



Health Outcomes Tool (HOT)

Theoretical framework and manager's user guide

HelpAge

International

HelpAge International is a global network of organisations promoting the right of all older people to lead dignified, healthy and secure lives.

Health Outcomes Tool (HOT) Theoretical framework and manager's user guide

Published by HelpAge International
PO Box 70156, London WC1A 9GB, UK

Tel +44 (0)20 7278 7778

info@helpage.org

www.helpage.org

Registered charity no. 288180

This guide is the product of a three-year process of developing and validating the Health Outcomes Tool (HOT). HelpAge International would like to thank Pfizer International and HelpAge USA for supporting the development process and implementation of the tool in several countries around the world.

The guide is based on field experience, piloting of the tool, and training events. We would like to thank the following institutions and individuals who have contributed to this work: University Javeriana/Institute of Ageing in Colombia; consultant Mauricio Chumacero; Fundación Horizontes and Suma PUNCHAY in Bolivia; the HelpAge Tanzania team; Prakash Tyagi and GRAVIS team in India; and project staff from the Better Health project. Thanks also to Giorgi Pkhakadze, Rachel Albone and Kristin Bodiford for valuable inputs and to Paul Ong, for initiating the development of this tool.

Authors; Jonna Bertfelt and Catherine Dusseau/ HelpAge International

Front on front page; Kristin Bodiford/ HelpAge US

Back photo by Jonas Wresch/ HelpAge International



@HelpAge



HelpAge International

Copyright © HelpAge International 2016
This work is licensed under a Creative Commons Attribution-NonCommercial
4.0 International License, <https://creativecommons.org/licenses/by-nc/4.0>

ISBN 9781-1-910743-19-5

Contents

04	1. Background
04	1.1 What is the Health Outcomes Tool (HOT)?
04	1.2 Why was HOT developed?
05	1.3 HelpAge's global strategy 2020 and the corporate indicators for health and care
05	1.4 Theory of change (ToC) and the HOT
07	1.5 What is healthy ageing and how do we measure it?
09	2. Theoretical framework
09	2.1 Definitions
10	2.2 Mixed methods approach
10	2.3 Scoring methods
11	2.4 The HOT: structure and contents
15	2.5 The validation process
16	3. How to implement the HOT
16	3.1 Evaluation design and sampling
21	3.2 Piloting the HOT in new contexts
23	3.3 Planning the training and survey work
25	4. Data analysis
25	4.1 HOT Database and analysis
29	Annex 1: Timeline of validation process
31	Annex 2: Sample size calculator
32	Annex 3: Decision tree for sampling strategy
37	Annex 4: Training schedule
39	Annex 5: Test
40	Annex 6: Test answers

1. Background

1.1. What is the Health Outcomes Tool (HOT)?

The Health Outcomes Tool is a monitoring and evaluation (M&E) tool to be used in health and care programmes for older people at community level. By collecting data on selected health and care indicators the HOT can be used to assess the current health status of older women and men **as they perceive it** as well as measure changes in health status and wellbeing over time.

By using this tool to monitor and evaluate all our health and care programmes globally, we can monitor progress and evaluate impact on older people's health and wellbeing in and across countries and use the aggregated data to follow changes at the regional and global levels.

The HOT is a tool that can be used to monitor and evaluate any health and care projects implemented by HelpAge or our partners. The HOT package includes 4 parts;

1. The HOT basic questionnaire
2. The HOT theoretical framework and manager's user guide (this document)
3. The HOT user's guide (for field staff and data collectors)
4. The HOT database, a data entry and data analysis tool (in Excel)

(Subsequent versions of the package will include a qualitative data collection tool along with guidelines for its implementation.)

1.2. Why was the HOT developed?

Before HelpAge decided to embark on developing our own M&E tool a review of existing and internationally used tools was done. Some of the tools reviewed (e.g. the World Health Organization Quality of Life (WHOQOL) tool, EQ-5D™ and Easy Care) are well-known tools and some of them are already being used across the HelpAge network, while others were new to us.

In order to assess whether any existing tools could meet the organisation's M&E needs, a list of criteria was developed. On the top of this list was **ability to measure healthy ageing** as well as **ability to measure change against HelpAge's corporate indicators**. The following criteria were also listed:

- Able to produce evidence that can feed into HelpAge's policy work
- Easy to implement
- Short data collection instrument
- Suitable for evaluating programmes/interventions in low- and middle-income countries
- Balance between brevity of instrument and a design that enables data collection and case-tracking capabilities
- Work as both evaluation and monitoring tool
- Amenable to quick modification and adaptation to a variety of cultural and social contexts
- Useful for providing a snapshot of current situation and perceived health status (assessment)
- Data collected can be aggregated to higher levels

However, when reviewed against the list of criteria above, no existing

tool¹ was able to fulfil all (or most) of our requirements. Some tools are too focused on a specific area (e.g. a disease) while others are too vague (tries to cover too much). Some tools are too basic, meaning that they do not include enough or appropriate “domains” to cover our wide variety of projects. Some tools are very comprehensive and therefore very time-consuming and costly to administer, and often requiring trained research staff and long implementation schedules (e.g. the WHO-SAGE).

Additionally, many of the tools reviewed were originally developed for research rather than for M&E. Another common problem is that many tools are designed for use in high income countries and for self-administration, which is difficult in settings characterised by low levels of education and literacy.

1.3. HelpAge’s Global Strategy 2020 and the corporate indicators for health and care

Over the past decade, it has become more and more important for non-government organisations (NGOs) to be able to show progress and present evidence that their interventions are achieving the intended outcomes. Thus, HelpAge wants to be able to demonstrate achievements (outcomes and/or impact) and efficiency in our health and care work. Furthermore, the need to monitor and evaluate our work is a clear demand from project participants and donors (accountability), as well as from a wide range of other agencies (credibility), particularly if we are to influence policy.

In this context, HelpAge decided to develop corporate indicators to assess and chart our own progress. These are the outcomes that the organisation has committed to working towards. The corporate indicators were developed in line with HelpAge’s Global Strategy 2020 (which is based on our theory of change – see below). Therefore, making progress on these indicators means working to deliver the HelpAge strategy. This means that every project implemented by HelpAge should contribute to and report on progress against the corporate indicators.

HelpAge’s corporate outcome for health is:

“More older people will report better health and wellbeing.”

And the corporate outcome for care is:

“More older people living in isolation, or with chronic conditions and frailty, will be supported to live well in a place of their choice.”

1.4. Theory of change (ToC) and the HOT

It is sometimes difficult to admit that our hard work in training, advocacy, education, assistance or support isn’t having the impact for older people that we had hoped. In addition, donors², authorities and international communities are increasingly requiring that NGOs demonstrate the impact of their implemented activities. There are many ways of doing this, but one common approach is to base the work of the organisation on a theory of change (ToC).

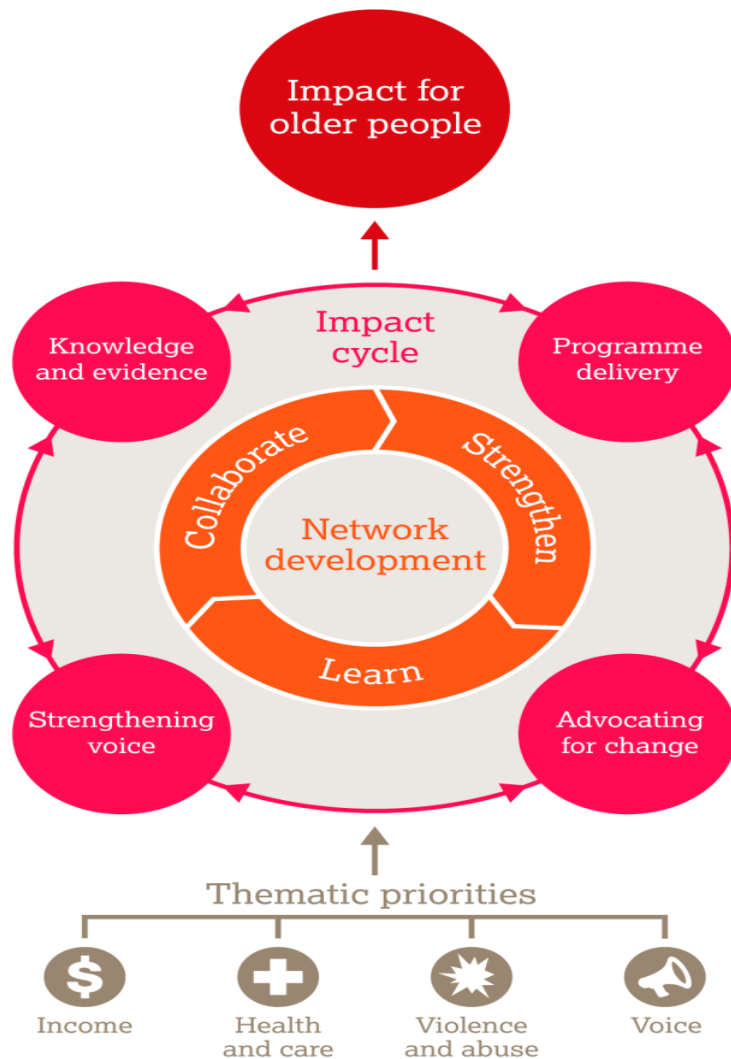
A ToC helps the organisation think through the outputs and outcomes that need to be delivered in order for our work to have an impact. It is the product of a series of critical-thinking exercises that provide a

1. List of tools reviewed: EQ-5D, WHO QOL battery, CAFOD battery tools, Easy Care, QOL AD and carers/families tests, PROMIS, Camberwell Assessments of need for the elderly, FACE, HART-Handicap Assessment and resources tool, interRAI, outcomes and assessment information set-OASIS.

comprehensive picture of the early and intermediate changes in a given community that are needed to deliver long-term impact. The planning cycle is a key step in defining indicators for each precondition that will be used to assess the performance of interventions.³ As the HOT has been developed to show progress on our corporate indicators (developed with the Global Strategy and our ToC in mind), using the tool to evaluate our health and care programmes is an excellent way of measuring progress towards our Global Strategy as well as making us accountable for the work we do.

