MONITORING AND EVALUATION FRAMEWORK

for WASH Market-Based Humanitarian Programming

GUIDANCE DOCUMENT

ANNEX 4 – GUIDANCE FOR SURVEY DESIGN

OXFAM

ANNEX 4: GUIDANCE FOR SURVEY DESIGN

4.1 DIRECT AND INDIRECT BENEFICIARIES

INTRODUCTION

Establishing the **absolute number of beneficiaries** in crisis situations is challenging in comparison with determining the **proportion (e.g. %) of the population**[1] fulfilling a given criteria.

To establish a population proportion one can take a sample of the population and determine the proportion of the sample that fulfils a certain criteria. Then infer that with some margin of error, the same proportion is valid for the whole population. This means that determining the proportion of a population can be assessed without knowing the absolute number of people in the population. Moreover and contrary to what people often sense even the sample size is independent of the population size in relatively large population.

Absolute population figures are usually obtained through civil registration of vital events or vital statistics (See Wikipedia). The United Nations Population Fund (UNFPA) recommends these figures to be checked every 10 years by a census (See Wikipedia) in countries where vital statistics might be less reliable. Methods using remote sensing (Henderson & Xia 1997) or used from population biology (Bostoen et.al. 2007) might be used were such data is not available, not reliable or not relevant (due to e.g. a crisis). However, these methods are not always easy to implement and fall outside the scope of programme monitoring.

This all to show that finding absolute population or beneficiary figures is not a trivial matter. Because of the disproportionate cost and effort of getting accurate beneficiary numbers organisations rely often on estimates. Estimates lead often to large and contested figures in particular for secondary beneficiaries. To avoid that, in this document we suggest a relative simple and practical approach for estimating beneficiaries for programmes which include cash transfers.

DEFINING BENEFICIARIES

Programmes often distinguish between direct (or targeted) and indirect beneficiaries. These definition can change between projects and will also vary depending of project purpose. Definitions expressing programmatic ambitions often differ from the measurable definitions used for practical monitoring. This is to avoid that programmatic ambitions are limited to measurable targets.

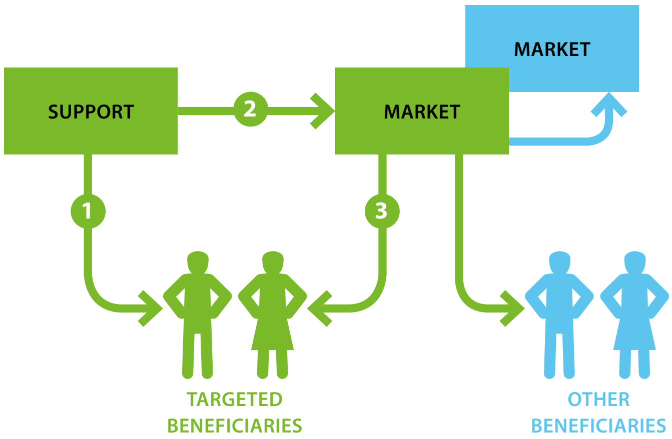
Direct beneficiaries in this document will be defined as those defined by programme as directly benefiting from project-funded activities, while **indirect beneficiaries** are those who also benefit as a result of improvements made to serve the direct beneficiaries. Although this classification may seem clear, different organisations can have different views regarding who is considered direct or indirect beneficiary.

A WASH installation can benefit a small number of users directly, but a market strengthening action directed towards a regulatory change (for example) could have a benefit to a much large number of people directly or indirectly.

The similar case is with MBPs because, while the activities are often at the market level and not directly towards the client within that market system, they are indirectly benefiting. MBP intervention aims at supporting the “traditional” primary (or targeted) beneficiary (as end user of the product or the activity - see flow ① in Figure 1 below), through activities that support the market. So it reaches end user indirectly through market support (flow ② and ③).

[1] Population here is used in its statistical sense as the union of all basic sampling units of interest which can be people, families, but also cars, institution or anything of interest.

Figure 1: Direct and indirect beneficiaries in MBPs



The **direct beneficiaries** remain the "traditional" primary beneficiary who need the products even though the flow of products is guaranteed through a market based approach as shown the flow ② in the figure above. While the market also receives this direct support, it should be seen also as means to provide goods to the targeted end-beneficiaries. Obviously the traders are also direct beneficiaries but their number will be smaller than the number of end user and so they can, in terms of numbers often be ignored.

The **indirect beneficiaries** are those that benefit from the project within the market system or even within the population, but are not directly targeted by the programme (drawn in orange color in the figure above). There are two effects related to complicate with this definition:

1 The mass-effect, best know from immunisation in which the whole population benefit from immunisation if the vaccination coverage is above a certain level. Market based programmes are based on the idea of a similar wider benefit, but is not clear yet if there is such a clear measurable effect as in vaccination.

