

3

COMPONENTS, METHOD AND APPROACH

The State of the Humanitarian System report is based on several different research components including: reviews of evaluations and research literature, interviews, case studies and questionnaires.

Components

Needs and funding

The section covering humanitarian needs and funding is based on a desk review of data from public sources and a survey of private funding. For more information on sources, see the methodology section below.

Composition of the system

This element of the report is based on Humanitarian Outcomes' database of humanitarian organisations. Information on data sources is included in the methodology section, below.

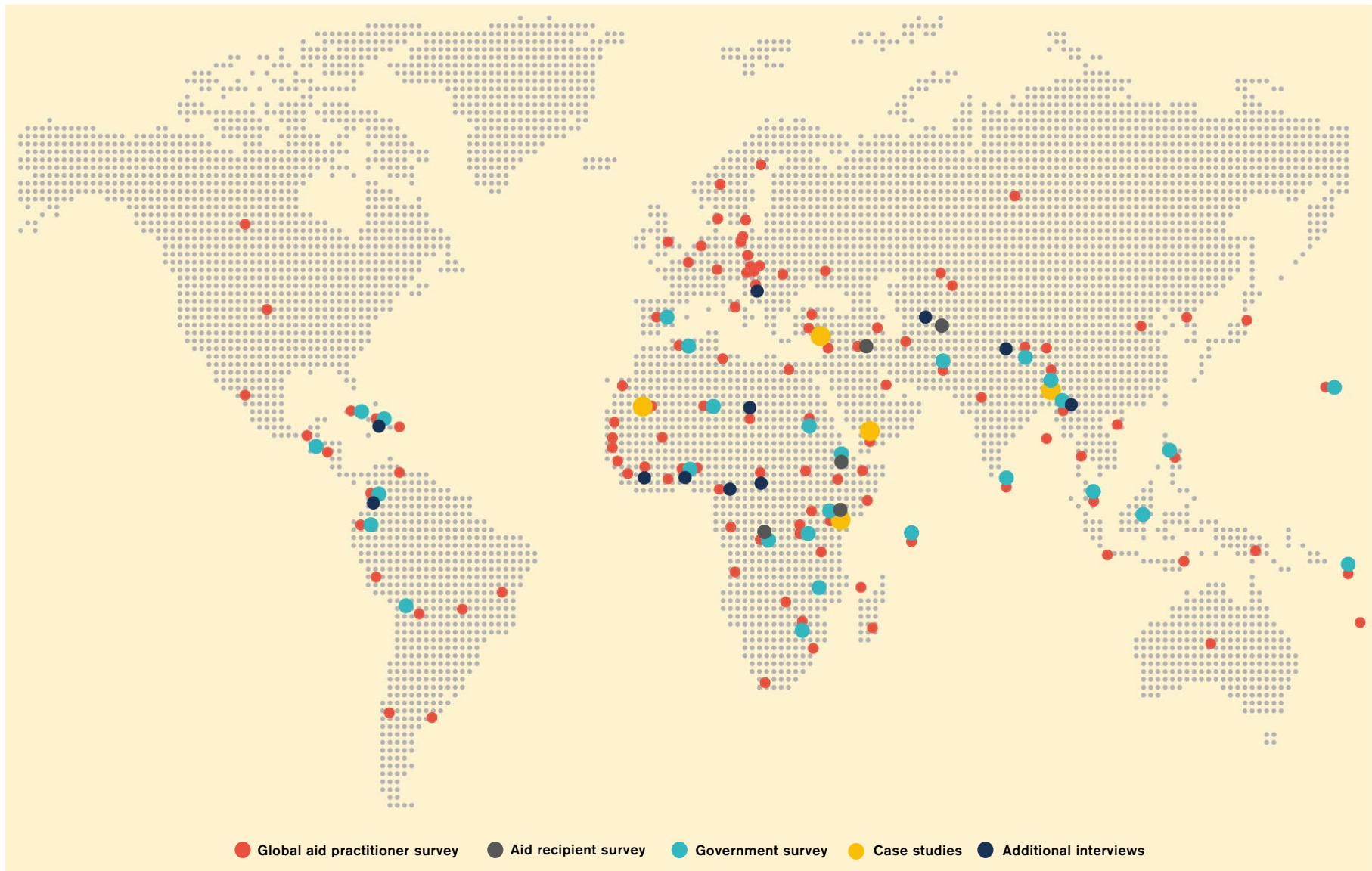
Performance of the system

The element of the report that assesses the performance of the system over the period 2015–17 is based on a synthesis of the following components, each of which uses distinct methods (see the methodology section below).