HelpAge’s theory of change (see Figure 1) is an attempt to describe, in simple language, what we do, why we do it, and how it contributes to achieving our vision of a world in which all older people can lead dignified, healthy and secure lives.

Figure 1: HelpAge’s theory of change



2. The UK Department for International Development (DFID) has produced useful guidance on using a theory of change approach: <https://www.gov.uk/government/news/dfid-research-review-of-the-use-of-theory-of-change-in-international-development>

3. There are numerous resources on theories of change: <https://www.theoryofchange.org/>; see also: <http://blogs.lse.ac.uk/jsrp/2013/05/28/reflections-on-theories-of-change-in-international-development/> <http://learningforsustainability.net/evaluation/theoryofchange.php>;

1.5. What is healthy ageing and how do we measure it?

Healthy ageing can be defined as the process of maintaining the functional ability that enables wellbeing in older age. Even within a “natural” ageing process, we will often see a decline in different functions. We might expect that, as life expectancy increases, there are an equivalent amount of years gained that can be lived in good health. However, this is not the case. Instead, for every year of life gained only 0.8 years of healthy life is gained.⁴ Health among an older population is a unique and somewhat complex concept. Consequently, many of the existing tools and indicators that are used to measure health and health outcomes are not applicable when we want to measure health and wellbeing among older people. Below is a list of indicators commonly used in public health M&E and research, with a brief explanation of why each of these is not applicable in measuring health outcomes among older people.

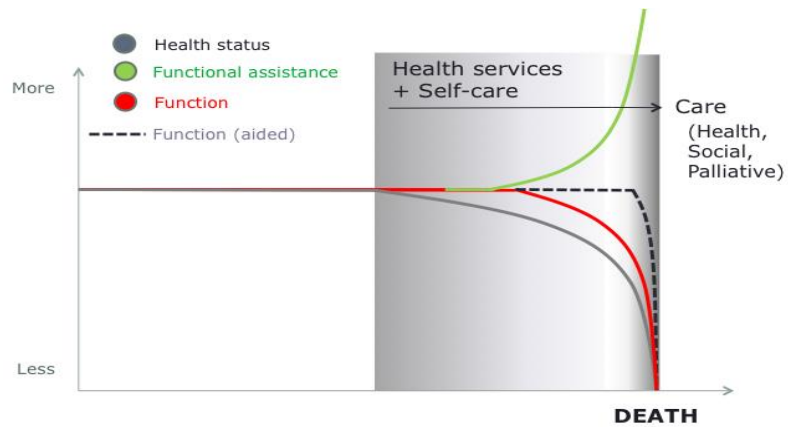
- **Mortality** (death) and **morbidity** (ill-health) alone are not, in general, good indicators for quality of life. Mortality and morbidity are crude ways of measuring health, not able to capture nuances. These indicators work better for younger populations but don't tell us much about healthy ageing.
- **Life expectancy** and **healthy life expectancy** (or healthy years lost to disability) are sound and comprehensive indicators that are used at policy, national and international levels. However, they are too long-term for the work that we are doing and for the kind of programmes that we usually want to evaluate. Change will not be measurable in these indicators after only 6 months or one year.
- Although many older people are affected by illness, specific measures of **prevalence** or **incidence** of disease (e.g. non-communicable diseases) are too focused or limited for the wide range of issues we work on. Moreover, healthy ageing is not restricted to having (or not having) a disease, according to the definition and theory in the WHO's new healthy ageing framework.
- **Quality of life** alone is too unspecific.

For people in older age, we need to include a wider range of criteria to evaluate health and wellbeing, as we can't expect health outcomes to keep improving. Health in older age is not about absolute improvements in health. In Figure 2, the black dotted line represents the aspirations we have with our work – for older people to achieve a level of functioning (assisted or not) that is being maintained for as long as possible. This is what we should be measured as a key outcome – whether our health and care programmes in developing countries are helping older people to maximise and maintain, for as long as possible, the functional ability and resilience that enables wellbeing in older age.

Measuring functionality is therefore key in assessing health of older people. It is critical as it emphasises the capabilities of older people instead of stressing continuous improvements in absolute health, which can be hard to achieve in older age. Enabling the “beings and doings” that older people treasure is the cornerstone of a function-oriented approach to health, in which quality of life is given priority over mere survival. This focus transcends cultural and geographic contexts we work in as well as the different types of projects we implement.

4. World Health Organization (WHO) new framework; see WHO, World report on healthy ageing, October 2015; www.who.int/ageing/publications/world-report-2015/en/

Figure 2: A function-oriented approach to health in older age



Furthermore, assessments of health are often medical and obtained by examinations conducted by health workers or through laboratory tests. But, by focusing on individuals' own perceptions of their wellbeing, the HOT provides a new perspective on ill-health among older people. For example, it is well-known and understood that diabetes involves poor body regulation of blood glucose, but the effect of the illness on a person's perception of their health, social relationships, working capacity, and financial status has received little attention in research.

2. Theoretical framework

The HOT has been developed with the aim of tracking and understanding the health status and quality of life of older people. It is a planning, monitoring, evaluation and learning tool. In this part of the guide, we discuss the theoretical framework on which the HOT is built. By explaining the theory, rationale and thinking behind the tool, we hope it will be easier to understand and implement effectively.

2.1. Definitions

Monitoring

Monitoring is the continuous process of assessing activities and results in relation to pre-set targets and objectives. It provides management and other stakeholders with information about ongoing interventions, the progress they are making towards achieving objectives, and how allocated funds are being spent.⁵

Evaluation

An evaluation is an exercise to systematically and objectively assess whether a project is meeting its objectives. It is developed to address specific questions or challenges so that all stakeholders can learn about the inputs, outputs, outcomes, impacts or processes associated with an intervention (project).

Indicator

An indicator is a quantitative or qualitative variable that provides a reliable means of measuring achievement, enabling stakeholders to reflect on the changes connected to an intervention.

Tool

A tool is a way of collecting information (data) that can be analysed to show progress against the selected indicators. Many tools use surveys or questionnaires. The HOT questionnaire is a tool to collect data that can be analysed at the project level but also at a higher (aggregate) level.

Results-based approach

The results-based approach to programme development and accountability is used by United Nations agencies and other bodies, including the Organisation for Economic Co-operation and Development (OECD).⁶ It is a project cycle approach, whereby planning and implementation (with integrated monitoring) are followed by evaluation, so that evidence and lessons learnt can be applied to plan future work. The concept of accountability is key in this approach.

A results-based approach assumes that we are accountable for our work and for the changes it brings. In order to be accountable, we **need to measure changes**, which the HOT is designed to do. It can, in an objective way, measure change (evaluate outcomes) of our programmes. Additionally, as we start using the HOT globally (across our programmes), it means that we can compare the work we are doing in different countries and come up with recommendations for the best solutions, based on internally produced evidence.

5. Stefan Molund Göran Schill, Looking back, moving forward, Sida Evaluation Manual, 2nd revised edition, Sida, 2007

6. United Nations, Results-based management handbook, United Nations Development Group, 2011.

2.2. Mixed methods approach

HelpAge's health and care indicators;

Percentage of older men and women in ACTIVE HelpAge projects reporting a better perception of their health

Percentage of older men and women in ACTIVE HelpAge projects reporting a better satisfaction with their life/wellbeing

Thus, as you understand from the way our corporate indicators are stated, any tool we use to collect information for these indicators should be able to produce quantitative data. The purpose is to provide data/basic metrics over time and measure change in key outcomes. Although the corporate indicators are expressed in percentages, they are based on people's perceptions and opinions – **i.e. we have to quantify subjective information.**

A scoring method has been found to be an effective way to transfer perceptions into quantitative data, and to capture changes over time. Several institutions such as the World Health Organization (WHO) and the Centre for Disease Control and Prevention (CDC) are using scales for recording respondent's answers – for instance, in the WHO Quality of Life (QOL) tool,^{7,8} the Self-Perceived Quality of Life (SPQL), and the Patient-Reported Outcomes Measurement Information System (PROMIS).⁹

In addition recoding the respondent's answers (as a score on a scale) the HOT questionnaire also allows the interviewer to leave a comment about the situation, to add important information to complement the answer given. This additional qualitative information is not part of the main HOT analysis but is extremely valuable in project evaluation and for future planning.

