUAC) and oedematous malnutrition characterised by the presence of bilateral pitting oedema (see Table 1). The United Nations Children's Fund's (UNICEF) latest report on the State of the World's Children[1] estimates that 10% of children under 5 years old in least developed countries have a low WFH. According to United Nations estimates 875 000 children under five deaths[2] are attributed to Low WFH annually. These estimates do not include oedematous malnutrition. Overall, prevalence estimates of Global Acute Malnutrition (GAM) are similar whether including oedematous malnutrition or not[3]. There is an increasing interest in MUAC-only nutrition programming [4-9] and throughout the paper, AM is based on MUAC assessment alone.

Table 1: Acute Malnutrition definition and classification

Case definition	
Severe Acute Malnutrition (SAM)	WFH < -3 SD *and/or* oedema *and/or* MUAC<115 mm
Global Acute Malnutrition (GAM)*	WFH < -2 SD *and/or* oedema *and/or* MUAC<125 mm

WFH: Weight-for-Height/Length; MUAC: Middle-Upper Arm Circumference

*WHO has not endorsed MUAC<125mm as being a measure of GAM but for the purpose of this study, MUAC<125mm will be referred to as GAM.

The assessment of the prevalence of acute malnutrition in children under five is widely used for the detection of nutritional emergencies, planning interventions, advocacy and programme monitoring and evaluation. Its estimation usually relies

on cross-sectional multi-stage random cluster sample surveys[10-12] which are labour and resource intensive (i.e. time, logistics, and finance) especially in difficult settings, remote areas or when wide areas need to be covered. Furthermore, surveys are not able to provide the frequency and geographic resolution of data that would assist in enabling swift detection and targeted response to crises before they are well-established [12-16].

The PROBIT Method has been proposed as a more feasible alternative to standard surveys. This method estimates the prevalence of GAM according to any a cut-off of interest by using the observed mean and standard deviation (SD) of anthropometric indices (e.g. MUAC or WFH) to construct a distribution, assumed to be normal in shape, and computing the percentage of the distribution that falls below the cut-off (see Figure 1).[17-20] The method treats nutritional indices as continuous variables, instead of transforming each child observation into a binary datum (below or above cut-off), and as such has the possible advantage of decreasing the sample size required to estimate prevalence, while maintaining the same precision. Previous work has suggested that the assumption of a normal distribution is reasonable for MUAC, rendering the PROBIT approach potentially suitable for this index.[21]

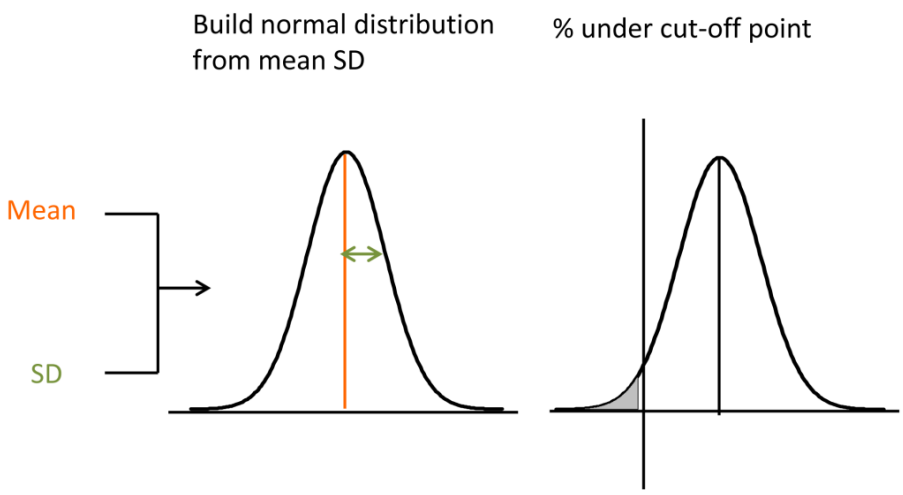


Figure 1: PROBIT Method

Previous studies have found that for simple random sample surveys, the PROBIT based prevalence estimate of acute malnutrition may have superior precision but can be subject to bias (Dale et al[18] and Blanton et al[22]). This study examined the use of the PROBIT Method for both simple random samples and two-stage cluster samples to estimate SAM (by MUAC alone) and/or GAM (by MUAC alone) in children under five. We assessed two methods: (i) PROBIT Method I, which takes the mean MUAC from the survey sample data and the MUAC SD from a database of previous surveys conducted within the same geographic stratum; and (ii) PROBIT Method II, which applies both the mean and SD of MUAC as observed in the survey sample. We assessed the performance of both methods for estimation (prevalence point estimate ± confidence interval) and classification (probability that prevalence is above/below a threshold of interest) purposes. To do so, we examined:

- the bias, precision (relative and absolute) and coverage (defined as the proportion of the 95%CIs from the test methods that contained the true prevalence value) of SAM and GAM prevalence estimates (based on MUAC alone) using both PROBIT Methods and the standard prevalence survey method (hereafter referred to as Classic Method);

- the probability of correctly classifying GAM prevalence (based on MUAC alone) for the different methods for according to programmatically important thresholds (5%, 10% and 15%).

Methods

Data sources

The study relied entirely on previously collected survey data. A total of 1068 cross-sectional cluster or exhaustive survey datasets from various settings were shared by six organizations (UNICEF, Food Security and Nutrition Analysis Unit (FSNAU), Epicentre/ (MSF), Action Against Hunger, Concern Worldwide and Goal). No formal sample size calculation was used. Instead, the study size was determined by availability of surveys and specific inclusion criteria. Eligible datasets had to: (1) include anthropometric data including MUAC, oedema, age, weight and height as well as meta-data on country, livelihood, residence, cluster (if cluster surveys) and date; (2) have a minimum of 25 clusters if cluster surveys[23, 24]. The last criterion aimed to minimize selection bias, which may be substantial with surveys featuring few clusters. Figure 2 gives the reasons for exclusion of datasets or records.

Of the 1068 surveys collected, 852 surveys were included in this secondary data analysis (55 exhaustive surveys and 797 cluster sample surveys). The 852 surveys contained 694 108 children of which 25 134 presented highly improbable values and were excluded from the analysis. The database included six variables for anthropometry (sex, MUAC, oedema, age, weight and height), six meta-data variables (organization, country, livelihood, residence, cluster [for cluster surveys only] and date) and three indices based on WHO standards (WFH, WFA and HFA) computed using the WHO's "Child Growth Standards" package[25] (see Figure 2). The sample size of surveys varied from 122 to 3491 with a median and mean of 885 and 907 respectively. Several regions were represented: East Africa with 554 surveys (65%), West Africa with 97 surveys (11.4%), Central and South Africa with 128 surveys (15%), Caribbean with 13 surveys (1.5%) and Asia with 60 surveys (7%).

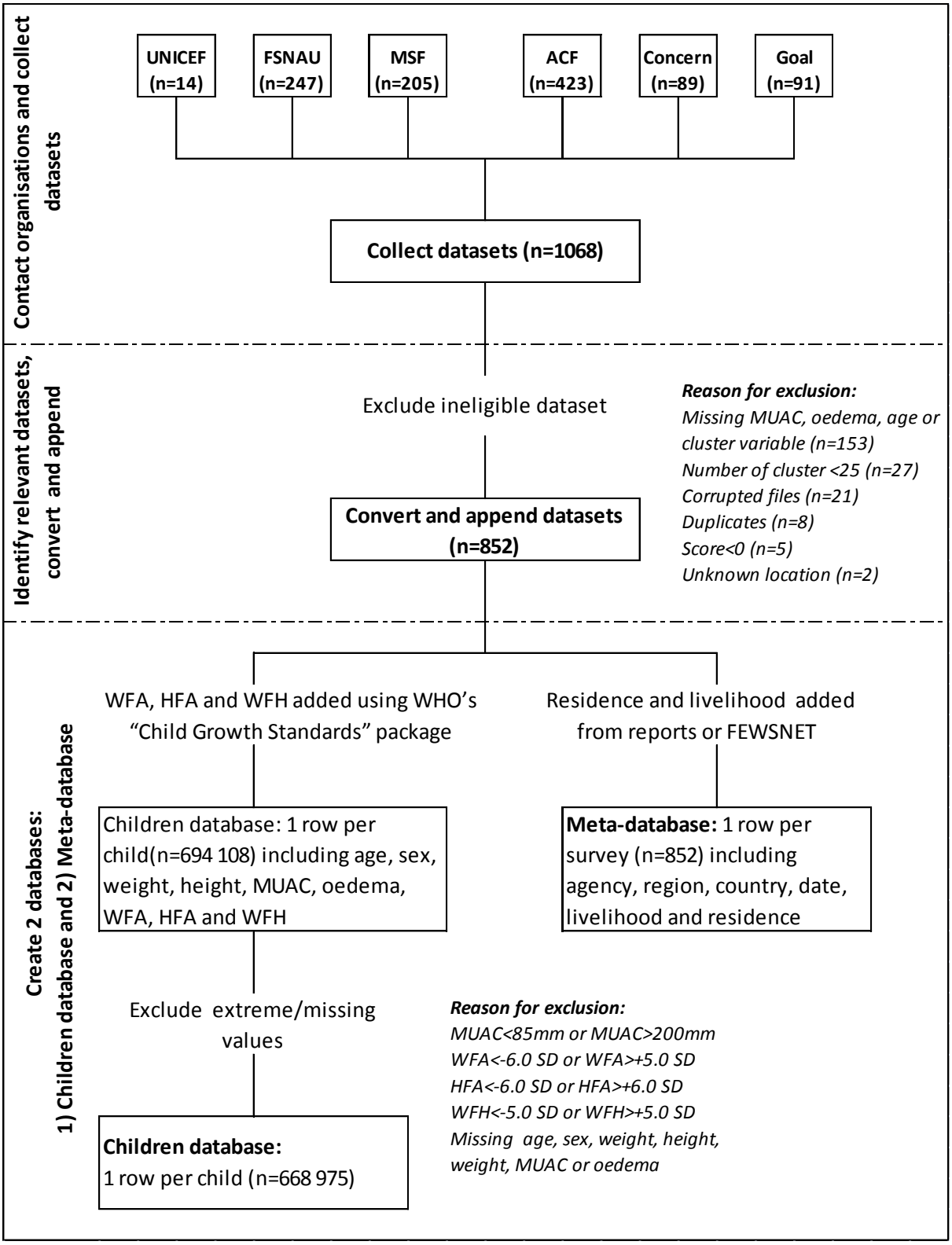


Figure 2: Data management

Simulation of small sample surveys

We tested the performance, for estimation and classification purposes, of Probit Method I, Probit Method II and the Classic Method, on simulated survey samples of varying size, drawn randomly from the larger survey database.

Specifically, we generated sub-samples from each of the 852 survey datasets, simulating 100 test surveys for each of eight different sample size scenarios (25, 50, 75, 100, 125, 150, 175, 200). To take into account the underlying clustered data structure when generating sub-samples from cluster survey datasets, we selected 25 clusters randomly, and, within each cluster, 1 to 8 child observations, again at random, to obtain sample sizes from 25 to 200. For non-cluster surveys, 25 to 200 children were randomly selected.

Overall therefore, the methods were tested on a total of 681 600 (852 source datasets x 8 sample size scenarios x 100 simulated sub-samples per sample size) simulated surveys.

Calculation of true prevalence

For each test of the method, we compared the estimate or classification yielded by the method to a measure of true prevalence. The prevalence point estimates were calculated from each of the 852 surveys and were taken as the true population prevalence measure against which to quantify the different methods' performance when applied to sub-samples drawn from that survey. This amounted to considering the source surveys as population-representative data.

Implementation of each method

i) Classic method

For each simulated sample, the prevalence was calculated as the proportion of children with MUAC below the given threshold: 125mm for GAM and 115mm for SAM . Confidence intervals for the prevalence were calculated using cluster adjusted standard errors for the proportion.

ii) PROBIT method I – Mean given simulated sample and SD from previous surveys

For each simulated sample, the PROBIT function was used to calculate the prevalence as the cumulative probability of MUAC less than the cut-off of interest, given a normal distribution of MUAC. The mean MUAC used to parameterise this distribution was the mean MUAC in the simulated sample, while the standard deviation (SD) of MUAC was the MUAC SD from previous surveys from the same geographical stratum (five regions: East Africa, West Africa, Central and South Africa, the Caribbean and Asia). No transformation was applied to the distribution in order the approximate a normal distribution as previous work has suggested that the assumption of a normal distribution is reasonable for MUAC[21].

The MUAC SD from previous surveys were weighted using effective sample size (sample size could not be used because of varying design surveys and differing numbers of clusters and children per cluster)[26], stratified by region (in order to minimise MUAC SD variability) and bootstrapped with 2000 replications in each region. The mean MUAC SD from each bootstrap (for each region) was then used with mean MUAC in simulated sample to parameterise the normal distribution. The

mean MUAC SD used were 12.3 mm for East Africa, 12.8 mm for West Africa, 13.0 mm for Central and South Africa, 12.8 mm for the Caribbean and 12.0mm for Asia.

The 95% confidence interval for the PROBIT prevalence was estimated using cluster bootstrapping with 2000 replications (2000 mean MUAC replications and 2000 weighted MUAC SD replications). For each replication the PROBIT Z score was calculated using MUAC SD randomly selected from the empirical distribution of MUAC SD of previous surveys in that geographic stratum and the mean MUAC from the bootstrap sample of the simulated survey. The standard error from the bootstrap distribution was used to calculate the confidence interval for the Z score, which was then transformed using the PROBIT function to calculate upper and lower confidence limits for the prevalence.

iii) PROBIT method II – Mean and SD from given simulated sample

For each simulated sample, the prevalence estimates were calculated using the PROBIT function to calculate the cumulative probability of MUAC less than the given threshold using mean MUAC and MUAC SD from the simulated sample. No transformation was applied to the distribution in order the approximate a normal distribution as previous work has suggested that the assumption of a normal distribution is reasonable for MUAC[21]. The same bootstrapping method as above was applied to compute confidence intervals, but this used both MUAC SD and mean MUAC from the bootstrap replications of the simulated survey to generate the bootstrap distribution of the PROBIT Z score.

Estimation approach: coverage, precision and bias

For each of the methods, we examined coverage, precision (absolute and relative) and bias overall, for different GAM prevalence categories (<5%, 5-9%, 10-14%, ≥15%) and per region to investigate possible characteristics that would confound the outcome of the methods.

Bias was defined as the average difference between the estimated prevalence generated by each test method and the true prevalence.

Absolute precision was defined as the average length of the 95%CIs generated by each test method (Abs [upper bound – lower bound] / 2).
Relative precision was defined as follow:
(Absolute precision x 100) /estimated prevalence

Coverage was defined as the proportion of the 95%CIs from the test method that contained the true prevalence value. If coverage is as expected, the nominal 95% CI of the proposed methods should contain the true value 95% of the time.

To further assess possible characteristics that would influence bias in particular, we used linear regression to explore associations between bias of GAM estimates as the dependent variable and the following independent variables: region, GAM categories based on MUAC (<5%, 5-9%, 10-14%, ≥15%), livelihood, residence, survey design (simple random sampling or clustered) and date (before 2006, after 2006. This date was chosen as the SMART Methodology[27], that brought rigour and standardization in the way surveys were conducted, was adopted in 2006). We did this regression for each proposed method.

Classification approach: probability of correctly classifying GAM prevalence

The different methods were assessed looking at a classification approach. For each survey, the true GAM (based on MUAC only) prevalence and the estimated GAM prevalence from the different methods were split into two categories according to different thresholds: GAM below 5%, 10% or 15% and GAM equal or above 5%, 10% or 15%. We then calculated the probability that the different methods correctly classify GAM prevalence ≥ threshold of interest[28].

Ethical approval

The project relied only on re-analysis of secondary data sources, none of which had uniquely identifiable information associated with each child-observation. Ethics approval for the project was sought and was obtained from the Ethics Committee of the London School of Hygiene and Tropical Medicine (LSHTM Ethics reference 6158).

Results

"True" prevalence observed in the database

The prevalence of GAM and SAM according to MUAC across surveys (n=852) varied from 1% to 47.7% and from 0% to 20.6% respectively. Median and mean GAM were 8.8% and 9.9% respectively while median and mean severe GAM prevalence were 1.7% and 2.2% respectively. Figure 3 shows the distribution of GAM and SAM prevalence measured with MUAC measure alone.