1. **Evaluation synthesis** – a synthesis of key points from 120 evaluations.
2. **Case studies** – visits to Bangladesh, Kenya, Lebanon, Mali and Yemen, combining observation with interviews with a total of 346 individuals: aid recipients, international NGO (INGO), UN and Red Cross/Crescent staff working in the country, members of national and local NGOs and government representatives.
3. **Key informant interviews** – 151 interviews with 153 interviewees at HQ level in donor organisations, UN agencies, INGOs and national NGOs (NNGOs), the International Red Cross and Red Crescent Movement, academia, the private sector and the military, plus another 150 interviews with representatives of organisations working in humanitarian operations in crisis-affected countries.
4. **Aid recipient survey** – a mobile phone questionnaire completed by 5,000 people who had received humanitarian assistance in Afghanistan, DRC, Ethiopia, Iraq and Kenya.
5. **Global aid practitioner survey** – an online survey aimed at humanitarian staff working in country programmes, which received 1,170 responses.
6. **Government questionnaire** – based on the practitioner questionnaire, a questionnaire sent to individuals working in the governments of affected states who are involved in humanitarian action, or who liaise with the international humanitarian system.
7. **Targeted literature review** – addressing specific information gaps not covered by the other methods.

In addition, the SOHS was supported by Ground Truth Solutions, who kindly shared recent data from their new Humanitarian Voice Index, and the Peer2Peer support team (formerly STAIT), who allowed the SOHS to access P2P reports for the period 2015–17.

Figure 2 / The geographical spread of SOHS 2018 components



Note: This map shows where evidence was gathered for the SOHS 2018. Where components recorded geographical data these have been visualised above. This map is not to scale.



The analysis in this edition of the SOHS draws on data reported to OECD DAC and FTS. These two sources use different criteria regarding what can be counted as international humanitarian assistance.

Methodology

Humanitarian needs and funding

Overview

The analyses on people in need, humanitarian funding and crisis-related financing are primarily collated from public sources; private funding figures were collected through a survey. The main reporting platforms for international humanitarian assistance are the OECD's Development Assistance Committee (DAC) and OCHA's Financial Tracking Service (FTS). It is mandatory for OECD DAC members to report their humanitarian contributions as part of their reporting on official development assistance (ODA), and in accordance with definitions set out by the DAC. Some non-DAC governments and multilateral institutions voluntarily report to the DAC. FTS is open to all humanitarian donors and implementing agencies to voluntarily report international contributions of humanitarian assistance, according to agreed criteria.

The analysis in this edition of the SOHS draws on data reported to OECD DAC and FTS. These two sources use different criteria regarding what can be counted as international humanitarian assistance. This has been taken into account when calculating aggregate volumes, and where necessary the analysis explains the data sources and methodologies used to reconcile figures or to prevent double counting. At the beginning of 2017 FTS released a new flow-based model which, along with a number of new functionalities, also makes it easier to trace funding, allowing us for the first time to compile figures beyond first-level recipients.

Crisis categories

For the analysis of crises by category, we determined the category into which each crisis fell by considering a series of indicators on conflict, displacement and losses associated with natural hazards. We used indicators in INFORM's Index for Risk Management and data from the Internal Displacement Monitoring Centre (IDMC) to identify countries affected by conflict; for countries affected by disasters associated with natural hazards, we used data from the CRED Emergency Events Database (EM-DAT) and Food and Agriculture Organisation (FAO) reports on El Niño; and to identify refugee-hosting countries, we used data from UNHCR and the UN Relief and Works Agency (UNRWA).

For a clearer visualisation, figure 12 shows 'Complex crises' as those that were marked as having scored the criteria for all three of the types of crisis above (conflict, refugee crisis and natural hazards).

International humanitarian assistance

Our estimate of total international humanitarian assistance is the sum of that from private donors and from government donors and EU institutions. Our calculation of international humanitarian assistance from government donors is the sum of:

- ‘Official’ humanitarian assistance (OECD DAC donors).
- International humanitarian assistance from OECD DAC donors to non-ODA eligible countries from FTS.
- International humanitarian assistance from donors outside the OECD DAC using data from FTS.