2.3. Scoring methods

Studying qualitative concepts such as people's opinions can be challenging. As our corporate indicators and the HOT tool is a quantitative tool **we have to translate qualitative information into quantitative data to record it**, a process that has to be a carefully guided to remain as objective as possible.¹⁰ One common method used to quantify people's perceptions is by a scale and a scoring technique, which is considered a quantitative method. This technique is commonly used in research (such as social sciences and marketing) especially when the aim is to assess people's perceptions or opinions.

To measure, is an attempt to identify the quantity, capacity or degree of something. *Measure* is formally defined as the act of assigning symbols or numbers to something according to a specific set of rules. Measurements can be categorised by the type of information that is communicated by the symbols or numbers assigned to the variables of interest.

Some accepted and widely used scoring methods use category scales, like the Likert Scale¹² can limit data analysis.¹³ Scales with categories, such as the Likert Scale are often not precise enough to identify smaller changes, as the answers could stay in the same category even though a small change has actually taken place. To be able to measure change, especially smaller changes, the ratio scale¹⁴ is often a better option.

7. Measuring quality of life, Division of mental health, WHO/MSA/MNH/PSF/97.4

8. World Health Organization. 'The World Health Organization Quality of Life assessment (WHOQOL), position paper from the World Health Organization'. Social Science & Medicine, 41 (10), 1995, pp. 1403–1409

9. PROMIS is a National Institutes of Health (NIH) Roadmap initiative designed to develop an electronic system to collect self-reported HRQOL data from diverse populations of individuals with a variety of chronic diseases and demographic characteristics. Currently, HHS monitors HRQOL in the United States by administering selected PROMIS and other HRQOL items on the Behavioral Risk Factors Surveillance System (BRFSS), the National Health and Nutrition Examination Survey (NHANES), and the National Health Interview Survey (NHIS). It is promoted by CDC (Centre for Disease Control and Prevention) and Healthy People 2020 initiative: It is based on the EQ-5D

10. Les Éditions de la Chenelière inc., 2006, Savoir plus: outils et méthodes de travail intellectuel, 2e éd. 2 (Raymond Robert Tremblay et Yvan Perrier)

11. Tim Unwin, presentation on 'Quantifying the subjective', www.gg.rhul.ac.uk/ict4d/M2LP2.pdf

12. McLeod SA, 'Likert Scale', 2008, www.simplypsychology.org/likert-scale.html

13. Acervo_bibliotecologia_escalas_Escala de Likert.pdf

14. WHO. 'The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization', Social Science & Medicine, 41(10), 1995, pp. 1403–1409

The number of points /categories in a scale should depend on:

- the analysis needed
- the level of sensitivity and size of the expected change.¹¹

Taking all of the above into account, we have chosen a scoring method for the HOT. A scale in form of a line, with numbering from 0 to 10 (or 0 to 100). So, for each question, respondents are asked to mark their position on the scoring line.¹⁵

One weakness of a 'scoring method' is that it is common to obtain many answers around the middle point. However, choosing the middle category means 'moderate agreement', not an indecisive answer. Indecisiveness, on the other hand, usually generates a more positive score.¹⁶ Additionally, it has to be taken into consideration that, in HOT, a respondent's choice is the result of a thought process that might end up in a moderate choice, particularly if the issue analysed is broad and complex.

2.4. The HOT: structure and content

Understanding "health"

The World Health Organization (WHO) defines health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity".¹⁷

However, health is not an end but a means to a much higher purpose such as **quality of life**, which WHO¹⁸ defines as: "An individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns."

Given that 'health' is such a complex concept, it is not surprising that there are differences in how we understand this concept (and this word). Our understanding could depend on our geographic and cultural context, educational background, age and gender etc. (See box below, with a case study from piloting of the HOT in India). Researchers agree that QOL (quality of life) and HRQOL (health related quality of life) are multidimensional and using only one question to find out people's perceptions of their health or life satisfaction might be too general to properly assess their situation. Thus, we need more information about several domains (areas) of what influences 'health' to draw a more complete picture.

15. Note that PROMIS is a basis for some of the HOT questions as well as the 10cm line. For a good review of data collection methods based on self-reporting of wellbeing, see www.cdc.gov/hrqol/wellbeing.htm

16. <http://www.southalabama.edu/coe/bset/johnson/lectures/lec5.pdf>

17. Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946.

18. See note 8

Case study: Difficulty in defining health, wellbeing and life satisfaction

In a HOT training session in India, trainees were asked to define some of the key concepts used in the questionnaire. The lists below give examples of the definitions they gave of “to be healthy” and “life satisfaction”.

The results show that even if there are some common meanings attached to “health”/“being healthy” (eg, “lack of diseases”), each person gives a little twist to the definition. For instance, “being able to do my things” or “feeling wellbeing”. In this case, most participants shared a similar background and socioeconomic status. Even then, the definitions they gave were not exactly the same. Here are some examples of trainers’ answers;

What is health?

- Health is good if life is good. A long life is related to health, good food is good health
- She/he can do their work, own routine without any support
- She/he doesn’t suffers from any seasonal/chronic disease
- Health is defined by how the individuals daily routine and the ability to carry out this routine
- If there is a serious problem like breathing problem, then we can say that this person has health issues
- Physical state of a person is health but a person can be healthy in many ways like physically, mentally...
- Ideally happy (mentally) while being free from sickness
- Health will be if we can do our daily work with so our basic needs covered
- Good health means: food, medicine, basic needs are met. Able to do their work on their own and daily.
- Behaviour[tasks and daily routines] is the same as it has always been
- Physical, mental, social and spiritual wellbeing.
- Free from health conditions along with physical abilities, functionality, social, mental and spiritual wellbeing
- No big chronic disease

What is life satisfaction?

- A happy life is being healthy
- Good at all levels: family, health, being happy, basic needs fulfilled
- In old age it is a normal condition to have some health problem affecting our daily life but how we manage these problems is what is important
- Quality of life is healthy life in every manner[way]. If a person is healthy physically and mentally, she/he will be having satisfaction all the ways
- We need each other every day at the village. Happy is living with family and having an income
- Living with a family is good for the person
- To have [life] satisfaction: happiness, functionality, social connections and status
- I have a job or work, my earning, family and no health problem

The use of domains

The HOT is collecting data in four domains which represent important parts of what health can be¹⁹ (see Figure 3, below). All domains are inter-related and all of them have an impact on the overall outcome, health. Separately they show us how we are performing in different areas connected to health and wellbeing.

In addition, there are many factors within and around our programmes that might affect the results of an evaluation. People might have conditions that our programmes are not designed to address – such as chronic illness or pain – which could affect the overall results. Our programmes might contribute to achieving some but not all of the intended outcomes, and external factors could counterbalance any overall positive change. How do we then make our positive impact visible? One way is to break down “health” into smaller pieces (domains) and collect data on all of them.

Figure 3: The four domains measured by the Health Outcomes Tool



These four domains are included in the HOT because:

- they relate to a person’s perception of their health and wellbeing
- they relate to the key issue of functionality and care/assistance
- they are common ground between all HelpAge projects and our 2020 Global Strategy
- they will become more common in our programmes in the future

The HOT questionnaire

The HOT questionnaire was developed to be:

- as short as possible
- easy for project staff to implement without needing research expertise
- applicable in all geographic and cultural contexts.

The HOT evaluation is not aiming at collecting data around every aspect of older people’s lives, instead we have identified the areas that are important to learn about in order to understand older peoples health and wellbeing. By keeping this focus we have developed a short core questionnaire of around 35 questions (depending on skips).

19. As an example, health-related quality of life (HRQOL) is a multidimensional concept that includes domains related to physical, mental, emotional, and social functioning and focuses on the impact health status has on quality of life. A 10-item global HRQOL scale was developed to assess selected physical and mental health symptoms, including functioning and

The HOT questionnaire is structured as follows:

1. Introduction and consent
2. Basic information and demographics
3. Two questions on health perception and life satisfaction
4. Eighteen questions and follow-up questions around the four domains:
 - o functionality
 - o dependency
 - o perception of services
 - o self-care
5. One objective test

Basic information and demographics

As well as collecting data on health, wellbeing and the four health-related domains (functionality, dependency, perception of services and self-care), the HOT questionnaire also collects some basic information and demographic data. This means that we can report disaggregated data (which is a requirement of our Sex, Age, Disability Disaggregated (SADDD) Data policy, see Table 1 below). It also allows us to better analyse the impact of our interventions in different sub-groups of the target population.