2 The multiplier effect (See Wikipedia) or the factor that describes the volume or size of the indirect economic activities that are made possible due to the direct market support. These are based on a Keynesian consumption model.

Estimating secondary beneficiaries using these effects is challenging and more an academic activity. The method below describe a practical approach of estimating direct and indirect beneficiaries, which can be used for different situations.

METHOD

As this document covers projects with a cash transfer component for NFIs we will assume all the direct beneficiaries receive the cash transfers. The other people buying similar objects as covered by the cash transfer but not recipients of a cash transfer are considered indirect beneficiaries.

The way to measure this is to:

- Go to all or, if there too many, a randomly selected number of shops for some consecutive days after the cash transfer,
- Register each person that buys a NFI which was part of the WASH basket used to determine the size of the cash transfer.
- Register for each of the people buying whether they received a cash transfer.
- Calculate the ratio of indirect beneficiaries to the number of beneficiaries.

A calculated example:

Data is collected from seven shops (shops 1-7) for three days (Day 1–3) as shown in the table below. For each day, each purchase of a WASH NFI' included in the programme is noted down and the fact that the buyer is included or not ('in' or 'out') of the cash transfer programme. The data can be collected by a surveyor or the trader him- or herself. The data can be grouped in the way as shown below.

Table 1: Data as collected in seven shops over three days

	Day 1		Day 2		Day 3		**Sub totals**	
CT prg.	in	out	in	out	in	out	in	out
Shop 1	9	31	12	23	12	17	33	71
Shop 2	7	25	15	26	13	21	35	72
Shop 3	13	18	9	30	11	24	33	72
Shop 4	14	18	15	25	10	29	39	72
Shop 5	15	18	17	30	9	33	41	81
Shop 6	7	29	17	17	13	18	37	64
Shop 7	8	31	17	27	14	24	39	82
Totals	73	170	102	178	82	166	**257**	**514**

For the calculation the line totals for 'in' and 'out' are calculated by adding the day values together. The sum of the line totals are then added together to obtain the totals over the three days and the seven shops. In the example it is 771 (257+514) shoppers bought WASH related NFIs part of the basket of supported products. Of the 771 roughly one third was part of the programme while two third was not, or for each person in the programme there are two beneficiaries (who also use the supported store) that are not part of the programme.

Imagine that the programme does cash disbursement of 3257 Households with and average household size of 4.6 people.

The direct beneficiaries are:	3257 X 4.6 = 14,982 people.
The indirect beneficiaries are:	14982 X (514 /257) = 29,964 people
The total number of beneficiaries is:	14982 + 29964 = 44,946 people

4.2 SAMPLING METHODS

This Annex outlines the possible sampling design and sampling methodology to be employed. Whilst it is important to use the same indicators in the various surveys so they are comparable, it is not necessary that identical sampling methods are used. What is important is that sample taken is **representative** for the overall population[2]. When a sample is taken and analysed the conclusions for the sample are assumed for the whole population it represented. This process is called statistical inference. The steps in sections below explain some of the possible methods to determine sample sizes.

DETERMINE BASIC SAMPLING UNIT AND THE TARGET POPULATION[3]

For the generic indicator households surveys are used which makes households the smallest unit of interest or the basic sampling unit[4]. To monitor performance over time in a comparative way the overall population and the population groups need to be clearly defined. For example when talking about an urban area it often not clear where the urban area stop and the peri-urban or rural areas starts. For comparison over time it is important that the same populations are used, which can be done by using streets, rivers and other physical boundaries to clearly delimit the area of interest.

SELECTING SAMPLE DESIGN

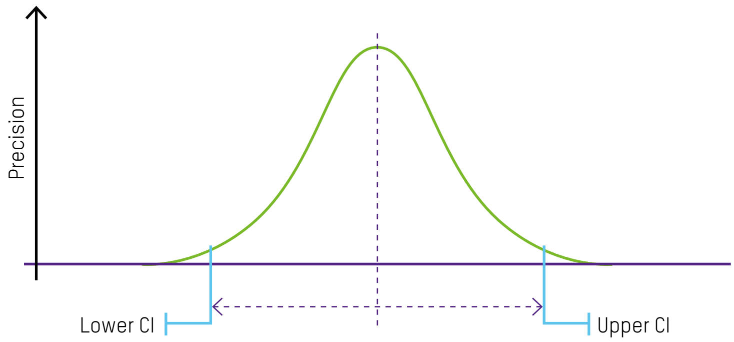
To explain why sample **size** is not the most crucial aspect in obtaining a representative sample we explain below the difference between accuracy and precision in statistics.