The calculation of ‘official’ humanitarian assistance comprises:

- The bilateral humanitarian expenditure of OECD DAC members, as reported to the OECD DAC database.
- The multilateral humanitarian assistance of OECD DAC members. This comprises:
 - The unearmarked ODA contributions of DAC members to nine key multilateral agencies engaged in humanitarian response: FAO, IOM, United Nations Development Programme (UNDP), UNHCR, OCHA, UNICEF, UNRWA, WFP and World Health Organization (WHO), as reported to the OECD DAC. We do not include all ODA to FAO, IOM, UNICEF and WFP, but apply a percentage to take into account that these agencies also have a development mandate. These shares are calculated using data on humanitarian expenditure as a proportion of the total core ODA spent by each multilateral agency.
 - The ODA contributions of DAC members to other multilateral organisations (beyond those already listed) which, although not primarily humanitarian-oriented, still report a level of humanitarian aid to OECD DAC. We do not include all reported ODA to these multilaterals, just the humanitarian element.
 - Contributions to the CERF that are not reported under DAC members’ bilateral humanitarian assistance. This data is taken from the CERF website (<https://cerf.un.org/>).

Official humanitarian assistance from OECD DAC countries that are also members of the EU includes an imputed calculation of their humanitarian assistance channelled through EU institutions, based on their ODA contributions to these institutions. This is not included in our total calculations for international humanitarian assistance and response to avoid double counting. Our estimate for official humanitarian assistance in 2017 is derived from preliminary DAC donor reporting on humanitarian aid grants. Turkey’s humanitarian assistance, which it voluntarily reports to the DAC,

largely comprises expenditure on hosting Syrian refugees within Turkey. This is not included in our total international humanitarian assistance and response calculations elsewhere in the report as these only include amounts directed internationally by donors.

Composition of the humanitarian system

The organisational mapping research for *The State of the Humanitarian System* reports encompasses the core actors in the international humanitarian system, defined as the UN humanitarian agencies, the International Red Cross and Red Crescent Movement and NGOs engaged in humanitarian relief efforts. Searching publicly available documents, the researchers gathered and compiled, to the extent available, the following figures for humanitarian organisations for the years 2014–2017:

- operational programme expenditure (OPE), i.e. excluding HQ and non-programme costs
- humanitarian expenditure (a subset of OPE, as distinct from development, religious or other work)
- total field staff
- national staff
- international staff.

Most of these figures were sourced from annual reports and financial statements, supplemented by website information and direct queries to the organisations concerned. UN agency information was also supplemented by figures from the databases of the UN Chief Executives Board (www.unsystem.org). The UN humanitarian agencies include the full members of the IASC,¹ plus the IOM and UNRWA.

The International Red Cross/Crescent Movement consists of the International Committee of the Red Cross (ICRC), the International Federation of Red Cross and Red Crescent Societies (IFRC) and 190 individual National Societies. Figures for the national societies come from IFRC's Databank and Reporting System (<http://data.ifrc.org/fdrs/data-download>), of which the most recent year available was 2016. The National Societies' data used for these measures exclude those in high-income countries, on the reasoning that they are unlikely to require an international humanitarian intervention in response to crisis, and their disproportionately large staff sizes would inflate the global estimate of humanitarian workers. International NGOs encompass those that operate in humanitarian response, as indicated by inclusion as a recipient or provider agency on FTS, registration with a major consortium or registry of international aid organisations, a past implementing partnership with a UN humanitarian agency or a recipient of humanitarian funds from a large government donor or the European Commission.

NGOs comprise the largest part of the Humanitarian Outcomes' database used to compile this information. For the largest, most significant humanitarian actors, and roughly 200 organisations of decreasing size, budget and staffing data were gathered manually, and their formally reported numbers entered into the database. Because of the large number of far smaller organisations, and the lack of publicly available annual reports or financial statements for most of them, Humanitarian Outcomes uses an algorithm for imputing the missing data, which improves in accuracy the more hard data is gathered directly from these organisations. In this model, NGOs are divided into tiers according to overseas programme expenditures and operational profiles. The tiers are used to calculate the mean values of staff and expenditures that are used to extrapolate missing data, on the logic that similarly sized and mandated organisations have similar operational configurations, presence levels, and staff-to-budget ratios (a process known as conditional mean imputation). Because data is largely unavailable for the lower tiers of smaller and local NGOs, these figures are the softest estimates with the most imputation used. However, given the great disparity in size between the largest and smallest organisations in the system, we expect much of this to fall within a rounding error.