"The table below represents the minimum data disaggregation that is expected. It covers internal policy and the requirements of DFID [UK Department for International Development] and ECHO (European Commission Humanitarian Aid & Civil Protection). USAID [United States Agency for International Development] /OFDA [Office of U.S. Foreign Disaster Assistance] also requires gender and age analysis but do not specify age cohorts." HelpAge SADDD policy, January 2015

Table 1: HelpAge SADDD policy

Several donors take 50+ as one age group. This is not sufficient for internal purposes. The 80+ group should be further disaggregated if appropriate (80-89 etc.)

Age	Women	Men	Other	Total	Persons with disabilities *
>5					
5-17					
18-49					
50-59					
60-69					
70-79					
80+					
Total					

**Where possible, disability data should be disaggregated by age and sex. This may be difficult at an early stage in projects due to lack of data but an overall percentage is the absolute minimum requirement.*

Data on age and gender

Data on age and gender should always be collected in any data collection by HelpAge. Data on age is collected as full years and data on gender is collected as male, female or other.

Data on poverty

By collecting data on poverty, we will be able to analyse and disaggregate our results (in a crude way) based on the level of poverty. Furthermore, these data are important to us as we try to understand how we are targeting our programmes and how poverty is interacting with the changes seen in the outcome indicators.

Globally there have been longstanding discussions about how to measure level of poverty. The methods developed (by different organizations and researchers) are all based on different philosophies, definitions of poverty, and methods of measuring. Some of the well-known poverty measures include the Multidimensional Poverty Assessment Tool (MPAT), the Multiple Indicator Cluster Survey (MICS), the Progress out of Poverty Index (PPI), and the Participatory Poverty Assessment (PPA). However, no method is perfect and the all have their pros and cons.

For the HOT, we have decided to base our data collection around poverty on the 'basic needs approach', which means that we want to understand if the respondent has access to basic needs (defined as shelter/housing, food, safe water and being able to keep a good hygiene) and the method for recording these data is through a 'counting approach'. By using this approach, we can find out which basic needs respondents can or cannot access.²⁰ This creates an index which gives a 'poverty score'.

Data on disability

The HOT questionnaire collects basic prevalence data on the following disabilities: impaired eyesight (Question C11a), impaired hearing (Question C11b), and communication difficulties (Question C11c). Data on memory problems (Question Q5) as well as physical disabilities (Question Q4) are collected as outcome indicators.

Furthermore, the questionnaire includes a question (to be answered by the interviewer), asking: 'In your opinion, does the respondent have a disability or impairment that could affect his/her answers in this interview?' We do not intend to exclude any older person from being surveyed but want to keep this answer in mind when analysing the data as it could explain outliers, incompleteness or other illogical errors in the data.

2.5. The validation process

The first iteration of the HOT tool was developed and used in June 2012. Since then, the tool has undergone some changes and an extensive validation process. This has included testing its **reliability** (*Are measurements stable in different environments?*) and **validity** (*Is the tool actually measuring what we want it to?*). This process has led to various iterations of the tool. The validation process has taken place on three continents (Latin America, Africa and Asia), and in eight countries (see Annex 1).

The validation process has been important in making sure that the questionnaire and the data collection methods are as non-intrusive as possible, thereby making the HOT as user-friendly as possible. However, we are aware that the questionnaire will be received differently in different contexts, it is therefore advisable to conduct pre-testing before rolling out the questionnaire in a new context (for more on this, see the next section).



20. Alkire S, Foster JE, Seth S, Santos ME, Roche JM and Ballon P, Multidimensional poverty measurement and analysis: Chapter 4. 'Counting approaches: definitions, origins, and implementations', Oxford, Oxford University Press, 2015

3. How to implement the HOT

If you are managing the implementation of a HOT evaluation, this is the most important part of this guide. It explains how to plan for and set up the evaluation in the best possible way in order to get credible results.

Tip! As you work through this section, it will be helpful to have access to the concept note and logframe for the programme that you will be evaluating.

3.1 Evaluation design and sampling

Deciding on an **evaluation design, sample size, sampling strategy** and **replacement strategy** is not always easy, but it is very important as the evaluation design and sampling will determine the quality of your evaluation (see box below).

The following section will guide you in how to decide what evaluation design and sampling strategy to use. First a short technical discussion around each of the four topics is provided, following are options/strategies you can choose in order to put together a good evaluation design that suits your context.

Why does sampling matter?

When we want to evaluate a programme for **all** older people in district X, we want the results to be representative of the entire **target group** of the intervention (i.e. **all** older people in the district) – not just the people that have been interviewed. This is why we need to think about how we design and sample.

Remember! Without proper sampling, the results will **only** be valid for the population we are interviewing.

3.1.1 Design

There are many ways to design project evaluations using quantitative methods. They span from very rigorous designs (such as randomised controlled trials (RCTs), the golden standard) to less rigorous designs, which make it harder to generalise findings. Choosing the right and most appropriate design can be difficult, and is usually determined by the resources available (such as time and money). Simply put: More rigorous (better) design is often more expensive.²¹ For a HOT evaluation, we suggest one of the following designs;

1. Post-test only design

- Gives you an understanding of the current (snapshot of) situation
- Not very rigorous. Should **ONLY** be used if no baseline data was collected.
- In this design you only collect data (i.e. interview respondents) at one point in time – after the intervention/ programme has been implemented. However, if you want, you could use a 'recall method' to collect baseline data in order to make comparisons (i.e. ask the same question twice: "how it was before?" and "how it is now?").

2. Pre-post test design

Preferred evaluation design for HOT evaluations.

- Evaluates change over time in outcomes. (I.e. change in outcomes between data collections)
- Good value for money and good quality
- When using this design you interview the same sample on two or more occasions – before the programme (baseline) and after the programme have been implemented. To make sure that results are on track, one or more *midlines* can also be added. This design allows you to compare your results before and after the intervention - you will be able to show change in outcomes.

3. Pre-post test design with control group

- Best rigour but expensive
- In this design you interview the sample on two or more occasions (before and after the programme has been implemented) but can also compare your results with a control (or 'non-intervention') group. This is a group that has not received the intervention but which has the same characteristics as your target group. Hence, if your programme has been effective, you would expect your target group to show better results in outcome indicators than your control group.
- With this design it is easier to claim that any change you see in outcome indicators between baseline and endline is actually due to the programme rather than due to natural processes or interventions by others.

Note that (for the last two options) you need to decide how many data collections to conduct. In some cases, this is part of the contract/agreement with the donor. If the evaluation aims to show what effects the project has had overall, two data collections (baseline and endline) are usually enough. To show the greatest effect, you need to make sure that the baseline data collection is done **before** or very early in the programme implementation. If you want to use the HOT to monitor progress made during the intervention, you might want to add one or more *midline* data collections, which can be useful for programmes that span over several years.

3.1.2. Sample size

Sampling is the process of selecting units (in this case, people) from the population of interest, as we usually do not have the resources to interview everyone who has taken part in the programme. Deciding on the right *sample size* and *sampling strategy* are very important steps in planning the evaluation. Correct sampling is important because we need to be confident that the results are representative of the target population. However, remember that a HOT evaluation will, always, only be representative of our target group – not the general population.

Calculating the sample size/ number of people to interview

The most important point to make about the sample size is that it should be decided through calculation, not estimation. Using a sample size calculator (see below) will help you, although you will still need to input some key information for the calculations to come out correct.

21. 'Types of evaluation designs', Measurement, Learning & Evaluation Project for the Urban Reproductive Health Initiative, www.urbanreproductivehealth.org/toolkits/measuring-success/types-evaluation-designs#Non-experimental

Calculating your sample size

- **Population size**

The first thing you need to figure out is how big your population size is (the programme's target population). The sample frame is the total population that you intend to attribute your results to. To help you identify this number, refer to the project's concept note – it should include an estimate of the target population.

- **Margin of error (confidence interval)**

No sample will be perfect, so you need to decide how much error to allow. For instance, if a result is 70%, the real value of the indicator is actually somewhere between 65% and 75%, with a margin of error of +/- 5%. **(5% is commonly used and also recommended for HOT evaluations).**

- **Sample size**

Once you know your population size and confidence interval, you can calculate your sample size. You can do this by using a sample size calculator (see links below) or by using the table in **Annex 2**.

- **Over-sampling - response rate**

Not every person selected will agree to be interviewed, so you might end up with a smaller sample than expected. To cater for this, over-sampling at baseline is good practice. The response rate is the ratio of respondents that accepts to be surveyed compared to the total number of respondents invited to participate in the survey. For instance, if you ask 400 people and only 200 accept to complete the interview, your response rate is 50%.²²

Previous HOT evaluations haven't had many refusals, but you should take into account that some people might not be present at the agreed time or place. In addition, as we want to follow the same people during the entire project (which could be a couple of years) we should take into account mortality rates for over 60s (which differ in different contexts). Keeping all this in mind, oversampling of 10% is advisable.

