# Navigating data quality

# Part 1: Basic of nutrition survey interpretation

## Introduction

**Slide 1:**

- A huge welcome to ENN's navigating data quality presentation series.
- My name is Eilise Brennen and I'm a nutritionist at ENN. I'm part of a team that has been developing a series of bite-size presentations to help nutrition practitioners and policymakers get to grips with some common topics regarding understanding nutrition data. In this series, we have selected three related topics that give us an overview of some of the important considerations when interpreting nutrition data, including survey and coverage data.

**Slide 2:**

- This series will comprise of three videos covering:
    1. Basics of nutrition survey interpretation
    2. Importance of understanding seasonality and its impact on nutrition outcomes
    3. Unpacking global estimates of severe acute malnutrition (SAM) treatment coverage
- I will be taking you through the first video of this series with my colleagues Phillip James and Stephanie Wrottesley taking you through topics two and three.
- So let's get straight into topic one, the basics of nutrition survey interpretation.

## Basics of nutrition survey interpretation

**Slide 3:**

- Nutritional status is a key indicator to monitor the overall welfare of a population.
- Nutrition surveys can inform programming and policymaking, by
    - monitoring changes in a population's nutritional status over time,
    - identifying at-risk groups or individuals,
    - helping inform when certain nutrition programmes should be initiated,
    - helping evaluate the nutritional impact of current nutrition programmes and
    - tracking progress towards nutrition and public health objectives.
- Subsequently, it is vital to be able to understand and correctly interpret nutrition survey data.
- In this video, I will take you through some of the basics of interpreting nutrition survey data, focusing on the difference between incidence and prevalence and how to interpret confidence intervals (C.I).
- This video will not cover survey design and methodologies, but further resources are provided in the final slide if you would like more information on these topics.

## Prevalence and incidence

**Slide 4:**

- It's important to understand the difference between incidence and prevalence, as cross-sectional nutrition surveys (which make up the majority of nutrition surveys) only collect prevalence data, which is reported as percentages or proportions.
- Point prevalence is the number of current cases e.g. the number of children who are wasted in a specific population at a particular time point.

- Or in other words, it provides a snapshot of your population of interest at a particular time point.
- Similarly, period prevalence is the number of current cases in a specific population but over a specified period.
- This could be on a different sample or the same sample and provides a 'snapshot' at different time points.
- While incidence is the number of new cases occurring in a specific population over a particular period of time.
- So, why is this important, and what does this mean when looking at wasting caseloads for example?
- I will explain this using a bathtub analogy.

**Slide 5:**

- So here is our bathtub, which represents our population of interest, which in this case is the number of children 6-59 months.
- The water in the tub represents the prevalence of wasted children 6-59 months.
- The water is drained away or the prevalence of wasting decreases when children are no longer wasted or they have died.
- Water can also be added to the tub via the tap – or new cases can be added to our population. This represents incidence.
- So, by only looking at the prevalence of wasting, you do not take into account the number of children that become wasted over time (the new cases) or those who recover, relapse or die over time.
- This is particularly important when estimating the burden of children who are wasted, as wasting is a relatively short-term condition and is affected by seasonality, which Philip will explain further in the next video.
- So, by only looking at the prevalence of wasting you underestimate the burden of children who are wasted, meaning policy makers and programme managers cannot reliably predict caseloads and resources needed for treatment.
- This is something Stephanie will pick up on in the third video and I will touch on in the next slide.

**Slide 6:**

- Recent studies such as this one highlights the extent to which prevalence estimates underestimate the burden of wasting.
- This study looked at severe wasting, but the same concept also applies to moderate wasting.
- The study analysed prevalence and incidence data from 352 sites in 20 countries and calculated country specific incidence correction factors.
- The incidence correction factor is used for practical calculations of the burden of wasting when incidence data is not available, for example, nutrition programmes needing to plan for wasting caseloads.
- It is a multiplication factor that converts the prevalence of wasting, into the burden of wasting over a period of time.
- The current globally recommended incidence correction factor for severe wasting is 1.6.

- This study's findings suggest that the overall burden of severe wasting could be underestimated by 4.6 times using prevalence data alone and by 1.8 times using the globally recommend incidence correction factor of 1.6.
- The country-specific incidence correction factors were also hugely varied.
- For example, estimates ranged from 1.3 in Niger to 30.1 in Burundi.
- This starkly highlights the limitations of only using prevalence data to estimate the burden of severe wasting.

## Sampling

**Slide 7:**

- Now let's move on to C.I.
- In order to understand C.I we first need to understand sampling and sampling error.
- It is not usually possible to conduct a survey of the entire population of interest, so instead, we take a sample of that population.
- However, our sample will never exactly match the characteristics of the overall population, with each sample being different.
- For example, the nutritionist, Maria, wants to know the proportion of children in her home town born low birth weight (LBW) (or the orange babies) in 2021.
- However, 1,000 children were born in Maria's home town in 2021, and Maria does not have the resources to collect data for all these infants, so she takes a sample of 100 infants.
- In Maria's home town, the true proportion of infants born LBW is 20%, but in Maria's sample, 25% of infants were born LBW.
- If Maria decided to take another sample of 100 infants, the proportion of infants born LBW would be different. For example, it could be 23% or 19%, as each sample is slightly different from the overall population by different amounts.
- The difference between the true population values and the values derived from the sample is called sampling error.