As part of the research assignment for the SOHS 2018, Humanitarian Outcomes recruited a data scientist/statistician to review and test the imputation methodology for validity and rigour, and to recommend improvements. The report found that the first-level imputations (estimating missing data based on previous actual numbers and staff-to-budget ratio from that organisation) were highly accurate, and second-level imputations (using tier averages) significantly less so, given the amount of missing data. The report recommended a modification to the algorithm based on this finding to increase the accuracy of second-level imputations, which Humanitarian Outcomes has adopted and used in this year's mapping. This involves taking forward previous years' first-level imputations to include as data points for the second-level imputations.

As a general caveat, while the model used produces rigorous, systematic estimates for the organisational mapping, they are still just estimates, and should be considered and cited as such.

Performance of the humanitarian system

Developments to the method for the 2018 edition

This edition of the SOHS, like previous editions, has been created through a synthesis of findings from separate research components using distinct methods. To better facilitate this synthesis and make the process more transparent, ALNAP developed a study matrix with indicators for each of the DAC criteria, used to ensure consistency across the different consultants and research components, and to ensure that all of the key issues were addressed, and that the different components addressed these issues in the same way using a common set of questions. This study matrix is available in annex 3.



In line with recommendations from the SOHS Methods Group, this edition made a specific attempt to increase the amount of information collected in humanitarian operations, and in particular to increase the amount of information collected from aid recipients.

For the first time, all of the interviews (key informant interviews and case studies) were recorded and transcribed verbatim. To ensure consistency between the various elements of the research, and to allow for comparison of all the data from different elements on the same topic, these interview transcripts and the evaluation synthesis were then coded in MaxQDA using a common coding framework (essentially a list of topics and sub-topics, based on the study matrix described above).

These improvements to the methodology are part of ALNAP's continuous attempts to ensure and improve the quality of evidence used in its research. They are particularly important, given the increased amounts of data gathered as part of the 2018 edition. In line with recommendations from The State of the Humanitarian System Methods Group, this edition makes a specific attempt to increase the amount of information collected in humanitarian operations, and in particular to increase the amount of information collected from aid recipients:

- This edition includes 346 interviews with individuals in five country case studies and 17 other countries (compared to 201 interviews in four countries in 2015).
- This edition includes responses from 5,000 aid recipients in five countries (compared to 1,189 aid recipients in four countries in 2015).

In addition, statistical regression was used for the first time on the aid recipient surveys to understand relationships between participants' responses to performance questions, as well as relationships between these responses and participants' characteristics (age, gender, type of crisis, status).

Evaluation synthesis

A synthesis of relevant evaluation findings from the period January 2015 to December 2017 forms one component of the evidence base for the SOHS 2018. Around 170 evaluations were considered, based primarily on a search of the ALNAP evaluation database. Of these, 121 of the most relevant were included in the synthesis process. These were then scored for quality and depth of relevant evidence (see below), and the synthesis process was organised in such a way as to prioritise findings from evaluations with the highest evidence scores.

Selection of evaluations

The sample of evaluations was purposive, designed to give a reasonable balance of evidence across the following parameters:

- subject matter by type of crisis: ‘natural’ disaster/health crisis, conflict, forced displacement or generic
- region/country
- commissioning agency

A bias was noted in the pool of available evaluations towards evidence from certain countries and regions, notably the Syria region and countries affected by the Ebola Epidemic in West Africa. This is reflected to some degree in the sample chosen, since these appear to represent particularly rich areas of learning for the humanitarian sector, and have to a significant degree reshaped thinking about the role of the system and appropriate programme approaches. Each evaluation included in the Evaluations Matrix produced as part of the evaluation synthesis process includes an evidence score for each report included in the sample. This is done on a scale of 1–3, with 3 representing the strongest evidence. These scores are based on the judgement of the researchers against two parameters, each with its own criteria:

- Evidence depth: the depth and extent of relevant analysis in the report (‘relevant’ here means that it relates to the themes highlighted in the coding system – see below). The related criteria included whether the work appeared to add significantly to the existing evidence base on the subject.
- Evidence quality: the quality of the analysis and the related evidence base. Here, we considered in particular how well argued and evidenced the conclusions of each evaluation were.