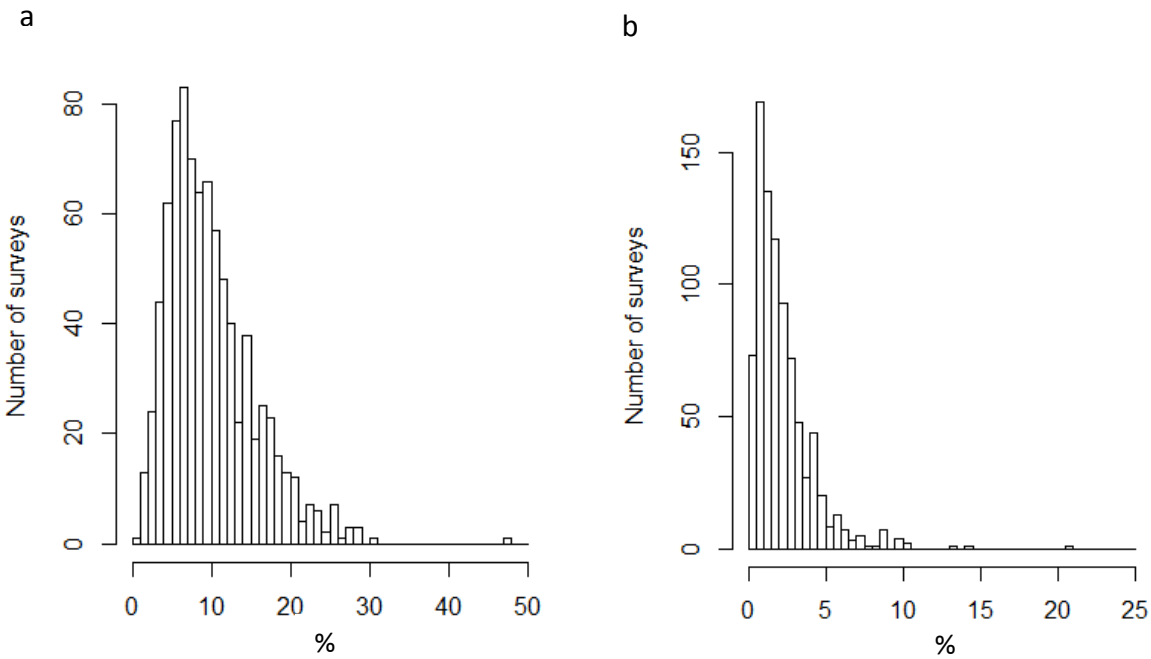


Figure 3: Distribution of GAM (MUAC<125mm) (a) and SAM (MUAC<115mm) (b)

Estimation approach

Bias

The mean bias in the estimations of GAM prevalence tended to be larger, the smaller the sample size. The estimates of GAM prevalence were practically unbiased using the classic method for sample sizea above 50. PROBIT method I's mean bias varied from 1.2% (sample size = 25) to 0.8% (200). PROBIT method II had lower mean bias, varying from 0.8% to 0.7% (Table 2 and Figure 4). On average, both PROBIT Methods overestimated the prevalence of GAM . Individual simulated surveys showed both positive and negative bias for all methods (see Supporting information 1).

Table 2: Bias in estimated GAM prevalence (estimated - true value)

Sample size	Classic method		PROBIT method I		PROBIT method II	
	Median	Mean	Median	Mean	Median	Mean
25	0.0	0.7	0.9	1.2	0.4	0.8
50	0.0	0.2	0.8	1.0	0.5	0.8
75	0.0	0.0	0.9	0.9	0.6	0.7
100	0.0	0.0	0.8	0.9	0.6	0.7
125	0.0	0.0	0.8	0.9	0.6	0.7
150	0.0	0.0	0.9	0.9	0.6	0.7
175	0.0	0.0	0.8	0.8	0.6	0.7
200	0.0	0.0	0.9	0.8	0.6	0.7

The bias in the estimation of SAM was minimal using PROBIT Methods (mean bias varied from 0.2% to 0.1% and 0.3% to 0.1% for PROBIT Method I and PROBIT Method II respectively). The Classic Method was biased for sample sizes under 75 (mean bias of 1% and 0.6% for sample sizes of 25 and 50 respectively) (see Table 3 and Figure 5). Individual simulated surveys showed both positive and negative bias for all methods (see Supporting information 1).

Table 3: Bias in estimated SAM prevalence (estimated - true value)

Sample size	Classic method		PROBIT method I		PROBIT method II	
	Median	Mean	Median	Mean	Median	Mean
25	0.0	1.0	0.1	0.2	0.0	0.3
50	0.0	0.6	0.1	0.1	0.0	0.2
75	0.0	0.3	0.1	0.1	0.0	0.1
100	0.0	0.2	0.1	0.1	0.0	0.1
125	0.0	0.2	0.1	0.1	0.0	0.1
150	0.0	0.1	0.1	0.1	0.0	0.1
175	0.0	0.1	0.1	0.1	0.0	0.1
200	0.0	0.1	0.1	0.1	0.0	0.1

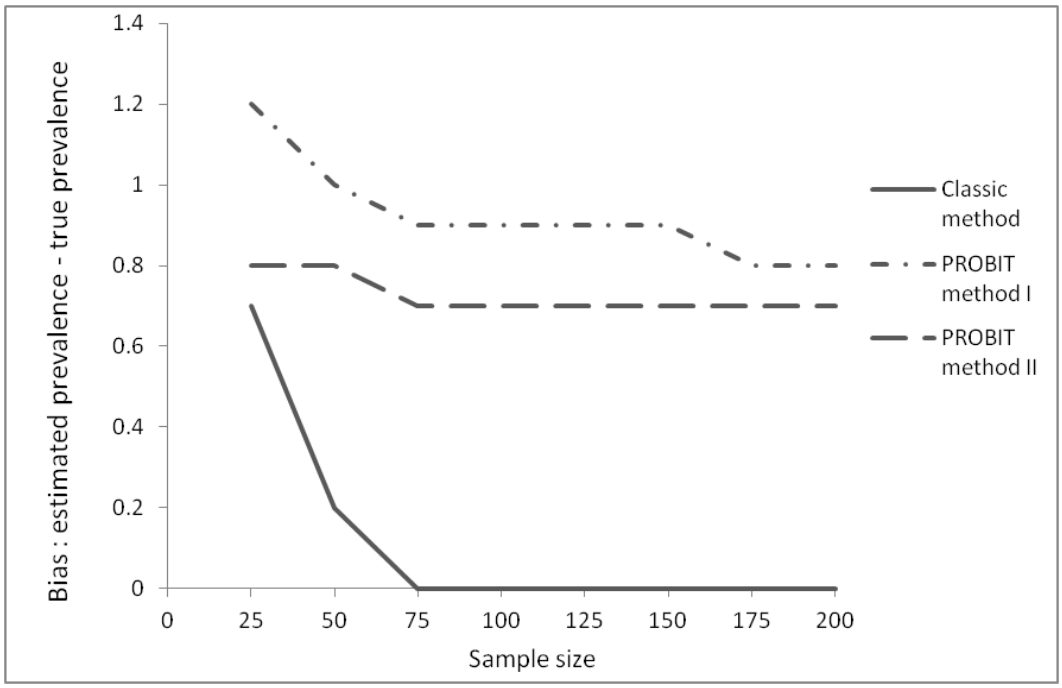


Figure 4: Bias in GAM estimates

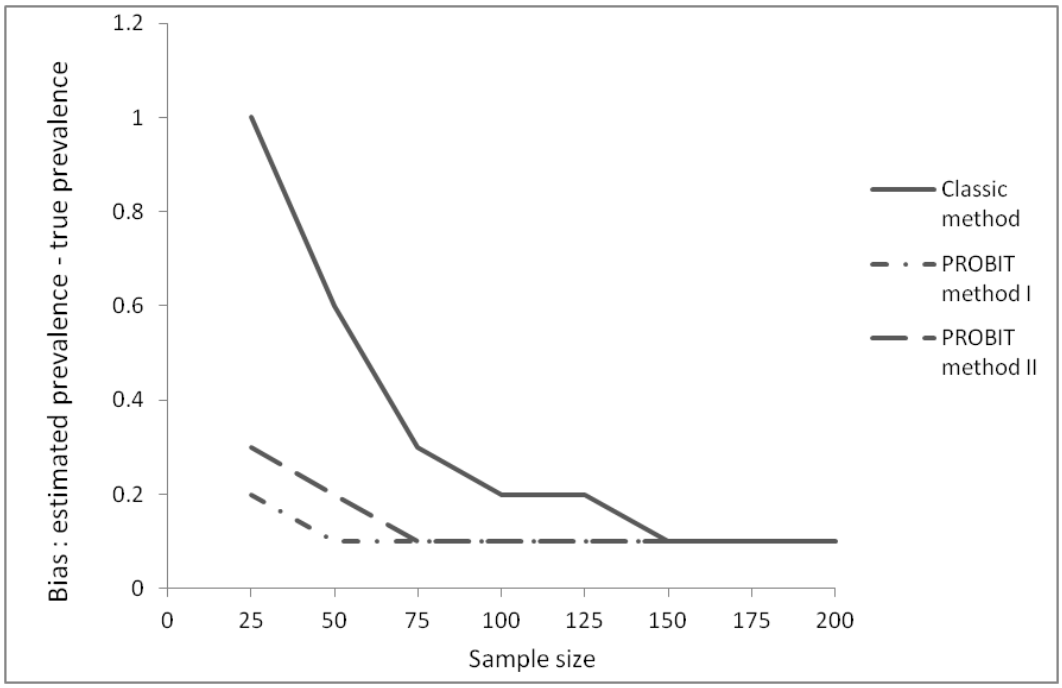


Figure 5: Bias in SAM estimates

The bias of GAM was smaller, the higher the GAM prevalence, and the bias of SAM was lower in low (<5%) and high (≥ 15%) GAM categories using the Classic Method. The bias of GAM using PROBIT Method I was lower for the last two GAM categories (10-14% and ≥ 15%) while using PROBIT Method II, it was lower in the first two GAM categories (<5% and 5-9%). The bias of SAM was higher in the highest GAM category for both PROBIT Methods (see Supporting information 2). For all methods, the bias was larger in East Africa and Asia and much lower or null for the Caribbean. The PROBIT methods had positive bias in all regions except Caribbean, where the prevalence tended to be underestimated or unbiased (Supporting information 3).

Precision

The precision of GAM and SAM prevalence estimates increased as the sample size increased for all methods. The classic method had the lowest precision for both GAM and SAM , varying from approximately 14.2% to 4.7% for sample sizes from 25 to 200 (see Table 4 and 5 and Figure 6 and 7).

The precision was higher using PROBIT methods for all sample sizes. PROBIT Method I yielded better precision for sample sizes < 75 compared to PROBIT method II. For sample sizes ≥ 75, the opposite pattern was observed. Similar results were observed for GAM and SAM (see Table 4 and 5; Figure 6 and 7).

Table 4: Precision of GAM estimates (half of 95% CI)

Sample size	Classic method		PROBIT method I		PROBIT method II	
	Median	Mean	Median	Mean	Median	Mean
25	13.7	14.2	7.7	7.9	9.8	9.9
50	9.0	9.3	6.2	6.2	6.7	6.7
75	7.7	7.4	5.5	5.5	5.4	5.4
100	6.1	6.4	5.2	5.1	4.6	4.6
125	5.6	5.8	5.0	4.9	4.2	4.1
150	5.1	5.3	4.8	4.7	3.8	3.8
175	4.8	5.0	4.7	4.6	3.5	3.5
200	4.5	4.7	4.6	4.5	3.3	3.3

Table 5: Precision of SAM estimates (half 95% CI)

Sample size	Classic method		PROBIT method I		PROBIT method II	
	Median	Mean	Median	Mean	Median	Mean
25	12.9	10.7	2.7	3.0	4.3	4.8
50	7.1	6.8	2.2	2.5	2.5	2.8
75	5.0	5.1	2.1	2.3	2.0	2.2
100	3.8	4.1	2.0	2.2	1.7	1.9
125	3.3	3.5	2.0	2.1	1.5	1.6
150	2.8	3.1	1.9	2.1	1.3	1.5
175	2.5	2.8	1.9	2.0	1.2	1.4
200	2.2	2.6	1.9	2.0	1.2	1.3

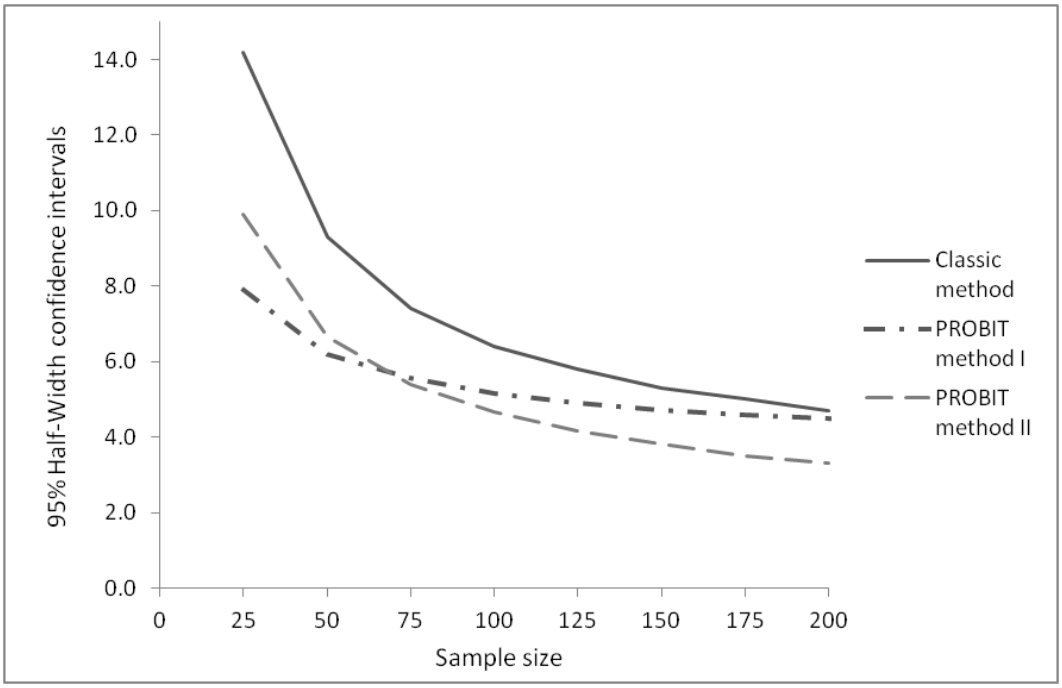


Figure 6: Precision of GAM estimates

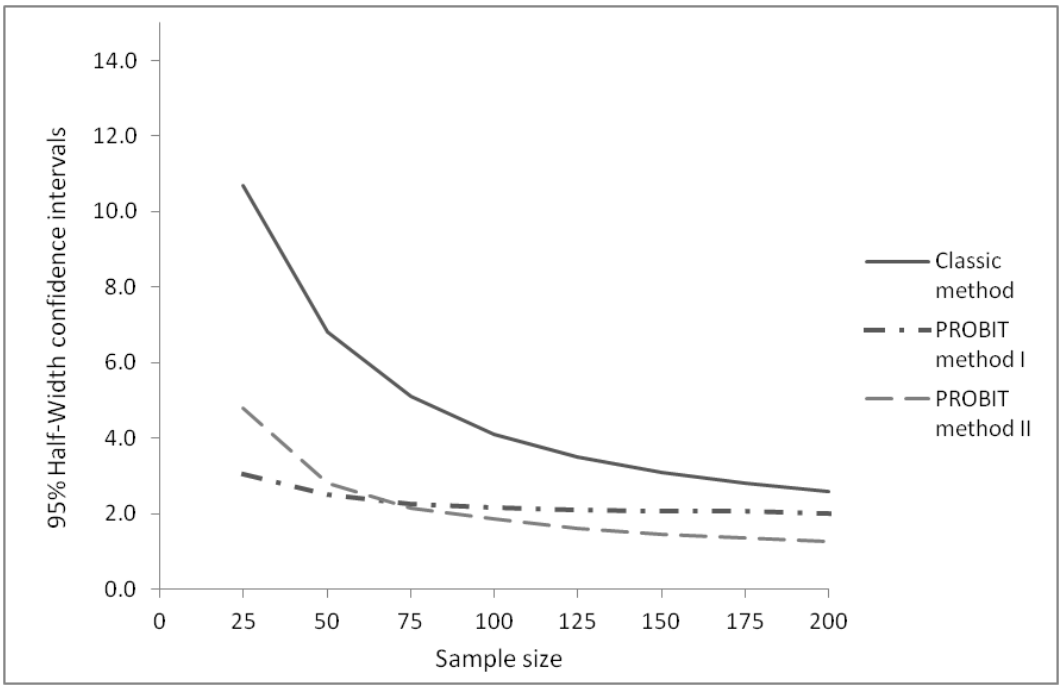


Figure 7: Precision of SAM estimates

For all methods, the precision of GAM and SAM estimates was lower the higher the level of GAM (see Supporting information 2). For all methods, the precision was superior in the Caribbean with larger differences using the PROBIT Methods compared to the classic Method (see Supporting information 3).