[2] Population is used in its statistical sense of the grouping of all the basic sampling units which for UWSS are mainly households
[3] Population is here used in its statistical sense of the group of all basic sampling units
[4] To calculate population figures it is good to collect the population size as well in the survey

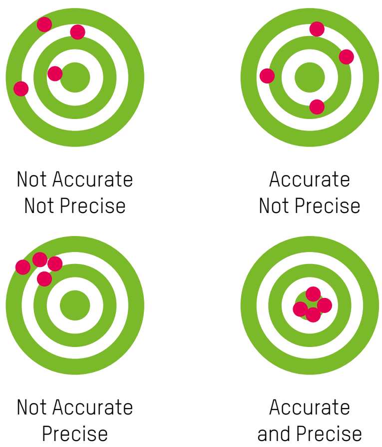
ACCURACY VERSUS PRECISION

The accuracy of a value is the degree to which the result of the measurement, calculation, or specification, conforms to the correct value or standard. In this case it means, for example, seeing how good the true water coverage is in comparison to the coverage measured in a survey. However in our survey the true accuracy can not be measured, but we can determine how well the accuracy is likely to be.

Precision is the extent to which we would obtain the same result if we repeated our measure as shown in Figure above. Precision is expressed in confidence intervals (CI), which give the probability of the measured value as shown below.



Ideally one seeks an accurate and precise estimate. Contrary to popular belief, small confidence intervals are no guarantee of an accurate estimate as is shown below. One can have small confidence intervals for an inaccurate measure.



While precision can be calculated from the dataset based on the sampling strategy, accuracy can not be calculated.

In short, accuracy is determined by how representative the sample is for the whole population, or how likely every person or household could have been selected. This is solely determined by the way the data is collected. Precision relates to the sample size and the sample design.

A simple example: If you have a bathroom scale which does not measure your correct weight but each time you stand on it, it displays the same weight, your measure is precise, but not accurate.

To put it simply:

Data collection process → Accuracy
Sample size → Precision

SIMPLE RANDOM SAMPLING (SRS)

Simple Random Sampling (SRS) is the basis of all probability sampling. Each member of the population has an equal and known chance of being selected. This minimises bias and simplifies analysis of results. The variance or uncertainty between individual results within the sample is a good indicator of variance in the overall population, which makes it relatively easy to estimate the accuracy of results. When there are very large populations, it is often difficult or impossible to identify every member of the population to ensure an equal and known probability of selection, so the pool of selected subjects risks becoming biased.

To obtain a simple random household sample a list of households has to be made and from this list a number of households randomly selected. There are various formulas for calculating the required sample size. These formulas require knowledge of the variance, proportion of the measure of interest in the population and the maximum acceptable error. To avoid having to use (and understand) these formulas Krejcie & Morgan (1970)[5] put the values in a table. The confidence level of 95%, used very commonly in research, is sufficient. For programmes that want to achieve a substantial change a degree of precision of 10% will suffice. When change is little a lower percentage or higher precision might be required. As a rule of thumb take a precision no lower than half of the change you expect in your programme. For instance, if the programme expect that 20% or more people will take up a improved sanitation take 20% / 2 = 10% as your degree of precision. In the table the sample size for a population of 10,000 and precision of 10% is 95.

[5] Tables are made for finite population and proportional errors

Required Sample Size

Confidence = 95%

Population size	Degree of precision or margin of error			
	10%	5%	2.5%	1%
10	9	10	10	10
20	17	19	20	20
30	23	28	29	30
50	33	44	48	50
75	42	63	72	74
100	49	80	94	99
150	59	108	137	148
200	65	132	177	196
250	70	152	215	244
300	73	169	251	291
400	78	196	318	384
500	81	217	377	475
600	83	234	432	565
700	85	248	481	653
800	86	260	526	739
900	87	269	568	823
1,000	88	278	606	906
1,200	89	291	674	1067
1,500	90	306	759	1297
2,000	92	322	869	1655
2,500	93	333	952	1984
3,500	93	346	1068	2565
5,000	94	357	1176	3288
7,500	95	365	1275	4211
10,000	95	370	1332	4899
25,000	96	378	1448	6939
50,000	96	381	1491	8056
75,000	96	382	1506	8514
100,000	96	383	1513	8762
250,000	96	384	1527	9248
500,000	96	384	1532	9423
1,000,000	96	384	1534	9512
2,500,000	96	384	1536	9567
10,000,000	96	384	1536	9594
100,000,000	96	384	1537	9603
264,000,000	96	384	1537	9603

Source: monitoring(4)change 2015, adapted from Krejcie & Morga, 1970.