Each of the two parameters was scored 1–3, with the overall value score being the average of the two scores (e.g. $3/2 = 2.5$). Reports that scored 2.5–3 were considered core reference documents for the purposes of the synthesis, and a primary source of evidence. Reports scoring 1.5–2 were considered a secondary source, and those scoring 1 were used to supplement the evidence base where appropriate. In-depth and multi-sector evaluations tended to score higher than lighter reviews, based on depth of coverage.

Analysis of information

Analysis was conducted according to the common coding system in MaxQDA 2018. The researchers coded and grouped extracts from the evaluations under the relevant themes. Decisions about which extracts to include were made by researchers based mainly on relevance. Results were then synthesised. The synthesis process involved two main elements:

1. Collating the material according to related findings on common themes.
2. Identifying common findings: findings that appeared to be (broadly) common across a range of evaluation evidence. In addition to this, some 'meta findings' were identified; in other words, findings that emerged not from any one source or sources but from a 'helicopter' view of the evidence as a whole. This included findings about the state of the evidence itself.

A 'first cut' synthesis was made based on around 30 evaluations from the sample that were judged to provide the greatest depth and quality of evidence. Provisional findings were identified as the process went on, with extracts grouped under sub-headings (provisional findings or hypotheses) within each theme, according to whether they tended to confirm (or contradict) the finding in question. These were then tested against the wider sample, with additional extracts added as appropriate. The process was thus both inductive and iterative.

The main findings from the synthesis were presented in a preliminary summary according to the common format (an Excel spreadsheet) agreed for the SOHS 2018 process, i.e. according to the criteria and indicators agreed in the Study Matrix. In the summary, evidence for each point was presented as strong, moderate or weak.

Constraints and limitations

Attempts to conduct a systematic, comparative review of evidence from humanitarian evaluations across the sector are hampered by several factors. One is the variability in the object of evaluation: most of the evaluation material is response- and organisation-specific. Related to this is the difficulty of controlling for contextual variables. A third factor is the variability in the methods of investigation adopted in the evaluations, and the way in which results are recorded. Most of the available evidence is qualitative; where quantitative results are available, the factors noted above tend to make comparison difficult or impossible. This rules out any statistically valid process of meta-analysis, and variations in the quality of evidence from different sources require some weighting of evidence that invariably involves a degree of subjectivity and hence bias. In short, evaluation synthesis does not meet the criteria for systematic review in the stricter sense. That said, a systematic approach to reviewing the available evidence, using evidence quality criteria to identify an analytical sample from a relatively large pool of evidence, can have greater validity than would be the case with a smaller sample,



The pool of publicly available humanitarian evaluations for the period 2015–17 is substantial, and allows a sample to be compiled covering a reasonable cross-section of different crisis and organisational types.

randomly selected and analysed in a non-systematic way. Although there is considerable reliance on the judgment of those compiling the synthesis – and some inevitable bias in this process – we believe that the evaluation synthesis conducted for the SOHS meets these criteria. Triangulating the results with those from the other methods of enquiry used for the SOHS also gives greater confidence in the findings.

The pool of publicly available humanitarian evaluations for the period 2015–17 is substantial, and allows a sample to be compiled covering a reasonable cross-section of different crisis and organisational types. However, enquiries suggest that a large amount of relevant material – particularly internal reviews on more sensitive issues such as accountability and remote management – is not publicly available, and so was not included in the sample used for the SOHS. This suggests that some more sensitive issues in particular may be under-represented in the sample.

Finally, as noted the sample is biased towards particular contexts (e.g. Syria region, Ebola Crisis) and tends to reflect the predominant concerns of the system in the period 2014–16, given the time lag in conducting evaluations. For example, there is more analysis of relevance to the UN's Transformative Agenda than there is to the agenda that emerged from the WHS in May 2016.