Annex VII presents relative presicion of GAM and SAM estimates. Both PROBIT Methods have better relative precision than the classic method for all sample sizes. The PROBIT Method I practically reaches the 30% of recommended relative precision[29] of GAM estimates for a sample size of 200. The relative precision of SAM estimates is very large for all sample sizes.

Coverage

Coverage was better for GAM compared to SAM and was generally higher for the PROBIT approach II but never reached 95% (see Table 6). The PROBIT Methods had a clear advantage in term of coverage compared to the classic method for sample sizes < 50 for GAM and for sample sizes <150 for SAM (see Table 6).

With the classic method, the coverage for GAM increased as the simulated sample size increased while for the PROBIT methods coverage was higher with smaller sample size. Similar trends were observed for SAM estimates. The classic method showed extremely low coverage of SAM estimates for small sample sizes, while coverage for both PROBIT methods was higher and more stable (see Table 6).

Coverage depended on bias and precision. The smaller the bias, the better the coverage and the larger the confidence intervals around the estimate prevalence the better the coverage.

Table 6: Coverage of different methods for GAM (a) and SAM (b)

(a)

Sample size	Classic method (%)	PROBIT method I (%)	PROBIT method II (%)
25	83.7	92.1	94.5
50	93.9	91.2	93.6
75	96.5	90.4	93.1
100	97.3	89.8	92.5
125	98.1	89.4	91.8
150	98.4	88.9	91.0
175	98.4	88.8	90.4
200	98.6	88.4	89.9

(b)

Sample size	Classic method (%)	PROBIT method I (%)	PROBIT method II (%)
25	35.4	89.9	92.7
50	55.3	88.8	91.1
75	67.1	88.0	89.7
100	75.2	87.4	88.7
125	80.4	87.3	87.7
150	84.2	87.1	86.5
175	87.3	87.0	85.7
200	89.6	86.8	84.5

The classic method had the largest differences by GAM categories; coverage of GAM and SAM estimates increased as GAM increased (see Supporting information 2). The coverage of GAM and SAM using PROBIT Method I increased as GAM increased and then decreased for GAM categories ≥ 15%. The coverage of GAM with PROBIT Method II increased as GAM increased (see Supporting information 2).

Coverage of GAM and SAM estimates varied substantially between regions using the classic method with higher coverage in East and Asia and lower coverage in the Caribbean. For GAM , the coverage using the PROBIT Method I was slightly higher in Asia compared to the other regions. The coverage of SAM using PROBIT Method I was much lower for the Caribbean and higher for Asia. The coverage of GAM estimates using PROBIT method II were similar in all regions whereas SAM coverage was lower in the Caribbean compared to the other regions (see Supporting information 3).

Potential sources of bias

We investigated possible sources of bias by computing univariable linear regression with mean bias in GAM estimates as dependent variable and the following independent variable: region, GAM categories, livelihood, residence, survey design and date.

The multivariable linear models built with variables showing univariable association are presented in table 7, 8 and 9 for the Classic Method, the PROBIT Method I and the PROBIT Method II respectively.

The differences observed by region in the classic method were only statistically significant between for Central and South Africa that presented a lower mean bias in the estimates of GAM . The difference in the mean bias between regions were all statistically significant for both PROBIT Methods. The differences between regions were similar to the ones observed in the descriptive analysis Supporting information 2 and 3. The differences observed by GAM category in the descriptive tables Annex 2 and 3 are still apparent and statistically significant in the multivariable models for all methods (Table 7, 8 and 9). The mean bias in GAM estimates decrease as sample size increased for all methods (Table 7, 8 and 9).

The mean bias in GAM estimates using the Classic Method was statistically higher in the “other” livelihood zone (included fishing, riverine and mixed livelihood zones) compared to the agriculture livelihood zone and for displaced populations compared to other residence status (Table 7).

The mean bias in GAM estimates using the PROBIT Method I was significantly higher in Agro-pastoral and Pastoral zones compared to agriculture and significantly lower in “other” category of livelihood zones. It was statistically lower for all residence types compared to rural residence. It was significantly lower for clustered surveys compared to simple random design and higher after 2006 (Table 8).

The mean bias in GAM estimates using the PROBIT Method II was significantly higher in agro-pastoral, pastoral and "other" livelihood zones compared to agriculture livelihood zone. It was significantly lower for all residence type compared to rural residence. It was significantly lower for clustered surveys compared to simple random design and higher after 2006 (Table 9).

Table 7: Multivariable regression of mean bias in GAM estimates using the classic method

Region	Coef	95% CI	t-statistic	P-value
East Africa	-	-	-	-
Asia	0.012	-0.019; 0.044	0.77	0.441
Caribbean	-0.017	-0.079; 0.045	-0.54	0.591
C & S Africa	-0.027	-0.050; -0.004	-2.3	0.022
West Africa	-0.016	-0.041; 0.009	-1.28	0.199
GAM (MUAC)				
<5%	-	-	-	-
5-9%	-0.133	-0.155; -0.111	-11.99	<0.001
10-14%	-0.215	-0.240; -0.191	-17.42	<0.001
≥ 15%	-0.273	-0.300; -0.247	-20.34	<0.001
Livelihood				
Agriculture	-	-	-	-
Agro-Pastoral	0.016	-0.003; 0.036	1.69	0.092
Other	0.073	0.050; 0.096	6.12	<0.001
Pastoral	-0.021	-0.046; 0.004	-1.67	0.094
Residence				
Rural	-	-	-	-
Displaced	0.029	0.009; 0.050	2.79	0.005
Other	-0.005	-0.030; 0.020	-0.4	0.691
Urban	-0.016	-0.045; 0.013	-1.1	0.270
Sample size				
25	-	-	-	-
50	-0.598	-0.627; -0.568	-40.17	<0.001
75	-0.704	-0.733; -0.675	-47.31	<0.001
100	-0.722	-0.751; -0.693	-48.53	<0.001
125	-0.740	-0.770; -0.711	-49.77	<0.001
150	-0.747	-0.776; -0.718	-50.2	<0.001
175	-0.739	-0.768; 0.710	-49.66	<0.001
200	-0.735	-0.764; -0.706	-49.41	<0.001
Date				
Before 2006	-	-	-	-
After 2006	-0.030	-0.046; -0.014	-3.64	<0.001

Table 8: Multivariable regression of mean bias in GAM estimates using the PROBIT Method I

Region	Coef	95% CI	t-statistic	P-value
East Africa	-	-	-	-
Asia	0.431	0.400; 0.462	27.26	<0.001
Caribbean	-1.772	-1.832; -1.711	-57.05	<0.001
C & S Africa	-0.094	-0.117; -0.071	-8.08	<0.001
West Africa	-0.626	-0.650; -0.602	-50.76	<0.001
GAM (MUAC)				
<5%	-	-	-	-
5-9%	-0.185	-0.207; -0.164	-17.04	<0.001
10-14%	-0.487	-0.511; -0.464	-40.09	<0.001
≥ 15%	-0.464	-0.490; -0.438	-35.14	<0.001
Livelihood				
Agriculture	-	-	-	-
Agro-Pastoral	0.454	0.435; 0.472	47.27	<0.001
Other	-0.466	-0.489; -0.443	-39.77	<0.001
Pastoral	0.139	0.114; 0.163	11.19	<0.001
Residence				
Rural	-	-	-	-
Displaced	-1.801	-1.822; -1.779	-166.35	<0.001
Other	-0.659	-0.683; -0.634	-52.27	<0.001
Urban	-0.994	-1.022; -0.966	-69.31	<0.001
Sample size				
25	-	-	-	-
50	-0.176	-0.204; -0.147	-12.03	<0.001
75	-0.240	-0.269; -0.212	-16.46	<0.001
100	-0.283	-0.312; -0.255	-19.4	<0.001
125	-0.290	-0.318; '-0.261	-19.84	<0.001
150	-0.303	-0.331; -0.274	-20.72	<0.001
175	-0.321	-0.350; -0.293	-21.98	<0.001
200	-0.324	-0.352; -0.295	-22.16	<0.001
Date				
Before 2006	-	-	-	-
After 2006	0.691	0.675; 0.707	85.03	<0.001
Survey design				
Simple random	-	-	-	-
Clustered	-0.807	-0.839; -0.775	-49.48	<0.001

Table 9: Multivariable regression of mean bias in GAM estimates using the PROBIT Method II

Region	Coef	95% CI	t-statistic	P-value
East Africa	-	-	-	-
Asia	0.239	0.210; 0.268	16.12	<0.001
Caribbean	-0.790	-0.847; -0.733	-27.14	<0.001
C & S Africa	0.048	0.026; 0.069	4.35	<0.001
West Africa	-0.293	-0.316; -0.271	-25.37	<0.001
GAM (MUAC)				
<5%	-	-	-	-
5-9%	0.176	0.156; 0.196	17.23	<0.001
10-14%	0.311	0.288; 0.333	27.25	<0.001
≥ 15%	0.280	0.255; 0.304	22.6	<0.001
Livelihood				
Agriculture	-	-	-	-
Agro-Pastoral	0.179	0.161; 0.197	19.88	<0.001
Other	0.288	0.266; 0.309	26.19	<0.001
Pastoral	0.545	0.522; 0.568	46.93	<0.001
Residence				
Rural	-	-	-	-
Displaced	-0.572	-0.592; -0.552	-56.39	<0.001
Other	-0.180	-0.203; -0.157	-15.25	<0.001
Urban	-0.284	-0.311; -0.258	-21.14	<0.001
Sample size				
25	-	-	-	-
50	-0.063	-0.090; -0.036	-4.61	<0.001
75	-0.094	-0.121; -0.067	-6.84	<0.001
100	-0.105	-0.132; -0.078	-7.69	<0.001
125	-0.109	-0.136; -0.082	-7.97	<0.001
150	-0.111	-0.138; -0.084	-8.12	<0.001
175	-0.110	-0.137; -0.083	-8.05	<0.001
200	-0.114	-0.140; -0.087	-8.29	<0.001
Date				
Before 2006	-	-	-	-
After 2006	0.026	0.011; 0.041	3.41	0.001
Survey design				
Simple random	-	-	-	-
Clustered	-0.039	-0.069; -0.009	-2.55	0.011

Classification approach

We assessed the classification approach looking at three GAM prevalence (based on MUAC) thresholds: 5%, 10% and 15%. The probability of correctly identifying the GAM prevalence as above these three thresholds decreased as the threshold used increased and increased as the sample size increased (see Table 10).

The PROBIT Methods had better outcomes for smaller sample size compared to the Classic Method. The probability of correctly identifying the GAM prevalence as above 5% was high while for a 10% and 15% threshold, it was quite low (see Table 10).

A probability of 90% means that one in every ten survey will wrongly classify the nutritional status of the population as “normal” instead of as “poor”, “serious” or “critical” depending on the threshold used (5%, 10% or 15%).

Table 10: Probability of correctly classifiying the true prevalence of GAM as exceeding a threshold of 5%, 10% or 15% for the different methods

Sample size	Probability of GAM ≥ 5% (%)			Probability of GAM ≥ 10% (%)			Probability of GAM ≥ 15% (%)		
	Classic method	PROBIT I	PROBIT II	Classic method	PROBIT I	PROBIT II	Classic method	PROBIT I	PROBIT II
25	87.7	91.2	92.4	56.0	69.7	69.6	32.7	55.7	52.1
50	91.4	91.5	93.4	67.1	72.8	74.8	53.1	62.4	61.4
75	92.7	91.7	93.7	77.8	74.2	77.4	66.5	65.9	66.8
100	93.2	91.6	94.0	78.8	75.0	79.3	68.1	68.0	69.8
125	95.8	91.8	94.3	84.0	75.4	80.6	73.2	68.6	70.9
150	95.7	91.7	94.4	83.7	75.8	81.4	77.4	69.4	72.6
175	95.5	91.8	94.5	86.9	75.9	82.1	81.0	70.1	73.7
200	95.5	91.8	94.7	86.2	76.1	82.3	79.3	70.6	74.5

Discussion

This study analysed an exceptionally large database of anthropometric surveys conducted mostly in emergency situations, so as to test the performance of two candidate methods for acute malnutrition prevalence estimation, compared to the current mainstay method. Importantly, this study suggests that two PROBIT Methods, relying on far lower sample sizes that is typically the case with classic anthropometric surveys, had better performance than the classic method for estimating prevalence of GAM and SAM in a simulation of small surveys based on real survey data. The PROBIT Method had far better precision than the classic Method for all sample sizes and a better coverage for smaller sample sizes, while having relatively little bias.

The mean bias in GAM estimates did not vary much between sample sizes and was slightly lower for PROBIT Method II (approximately 0.9 for PROBIT Method I and 0.7 for PROBIT Method II); it was very low for SAM estimates. Although the PROBIT methods overestimated the prevalence of GAM , individual simulated surveys showed both negative and positive bias and it would therefore be difficult to apply a systematic downward correction. The precision of PROBIT Methods was systematically greater than the precision from the Classic Method. It was reasonable starting at a sample size of 50 for PROBIT Method I (6%) and from a sample size of 75 for PROBIT Method II (5.5%).

Blanton and Bilukha[22] had concluded that bias from PROBIT method is population dependent, which is supported by our results. We found the PROBIT methods had smaller bias with higher level of GAM . The PROBIT methods also had minimal mean bias in the Caribbean and higher bias in Asia; different sample sizes might be required depending on the region (e.g. 25 in the Caribbean and 50 or more in Asia). The PROBIT method also showed differences in bias between different livelihood zones and residence status. Mean bias also seemed to be higher after 2006 for both PROBIT Methods.

As would be expected, the precision of all three methods depended on sample size and GAM prevalence, with worse precision (on the proportion scale) as prevalence became closer to 50%, and with smaller sample size. However, relative and absolute precision was better for the PROBIT methods than the Classic method for all sample sizes.

The PROBIT Methods had better classification performance for smaller sample sizes for all cut-off points assessed (5%, 10% and 15%). The 10% and 15% thresholds yield low probabilities of classifying the situation correctly while the 5% threshold classified the situation correctly over 90% of the time with a sample size as small as 25 for both PROBIT Methods.

This paper had several strengths:

i) It showed the PROBIT Methods can produce robust estimates for sample sizes as small as 50. This important finding opens up avenues for the use of either PROBIT method as part of nutrition surveillance systems and in particular for early warning, since the small sample size requirement would enable regular data collection and timely generation of information.

ii) It examined the performance of the PROBIT Method in different regions: the sample size required may differ depending on the region.

iii) It assessed the PROBIT Method for a range of survey designs and is the first to include analysis for clustered sample surveys.

iv) It focuses on the assessment of AM with MUAC which is more feasible in the community and allows for rapid screening. The use of MUAC for community surveillance is best to assess the number of children in need of treatment as there is an increasing interest in MUAC-only nutrition programming [4-8]

v) It explored the possibility of using PROBIT with a classification approach.

We also recognise this analysis some limitations:

i) Surveys were used as proxies for true populations. Although it is hard to quantify the impact on the bias and precision of the proposed methods, we do not believe it would outweigh the advantages of the methods.

ii) We did not factor in oedematous malnutrition. However, estimates of GAM are generally similar whether including oedematous malnutrition or not, with the exception of areas with very high kwashiorkor burden[3].

iii) We did not transform or smooth our data from the simulated sample to ensure the MUAC distribution approximated normality. Previous work has suggested that the assumption of a normal distribution is reasonable for MUAC. That study also showed that different transformations or smoothing techniques may be required for "non-normal" distributions to reach normality which would render this method more complicated[21]. Furthermore, the bias in prevalence estimates was only very slightly reduced when Dale et al[18] assessed normal transformed data compared to non-transformed data. We therefore do not believe this may have significantly impacted the outcomes of the methods assessed.

iv) The design of the 852 surveys was taken into account when simulating samples but the design effect was not further factored in when calculating the confidence intervals of both PROBIT methods which may have underestimated the width of the confidence intervals for both methods.

v) Although assessment of acute malnutrition is traditionally done by estimating the prevalence, it should be assessed using incidence[30, 31]. The PROBIT methods may be inappropriate for this purpose as cases are not directly identified but their numbers are estimated. As a result, it is impossible to say at each assessment whether a case is a new case or a case which has already been identified.