Example for the calculation using the table above for a given population:

In an area with an estimated 13,783 people and 3,838 household a household survey is planned. The survey serves as a baseline to measure an increase in the number of households with access to critical WASH service. The ambition is to increase the number of households with access from 20% to 25% points. The sampling unit for a possible survey will be the household. This means that the population[6][size for the example is 3,838 households. In the table we look at the first column with population sizes and find either 3500 or 5000. The minimum improvement expected is 20%, which divided in two as a rule of thumb makes 10% precision.

Looking in the table we can see that for a population of 3500 and a precision of 10% the sample should be 93 while for a population of 5000 and a precision of 10% the sample size is 94. From the two figures take the highest as the sample size to be selected.

4.3 LIKERT-TYPE SCALES

A Likert-type scale is a psychometric scale commonly used for scaling responses in survey questionnaires. It is often used interchangeably with rating scale, even though the two are not synonymous. The scale is named after its inventor, psychologist Rensis Likert and uses a format in which responses are scored along a range as means of capturing variations. When responding to a Likert item, respondents specify their level of agreement or disagreement on a symmetric agree-disagree scale for a series of statements. Thus, the range captures the intensity of their feelings for a given item and helps to convert qualitative information into quantitative.

While the scale is ordinal not each step can be considered of the same value so it is difficult to give values to each step as was often done in the past.

For the WASH MBP M&E framework we consider this method for several indicators.

WITHIN A COMPOSITE INDICATOR

If the indicator has three conditions that need fulfilling e.g.:

1 CONDITION A

☐	■	☐	☐	☐
Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree

2 CONDITION B

☐	☐	■	☐	☐
Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree

3 CONDITION C

☐	☐	☐	■	☐
Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree

The overall response is the lowest most right answer of the three questions

[6] Population is used here as defined in statistical terms as the count of all basic sampling units

PERCENTAGES OF MULTIPLE ANSWERS TO ONE INDICATOR

Calculation example.

Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
73 HH	32 HH	42 HH	15 HH	27 HH
$\frac{73}{73+32+42+15+27}$ =39%	$\frac{32}{73+32+42+15+27}$ =17%	$\frac{42}{73+32+42+1}$ =22%	$\frac{15}{73+32+42+15+27}$ =8%	$\frac{27}{73+32+42+15+27}$ =14%
If the base line was as below				
37%	13%	25%	8%	17%
The difference between the follow up measurement and the baseline becomes				
39-37=+2%	17-13=+4%	22-25= -3%	8-8=0%	14-17=-3%
Total of Agree		Neither	Total of Disagree	
+6%		-3%	-3%	

MEDIAN ANSWER OF MULTIPLE ANSWERS

The median is the value separating the higher half of a series of values from the lower half. In simple terms, it may be thought of as the "middle" value of a data set. So if the indicator is collected at three household that provide a reply then the middle category would be the value "Neither Agree or Disagree" as household A has one value higher and household C has one value lower.

HOUSEHOLD A

☐ Strongly Agree	■ Agree	☐ Neither Agree or Disagree	☐ Disagree	☐ Strongly Disagree

HOUSEHOLD B

☐ Strongly Agree	☐ Agree	■ Neither Agree or Disagree	☐ Disagree	☐ Strongly Disagree

HOUSEHOLD C

☐ Strongly Agree	☐ Agree	☐ Neither Agree or Disagree	■ Disagree	☐ Strongly Disagree

If an extra household D would have a value as below:

HOUSEHOLD D

☐ Strongly Agree	☐ Agree	☐ Neither Agree or Disagree	■ Disagree	☐ Strongly Disagree

The middle value could be either "Neither Agree or Disagree" or "Disagree" as both could be considered middle values. For the WASH M&E framework the lowest (most to the right) value will be taken in such cases, so the median becomes "Disagree"

When there are many values as in the case of the measurement above we look in which the 50% value falls. So using the same example we get …

Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
73 HH	32 HH	42 HH	15 HH	27 HH
39%	17%	22%	8%	14%
0–39%	39–56%	56–78%	78–86%	86–100%

The middle value or 50% value is in the agree category so the median value is "Agree".

For the baseline used above the values are:

Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
37%	13%	25%	8%	17%
0–37%	37–50%	50–75%	75–83%	83–100%

Again the 50% value is the middle value but in case of doubt between "Agree" or "Neither Agree or Disagree" we choose by convention the lowest value so in this case the median value goes from "Neither Agree or Disagree" in the baseline to "Agree" in a follow up measurement.

Commissioned by **Oxfam**

Produced by

FlowNet Sustainable Water Solutions

monitoring [4] ch∆nge

Funding from

USAID FROM THE AMERICAN PEOPLE

OXFAM

Oxfam GB, Oxfam House,
John Smith Drive, Cowley,
Oxford OX4 2JY, UK