Sample size calculators

Qualtrics, www.qualtrics.com/blog/determining-sample-size/

Creative Research Systems, www.surveysystem.com/sscalc.htm

Calculator.net, www.calculator.net/sample-size-calculator.html

3.1.3. Sampling strategy

As we would like the results from the sample to be valid for the whole target population, the most accurate strategy for sampling is to use a **probability sampling** (i.e. every person in the target population has the same, or a known, chance of being chosen to be part of the evaluation). You use probability sampling to minimise the risk of a biased sample – to make sure that as wide a variety of people as possible are represented. If your aim is not to generalise your findings to the whole target population, or if this way of sampling is not possible, **non-probability sampling** is another option.²³

Whichever sampling strategy you choose, you need to follow it through for the duration of the evaluation. Once you have completed the data collection and analysed the data, your results will always be looked at in the light of your sample and sampling strategy.

Good results will mean nothing if the sampling wasn't done correctly, while *mediocre or bad* results in an evaluation where sampling was correct are much more useful.

Below is a list of sampling methods (not exhaustive) that are useful in a HOT evaluation. The method you choose should be based on what fits your context best. It will probably be determined by the nature of the project being evaluated, how big the target population is and the sample size needed as well as the resources available.

Every time a HOT evaluation is implemented, the situation will look different. It is therefore difficult to decide on one way of drawing a sample that can be used in every situation. However, we envisage that most projects will fall under one of the following three scenarios:

Sampling Strategy 1: Randomized and stratified sampling

- When? If target population is well-known (often smaller samples)

Sampling Strategy 2: Cluster and random sampling

- When? If target population is less well-known (often bigger samples)

Sampling Strategy 3: Sampling through screening and convenience sampling

- When? If samples has a specific characteristic (e.g. older persons with a specific disease) OR when information about the sample is limited.

The decision tree and detailed notes on each of the suggested sampling strategies in **Annex 3** will help you make a decision of which sampling strategy to use.

Sampling for a control group

For a pre-post test design with a control group, you also have to choose the right control group and apply the same method for sampling. For the comparison between the two groups to be meaningful, they should be as similar as possible. By similar we mean in terms of age, gender, level of poverty, geographic location, or anything else that might affect the outcome indicators. If you're working with a specific group in a community (for example, people with NCDs), the control group can be selected from the same community, but only if the community is sufficiently large. If the programme is working with all (or most) people in a community, the control group should be selected from a similar community that hasn't benefited from the intervention.

3.1.4 Replacement strategy

As the HOT methodology is designed as a pre-post test study, the same sample (same respondents) should be surveyed at every data collection. This also allows us to follow up on specific cases.

Depending on how you decide to analyse the data you collect, you might want to replace respondents that are not available at a subsequent round of data collection. The basic HOT data analysis looks at two things: change in means, and percentage of people reporting an improved situation (scoring higher) in any of the outcome indicators. If you are interested in the former, we recommend that you replace any respondents that are lost to follow up, whether at midline or endline. This is so that you have a sufficient sample at all data collections.

22. 'How to determine population and survey sample size?' Blog, Check Market, www.checkmarket.com/2013/02/how-to-estimate-your-population-and-survey-sample-size/

23. Better Evaluation, <http://betterevaluation.org/plan/describe/sample>

If a selected respondent cannot participate in an interview (at baseline or at follow-up), make sure to replace this person with another person that has similar characteristics. Try to find someone with as many of the same characteristics as possible. Mentioned in the list below are some of the characteristics to take into account when you make replacements.

Any replacement respondent should be:

- a beneficiary of the programme being evaluated (if not in the control group)
- the same gender as the person who is not available for interview
- the same age group as the person who is not available for interview
- living in the same geographical location as the person who is not available for interview.
- have the same programme specific characteristics, if any

3.2 Piloting the HOT in new contexts

The HOT has been tested and validated in different countries, cultures and contexts. We have, through this process, developed a tool that is short, non-intrusive, and easy to implement in most contexts. However, you should always pre-test the tool (test the questionnaire in the local community), particularly in new contexts. As a minimum, you should pre-test it with the following points in mind.

Translation to vernacular languages

If the tool needs to be translated into the local language or languages, make sure to do this before you pre-test it and make sure that translations are done by a professional translator. We think of the questions in the HOT questionnaire as *concepts* that should be explored. In the English version of the survey, the questions have been phrased in a way that makes most sense when conducting the survey in English. However, when translating the survey, do not translate it word by word. (Make sure the translator has access to the table in Annex 1 in the accompanying User's Guide.) This table provides detailed guidance on how to interpret (and therefore translate) the questions, words and phrases used in the HOT questionnaire.

The tool should also be back-translated by a different person as a way of cross-checking that the questionnaire has been translated correctly. When resources are limited, back-translations are often not prioritised. However, this is a crucial step in the adaption of the tool and should not be neglected. One option could be to back-translate around 15% of the tool and check the error rate. If it's high, this would suggest that more work is needed on the translations. This work requires time, so the translation should be done well in advance of the data collection beginning, and always before pre-testing.

The scoring method

The scoring method can be new to many people, both interviewers and respondents. During the development of the tool we have, at times, encountered situations where respondents found it difficult to understand the scoring method. However, we have learnt that this can be solved by training interviewers well. When pre-testing the tool in your local context, you will get a feel for whether the scoring technique is easily understood by respondents, or not. If the scoring method is complicated for respondents and interviewers alike, then you need to think about how you can address this during the training

for the data collection team – practise explaining the scoring method for respondents and share experiences on who to do it well.

The scale / 'scoring line' can be explained and visually presented in different ways. During the pre-test, find out which why of presenting it people seem to understand most readily. The scoring line has been presented as a 10cm line on a piece of paper (works better for literate people) or as a 1 metre ruler (seems to work better for illiterate people). Other ideas include using a vertical line (something similar to a thermometer). You can find out more about this in the User's Guide (pages 16-23).

Length of the survey

To help you with planning, it's important to find out how long it takes to complete an interview. You also need to let respondents know how much time you are asking them to give. In general, the HOT interview takes between 20 and 40 minutes, but this can differ depending on the situation or if you have added some questions to the questionnaire.

Cultural sensitivity

The survey has been developed to be as non-intrusive as possible, but problems can arise when the HOT is implemented in a new cultural context. You should therefore keep cultural sensitivity in mind when you are pre-testing the questionnaire. A good way to handle potentially upsetting questions is to train interviewers in building good rapport with respondents, or adding scripts to the questionnaire that clearly explain, at the start, what kind of questions that will be asked and why. If some questions in the survey are unacceptable in the local context, they might need to be removed or changed.

Pre-test of additional questions

If you have decided to add some questions to the HOT questionnaire, these also need to be pre-tested, keeping all the above-mentioned considerations in mind.

Adapting the HOT questionnaire

The basic set of questions in the HOT questionnaire is aimed at collecting data on and measuring change in HelpAge's corporate indicators. However, a HOT evaluation could be an opportunity to collect additional data when it is needed, and if capacity is available. You shouldn't remove or change any question in the basic version of the questionnaire but you can add new questions. The list below outlines possible ideas;

- Find out more about an existing domain, e.g. add more questions on NCDs or risk factors for NCDs
- Include additional objective tests or screening for nutritional status
- Increase the precision or detail, e.g. adding aspects of access to services, like distance and financial barriers.

Recommendations for changing questions in the questionnaire

When designing a questionnaire, it's easy to think of additional questions you could ask to collect more data of interest to your programme. What is harder is deciding which questions are not needed, as a questionnaire can easily become too long. Long questionnaires lead to interviews that are too long and generate poor-quality data. Our advice is: **Think about how you will use the data in advance. What analysis do you want to make?**

If you do add new questions, use a similar format to the existing questions so as not to confuse respondents.

3.3 Planning the training and survey work

Planning data collection

Start by deciding how many data collections you want to conduct and when they will take place. Before you use the HOT for the first time, allow sufficient time for translation, adaptation and pre-testing. Before **every** round of data collection, you need to allow for enough time to hire (or identify) and train the data collection team. After every data collection round, you need to allocate time for data entry, analysis and report writing.

Identifying the data collection team

The quality of your data will very much depend on how good your data collection team is and how well they are trained in using the HOT questionnaire. In addition to interviewers, the team will need supervisors (we recommend one for every eight interviewers), data entry staff, and translator(s) as well as back-translator(s) (see box below for example).

Interviewers will obviously play a key role, so choose your team members wisely.

Qualities of a good interviewer:

- **Literacy** Interviewers need to know how to read, write and make basic calculations.
- **Languages skills** Interviewers should be able to speak the language/s the interviews will be conducted in, or know how to work with a translator
- **Local knowledge** Interviewers should have a good understanding of the local context.
- **Interpersonal and communication skills** Interviewers must be able to relate well to other people and their situation. Effective communication involves active listening to what respondents are expressing, verbally and non-verbally.
- **Honesty** This is important for building rapport with respondents and for gaining their trust.²⁴

Training the data collection team

Although the HOT is not a complicated tool, it is extremely important that it is well understood by the data collection team, and training plays a key part in this. The HOT documentation provides the basis for the training. You can find supporting materials in the Annexes to this guide. They include a template for a three-day training schedule (**Annex 4**) and a test to check that trainees have understood the tool (**Annexes 5 and 6**). Additionally, all interviewers should be given a hard copy of the *HOT User Guide*, which should be referred to during the training. The interviewers should be allowed to keep their copy until the end of the data collection.