Future work could assess:

- Which of the two PROBIT methods is better for routine field use (e.g. considering usability, practicability). The PROBIT Method I requires previous surveys data from the region where the assessment is taking place. We may be able to use the SD MUAC used in the present work or use country level SD.

- Other approached in the Probit estimation. The paper describes one approach to applying probit to estimating prevalence. There are other approaches that might be considered e.g. more non-parametric compared to the semi-parametric approach described here or a Bayesian approach to the PROBIT Method, incorporating the prior information from previous surveys as a way to potentially increase precision and decrease bias. This method could for example use the MUAC SDs from surveys to inform a prior for the distribution of MUAC SD for use in a Bayesian estimate of PROBIT prevalence from small sample surveys.

- User-friendly software package/mobile phone platform where once raw-data entered, results would be available immediately (e.g. ENA[32])

Conclusion

PROBIT Methods have a clear advantage in the assessment of acute malnutrition prevalence based on MUAC, compared to the classic prevalence based method. Their use would require much lower sample sizes, and would thus enable great time and resource savings. There is great potential in their use in surveillance systems in order to produce timely and/or locally relevant prevalence estimates of acute malnutrition and enable a swift and well-targeted response/intervention.

Acknowledgements

We would like to thank the following people and organisations for sharing the datasets used for this research: Grainne Moloney and Elijah Odundo from FSNAU, Mara Nyawo from UNICEF Khartoum, Dr. Sheila Isanaka (Nutritional epidemiologist) from Epicentre/MSF Paris, Dr. Benjamin Guesdon and Cécile Salpeteur from Action Against Hunger-Paris, Dr. Anne-Marie Mayer and Gudrun Stallkamp from Concern Worldwide and Claudine Prudhon for sharing data from Goal. We would also like to thank Jane Bruce for supporting the PhD from which this study arose.

References

1. UNICEF, *The State of the World's Children*. 2014.
2. Black, R., et al., *Maternal and child undernutrition and overweight in low-income and middle-income countries.* The Lancet, 2013. **382**(9890): p. 427-451.
3. Frison, S., F. Checchi, and M. Kerac, *Omitting edema measurement: How much acute malnutrition are we missing?* American Journal of Clinical Nutrition, 2015. **102**(5): p. 1176-1181.
4. Collins, S., et al., *Key factors in the success of community-based management of severe malnutrition.* Food and nutrition bulletin, 2006: p. 49-79.
5. Dale N.M., et al., *Using Mid-Upper Arm Circumference to End Treatment of Severe Acute Malnutrition Leads to Higher Weight Gains in the Most Malnourished Children.* PLoS ONE, 2013. **8**(2): p. no pagination.
6. en-net, *Mid-Upper Arm Circumference and Weight-for-Height Z-Score as Indicators of Severe Acute Malnutrition: A Consultation of Operational Agencies and Academic Specialists to Understand the Evidence, Identify Knowledge Gaps and to Inform Operational Guidance. .* Emergency Nutrition Network, 2012.
7. Goossens, S., et al., *Mid-Upper Arm Circumference Based Nutrition Programming: Evidence for a New Approach in Regions with High Burden of Acute Malnutrition.* PLoS ONE, 2012. **7**(11): p. no pagination.
8. WHO, WFP, and UNICEF, *Community-based management of severe acute malnutrition. Joint statement*. 2007, WHO, WFP, UNICEF.
9. Briend, A., et al., *Low mid-upper arm circumference identifies children with a high risk of death who should be the priority target for treatment.* BMC Nutrition, 2016. **2**(63).
10. Young, H., et al., *Public nutrition in complex emergencies.* Lancet, 2004. **364**(9448): p. 1899-909.

11. Young, H.J., S, *The meaning and measurement of acute malnutrition in emergencies: a primer for decision-makers.* Humanitarian Practices Network, 2006. **56**.
12. Shoham, J., F. Watson, and C. Dolan, *The use of Nutritional Indicators in Surveillance Systems.* International Public Nutrition Resource Group, 2001.
13. Darcy, J.a.C.H., *According to need? Needs assessment and decision-making in the humanitarian sector*, in *Humanitarian Policy Group*. 2003.
14. Tuffrey, V., *A perspective on the development and sustainability of nutrition surveillance in low-income countries.* BMC Nutrition, 2016. **2**(15).
15. Tuffrey, V., *A review of nutritional surveillance systems, their use and value.* Briefing paper, Save the Children UK and Transform Nutrition Research Consortium 2016.
16. Tuffrey, V. and A. Hall, *Methods of nutrition surveillance in low-income countries.* Emerging Themes in Epidemiology, 2016. **13**(4).
17. Briend, A., et al., *Measuring change in nutritional status: a comparison of different anthropometric indices and the sample sizes required.* European journal of clinical nutrition, 1989. **43**(11): p. 769-78.
18. Dale, N.M., et al., *Assessment of the PROBIT approach for estimating the prevalence of global, moderate and severe acute malnutrition from population surveys.* Public Health Nutrition, 2013. **16**(5): p. 858-863.
19. Mude A, B.C., McPeak JG, Kaitho R, Kristjanson P, *Empirical Forecasting of Slow-Onset Disasters for Improved Emergency Response: an Application to Kenya's Arid North.* Food Policy, 2009. **34**(4): p. 329-339.
20. WHO, *Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee*, in *World Health Organization Technical Report Series*. 1995. p. 1-452.
21. Frison, S., et al., *Is Middle-Upper Arm Circumference "normally" distributed? Secondary data analysis of 852 nutrition surveys.* Emerging Themes in Epidemiology, 2016. **13**(1): p. no pagination.
22. Blanton, C. and O. Bilukha, *The PROBIT approach in estimating the prevalence of wasting: Revisiting bias and precision.* Emerging Themes in Epidemiology, 2013. **10**(1).
23. Binkin, N., et al., *Rapid nutrition surveys: how many clusters are enough?* Disasters, 1995. **16**: p. 99-103.
24. Spiegel, P.B., et al., *Quality of malnutrition assessment surveys conducted during famine in Ethiopia.* JAMA, 2004. **292**(5): p. 613-8.
25. SCN, *United Nations Conference on Climate Change. Implications of climate change on undernutrition. SCN statement*. 2009, Standing Committee on Nutrition.
26. Furlow-Parmley C. , et al., *Combining estimates from two surveys: An example from monitoring 2009 influenza A (H1N1) pandemic vaccination.* Statistics in Medicine, 2012. **31**(27): p. 3285-3294.
27. SMART. [cited; Available from: http://smartmethodology.org/.
28. Bilukha, O.O.a.C.B., *Interpreting results of cluster surveys in emergency settings: is the LQAS test the best option?* Emerging Themes in Epidemiology, 2008(5): p. 25.

29. Prudhun, C. and P.B. Spiegel, *A review of methodology and analysis of nutrition and mortality surveys conducted in humanitarian emergencies from October 1993 to April 2004.* Emerging Themes in Epidemiology, 2007. **4**: p. no pagination.
30. Garenne, M., et al., *Incidence and duration of severe wasting in two African populations.* Public Health Nutrition, 2009. **12**(11): p. 1974-82.
31. Isanaka, S., et al., *Estimates of the duration of untreated acute malnutrition in children from niger.* American Journal of Epidemiology, 2011. **173**(8): p. 932-940.
32. Software, *ENA for SMART - Software for Emergency Nutrition Assessment*.

PART III

DISCUSSION – CONCLUSION

Chapter 8 – Discussion and Conclusion

8.1 Summary of research findings

The aim of this thesis was to develop a new method for nutritional surveillance to assess acute malnutrition prevalence using PROBIT Methods based on MUAC. Objective one, to compare the appropriateness of MUAC versus other anthropometric measurements or indices to assess change in the nutritional status of a population, identified MUAC as the best in the detection of short-term change. Objectives 2, 3, 4 and 5 explored the assumption behind the proposed Methods as well as the outcome and performances of the Methods. These four objectives were carried out on a database of 852 nutritional surveys including 668 975 children from 6 to 59 months old. Table 1 below summarises the main findings for each objective.

The PROBIT Methods presented in this thesis perform well. Both the PROBIT Method I (using mean from small sample size surveys and SD from pooled SD from surveys previously conducted in the geographic stratum) and the PROBIT Method II (using mean and SD from the small sample survey) have good outcomes using estimation and classification approaches. The advantage of PROBIT Methods over the classic method is particularly notable for very small sample sizes (under 75) which would allow considerable time and resource (i.e. logistics, finance and human resources) savings.

Table 1: Main findings for each objective

Objectives	Main findings
Identify appropriate index/measure to assess change in nutritional status of a population	- Measures and indices such as weight, WFH, TSF and MUAC perform well - After applying a set of criteria (simplicity, acceptability, cost, independence of age, reliability and accuracy), MUAC stands out as the best measure
Explore normality of MUAC	- MUAC distribution showed no departure from normality in 38% of the surveys - MUAC distribution showed no departure from normality in 75% of the surveys after applying LOESS or Spline smoothing techniques - Applying Box-Cox transformation on surveys showing departure from normality after smoothing resulted in over 80% surveys approximating a normal distribution
Explore MUAC-oedema association	- 60% of oedema cases are missed by measuring GAM only and over 80% by assessing SAM using MUAC - The difference in prevalence between GAM and GAM and between SAM and SAM are marginal overall, but the picture is different at regional level
Assess assumptions linked to MUAC SD	- Once stratified by region, MUAC SD fell within a reasonably narrow range - The variability of MUAC SD varied slightly over time - MUAC SD varied slightly by GAM category
Assess outcome of PROBIT Methods	- PROBIT Methods have clear advantage over the Classic Method for sample sizes under 75 using classification and estimation approaches - The PROBIT approach had better precision than the classic approach for all sample sizes and a better coverage for smaller sample sizes, while having relatively little bias

Estimation approach

The PROBIT Methods slightly overestimate the prevalence of GAM (slight bias), but the precision is systematically greater than the precision from the Classic Method. Furthermore, GAM precision was reasonable starting at a sample size of 50 for PROBIT Method I (half 95% CI of 6.2%) and from a sample size of 75 for PROBIT Method II (half 95% CI of 5.5%). The mean bias is minimal for SAM estimates using the PROBIT Methods and although the precision is higher than for the Classic Method, it is not very satisfactory. A precision of approximately +/- 2% is reached for a sample size of 100 for both PROBIT Method (compared to a +/- 5% for the Classic Method).

Classification Approach

The PROBIT Methods have higher probability to classify GAM prevalence correctly than the Classic Method for very small sample sizes; and the PROBIT Method II has a slightly better outcome using the classification approach than PROBIT Method I. The probability of correctly classifying the GAM prevalence was low for 10% and 15% thresholds, but was over 90% for a 5% threshold for both PROBIT Methods for sample sizes as small as 25.

Performance by region

Overall, the mean bias is slightly lower with PROBIT Method II while the mean precision is finer with PROBIT Method I. In the Caribbean, both PROBIT Methods have minimal bias, but the precision obtained with PROBIT Method I would allow using sample sizes as small as 25 while the PROBIT Method II would likely require a sample size of 50. Both PROBIT Methods have reasonable bias and precision in West Africa and Central and South Africa with a sample size of 50, while East Africa and Asia may require a sample size of 75 for better precision.

Assumptions behind the Methods

Both PROBIT Methods entail the following two common assumptions: firstly, that MUAC distribution is "normal" or can be transformed to take a normal distribution; and secondly, that SAM/GAM MUAC cut-offs are sensitive enough to oedema (i.e. there is an overlap between children with low MUAC and children with oedema). These two assumptions were verified as follows:

- Over a third of MUAC distributions in the database were normally distributed. MUAC distributions can easily be normalised applying simple smoothing techniques if the distribution is noisy or displays digit preference, and then by applying Box-Cox transformation if indicated (i.e. if data are skewed). Both Spline and LOESS smoothing techniques increased the proportion of "normal distribution" to three quarters; and this proportion reached over 80% after applying Box-Cox transformation on surveys showing departure from normality after smoothing (Chapter 4).

- Two thirds of oedema cases are missed by measuring GAM only, and over 80% are missed by assessing SAM using MUAC. The difference in prevalence between GAM and GAM and between SAM and SAM are not significant overall. In Central and South Africa where the largest number and prevalence of oedema cases was observed, the mean difference between global estimates was under 1%. The difference between SAM and SAM was 1%. These results indicate that using MUAC alone for global estimates is reasonable worldwide, while where kwashiorkor accounts for a non-negligible proportion of SAM, using SAM alone significantly underestimates the estimate of SAM prevalence.

The PROBIT Method I entailed other assumptions: (i) the variability in MUAC SD from our database of nutritional surveys is representative of the variability that we can expect in the future; (ii) MUAC SD is itself not strongly associated with average

nutritional status; and (iii) that within a given stratum, there is little variability in MUAC SD.

- MUAC SD varies slightly over time. The methods outcomes were therefore examined before 2006 and after 2006. This date was picked as SMART methodology was implemented in 2006, which may have impacted on the quality of the surveys. The bias in the PROBIT Method I is significantly lower after 2006 in a multivariable regression, but it was also the case for the PROBIT Method II that did not rely on MUAC SD from previous surveys. Furthermore, the mean bias is significantly higher for the Classic Method after 2006.

- MUAC SD varied a little between GAM categories based on MUAC (<5%, 5-9%, 10-14%, ≥15%). The Methods were assessed for different GAM categories. PROBIT Method I showed lower mean bias and lower mean precision for GAM categories 5-9% and ≥15%, while the PROBIT Method II had lower bias and higher precision for lower GAM categories (<5% and 5-9%), and the Classic Method had a higher bias the higher the GAM category. Differences in mean bias were significant in multivariable regressions for all Methods.

- There is reasonable homogeneity in MUAC SD across the entire dataset. MUAC SD varies between region, livelihood zones and residence status. The database was stratified by region as it most allows for minimising the variability of MUAC SD.

8.2 Overall applications of the thesis

The proposed methods are not a stand-alone system; rather, they are designed to complement and feed into existing nutrition surveillance systems. It is essential that the findings are made available on a real-time basis and can be fed into national early warning systems for food security and nutrition. This method is field-practical and requires low levels of resources (i.e. time, human resources, and finances), which would enable more routine and timely estimations of SAM/GAM prevalence. It is particularly suitable in resource-constrained and crisis-affected settings.

As part of the nutrition surveillance systems, some of the Method's general applications include:

i) better planning for nutrition programmes where the estimation of GAM prevalence would enable planning needs depending expected caseloads

ii) better targeted interventions, since an independent estimate of GAM could be computed for small spatial or population strata of a large region

iii) Better monitoring and evalution of nutrition programmes and justify more rapid and responsive scale-up/scale down of intervention

iv) More timely intervention as PROBIT Methods could be used for rapid assessment

Estimation and classification approach

The main advantage of estimation is that trends can be monitored over time, whereas the main advantage of classification is that the information is simple and easy to communicate, and the sample size requirement is usually reduced[1].

The estimation approach could be used to follow trends and enable interpretation as compared to what it to be expected at the same time of the year, and anticipate/put in place intervention before the situation deteriorates. In the absence of baseline data, arbitrary benchmarks for gravity of the nutritional situation could be used such as a MUAC adaptation of the widely used WFH GAM classification[2-4] (see Table 2). The same thresholds could be used with MUAC. Both PROBIT Methods proposed yield GAM estimates with reasonable precision using a sample size of 50 (different regions may use different sample sizes, see Research Findings section above).

Although the precision of SAM estimates produced by the PROBIT Methods is not very high, SAM estimates could be particularly useful to infer the SAM caseload to expect in programmes treating children with SAM. However, in places where kwashiorkor accounts for a non-negligible proportion of SAM, an upward correction factor will need to be applied to the SAM estimates generated.

The PROBIT Methods had good classification outcomes for the 5% threshold. The probability of correctly classifying GAM prevalence as ≥ 5 % was over 90%. This could be used to decide whether further investigation is needed. In a situation where the GAM prevalence is ≥ 5 % with aggravating factors (see paragraph below), additional, more costly and resource intensive assessments could then be conducted.

Whether using the estimation or classification approach, the proposed PROBIT Methods suggest the use of very small sample size for the assessment of acute malnutrition. This would allow substantial resource savings (i.e. time, financial, human resources) and would enable the frequency required for a swift response, if required. Furthermore, the use of MUAC alone would make training, supervision and analysis easier. It is also feasible and acceptable in the community and the MUAC tool is cheap, easy to transport and easy to use[5-7]. This would also allow for better monitoring and evaluation of nutrition programme as these method would allow to measure impact of programme on acute malnutrition prevalemce on a regular basis and at the programme geographic resolution.