Case studies

Selection of countries and of interviewees

Full case studies were conducted in five countries (Bangladesh, Kenya, Lebanon, Mali and Yemen). In addition, team members interviewed individuals and focus groups in a number of other countries (Afghanistan, Cameroon, Central African Republic (CAR), Chad, Colombia, Côte d'Ivoire, DRC, Greece, Haiti, Nepal, Nigeria). In total, 346 people were interviewed through 171 bilateral or focus group meetings.

Case study countries were chosen to provide a sample with geographical diversity (across regions) and contextual diversity (across the three main contexts considered in the report).

Table 3 / Number of interviewees per country

Afghanistan	5	Haiti	55
Bangladesh*	31	Kenya*	42
Cameroon	5	Lebanon*	44
CAR	1	Mali*	52
Chad	6	Myanmar	2
Colombia	16	Nepal	40
Côte d'Ivoire	9	Nigeria	8
DRC	9	Yemen*	17
Greece	4		
Total			346

* Case studies produced for this research

Of the total, 38% of interviewees were from Africa, 18% from the Middle East and North Africa, 23% from Asia, 20% from the Americas and 1% from Europe.

As noted above, it is not entirely accurate to say that any country represents only one crisis context. However, we can say, broadly, that 44% of interviewees were from disasters linked to natural phenomena (Colombia,² Haiti, Kenya and Nepal), 33% from conflict-affected areas (Afghanistan, Cameroon, CAR, Chad, Côte d'Ivoire, DRC, Mali, Myanmar, Nigeria and Yemen) and 23% from refugee-hosting contexts (Bangladesh, Greece and Lebanon).

The interviewees were selected to be representative of the diversity of organisations taking part in humanitarian action and its coordination. However, as the key informant section of the research was oriented towards HQ staff and international organisations (see below), the case study interviews aimed to give a higher profile to national organisations and to aid recipients. As a result, local actors (national and local NGOs, national and local authorities) represent the larger group, with 36% of interviewees, aid recipients 34% and international actors (international organisations and INGOs) 30%.

Table 4 / Number of interviewees per category

Category	Number of interviewees	Percentage of the total
Aid recipients	117	34%
NNGO	31	9%
LNGO	17	5%
Local and national authorities	75	22%
International organisation	35	10%
INGO	71	20%
Total	346	100%

Interviewee selection was partially purposive (inasmuch as interviewers attempted to interview a certain number of people from each category) and partially by convenience (interviewers were working on a short timeframe, and within categories tended to interview people who were available and qualified to speak on the situation). Aid recipients interviewed were selected on a convenience basis. Although the study team tried to achieve an appropriate level of diversity in terms of gender and age criteria this was not always possible and, in some cases, very few or no women were reached. Overall, 42% of interviewees were women and 58% men.

Table 5 / Gender ratio of interviewees

Country	Total interviewees	Women	Men
Afghanistan	5	0	5
Bangladesh	31	18	13
Cameroon	5	0	5
CAR	1	0	1
Chad	6	1	5
Colombia	16	5	11
Côte d'Ivoire	9	2	7
DRC	9	5	4
Greece	4	3	1
Haiti	55	34	21
Kenya	42	14	28
Lebanon	44	28	16
Mali	52	12	40
Myanmar	2	1	1
Nepal	40	15	25
Nigeria	8	4	4
Yemen	17	0	17
Total	346	142	204

Interview structure

Interviews were semi-structured. The interview protocol was constructed using the common study matrix (see annex 3). Interviews were conducted in local languages wherever possible (English, French and Spanish, as well as Arabic, Dari, Creole, Turkana, Rohingya and Bangla). Interviews followed ODI's research ethics policy, which covers all ALNAP research, and informed consent was obtained either in writing or, where written consent was deemed contextually inappropriate, verbally. Interviews were recorded and transcribed. Interviews in languages other than English, French or Spanish were then translated into English for coding.

Analysis

The interview transcripts were coded by two team members, using the common coding framework, in MaxQDA 12. This exercise resulted in the identification of 4,329 coded sections of text. These were then assessed to identify key findings under each code. To determine the strength of evidence for each code, team members used two main and combined criteria: quantity and consistency of quotes.

- If a large number of interviewees reported the same response/issue/perception for that topic, and there were no significant or alternative findings, the finding was recorded and classed as 'strong'.