Planning data collection in the field

The length of the data collection round is, of course, entirely dependent on the sample size, distances to cover, logistics, and the size of your data collection team. Listed below are some parameters you should take into consideration:

- Ideally, each interview should be conducted by a pair of interviewers. This is partly for security reasons but also because they can divide up the tasks: one person asks the questions (engages in the conversation) while the other one takes notes.

- Six interviews per day is a good average.
- We recommend that you have one supervisor for eight interviewers (or four pairs).
- Allow some time for replacement interviews or tracking of respondents.

Example of staff and time needed for a round of data collection

Sample size: 400

Number of interviewers: 8 (4 pairs)

Number of supervisors: 1

Number of surveys/day/interview pair: 6

Tracking days: 10%

$(400/(4*6))*1.1 = 18.3 \rightarrow 19$ working days

Confidentiality and tracking numbers

You must ensure that information received from respondents is kept confidential. This means that interviewers are not allowed to share information they receive in interviews. The identities of respondents should be kept confidential at all times.

You can ensure confidentiality by using tracking numbers instead of respondents' names in the questionnaires and in the database. A tracking number is assigned to every person in the sample and used for the whole duration of the evaluation, from baseline to endline. Replacements should be given a different tracking number.

Tracking numbers should be entered and stored in the master list (in the database). This list is the **only** place where names are linked to numbers and it should only be accessible to the project manager(s) or other authorised personnel. The master list should never be shared internally or externally, except for rational operative reasons. The completed questionnaires should **never** be shared externally.

Donors, universities or national ethical review boards sometimes require that protocols describe how confidentiality will be insured. This is not always the case for M&E activities but should still be worked on and taken seriously.

How tracking numbers are generated will vary depending on the sampling strategy. If you have a list of the names of the people in the sample, tracking numbers should be generated before data collection starts. However, be sure to plan how to assign new tracking numbers to replacement respondents. If you don't have a list of names before you start, the tracking numbers have to be created as the data collection is taking place.

Research approvals / approval from local authorities

Before you start a HOT evaluation, contact the local authorities to find out if you need any research approvals to conduct this kind of M&E activity. If you do, keep in mind that the process of acquiring one may be lengthy, and there might be additional rules around confidentiality and safety of respondents that you need to adhere to.

4. Data analysis

4.1 The HOT database and analysis tool

To facilitate data analysis, a user-friendly HOT database (in Excel) has been developed. It can be used to enter and store data, and automatically analyses any change in relation to the HelpAge corporate indicators. (Most statistical packages are compatible with Excel and any package can be used for analysing HOT data, if you opt for not using the HOT database.) How you conduct the analysis will depend on what you want to know and the knowledge and skills available in the project team.

The HOT database has the following components:

- A master list. A list of all the respondents in the evaluation and their assigned tracking numbers.
- A data entry format with automatic logical checks to prevent mistakes being made in data entry.
- A database. A spread sheet that contains and stores all data, from all rounds of data collection.
- Automatic basic analysis:
 - A profile of your sample by data collection round, with tables and graphs
 - Statistics on the most important outcome indicators (cross-sectional analysis)
 - Change analysis. Statistics of the change in outcome indicators between different rounds of data collection

Measuring change

The main objective of the HOT is to measure changes in perception in the sample between data collections. We would hope to see positive changes after our programmes has been implemented (see box).

Example of positive change: a respondent's score across three different data collections

Question; 'Overall, how would you rate your health during the past 3 months?'

0  100
Baseline Score: 38

0  100
Midline Score: 44

0  100
Endline Score: 56

The person in the example above has experienced an improvement in his/her general health status between data collections.

The database will use **all** respondents answers to calculate:

- The average score by round and how it has changed. (E.g. The mean score in question 1 in data collection 1 was 45 and increased to 57 in data collection 2.
- The number of people who have expressed positive change between data collections

By conducting this kind of analysis we will be able to report progress against our corporate indicators and most of the outcome indicators we use in our programmes. The change in people's perception can also be negative or not change at all.

Three possible situations: (each colour represents one person's answers)

1. Positive change (green)
2. Negative change (blue)
3. No change (red)

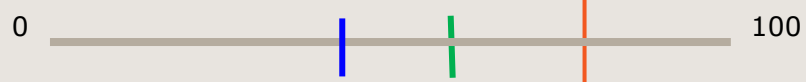
Baseline:



Midline



Endline



Outcome indicator in the automatic data analysis

The following outcome indicators are automatically calculated and analysed in the HOT database. It is however possible to use the HOT data to do any additional analysis that might be of interest for the project, or is a requirement for donor reporting.

- Q1. General health
- Q2. Satisfaction
- Q3a. Social and/or daily activities
- Q3b. Work activities
- Q4. Mobility
- Q5b. Impact of memory problem
- Q6. Social and/or daily activities
- Q7. Access care/help easily
- Q10. Access to health care
- Q11. Quality of health care
- Q12. Affordability of health care
- Q14. Actions taken for self-care

Limitations of data from the HOT

Due to limitations of resources and capacity, our sampling methods and sample size will not make our results representative of the entire older population (national level). We need to be prudent about how we present the results and be clear about our sampling methods.

Furthermore, HOT doesn't collect data on disease prevalence. A question around presence of disease, specifically diagnosed NCDs is included in the questionnaire. However, as the aim of this question is to understand how respondents perceive the healthcare system's response once they have been diagnosed, this data cannot be used as prevalence data. As we only ask about diagnosed disease it is likely that the actual prevalence is higher than what we can see in our HOT data, as many older people with NCDs in LMIC are never diagnosed. Therefore, if any data is presented using this information, please keep this in mind and be open about the limitations of the data in this regard.

Once data has been collected, you need to analyse the sample to find out if it is representative of the intended population. For instance, if your target population consists of 20 per cent of men and 80 per cent of women, your sample should follow a similar ratio. If it doesn't, this needs to be mentioned as a limitation when you communicate the results from the evaluation.

Additional analysis

As explained, the HOT database and analysis tool analyses data for programme evaluation by looking at any changes in Helpage health and care indicators. You can also conduct further analysis, but in order to do so it's advisable to think through what additional questions you want your analysis to answer and why (see examples in Table 2 below).

Table 2: Additional analysis table

Potential use of data	Purpose	Analysis that can be done
Reports to donors	Impacts of your intervention Indicators in logframe or agreed in proposals Partly covered in HOT database	Changes in different variables (e.g. access, perception of health services) How many people (and/or percentages of people) perceived a positive change? Averages in the group as additional information: how do people assess the situation? What were the average scores at baseline and endline?
Corporate indicators	See corporate indicators, expressed as how many people report improvement in health status Covered in HOT	Changes in 2 different variables; OP perception of their health and satisfaction with their life/wellbeing Presented in two different ways; How many people (absolute numbers) and percentages of people

24. Based on WHO World health survey, survey manual, World Health Organization, 2002

	database	perceived a positive change?
Local advocacy	Information on situation and change over time Performance of services as judged by users	Cross-sectional analysis of outcome indicators. Look at / reported mean scores for outcome indicators of interest and how they are changing Change in perception response by health services
Adjusting our work	Monitoring Adjustment work based on baseline or midline data collection and analysis Case tracking: specific cases have been identified through the baseline data collection round and can be followed up	Crosse-sectional analysis of outcome indicators in relation to investments made / planned Qualitative data analysis can inform quantitative results
National advocacy and policy development	Recommendations for policies new or exciting and programmes or policies	Information on situation and how it has/is evolved Performance of services as judged by users. Comparative analysis (See box analysis for 'International advocacy')
International advocacy	At global level, explore the potential to use findings in global advocacy	In-depth analysis Comparative analysis Determinants of good health perceptions Analyse relationships among variables Correlation analysis to understand how indicators correlate with each other. Regression analysis to learn which indicators that impact over all outcomes most.