Table 2: WHO decision chart for the implementation of selective feeding programmes (based on WFH GAM)

Finding	Action required
Malnutrition rate ≥15% or 10-14% with aggravating factors	*Serious situation:* -General rations (unless situation is limited to vulnerable groups); plus -Supplementary feeding generalised for all members of vulnerable groups, especially children and pregnant and lactating women - Therapeutic feeding programme for severely malnourished individuals
Malnutrition rate 10-14% or 5-9% with aggravating factors	*Risky situation* - No general rations; but - Supplementary feeding targeted to individuals as malnourished in vulnerable groups - Therapeutic feeding programme for severely malnourished individuals
Malnutrition rate < 10% with no aggravating factors	*Acceptable situation:* - No need for population interventions - Attention for malnourished individuals through regular community services

Adapted from WHO The management of nutrition in major emergencies (2000)[8]

In settings considered at risk, data could be collected on a quarterly basis as a starting point. The time of data collection will have to line up good (e.g. post-harvest) and bad (e.g. lean season) period of the year, as well as agricultural and disease calendars.

The same GAM estimates/or classification can mean different things in different contexts or time of the year. For example, a GAM prevalence of 10% cannot be interpreted the meaning the same thing post-harvest or during the lean season. GAM Estimates or classification need to be put into context, and other available data such as food security, morbidity and mortality, markets prices, access to food should be taken into account in order to interpret results (see “aggravating” factors in Table 2).

Although widely used, the evidence base behind the classification Table 2 is very weak and has limitations. Another more flexible apporach was described in a field exchange article[9] that would be based on the number of children suffereing from acute malnutriton the health system can treat and the expected number of cases. Intervention would then be planed when expected number of cases exceed the health system’s capacity.

Sampling design

Different sampling designs will be considered depending on various settings (e.g. individual camps; large regions; rural districts). The PROBIT Methods performed well using simple random sampling and multi-stage cluster sampling.

A simple random sampling could be used when a list of households is available or where the population is geographically concentrated and households are arranged in a regular pattern. Such a situation may occur in a camp or in urban areas (e.g.

blocks of flats). The basic sampling units (i.e. household) are numbered and the desired number of units are then randomly selected from a random number table.

Cluster sampling is the most common form of sampling and is done in two stages:

1. The entire population is divided into smaller discrete geographical areas (i.e. villages) whose population is known or can be estimated. Clusters are then randomly selected from these villages (using probability to population size (PPS) approach: each person in the whole area has an equal chance of being selected). For both PROBIT Methods, 25 clusters will be selected.
2. A simple random sampling can be used if possible to select units (i.e. households) within clusters. Other options are to use systematic random sampling when an updated and exhaustive list of households in the cluster is available or possible to make. In that case the first household is randomly selected and the subsequent households are visited systematically using a "sampling interval" (total number of households/ sample size required). Otherwise, the first household to visit is selected randomly and the subsequent households to visit is the next to the right until the number of children to measured reached.

Technology and mapping

The use of tablets and mobile phones for data collection is increasingly used for data collection and could be used to conduct the surveys using PROBIT Methods. Data collected would be automatically checked and compiled in order to produce results instantaneously. Mapping prevalence estimates or degree of seriousness of a situation using software such as Geographic Information System (GIS) could also be a useful tool to easily communicate outcomes to humanitarian actors and policy makers.

8.3 Limitations

I recognise this thesis has several limitations:

Although the main assumption behind the PROBIT Methods were validated, the following two assumptions behind the PROBIT Method II were not: (i) the variability in MUAC SD from our database of nutritional surveys is representative of the variability that we can expect in the future; and (ii) MUAC SD is itself not strongly associated with average nutritional status. Specifically:

i) MUAC SD varies slightly over time, and the out the mean bias of PROBIT Method II showed significant differences between time periods once in a multiple linear regression. However, this was the case for all methods assessed and is therefore not necessarily linked to MUAC SD's variability over time.

ii) Similarly, although MUAC SD varies slightly between GAM categories based on MUAC (<5%, 5-9%, 10-14%, ≥15%), the outcomes of all Methods assessed showed significant differences by GAM categories.

Another important limitation is the fact that surveys were used as proxies for true populations. We do not know what implication this may have on the bias and precision of the proposed Method.

Other limitations are discussed in each of the four research papers.

8.4 Future Research

Findings from this thesis raise additional questions, which could be addressed in the following future research:

i) The development of a user-friendly data collection tool using tablets or smart phones. The device or software could include a system to flag extreme values in order to avoid errors in data entry;

ii) The development of a user-friendly tool to enter and analyse data using the PROBIT Methods. The tool would calculate the estimates of SAM/GAM or classify GAM instantaneously once MUAC values collected. This would allow results to be available as soon as data collection is over. Digit preference could be explored as possible way to assess data quality;

iii) The piloting of the PROBIT methods presented in this thesis could be conducted in one or more field settings, such as in urban or rural areas and in different regions, selected based on findings from this thesis. Different sampling strategies could also be tested. The pilot studies would mainly focus on the feasibility and usefulness of the Methods;

iv) Although the Methods presented in the thesis perform well, there is potential for improvement. Future research could examine a Bayesian approach to the PROBIT Method, incorporating the prior information from previous surveys as a way to potentially increase precision and decrease bias. For example, this method could use MUAC SDs from surveys to inform a prior for the distribution of MUAC SD for use in a Bayesian estimate of PROBIT prevalence from small sample surveys; and

v) This thesis highlighted that in countries where kwashiorkor accounts for a large proportion of SAM cases, the estimation of SAM based on MUAC alone would underestimate the prevalence of SAM. Another potential area for research would be to establish the necessary upward correction needed for SAM in these high prevalence kwashiorkor countries.

8.5 Conclusions

This thesis explores the use of PROBIT Methods with MUAC for the assessment of acute malnutrition. It shows that these Methods have a clear advantage compared to the classic prevalence based method. The use of PROBIT Methods would require much lower sample sizes, which would allow substantial resource savings (i.e. time, financial, and human resources) and would enable the frequency required for a swift response. The use of MUAC alone would make the training, supervision and analysis easier. Furthermore, MUAC is the best measure to detect short term changes in the nutritional status of the population and is the most suitable measure at community level. Finally, the proposed PROBIT Methods fit in well with the increasing interest in MUAC-only nutrition programming. There is great potential in their use in surveillance systems in order to produce timely and/or locally relevant prevalence estimates of acute malnutrition, to enable timely and well-targeted responses and interventions.

References

1. Blanton, C. and O.O. Bilukha, *The PROBIT approach in estimating the prevalence of wasting: Revisiting bias and precision.* Emerging Themes in Epidemiology, 2013. **10**(1): p. no pagination.
2. Mason, J.B. and J.T. Mitchell, *Nutritional surveillance.* Bulletin of the World Health Organization, 1983. **61**(5): p. 745-55.
3. Young, H.J., S, *The meaning and measurement of acute malnutrition in emergencies: a primer for decision-makers.* Humanitarian Practices Network, 2006. **56**.
4. Young, H.J., S, *Review of Nutrition and Mortality Indicators for the Integrated Food Security Phase Classification (IPC): Reference Levels and Decision-Making*. 2009.
5. Briend, A., et al., *Mid-upper arm circumference and weight-for-height to identify high-risk malnourished under-five children.* Matern Child Nutr, 2012. **8**(1): p. 130-3.
6. Myatt, M., T. Khara, and S. Collins, *A review of methods to detect cases of severely malnourished children in the community for their admission into community-based therapeutic care programs.* Food and nutrition bulletin, 2006. **27**(3 Suppl): p. S7-23.
7. Prendiville, N., *Nutrition Surveillance in Somalia.* Field Exchange, 2001(14): p. 14.
8. WHO, *The management of nutrition in major emergencies*. 2000, World Health Organization: Geneva.
9. Hailey, P. and D. Tewoldeberha, *Suggested New Design Framework for CMAM Programming. .* Field Exchange 2010. **39**.

ANNEXES

Annex I: MUAC versus MUAC-for-age

MUAC grows continuously with age, and when a fixed cut-off is used for identifying SAM children, more young children are selected than with an index independent of age. In 1993, a WHO Expert Committee reviewed the scientific evidence underlying the use and interpretation of MUAC [1]. The Committee examined mean MUAC data across ages from the NCHS sample of children in the USA, and for a cohort of Malawian children. For both populations, MUAC increased by approximately 2 cm between 6 and 59 months of age. A WHO expert Committee therefore recommended a new MUAC-for-age set of reference data for children aged 6-59 months [2] that was later included in the 2006 WHO growth standards [3]. While accurate assessment of age is problematic in many developing countries [4-6], it allows for better comparison of anthropometric status across populations with varying age structure, and could potentially be more sensitive to alterations in anthropometric status than MUAC alone.

Despite the availability of MUAC-for-age references, MUAC-for-age is not used or very little while uncorrected MUAC is increasingly recognised as a very useful measure of anthropometric status (see chapter 1 and 2). There are numerous reasons for this:

- The need to determine age is avoided[7]. This is widely acknowledged as problematic in many developing countries [4-6]. Any nutritional indices which involve the collection of age data can be difficult, especially in emergency settings. There are numerous advantages associated with not needing age:
    - Simplicity: there is also no need to use complex look-up tables to determine whether MUAC-for-age is normal or low.
    - Cost and time savings: assessment is very quick and staff can spend their time on other more urgent issues.

- Uncorrected MUAC biases younger children. This is a desirable bias in terms of admission to child health/nutrition programmes since these are at higher risk of mortality than older children. Hence is this improves the selection of a high risk group.

When looking at age distributions in our database (see chapter 3 for details on database), there was strong evidence of rounding of the age variable, suggesting poor quality, inaccurate data. Surveys included in the database were conducted in settings where dates of birth are usually unknown. Age was approximated with a calendar of event and recorded in months. Table 1 presents age preference for each year (12 months, 24 months, 48 months and 59 months old) with the expected/observed proportion of children per age. Overall, there were 2.2 times more children with an age rounded to the nearest year than expected. The age preference was particularly high for children around 3 and 4 years old. For this reason of inaccurate age, accurate MUAC-for-age is unlikely and risks errors of interpretation and incorrect (or missed) referrals if it were to be used in any field programmes.

Table 1: Age preference across all surveys

Age in months	Number of children	Observed %	Expected %
12	18646	2.8	1.9
24	30686	4.6	1.9
36	38785	5.8	1.9
48	34768	5.2	1.9
59	18101	2.7	1.9
All	140986	21.1	9.5

Despite the availability of references, since early 2000s, the use of MUAC-for-age has not been. It is emphasised in numerous international guidelines:

- A 2007 Joint Statement by WHO, WFP and UNICEF on community-based management of SAM[8].
- In a 2009 WHO/UNICEF Joint statement on "WHO child growth standards and the identification of severe acute malnutrition in infants and children" [9]
- Recently up-dated 2013 WHO guidelines on the management of SAM recommend uncorrected MUAC and not MUAC-for-Age[10].

Other recent publications discussing methodological approached to nutrition surveillance, the measurement of acute malnutrition and the use of anthropometric indicators all focus on unadjusted MUAC and did not mention MUAC-for-Age [11-16]. Table 2 below was borrowed from a review of methods to detect cases of severely malnourished children in the community[17]. MUAC was clearly identified as the most adequate measurement for the screening and detection of malnutrition in the community.

MUAC is clearly identified as most appropriate when compared to MUAC-for-age. For all the above reasons, this study focuses on unadjusted MUAC as it is dominant in international policy and is likely to remain so for the foreseeable future.

Table 2: Capacity of common indicators with regard to key properties of case-detection methods for screening and case detection of malnutrition in the community (adapted from Myatt et al)

Property	Indicator						
	Clinical	WFA	HFA	WFH	MUAC	MUAC/A	MUAC/H
Simplicity	No	No	No	No	Yes	No	Yes (by quick stick only)
Acceptability	No	No	No	No	Yes	Yes	Yes (by quick stick only)
Cost	No	No	No	No	Yes	Yes	Yes (by quick stick only)
Objective	No	No	No	Yes	Yes	No	Yes
Quantitative	No	Yes	Yes	Yes	Yes	Yes	Yes
Independent of age	Yes	No	No	No	Yes	No	Yes
Precision(reliability)	No	Yes	No	No	Yes	Yes	Yes (by quick stick only)
Accuracy	No	No	No	No	Yes	No	Yes
Sensitivity	NA	Yes	No	No	Yes	Yes	Yes
Specificity	NA	Yes	No	No	Yes	Yes	Yes
Predictive value	NA	YEs	No	No	Yes	Yes	Yes

MUAC/A: MUAC-for-Age, MUAC/H: MUAC-for-Height

References

1. *Workshop 1: The physical status: The use and interpretation of anthropometry in adolescence: Discussion.* Hormone Research, 1993. **39**(SUPPL. 3): p. 68-70.
2. de Onis M and J.P. Habicht, *Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee.* 1996. **64**: p. 650-658.
3. de Onis, M., R. Yip, and Z. Mei, *The development of MUAC-for-age reference data recommended by a WHO Expert Committee.* Bull World Health Organ, 1997. **75**(1): p. 11-8.
4. Hamer, C., et al., *Detection of severe protein-energy malnutrition by nurses in The Gambia.* Arch Dis Child, 2004. **89**(2): p. 181-4.
5. Jelliffe EFP, J.D., *The arm circumference as a public health index of protein-calorie malnutrition of early childhood.* Journal of tropical pediatrics, 1969. **15**(4): p. 177-260.
6. Bairagi, R. and R.I. Ahsan, *Inconsistencies in the findings of child nutrition surveys in Bangladesh.* American Journal of Clinical Nutrition, 1998. **68**(6): p. 1267-71.
7. De Onis, M., R. Yip, and Z. Mei, *The development of MUAC-for-age reference data recommended by a WHO Expert Committee.* Bulletin of the World Health Organization, 1997. **75**(1): p. 11-18.
8. WHO, WFP, and UNICEF, *Community-based management of severe acute malnutrition. Joint statement*. 2007, WHO, WFP, UNICEF.

9. WHO and UNICEF, *WHO child growth standards and the identification of severe acute malnutrition in infants and children. A Joint Statement by the World Health Organization and the United Nations Children's Fund.* 2009.
10. WHO, *Updates on the management of Severe Acute Malnutrition in infants and children*. 2013, WHO.
11. Bilukha, O., et al., *Measuring anthropometric indicators through nutrition surveillance in humanitarian settings: Options, issues, and ways forward.* Food and Nutrition Bulletin, 2012. **33**(2): p. 169-176.
12. Shoham, J., F. Watson, and C. Dolan, *The use of Nutritional Indicators in Surveillance Systems.* International Public Nutrition Resource Group, 2001.
13. Young, H.J., S, *The meaning and measurement of acute malnutrition in emergencies: a primer for decision-makers.* Humanitarian Practices Network, 2006. **56**.
14. Tuffrey, V., *A perspective on the development and sustainability of nutrition surveillance in low-income countries.* BMC Nutrition, 2016. **2**(15).
15. Tuffrey, V., *A review of nutritional surveillance systems, their use and value.* Briefing paper, Save the Children UK and Transform Nutrition Research Consortium 2016.
16. Tuffrey, V. and A. Hall, *Methods of nutrition surveillance in low-income countries.* Emerging Themes in Epidemiology, 2016. **13**(4).
17. Myatt, M., T. Khara, and S. Collins, *A review of methods to detect cases of severely malnourished children in the community for their admission into community-based therapeutic care programs.* Food and nutrition bulletin, 2006. **27**(3 Suppl): p. S7-23.