## Confidence intervals

**Slide 8:**

- We use C.I to calculate this uncertainty around a sample to make inferences about the true population from the sample population.
- A C.I is expressed as an upper and lower interval, and represents the range of values we would expect the true population values to fall between if we resampled from that population, within a certain level of confidence.
- A 95% C.I is most often used. But 90% and 99% are sometimes used.
- For a 95% C.I, if you repeatedly took a sample 100 times and calculated a sample proportion and a 95% C.I for each sample, we would expect the true population value to be found within 95% of these C.I's.

**Slide 9:**

- Let's look at an example, Mohammed carried out a study in an urban settlement in Kenya. He wants to find out what percentage of children 6-59 months are wasted, or in other words, the percentage of children with weight-for-height z scores below -2. Out of a population of 10000 children 6-59 months, Mohammed samples 1000 children and finds 5% of these children are wasted.

- In order to make inferences about the true percentage of wasted children, Mohammed calculates a 95% C.I.
- It is 3.7% to 6.5%
- So, we can say that we are 95% confident that the true percentage of wasted children is between 3.7% and 6.5% in this urban settlement in Kenya.
- However, it is always important to remember that we are still 5% confident that the C.I, 3.7% to 6.5% does not contain the true percentage of wasted children.

**Slide 10:**

- The confidence level, sample size and variability (the spread of the data) will affect the width of the C.I.
- A 99% CI will be wider than 95% CI for the same sample.
- The greater the variability or the greater the spread of the data then the wider the C.I.
- The larger the sample size, the narrower the C.I.

**Slide 11:**

- For example, Maria is now in Cox Bazar, Bangladesh, and also wants to find out the prevalence of wasted children. There were 205,000 children 6-59 months in Cox Bazar. Maria takes two samples.
- In the first sample she selects 14000 children.
- In the second sample she only selects 100 children out of the population of 205,000 children.
- In sample 1 the prevalence of wasted children is 12%.
- In sample 2 the prevalence of wasted children is 13%.
- As data was collected from a smaller number of children in sample 2, sample 2 will have a very wide C.I compared to sample 1.
- So Maria now calculates her C.I for each sample in order to make inference about the true prevalence of wasting in Cox Bazar.
- From sample 1, Maria can say that she is 95% confident that the true proportion of wasted children is between 11.5% and 12.6% in Cox Bazar.
- From sample 2, Maria can say that she is 95% confident that the true proportion of wasted children is between 7.1% and 21.2% in Cox Bazar.
- It is clear that sample 1 gives us a more precise picture of the prevalence of wasting in Cox Bazar compared to sample 2.

**Slide 12:**

- It is important to take C.I into account when comparing data or looking at changes over time.
- For example, the paper drought, conflict and child undernutrition in Ethiopia looks at the prevalence of wasting between 2000 and 2013.
- On the X axis you have the prevalence of wasting, while on the Y axis you have the survey year.
- The dots represent the prevalence of wasting in the sample population, while the horizontal lines on either side of the dots represents the 95% C.I.
- You will see that there is huge variability in the width of the C.I between 2000 and 2009 and 2010, with very narrow C.I or more precise estimates of the prevalence of wasting achieved in 2009 and 2010.

- This may be because of a variety of reasons, including increased survey quality and sample size.
- In the year 2000 over 11,000 people were surveyed while in the years 2009 and 2010 over 21,000 and nearly 20,000 people were surveyed respectively.
- Between 2000 and 2001 the prevalence of wasting went from approximately 18 to 10%. However, the 95% C.I overlap.
- In other words, the range where we are confident that the true prevalence of wasting lies, could contain some of the same numbers in both 2000 and 2001.
- Therefore, it is possible that the true prevalence of wasting decreased between 2000 and 2001, but it is also possible that there was no change in prevalence or that the drop in prevalence was not as dramatic as suggested by the prevalence figures for the sample population.
- As a general rule, we can be more confident saying that there is a true difference when the C.I do not overlap.
- For example, we can be more confident in saying there was a decrease in the prevalence of wasting between 2000 and 2012, as the C.I do not overlap.
- This is just one example of why it is so important for us as nutrition practitioners to not just compare means but also interpret C.I when comparing data.

## Conclusion

**Slide 13:**

- I'd like to thank our donors who generously supported us to make this presentation: Irish Aid and the Eleanor Crook Foundation.
- Thank you so much for tuning in and I hope this video helped unpack some of the considerations to take into account when you are next interpreting survey data. Do have a listen to the next two videos in the series to learn more about data considerations regarding seasonality and estimation of global coverage figures.
- We would really appreciate feedback on this presentation and any others you have watched in this series. On the MediaHub page where you found this video there is a link to a very short survey that will only take a couple of minutes to fill in. We would love to know how useful you have found these short presentations, what we could improve on, and your thoughts on future topics. We would be very grateful if you could make time to help us improve similar work going forwards.

**Slide 14:**

- To finish I will share these additional resources, where you can find more information on interpreting survey data. Thank you for your time.