Annex 1: Timeline of validation process

Two years of validation in different contexts show that the tool has fulfilled the key requirements

Momentum	People and locations involved	Result
June 2012, design of the first version of HOT	Paul Ong, Global Health Adviser and Catherine Dusseau, Regional Health Adviser for Latin America	First version and rationale paper drafted
April 2013, application of first version for DFID report at global level	Paul Ong and Tim Barker – tool implemented in Bolivia, Colombia, Tanzania and Cambodia	Data for reports coherent with project approaches and teams' interpretation; first field validation
2014, agreement with Pfizer International to design and validate a new version of HOT	HelpAge International and HelpAge USA	Project funded to support development of tool
May 2014 meeting of experts to review a new version, London	With Paul Ong, Catherine Dusseau, Dr Enrique Vega (Pan American Health Organization (PAHO)), Peter Lloyd Sherlock, Jeremy Dale, Sara Gallardo (IMEDER)	2nd version of the tool reviewed and plans for validation established
June 2014 to April 2015, one year of validation work	Research teams in Bolivia and Colombia Tanzania team using in project monitoring Better Health project (4 countries in Africa) using HOT for baseline	Validation of the tool against coherence, pertinence and validity. Feasibility and cultural adaptation First analysis on longitudinal collection
May 2015, 2nd experts meeting, Washington DC	With Paul Ong, Catherine Dusseau, Charlotte Aberdein and Susan Riker Dr Enrique Vega (PAHO), Loic Garcon (WHO Kobe centre), Dr Luis Miguel Gutierrez (Institute of Ageing Mexico), Kristin Bodiford (Community Strengths), Emi Kyiota	3rd version of HOT agreed by group of experts

	(Ibasho), Prakash Tyagi (partner GRAVIS)	
July 2015, first presentation of HOT with London office's directors		Plan to extend and strengthen HOT as institutional monitoring tool
October 2015 to September 2016, 2 nd year of support from Pfizer to extend use of the tool	Research teams in Bolivia and Colombia Tanzania team using in project monitoring Better Health project (4 countries in Africa) Extension to Uganda and to India	Validation as a longitudinal tool New contexts identified Validation of 3rd version
Training and guidelines		
2014	Training of local teams in Bolivia and Colombia	
February 2015	Training Better Health project, East, West and Central Africa (EWCA) region and Tanzania partners	
November 2015	2nd training in Ethiopia - included EWCA and southern Africa regional offices, Uganda, London policy team, representatives from Middle East/Eastern Europe and Myanmar offices	
January 2016	Training in India for GRAVIS team and Nepal office	

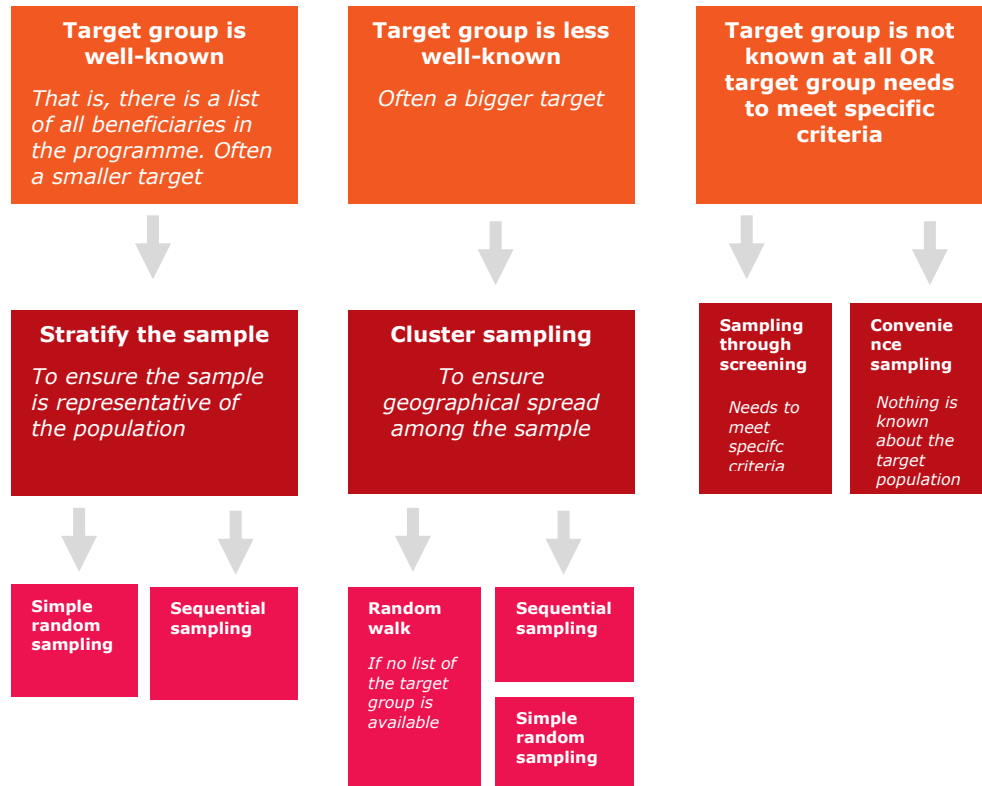
Annex 2: Sample size calculator

<https://www.qualtrics.com/blog/determining-sample-size/>

Required Sample Size								
Population Size	Confidence = 95%				Confidence = 99%			
	Margin of error				Margin of Error			
	5.0%	3.5%	2.5%	1.0%	5.0%	3.5%	2.5%	1.0%
10	10	10	10	10	10	10	10	10
20	19	20	20	20	19	20	20	20
30	28	29	29	30	29	29	30	30
50	44	47	48	50	47	48	49	50
75	63	69	72	74	67	71	73	75
100	80	89	94	99	87	93	96	99
150	108	126	137	148	122	135	142	149
200	132	160	177	196	154	174	186	198
250	152	190	215	244	182	211	229	246
300	169	217	251	291	207	246	270	295
400	146	265	318	384	250	309	348	391
500	217	306	377	475	285	365	421	485
600	234	340	432	565	315	416	490	579
700	248	370	481	653	341	462	554	672
800	260	396	526	739	363	503	615	763
1,000	278	440	606	906	399	575	727	943
1,200	291	474	674	1,067	427	636	827	1,119
1,500	306	515	759	1,297	460	712	959	1,376
2,000	322	563	869	1,655	498	808	1,141	1,785
2,500	333	597	952	1,984	524	879	1,288	2,173
3,500	346	641	1,068	2,565	558	977	1,510	2,890
5,000	357	678	1,176	3,288	586	1,066	1,734	3,842
7,500	365	710	1,275	4,211	610	1,147	1,960	5,165
10,000	370	727	1,332	4,899	622	1,193	2,098	6,239
25,000	378	760	1,448	6,939	646	1,285	2,399	9,972
50,000	381	772	1,491	8,056	655	1,318	2,520	12,455
75,000	382	776	1,506	8,514	658	1,330	2,563	13,583
100,000	383	778	1,513	8,762	659	1,336	2,585	14,227
250,000	384	782	1,527	9,248	662	1,347	2,626	15,555
500,000	384	783	1,532	9,423	663	1,350	2,640	16,055
1,000,000	384	783	1,534	9,512	663	1,352	2,647	16,317
2,500,000	384	783	1,536	9,567	663	1,353	2,651	16,478
10,000,000	384	784	1,536	9,594	663	1,354	2,653	16,560
100,000,000	384	784	1,537	9,603	663	1,354	2,654	16,584
300,000,000	384	784	1,537	9,603	663	1,354	2,654	16,586

Annex 3: Decision tree for sampling strategy

How well do you know the target group?



Sampling Strategy 1: Randomized and stratified sampling

Sampling strategy 1 can be implemented when the following criteria are met:

- You have a list/register with information on programme participants. The list includes (at a minimum) name and gender for of each person.
- You know the composition of the target population. (This analysis is often done during project planning and should be documented in the project proposal.)

Carry out the sampling in two steps:

- Stratify the sample
- Sample the units

i) Stratify the sample

Stratifying your sample means making sure that all well-known sub-groups are represented in the sample in the fairest way possible. You want to try to have a fair representation of different of people with different characteristics in your sample – in particular, ensuring that women and men are represented in accordance with demography or the distribution in your target population. You also need to ensure that people of different ages are represented (this can be done by dividing the sample into age groups of 10 years). Additionally, if there are any other distinct sub-groups in the target population, stratification could also be done for these groups (e.g. rural and urban or abled and disabled). However, remember that you need to know in advance how people fit into the different sub-groups. **The stratification of your sample should reflect the sample.**

Stratification is often multi-layered, as in the example below (see box). However, you can only stratify your sample when you know something about its composition. You can do a single-layered stratification if you don't have a lot of information about your target group (e.g. stratify by gender only).

Example of stratification

A stratification process will usually be multi-layered: e.g., on area, gender and age group

A 100-people sample might look like this:

Location distribution: 1 urban district and 3 rural districts

Gender distribution: 54% women and 46% men

Age distribution: Under 70 years: 50%; over 70 years: 50%

Location	75 Urban				25 Rural			
Gender	41 women		34 men		13 women		12 men	
Age group	20 (<50)	21 (<50)	17 (>50)	17 (>50)	6 (<50)	7 (<50)	6 (>50)	6 (>50)

ii) Sample the units

Simple random sampling from a list

The next step is to sample the units – the people you want to interview. For best rigour, use **simple random sampling** (i.e. pick people randomly). An easy way of doing this is to type your list of potential participants into a spreadsheet (preferably Excel), divide the list according to the stratification (e.g. location, gender, age) and then randomly pick the required number of people from each group/strata. Simple random sampling is a type of probability sampling and is therefore a very good way of drawing a sample.