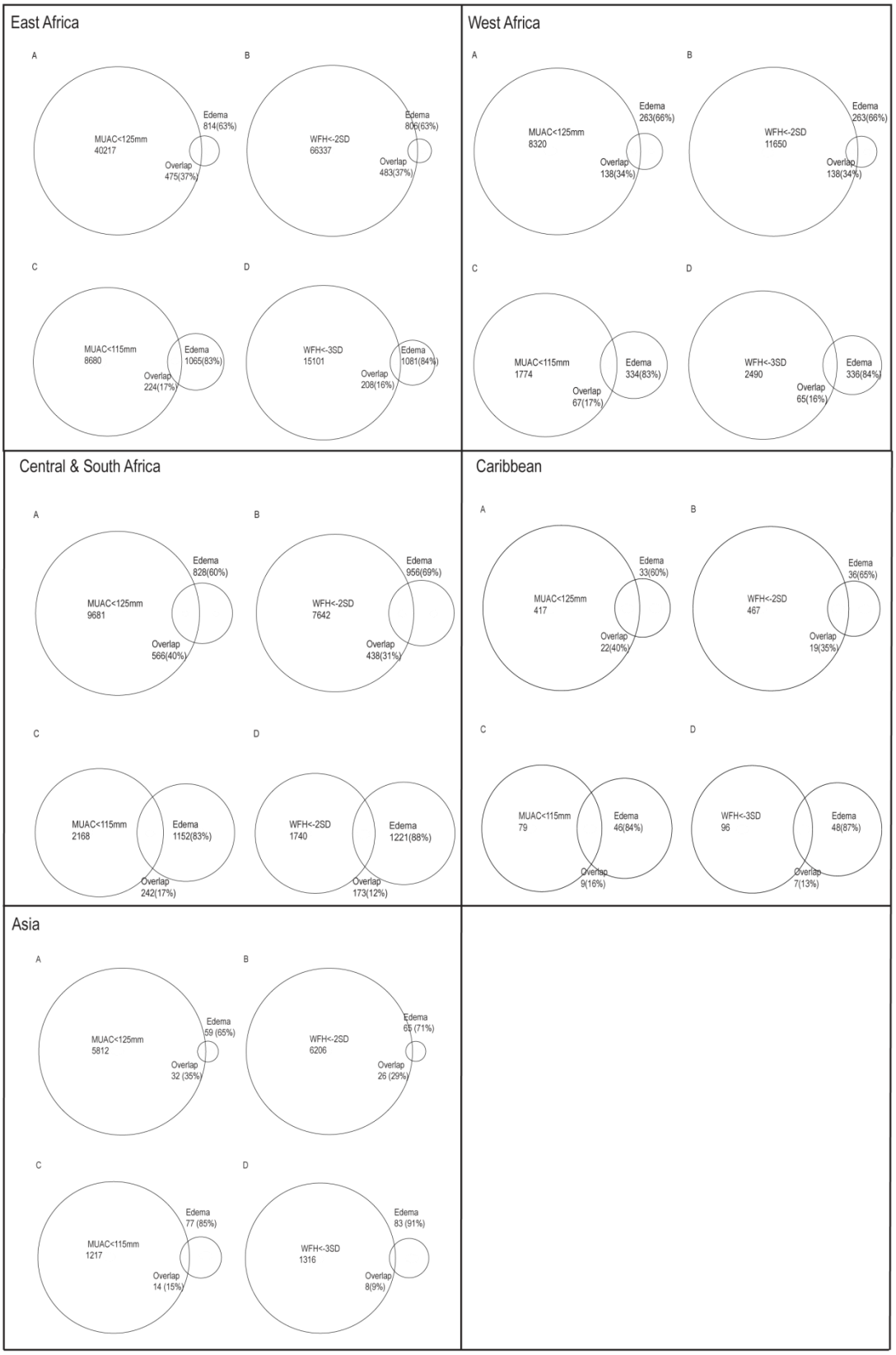
Annex II: Number of surveys per country

Table 1: Number of surveys per country

Country	Number of surveys
Afghanistan	19
Angola	18
Bangladesh	5
Burkina Faso	2
Burundi	10
CAF	5
Cameroun	1
Chad	19
DRC	72
Ethiopia	107
Guinea	5
Haiti	13
India	2
Kenya	16
Liberia	3
Madagascar	1
Malawi	8
Mali	6
Mauritania	2
Mozambique	1
Myanmar	9
Nepal	8
Niger	44
Nigeria	4
Pakistan	11
RCA	2
Rwanda	9
Sierra Leone	11
Somalia	192
South Sudan	122
Sri Lanka	1
Sudan	100
Tajikistan	1
Tanzania	1
Thailand	2
Uganda	16
West Timor	2
Zambia	2

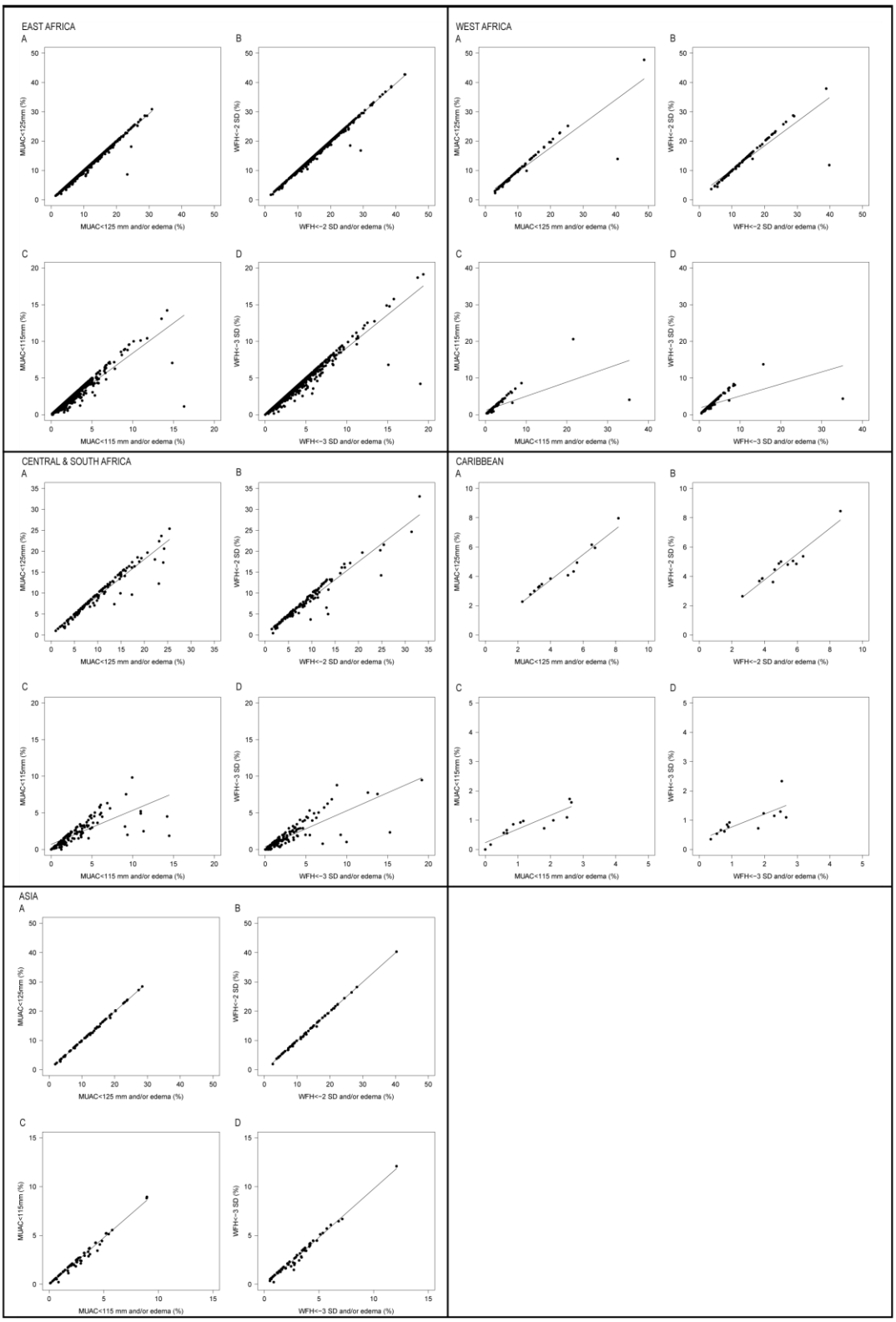
Annex III: Supplemental Online Figures

Online Supplemental Material



Supplemental Figure 1: Overlap between bilateral edema and wasting based on Middle-Upper Arm Circumference (MUAC) and Weight-For-Height (WFH) and between edema bilateral and severe wasting based on MUAC and WFH in different regions

Online Supplemental Material



Supplemental Figure 2: Wasting versus acute malnutrition based on Middle-Upper Arm Circumference (MUAC) and Weight-For-Height (WFH) and severe wasting versus severe acute malnutrition with MUAC and WFH in different regions

Annex IV: Supplemental Online Tables

Supplemental Table 1: Summary statistics of oedema prevalence in each region

Region	Minimum (%)	Lower quartile (%)	Median (%)	Mean (%)	Upper quartile (%)	Maximum (%)
Asia (n=60)	0	0.0	0.1	0.2	0.3	1.2
Caribbean (n=13)	0	0.1	0.2	0.5	1.1	1.6
Central & South Africa (n=128)	0	0.2	0.6	1.2	1.4	13.3
East Africa (n=554)	0	0.0	0.1	0.3	0.4	15.2
West Africa (n=97)	0	0.0	0.0	0.6	0.3	32.9

Supplemental Table 2: Summary statistics of the differences between estimates of GAM and GAM and SAM and SAM using MUAC or WFH per region

East Africa (n=554)	Minimum (%)	Lower quartile (%)	Median (%)	Mean (%)	Upper quartile (%)	Maximum (%)
GAM-GAM (MUAC)	0.0	0.0	0.0	0.2	0.2	14.7
GAM-GAM (WFH)	0.0	0.0	0.0	0.2	0.2	12.5
SAM-SAM (MUAC)	0.0	0.0	0.1	0.3	0.3	15.2
SAM-SAM (WFH)	0.0	0.0	0.1	0.3	0.3	14.8
West Africa (n=97)	**Minimum (%)**	**Lower quartile (%)**	**Median (%)**	**Mean (%)**	**Upper quartile (%)**	**Maximum (%)**
GAM-GAM (MUAC)	0.0	0.0	0.0	0.4	0.1	26.7
GAM-GAM (WFH)	0.0	0.0	0.0	0.4	0.1	27.9
SAM-SAM (MUAC)	0.0	0.0	0.0	0.5	0.2	31.3
SAM-SAM (WFH)	0.0	0.0	0.0	0.5	0.2	30.8
Central & South Africa (n=128)	**Minimum (%)**	**Lower quartile (%)**	**Median (%)**	**Mean (%)**	**Upper quartile (%)**	**Maximum (%)**
GAM-GAM (MUAC)	0.0	0.0	0.2	0.7	0.8	30.8
GAM-GAM (WFH)	0.0	0.1	0.3	0.8	0.9	12.6
SAM-SAM (MUAC)	0.0	0.1	0.4	1.0	1.1	12.6
SAM-SAM (WFH)	0.0	0.2	0.5	1.1	1.1	13.0
Caribbean (n=13)	**Minimum (%)**	**Lower quartile (%)**	**Median (%)**	**Mean (%)**	**Upper quartile (%)**	**Maximum (%)**
GAM-GAM (MUAC)	0.0	0.0	0.2	0.3	0.7	1.1
GAM-GAM (WFH)	0.0	0.0	0.2	0.4	0.7	1.1
SAM-SAM (MUAC)	0.0	0.0	0.2	0.5	1.0	1.4
SAM-SAM (WFH)	0.0	0.0	0.2	0.5	1.1	1.6
Asia (n=60)	**Minimum (%)**	**Lower quartile (%)**	**Median (%)**	**Mean (%)**	**Upper quartile (%)**	**Maximum (%)**
GAM-GAM (MUAC)	0.0	0.0	0.0	0.1	0.2	1.0
GAM-GAM (WFH)	0.0	0.0	0.0	0.1	0.2	1.2
SAM-SAM (MUAC)	0.0	0.0	0.0	0.2	0.3	1.0
SAM-SAM (WFH)	0.0	0.0	0.1	0.2	0.3	0.2

GAM, Global Acute Malnutrition; MUAC, Middle-Upper Arm Circumference; SAM, Severe Acute Malnutrition; WFH, Weight-For-Height

Annex V: Specificity and sensitivity of MUAC in the detection of oedema cases

Different receiver operating characteristic (ROC) curves were built to look at the specificity and sensitivity of MUAC in the detection of oedema cases. These curves are plots of the sensitivity versus one minus the specificity of MUAC in the detection of oedema cases as MUAC thresholds increase. Figure 1 shows the ROC curves of different models from one with all confounders identified with the multiple logistic regression (Chapter 5 table 3) to a simple model with MUAC and oedema only. The curves are very similar until the variable region is taken out of the model. It seemed therefore sensible to look at the relationship between MUAC and oedema in each region.

Table 1 looked at sensitivity and specificity of MUAC in the detection of oedema cases in each region. Sensitivity doesn't vary much; for a MUAC threshold of 110mm, sensitivity is around 60% (from 57.28 to 65.95%). Specificity varies more, from 63.26 to 84.55%. Figure 2 presents the corresponding ROC curves as MUAC threshold increased in each region.

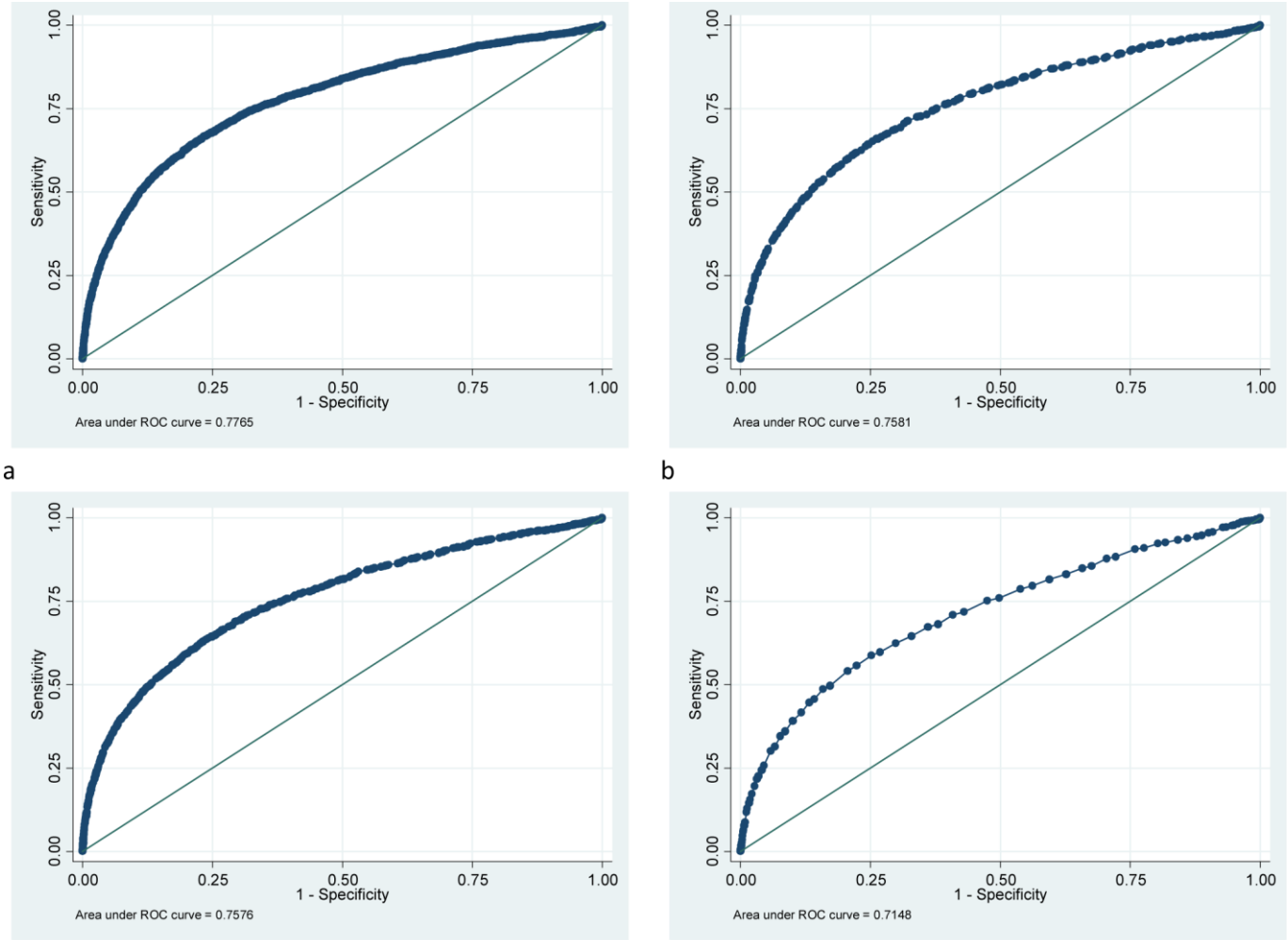


Figure 1: ROC curves of MUAC (continuous) in the detection of oedema cases in different models including adjusted for region, livelihood and age (a), for region and age (b), for region (c) and unadjusted (d)

Table 1: Sensitivity and specificity of different cut-off points of MUAC for oedema in each region

	MUAC cut-off point					
	110 mm	**115 mm**	**120 mm**	**125 mm**	**130 mm**	**135 mm**
East Africa						
Sensitivity	66.0	50.6	42.2	27.0	18.2	12.6
Specificity	66.0	82.4	88.0	95.4	97.7	98.9
West Africa						
Sensitivity	57.3	45.9	36.3	24.0	16.5	11.9
Specificity	72.3	84.0	89.1	95.4	97.7	98.8
Central & South Africa						
Sensitivity	63.1	50.3	42.6	30.5	16.9	7.6
Specificity	71.9	83.7	89.0	94.3	97.8	88.2
Caribbean						
Sensitivity	62.7	50.8	41.8	25.4	17.9	11.9
Specificity	84.6	92.2	95.5	98.3	99.1	99.6
Asia						
Sensitivity	61.7	46.8	34.0	28.7	17.0	8.5
Specificity	63.3	77.7	87.1	92.6	96.7	99.0

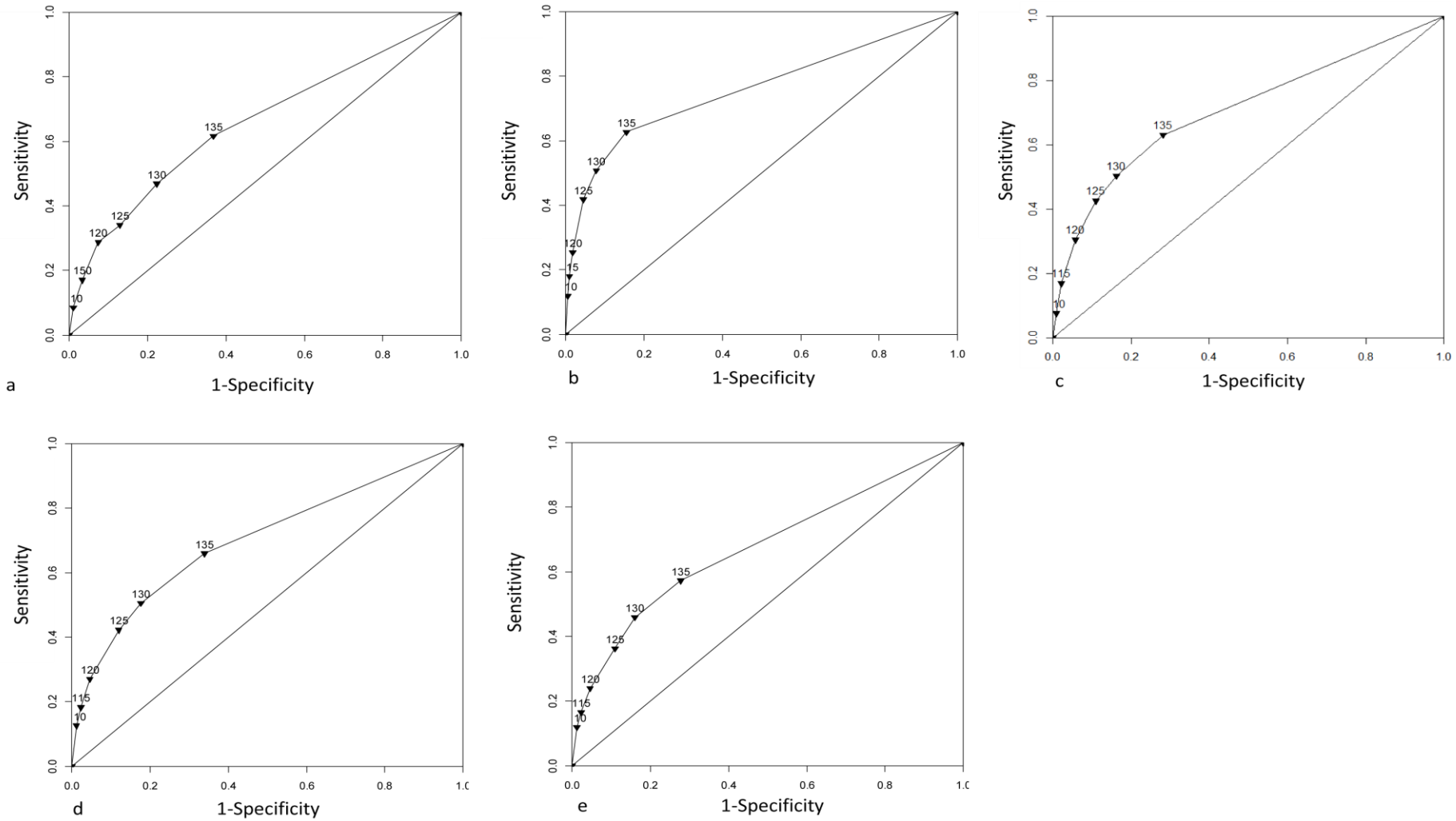


Figure2: ROC curve of oedema for different MUAC cut-offs in Asia (a), the Caribbean (b), Central and South Africa (c), East Africa (d) and West Africa (e)

Annex VI: Supporting information

Supporting Table 1: Bias in GAM and SAM estimates

Table 1: Bias in GAM and SAM estimates

Sample size 25	Minimum	Lower	Median	Mean	Upper	Maximum
Classic Method						
SAM	-16.6	0.0	0.0	1.0	1.6	23.2
GAM	-31.7	-2.1	0.1	0.7	3.3	33.1
PROBIT Method I						
SAM	-12.2	-0.7	0.1	0.2	1.0	17.0
GAM	-21.9	-1.5	0.9	1.2	3.6	25.4
PROBIT Method II						
SAM	-17.2	-0.8	0.0	0.3	1.2	16.0
GAM	-26.3	-2.3	0.4	0.8	3.6	27.5
Sample size 50	**Minimum**	**Lower**	**Median**	**Mean**	**Upper**	**Maximum**
Classic Method						
SAM	-12.6	0.0	0.0	0.6	1.1	17.4
GAM	-21.7	-2.3	0.0	0.2	2.3	21.9
PROBIT Method I						
SAM	-12	-0.6	0.1	0.1	0.9	11.2
GAM	-19.8	-1.1	0.8	1.0	3.0	20.1
PROBIT Method II						
SAM	-11.1	-0.6	0.0	0.2	0.3	14.5
GAM	-19.0	-1.5	0.5	0.8	2.8	19.0
Sample size 75	**Minimum**	**Lower**	**Median**	**Mean**	**Upper**	**Maximum**
Classic Method						
SAM	-8.8	0.0	0.0	0.3	0.8	12.8
GAM	-16.2	-2.0	0.0	0.0	1.9	17.3
PROBIT Method I						
SAM	-11.4	-0.6	0.1	0.1	0.8	8.7
GAM	-17	-1	0.9	0.9	2.8	17.3
PROBIT Method II						
SAM	-9.8	-0.5	0.0	0.1	0.7	10.5
GAM	-15.5	-1.1	0.6	0.7	2.4	16.2
Sample size 100	**Minimum**	**Lower**	**Median**	**Mean**	**Upper**	**Maximum**
Classic Method						
SAM	-7.6	-0.2	0.0	0.2	0.7	9.9
GAM	-13.6	-1.7	0.0	0.0	1.7	14.2
PROBIT Method I						
SAM	-11	-0.6	0.1	0.1	0.8	8.1
GAM	-15.4	-0.9	0.8	0.9	2.7	15.0
PROBIT Method II						
SAM	-8.4	-0.5	0.0	0.1	0.6	8.1
GAM	-12.4	-0.9	0.6	0.7	2.2	14.4

Sample size 125	Minimum	Lower	Median	Mean	Upper	Maximum
Classic Method						
SAM	-7.5	-0.3	0.0	0.2	0.6	10.9
GAM	-12.5	-1.5	0.0	0.0	1.5	13.7
PROBIT Method I						
SAM	-11	-0.6	0.1	0.1	0.8	7.6
GAM	-15.4	-0.9	0.8	0.9	2.7	15.2
PROBIT Method II						
SAM	-8.4	-0.5	0.0	0.1	0.6	8.3
GAM	-10.2	-0.7	0.6	0.7	2.1	13.7
Sample size 150	**Minimum**	**Lower**	**Median**	**Mean**	**Upper**	**Maximum**
Classic Method						
SAM	-6.9	-0.4	0.0	0.1	0.6	8.9
GAM	-11.1	-1.3	0.0	0.0	1.3	12.5
PROBIT Method I						
SAM	-10.7	-0.6	0.1	0.1	0.8	7.6
GAM	-14.5	-0.9	0.9	0.9	2.6	13.9
PROBIT Method II						
SAM	-7.8	-0.4	0.0	0.1	0.6	7.3
GAM	-9.5	-0.6	0.6	0.7	2.0	11.6
Sample size 175	**Minimum**	**Lower**	**Median**	**Mean**	**Upper**	**Maximum**
Classic Method						
SAM	-7.0	-0.4	0.0	0.1	0.5	7.1
GAM	-9.9	-1.2	0.0	0.0	1.2	12.0
PROBIT Method I						
SAM	-10.7	-0.6	0.1	0.1	0.7	7.3
GAM	-14.3	-0.8	0.8	0.8	2.6	13.6
PROBIT Method II						
SAM	-7.4	-0.4	0.0	0.1	0.5	7.2
GAM	-8.8	-0.5	0.6	0.7	1.9	13.2
Sample size 200	**Minimum**	**Lower**	**Median**	**Mean**	**Upper**	**Maximum**
Classic Method						
SAM	-5.4	-0.4	0.0	0.1	0.5	6.9
GAM	-9.7	-1.1	0.0	0.0	1.1	10.6
PROBIT Method I						
SAM	-10.8	-0.6	0.1	0.1	0.7	6.9
GAM	-14.8	-0.8	0.9	0.8	2.6	12.9
PROBIT Method II						
SAM	-7.2	-0.4	0	0.1	0.5	7.5
GAM	-8.7	-0.5	0.6	0.7	1.9	10.9

Supporting Table 2: **Bias, precision and coverage of the different methods by GAM categories**

Table 1: Bias of GAM (a, b, c) and SAM (d, e, f) estimates by GAM categories (based on MUAC) (%)

(a) Classic Method

Sample size	All Mean (%)	GAM <5% Mean (%)	5-9% Mean (%)	10-14% Mean (%)	≥15% Mean (%)
25	0.7	1.3	1.0	0.4	0.0
50	0.2	0.5	0.1	0.1	0.0
75	0.0	0.2	0.0	0.0	0.0
100	0.0	0.1	0.0	0.0	0.1
125	0.0	0.0	0.0	0.0	0.0
150	0.0	0.0	0.0	0.0	0.0
175	0.0	0.0	0.0	0.0	0.0
200	0.0	0.0	0.0	0.0	0.0

(b) Probit Method I

Sample size	All Mean (%)	GAM <5% Mean (%)	5-9% Mean (%)	10-14% Mean (%)	≥15% Mean (%)
25	1.2	1.4	1.3	0.9	1.1
50	1.0	1.2	1.1	0.7	0.9
75	0.9	1.2	1.0	0.6	0.8
100	0.9	1.2	1.0	0.6	0.7
125	0.9	1.2	1.0	0.5	0.7
150	0.9	1.1	1.0	0.6	0.7
175	0.8	1.1	1.0	0.5	0.7
200	0.8	1.1	1.0	0.5	0.7

(c) Probit Method II

Sample size	All Mean (%)	GAM <5% Mean (%)	5-9% Mean (%)	10-14% Mean (%)	≥15% Mean (%)
25	0.8	0.8	0.9	0.9	0.7
50	0.8	0.6	0.8	0.9	0.8
75	0.7	0.5	0.7	0.8	0.9
100	0.7	0.5	0.7	0.8	0.9
125	0.7	0.5	0.7	0.8	0.9
150	0.7	0.5	0.7	0.8	0.9
175	0.7	0.5	0.7	0.8	0.9
200	0.7	0.5	0.7	0.8	0.9

(d) Classic Method

Sample size	All Mean (%)	SAM <5% Mean (%)	5-9% Mean (%)	10-14% Mean (%)	≥15% Mean (%)
25	1.0	0.5	0.9	1.3	1.3
50	0.5	0.4	0.6	0.6	0.4
75	0.3	0.3	0.4	0.3	0.2
100	0.2	0.3	0.3	0.2	0.0
125	0.2	0.2	0.2	0.1	0.1
150	0.1	0.2	0.1	0.1	0.0
175	0.1	0.1	0.1	0.1	0.0
200	0.1	0.1	0.1	0.0	0.0

(e) Probit Method I

Sample size	All Mean (%)	SAM <5% Mean (%)	5-9% Mean (%)	10-14% Mean (%)	≥15% Mean (%)
25	0.2	0.1	0.2	0.2	0.5
50	0.1	0.0	0.1	0.1	0.3
75	0.1	0.0	0.1	0.1	0.3
100	0.1	0.1	0.1	0.0	0.2
125	0.1	0.1	0.1	0.0	0.2
150	0.1	0.1	0.0	0.0	0.2
175	0.1	0.1	0.0	0.0	0.2
200	0.1	0.1	0.0	0.0	0.2

(f) Probit Method II

Sample size	All Mean (%)	SAM <5% Mean (%)	5-9% Mean (%)	10-14% Mean (%)	≥15% Mean (%)
25	0.3	0.1	0.2	0.4	0.6
50	0.2	0.0	0.1	0.3	0.5
75	0.1	0.0	0.0	0.2	0.5
100	0.1	0.1	0.0	0.2	0.4
125	0.1	0.1	0.0	0.1	0.4
150	0.1	-0.1	0.0	0.1	0.4
175	0.1	-0.1	0.0	0.1	0.4
200	0.1	-0.1	0.0	0.1	0.4

Table 2: Precision (Half 95% CI) of GAM (a, b, c) and SAM (d, e, f) estimates by GAM categories(based on MUAC) (%)

(a) Classic Method

	GAM				
Sample size	All Mean (%)	<5% Mean (%)	5-9 % Mean (%)	10-14 % Mean (%)	≥15% Mean (%)
25	14.2	11.7	13.6	14.9	16.5
50	9.3	7.3	8.6	10.1	11.7
75	7.4	5.5	6.8	8.2	9.7
100	6.4	4.6	5.8	7.2	8.6
125	5.8	4.0	5.2	6.5	7.9
150	5.3	3.6	4.8	6.0	7.4
175	5.0	3.3	4.5	5.6	6.9
200	4.7	3.1	4.2	5.3	6.6

(b) Probit Method I

	GAM				
Sample size	All Mean (%)	<5% Mean (%)	5-9 % Mean (%)	10-14 % Mean (%)	≥15% Mean (%)
25	7.9	4.9	7.0	9.1	11.4
50	6.2	3.9	5.6	7.1	8.7
75	5.5	3.6	5.1	6.3	7.6
100	5.1	3.4	4.7	5.8	6.9
125	4.9	3.3	4.6	5.5	6.4
150	4.7	3.2	4.4	5.3	6.1
175	4.6	3.2	4.3	5.2	5.9
200	4.5	3.1	4.3	5.1	5.7

(c) Probit Method II

	GAM				
Sample size	All Mean (%)	<5% Mean (%)	5-9 % Mean (%)	10-14 % Mean (%)	≥15% Mean (%)
25	9.9	6.6	9.2	11.2	13.0
50	6.7	4.2	6.1	7.7	9.0
75	5.4	3.3	4.9	6.2	7.4
100	4.6	2.9	4.2	5.4	6.4
125	4.2	2.5	3.8	4.8	5.7
150	3.8	2.3	3.5	4.4	5.2
175	3.5	2.1	3.2	4.1	4.9
200	3.3	2.0	3.0	3.8	4.6

(d) Classic Method

	SAM				
Sample size	All Mean (%)	<5% Mean (%)	5-9 % Mean (%)	10-14 % Mean (%)	≥15% Mean (%)
25	10.7	6.0	9.7	11.7	12.9
50	6.8	4.5	6.4	7.2	8.0
75	5.1	3.6	4.7	5.4	6.1
100	4.1	3.0	3.8	4.4	5.1
125	3.5	2.6	3.2	3.7	4.5
150	3.1	2.2	2.8	3.3	4.1
175	2.8	2.0	2.5	3.0	3.8
200	2.6	1.8	2.3	2.8	3.6