Example of simple random sampling

One way of making a random selection is by using the “random” function in Excel. Allocate a random number to everyone on the list and pick your units according to a rule decided on beforehand. (However, be aware that the random numbers change at every action in the Excel sheet. Therefore, generate the numbers, then copy and paste them in the next column.)

Below is the list of the target population. Each person has received a randomly selected number. The rule (which was decided beforehand) is to pick the two people with the highest values. (Tip! Use conditional formatting to find them.)

Ana and Maya will be selected to participate in the evaluation.

Lisa	0.81
Ana	0.97
Maya	0.96
Hanna	0.45
Amy	0.20
Mathilda	0.75

Sequential sampling

Alternatively you could use “sequential sampling” (see example in box below). Decide on a rule (e.g., pick every third person on the list). If you have stratified your sample it would be a good idea to divide up your lists according to the characteristics you have stratified your sample (e.g. gender, location and age group).

Example of sequential sampling

Below is the list of target population. Each person has been assigned a number. The rule is to pick every third person in the list. Steven and Philip will therefore be selected for interview.

1	Lisa
2	Ana
3	Steven
4	Tom
5	Greg
6	Philip

Sampling Strategy 2: Cluster and random sampling

For bigger (more than few hundred) or geographically dispersed target groups, a stepwise approach starting with a cluster sampling by location can be used. It is also known as cluster sampling, meaning having the big population divided into smaller, more manageable size groups.

Clusters sampling

The first step will be to define your clusters. It is commonly done by dividing into geographic areas, e.g. neighbourhoods or communities. To select the cluster, you will need the list of communities, neighbourhoods or areas. You can choose to have a cluster in each community of work or select communities (e.g. 1 or 2 communities by province).

The number of units chosen in each cluster should be in proportion of number of people that are targeting in the same area. E.g. If your programme has half of its target group in area A – half of the sample should also come from area A.

In each cluster you will have to make sure that the sample is representing gender and age groups appropriately (stratification).

Select the units (individuals) in the clusters

The same method of selection has to be used for all clusters.

A random sampling is always the preferred choice. Thus, try to procure a list of people in the group/community through the project or local leaders. From this list apply the random sampling method as explained in the box for Strategy 1.

If you don't have a list, the sampling has to take place on the spot using field household sampling (see below).

Simple random sampling

Random sampling is always preferred. If you don't have a list of the people in the programme a random walk can be used for sampling your respondents.

This is how you sample households through a random walk:

- 1.** Draw a map of the village. Divide the village into 4 areas
- 2.** Start in 1 area and walk through the village, turn right at the first crossroads and left at the second crossroads etc.
- 3.** Decide a rule for household selection (e.g. pick every n^{th} household on your right hand side)
- 4.** If there is no older person in the household, go to the next household
- 5.** Always start your walk from a central point in the community, put a bottle or pen on the floor and make it spin. When it stops, go in whatever direction it points to.
- 6.** Repeat for all 4 areas of the village

Sampling Strategy 3: Sampling through screening and convenience sampling

If neither of the two strategies above is applicable to your situation, you could try convenience sampling or sampling through screening. Both techniques are described below.

Sampling through screening

If your target population is not the general older population but people with a non-frequent specific characteristic (such as a certain disease or disability, or if your programme is targeting people in need of care), screening might be the best way to conduct sampling.

If you expect more than 50 people with the criteria to enter the programme, choose every *n*th person for the evaluation. If you expect a smaller group (less than 50), all people in the programme can be part of the sample.

Convenience sampling

Convenience sampling (a non-probability sampling) is not a very rigorous technique so is generally only used when resources are very limited or there is no other way of sampling. The method is based on the ease or "convenience" of gaining access to a sample. Data is gathered from people who are readily available (such as people visiting a health centre). A convenience sample can easily become biased, so do give some thought to whether there are sub-groups (e.g. gender, age, geographic location, health status, level of poverty, level of education, etc.) that could get over- or under-represented with this sampling method, and take steps to address that.

This form of sampling can also be done by invitation – for instance, inviting all members of an older people's organisation or a service provider.

Annex 4: Training schedule

This suggested training schedule.

Day 1	
Time	Activity
9.00 – 9.30	Introduction of training and training participants
9.30 - 10.00	Introducing the Health Outcomes Tool
10.00 -10.30	Break
10.30– 12.00	<ul style="list-style-type: none"> Introducing the HOT evaluation in X project. Number and timing of data collection rounds The HOT user’s guide and how it should be used What is the HOT?
1.00 – 2.00	Lunch break
2.00 - 3.30	<ul style="list-style-type: none"> Discuss poverty question Discuss gender question
3.30 - 4.00	Break
4.00 – 5.30	<ul style="list-style-type: none"> Your role as an interviewer What to do if a respondent cannot part take in the survey

Day 2	
Time	Activity
9.00 – 9.30	Re-cap / Discussion of day 1
9.30 - 10.30	HOT scoring method
10.30 -1.00	The HOT survey – in detail
1.00 – 2.00	Lunch
2.00 - 3.30	Practise interviewing each other
3.30 - 4.00	Break
4.00 – 5.30	Exercise: <ul style="list-style-type: none"> Compile list of healthcare services in your context – see question Q10

Day 3	
Time	Activity
9.00 – 10.00	Re-cap / Discussion of days 1 & 2
10.00 - 11.30	Interview in front of class/role play *
11.30 - 12.00	Break
12.00 – 1.00	Test
1.00 – 2.00	Lunch
2.00 - 5.30	Practise interviewing each other

- **Role play:** In this exercise you randomly choose two participants from the group to interview each other in front of the group. Stop the interview after the questions have been asked for each domain, and discuss in the group what the interviewer did well and anything they could have done differently to get the best out of the interview.

Annex 5: Test

You are allowed to have paper version of the HOT survey open. But please put away the user's guide.

Maximum score for this test is 15 points.

1. What is the Health Outcomes Tool (HOT) and what is it for? (1 point)
2. What are the 4 domains/ topics in the HOT survey? (2 point)
3. What is the user's guide and how should it be used? (1 point)
4. Explain how the scoring method works. (5 points)
5. List at least 3 qualities that a good interviewer should have. (1 point)
6. Look at question Q3. What do we mean by 'social and daily tasks'? (1 point)
7. Look at question Q4. If your respondent is bed-ridden, do you think that his/her answer will end up closer to 0 or closer 100? (1 point)
8. Look at Q9. There is a skip pattern in question 9. Explain how it works. (1 point)
9. Look at Q10. If your respondent says that he/she receives all the help they need, where do you think they will be likely to score on the line? Close to 0 or close to 100? (1 point)
10. Are you supposed to read out the answer options in the questions? (1 point)

Annex 6: Test answers

You are allowed to have a paper version of the HOT survey open. But please put away the user's guide.

Maximum score for this test is 15 points.

1. What is the Health Outcomes Tool (HOT) and what is it for? (1 point)

The HOT is a monitoring and evaluation tool that can be used to monitor progress or evaluate the impact of projects or programmes.

2. What are the 4 domains/ topics in the HOT survey? (2 point)

1. Functionality
2. Dependency
3. Perception of services
4. Self-care

3. What is the user's guide and how should it be used? (1 point)

The user's guide is for anyone collecting data as part of a HOT evaluation. It explains how to understand, ask and answer the questions in the HOT questionnaire. Everyone working on the data collection should have a copy of this guide with them at every interview. In case you have forgotten how to ask a question or how to score the answer, you can refer to the user guide.

4. Explain how the scoring method works. (5 point)

The scoring method is a line from 0 to 100. This line is used for the respondents to answer the questions. As a rule, "0" is always the worst situation and "100" is always the best situation.

The respondent will point to where they are on the line and the interviewer will mark this point with a pen (if using a paper line). If you know the number of this point you can also record it, but that is less important. Respondents are also allowed to answer with a number; the interviewer should then mark that number on the line and record the number in the box.

5. List at least 3 qualities that a good interviewer should have. (1 point)

Be courteous, listen well, keep information about the respondent confidential, build rapport, stay objective, probe if needed, don't rush through the questions, etc.

6. Look at question Q3. What do we mean by 'social and daily tasks'? (1 point)

We mean feeding, bathing, clothing, walking in and around the house, toileting, maintaining continence.

7. Look at question Q4. If your respondent is bed-ridden, do you think that his/her answer will end up closer to 0 or closer 100? (1 point)

0

8. Look at Q9. There is a skip pattern in question 9. Explain how it works. (1 point)
If the respondent says 'NO' to question Q9b – do not ask question Q9c & Q9d
9. Look at Q10. If your respondent says that he/she receives all the help they need, where do you think they will be likely to score on the line? Close to 0 or close to 100? (1 point)
100
10. Are you supposed to read out the answer options in the questions? (1 point)
NO, never.



Find out more:

www.helpage.org/health

Published by HelpAge International 2016
PO Box 70156
London WC1A 9GB, UK
Tel +44 (0)20 7278 7778

info@helpage.org
www.helpage.org

Registered charity no. 288180

 @HelpAge  HelpAge International