(e) Probit Method I

	SAM				
Sample size	All Mean (%)	<5% Mean (%)	5-9 % Mean (%)	10-14 % Mean (%)	≥15% Mean (%)
25	3.0	1.4	2.4	3.6	5.5
50	2.5	1.2	2.0	2.9	4.4
75	2.3	1.1	1.8	2.7	4.0
100	2.2	1.0	1.8	2.6	3.8
125	2.1	1.0	1.7	2.5	3.6
150	2.1	1.0	1.7	2.4	3.6
175	2.0	1.0	1.7	2.4	3.5
200	2.0	1.0	1.6	2.4	3.4

(f) Probit Method II

	SAM				
Sample size	All Mean (%)	<5% Mean (%)	5-9 % Mean (%)	10-14 % Mean (%)	≥15% Mean (%)
25	4.8	2.6	4.1	5.7	7.3
50	2.8	1.3	2.3	3.4	4.7
75	2.2	0.9	1.7	2.7	3.8
100	1.9	0.8	1.5	2.3	3.2
125	1.6	0.6	1.3	2.0	2.9
150	1.5	0.6	1.1	1.8	2.6
175	1.4	0.5	1.0	1.7	2.4
200	1.3	0.5	1.0	1.6	2.3

Table 3: Coverage of GAM (a, b, c) and SAM (d, e, f) estimates by GAM categories (based on MUAC) (%)

(a) Classic Method

Sample size	GAM All (%)	<5% (%)	5-9 (%)	10-14% (%)	≥15% (%)
25	83.7	56.9	82.7	94.4	97.8
50	93.9	80.6	95.4	97.9	97.9
75	96.5	90.2	97.8	97.8	97.6
100	97.3	94.2	98.2	97.5	98.0
125	97.8	96.3	98.2	97.9	98.3
150	98.1	97.8	98.2	98.0	98.4
175	98.4	98.2	98.4	98.3	98.6
200	98.6	98.6	98.6	98.4	98.8

(b) Probit Method I

Sample size	GAM All (%)	<5% (%)	5-9 (%)	10-14 (%)	≥15% (%)
25	92.1	89.7	92.1	93.0	93.0
50	91.2	88.4	91.4	92.3	92.2
75	90.4	87.1	90.6	91.9	90.9
100	89.8	86.8	90.2	90.9	90.2
125	89.4	86.3	89.7	90.4	89.6
150	88.9	86.1	89.5	89.5	88.9
175	88.8	86.0	89.4	89.8	88.5
200	88.4	85.6	89.2	89.1	87.9

(c) Probit Method II

Sample size	GAM All (%)	<5% (%)	5-9 (%)	10-14% (%)	≥15% (%)
25	94.5	93.7	94.3	94.9	95.5
50	93.6	92.7	93.2	93.8	95.0
75	93.1	91.4	93.0	93.7	94.1
100	92.5	90.8	92.2	93.2	93.8
125	91.8	89.9	91.6	92.4	93.6
150	91.0	88.6	90.7	91.8	93.0
175	90.4	88.1	90.0	91.4	92.2
200	89.9	87.5	89.5	90.8	91.9

(d) Classic Method

Sample size	SAM All (%)	<5% (%)	5-9 (%)	10-14% (%)	≥15% (%)
25	35.4	10.4	26.7	46.5	66.4
50	55.3	22.4	47.8	70.2	85.8
75	67.1	33.1	61.5	82.6	93.3
100	75.2	42.5	71.8	89.0	96.6
125	80.4	49.9	78.4	93.2	97.5
150	84.2	55.9	83.6	95.2	98.4
175	87.3	61.9	87.7	96.5	98.6
200	89.6	66.8	90.7	97.4	98.7

(e) Probit Method I

Sample size	SAM All (%)	<5% (%)	5-9% (%)	10-14% (%)	≥15% (%)
25	89.9	86.5	91.2	91.2	88.5
50	88.8	85.1	90.8	89.8	86.2
75	88.0	83.9	90.2	89.3	84.8
100	87.4	83.7	89.9	88.3	83.7
125	87.3	83.2	89.8	88.3	83.5
150	87.1	82.7	89.9	87.8	83.5
175	87.0	82.5	89.8	87.8	83.2
200	86.8	82.4	89.8	87.6	82.7

(f) Probit Method II

Sample size	SAM All (%)	<5% (%)	5-9 (%)	10-14% (%)	≥15% (%)
25	92.7	89.1	93.2	93.7	93.7
50	91.1	86.5	91.7	92.1	92.6
75	89.7	84.3	90.4	91.4	90.8
100	88.7	83.1	89.6	90.2	89.7
125	87.7	81.9	88.7	89.6	88.5
150	86.5	80.5	87.5	88.4	87.3
175	85.7	79.0	86.7	87.9	86.4
200	84.5	77.4	85.6	87.3	85.0

Supporting Table 3: **Bias, precision and coverage of the different methods by region**

Table 1: Bias of GAM (a, b, c) and SAM (d, e, f) estimates by region (based on MUAC) (%)

(a) Classic Method

GAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	0.7	0.7	0.8	0.7	1.3	0.7
50	0.2	0.1	0.2	0.2	0.4	0.1
75	0.0	0.0	0.0	0.1	0.2	0.1
100	0.0	0.0	0.0	0.0	0.0	0.0
125	0.0	0.0	0.0	0.0	0.0	0.1
150	0.0	0.0	0.0	0.0	0.0	0.0
175	0.0	0.0	0.0	0.0	0.0	0.1
200	0.0	0.0	0.0	0.0	0.0	0.0

(b) Probit Method I

GAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	1.2	1.3	0.8	1.1	0.0	1.4
50	1.0	1.1	0.6	0.9	-0.1	1.1
75	0.9	1.0	0.6	0.8	-0.2	1.1
100	0.9	1.0	0.5	0.8	-0.2	1.0
125	0.9	1.0	0.5	0.8	-0.2	1.0
150	0.9	1.0	0.5	0.8	-0.2	1.0
175	0.8	0.9	0.5	0.7	-0.2	1.0
200	0.8	0.9	0.5	0.7	-0.2	1.0

(c) Probit Method II

GAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	0.8	0.9	0.5	0.8	0.1	0.8
50	0.8	0.8	0.5	0.8	-0.1	0.8
75	0.7	0.8	0.4	0.7	-0.1	0.8
100	0.7	0.8	0.4	0.7	-0.1	0.8
125	0.7	0.8	0.4	0.7	-0.2	0.8
150	0.7	0.8	0.4	0.7	-0.1	0.8
175	0.7	0.8	0.4	0.7	-0.2	0.8
200	0.7	0.8	0.4	0.7	-0.2	0.8

(d) Classic Method

SAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	1.0	1.0	0.9	1.0	0.6	1.1
50	0.6	0.5	0.6	0.5	0.5	0.5
75	0.3	0.3	0.4	0.3	0.4	0.3
100	0.2	0.2	0.2	0.2	0.3	0.2
125	0.2	0.2	0.2	0.2	0.2	0.1
150	0.1	0.1	0.1	0.1	0.2	0.1
175	0.1	0.1	0.1	0.1	0.1	0.1
200	0.1	0.1	0.1	0.1	0.1	0.1

(e) Probit Method I

SAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	0.2	0.2	0.2	0.2	-0.2	0.5
50	0.1	0.1	0.1	0.1	-0.3	0.4
75	0.1	0.1	0.0	0.1	-0.3	0.4
100	0.1	0.1	0.0	0.0	-0.3	0.3
125	0.1	0.1	0.0	0.0	-0.2	0.3
150	0.1	0.1	0.0	0.0	-0.3	0.3
175	0.1	0.1	0.0	0.0	-0.3	0.3
200	0.1	0.1	0.0	0.0	-0.3	0.3

(f) Probit Method II

SAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	0.3	0.3	0.2	0.3	-0.1	0.4
50	0.2	0.2	0.2	0.2	-0.2	0.3
75	0.1	0.1	0.1	0.1	-0.2	0.3
100	0.1	0.1	0.1	0.1	-0.3	0.2
125	0.1	0.1	0.1	0.1	-0.2	0.2
150	0.1	0.1	0.1	0.1	-0.2	0.2
175	0.1	0.1	0.0	0.1	-0.3	0.2
200	0.1	0.1	0.0	0.1	-0.3	0.2

Table 2: Precision (Half 95% CI) of GAM (a, b, c) and SAM (d, e, f) estimates by region (based on MUAC) (%)

(a) Classic Method — GAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	14.2	14.7	14.5	12.1	13.5	15.1
50	9.3	9.5	9.2	8.6	7.9	9.8
75	7.4	7.5	7.3	7.1	6.0	7.8
100	6.4	6.5	6.2	6.2	5.0	6.7
125	5.8	5.8	5.5	5.5	4.4	6.0
150	5.3	5.4	5.0	5.2	4.0	5.7
175	5.0	5.0	4.7	4.7	3.7	5.2
200	4.7	4.8	4.4	4.5	3.5	4.9

(b) Probit Method I — GAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	7.9	8.1	7.6	7.4	3.7	8.6
50	6.2	6.4	6.0	5.7	2.7	6.8
75	5.5	5.7	5.4	5.0	2.3	6.0
100	5.1	5.3	5.0	4.6	2.0	5.5
125	4.9	5.1	4.8	4.3	1.9	5.3
150	4.7	4.9	4.7	4.2	1.7	5.1
175	4.6	4.8	4.6	4.0	1.7	4.9
200	4.5	4.7	4.5	3.9	1.6	4.8

(c) Probit Method II — GAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	9.9	10.0	9.5	9.5	6.9	10.5
50	6.7	6.8	6.3	6.4	4.3	7.2
75	5.4	5.5	5.1	5.2	3.5	5.8
100	4.6	4.7	4.4	4.5	3.0	5.0
125	4.1	4.2	3.9	4.0	2.6	4.5
150	3.8	3.9	3.6	3.6	2.4	4.1
175	3.5	3.6	3.3	3.4	2.2	3.8
200	3.3	3.4	3.1	3.1	2.0	3.5

(d) Classic Method — SAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	10.7	13.3	13.2	5.1	13.0	13.3
50	6.8	7.6	7.6	4.3	7.2	7.5
75	5.1	5.5	5.4	3.7	5.0	5.4
100	4.1	4.3	4.3	3.3	3.9	4.3
125	3.5	3.7	3.6	2.9	3.2	3.7
150	3.1	3.2	3.1	2.7	2.7	3.2
175	2.8	2.9	2.8	2.4	2.4	2.9
200	2.6	2.7	2.5	2.3	2.1	2.7

(e) Probit Method I — SAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	3.0	3.1	3.0	2.8	0.9	3.5
50	2.5	2.6	2.5	2.2	0.6	2.8
75	2.3	2.3	2.3	2.0	0.5	2.6
100	2.2	2.2	2.2	1.9	0.5	2.5
125	2.1	2.2	2.1	1.8	0.4	2.4
150	2.1	2.1	2.1	1.7	0.4	2.4
175	2.0	2.1	2.0	1.7	0.4	2.4
200	2.0	2.1	2.0	1.7	0.4	2.3

(f) Probit Method II — SAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	4.8	4.8	4.5	4.7	2.9	5.1
50	2.8	2.8	2.7	2.8	1.5	3.1
75	2.2	2.2	2.1	2.1	1.0	2.4
100	1.9	1.9	1.7	1.8	0.8	2.0
125	1.6	1.6	1.5	1.6	0.7	1.8
150	1.5	1.5	1.4	1.4	0.7	1.6
175	1.4	1.4	1.3	1.3	0.6	1.5
200	1.3	1.3	1.2	1.2	0.6	1.4

Table 3: Coverage of GAM (a, b, c) and SAM (d, e, f) estimates by region (based on MUAC) (%)

(a) Classic Method — GAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	83.7	84.6	83.3	81.3	61.6	86.1
50	93.9	94.5	94.1	92.3	83.5	94.1
75	96.5	96.8	96.6	95.4	91.9	96.3
100	97.3	97.4	97.6	96.9	96.1	97.1
125	97.8	98.0	97.9	97.3	97.1	97.3
150	98.1	98.3	97.7	97.5	97.9	97.6
175	98.4	98.5	98.2	98.1	98.1	98.2
200	98.6	98.7	98.4	98.4	98.5	98.4

(b) Probit Method I — GAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	92.1	92.1	91.7	91.7	91.2	93.1
50	91.2	91.2	90.5	91.3	91.8	92.3
75	90.4	90.4	89.5	90.7	89.4	91.3
100	89.8	89.9	88.4	90.1	89.2	90.2
125	89.4	89.4	88.0	89.5	87.0	90.2
150	88.9	89.0	87.7	88.6	88.2	89.5
175	88.8	88.9	87.4	89.0	88.5	89.1
200	88.4	88.7	86.7	88.6	86.7	88.1

(c) Probit Method II — GAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	94.5	94.7	94.4	94.2	94.5	94.3
50	93.6	93.6	93.0	93.7	93.9	93.7
75	93.1	93.2	92.8	93.2	92.4	93.0
100	92.5	92.6	92.0	92.5	92.6	91.9
125	91.8	91.9	91.4	91.7	91.5	91.9
150	91	91.1	90.6	91.0	90.4	90.1
175	90.4	90.6	89.7	90.7	89.9	89.2
200	89.9	90.2	89.0	90.2	89.2	88.2

(d) Classic Method — SAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	35.4	36.0	32.9	34.7	14.4	39.7
50	55.3	56.0	52.6	54.1	31.7	60.8
75	67.1	68.2	64.3	65.5	40.5	71.4
100	75.2	76.1	73.2	73.1	53.8	78.8
125	80.4	81.5	78.3	77.6	59.9	83.7
150	84.2	85.2	83.7	81.1	64.6	86.8
175	87.3	88.5	86.5	83.8	70.2	89.2
200	89.6	90.6	89.9	86.4	73.0	90.4

(e) Probit Method I — SAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	89.9	90.1	91.5	87.7	78.4	93.4
50	88.8	88.8	90.5	86.6	72.9	94.4
75	88.0	88.0	90.3	85.8	68.2	93.8
100	87.4	87.4	89.7	85.0	64.3	94.1
125	87.3	87.2	89.8	85.1	62.0	94.6
150	87.1	87.0	89.6	85.0	58.8	94.7
175	87.0	86.7	89.4	85.3	57.2	94.9
200	86.8	86.7	89.3	85.0	55.7	95.0

(f) Probit Method II — SAM

Sample size	All Mean (%)	EA Mean (%)	WA Mean (%)	CSA Mean (%)	C Mean (%)	A Mean (%)
25	92.7	92.9	94.0	91.1	85.9	93.2
50	91.1	91.2	92.8	89.5	85.2	92.2
75	89.7	89.7	91.5	88.4	81.6	91.6
100	88.7	88.6	90.8	87.3	80.7	90.3
125	87.7	87.5	90.4	86.4	80.9	89.6
150	86.5	86.4	89.1	85.0	79.8	88.1
175	85.7	85.3	89.1	84.6	78.2	87.3
200	84.5	84.2	87.7	83.6	74.7	86.4

EA: East Africa; WA: West Africa; CSA: Central and South Africa; C: Caribbean; A: Asia

Annex VII: Relative precision of the classic method and both PROBIT methods

Table 4: Relative precision of GAM estimates

Sample size	Classic method		PROBIT method I		PROBIT method II	
	Median	Mean	Median	Mean	Median	Mean
25	170.7	176.3	105.0	129.6	81.0	84.9
50	105.8	130.3	70.5	77.5	64.6	67.2
75	84.6	103.2	57.0	60.9	57.7	60.1
100	71.1	86.4	49.1	51.8	53.9	56.2
125	64.2	76.1	43.9	46.0	51.4	53.7
150	58.6	68.5	40.1	41.8	49.7	52.1
175	55.1	63.0	37.1	38.7	48.5	50.9
200	52.8	58.9	34.7	36.2	47.6	49.9

Table 5: Relative precision of SAM estimates

Sample size	Classic method		PROBIT method I		PROBIT method II	
	Median	Mean	Median	Mean	Median	Mean
25	323.3	272.4	254.1	881.7	152.6	163.5
50	352.8	260.1	149.5	178.7	127.2	134.5
75	187.5	239.5	115.3	128.6	117.8	124.2
100	189.7	221.5	97.7	105.5	113.2	119.0
125	191.1	204.2	85.9	91.6	110.2	115.7
150	143.7	191.1	77.6	82.1	108.2	113.6
175	138.0	179.8	71.5	75.3	106.9	112.1
200	138.3	170.0	66.7	70.1	105.8	111.0