Five country case studies fed into the analysis for the SOHS 2018 report:

Bangladesh
Kenya
Lebanon
Mali
Yemen

- If a smaller but homogenous group reported the same response/issue/perception, the finding was recorded as strong, but attributed only to one group of respondents (category of key informant, or one country, or one category of context).
- If interviewees reported diverging responses/issues/perceptions, the finding may be ‘moderate’ (where a significant number of interviewees raised the issue and the majority agree) or ‘weak’ (where a small number of interviewees raise the issue).
- If an issue is only discussed by a few interviewees, no finding was recorded (however, where this was the case for findings from aid recipients, a lower threshold was applied – so that weak findings from recipients had a higher chance of being included in the final analysis).

These findings were then presented, by OECD DAC criterion, in a preliminary summary. During the writing process, the lead author returned to the codes to check the strength of findings for particular points.

Constraints and limitations

Despite attempts to select countries that broadly mirrored the distribution of activities conducted by the international humanitarian system, the interviews tend to over-represent disasters linked to natural phenomena, and to over-represent some regions (particularly Latin America) when compared to the size of humanitarian activities in that context/region (as measured by humanitarian expenditure).

The convenience nature of the sample, particularly with respect to interviewees from the affected population, is a significant constraint. As described above, recipient interviewees were selected mainly on an opportunistic basis (communities available during a short field visit, beneficiaries living in urban areas near aid agencies’ offices), introducing potential bias into the results. People further from towns and roads were not adequately represented. As noted, women and girls were under-represented.

A further constraint was the limited amount of time available for data collection. The researchers had only ten days in each country to meet stakeholders and run the interviews. This was, to a degree, compensated by the fact that the international researchers worked with a local counterpart, who was able to provide information on the context and situation. Finally, case study findings are based almost exclusively on perceptions rather than objectively verifiable data. To a degree, this constraint can be addressed by triangulating the perceptions of different types of respondent – recipients, local authorities and international NGOs, for example.

Key informant interviews

Selection of interviewees

The key informant interviews were designed to be as representative as possible. The team aimed to cover all of the major types of actor within the sector: UN agencies, the Red Cross and Crescent Movement (RCRC), international NGOs, national NGOs, donors, development banks and other multilaterals, think-tanks, academia, the media, affected governments and commentators. The team also sought out respondents at different levels of the system and of the organisations and bodies outlined above – from senior leaders to those working at functional, operational or operational coordination levels in humanitarian programmes. The team also used a snowball approach, asking interviewees to recommend people who had differing views or who represented a particular aspect of a discussion, or who had specific technical or geographic expertise. In all, 153 people were interviewed. The breakdown of interviewees by type of agency is given in Table 6.

Table 6 / Breakdown of interviewees by type of agency

Category	Male	Female	Total
Academic/analysis	18	10	27 (17.7%)
Donors	9	11	20 (13%)
UN	25	10	35 (23%)
RCRC	5	5	10 (6.6%)
INGOs	21	9	30 (19.8%)
NNGOs	2	1	3 (2%)
Humanitarian networks	7	6	13 (8.6%)
Regional organisations	3	3	6 (4%)
Other (media, private sector, peace orgs, etc.)	3	5	8 (5.3%)
Total	93 (60.8%)	60 (39.2%)	153

Interviews were semi-structured, based on a protocol derived from the common study framework. Interviewees generally took a global, rather than an operation-specific, view of the performance of the system. Interviews were conducted in English, and followed ODI's research ethics policy.

Analysis

Interviews were either noted or recorded and transcribed, and the resulting documents were coded in MaxQDA 12 using the common coding framework. The coding exercise resulted in 3,320 coded sections of text. The team analysed the coded texts against the common study matrix to identify evidence relating to each element of the matrix and determine the strength of this evidence. Evidence strength was assessed using a scoring



Interviews with around

500

PEOPLE

fed into the analysis for
the SOHS 2018.

system. If an issue recurred in a significant number of interviews (typically 10–20-plus) in largely the same way then it was assessed as ‘strong’. If an issue recurred in a similar way in a number of interviews, either numerically (5–10) or in a very similar way among a smaller number of people with expertise in that particular topic, it was assessed as ‘moderate’. If the issue recurred in a handful of interviews, or in a number of interviews but in a different way (for instance a wide range of divergent views on a particular issue), the evidence was assessed as ‘weak’. The findings were presented, by OECD DAC criterion, in a preliminary summary.

Constraints

The major constraint, as might be expected with an exercise of this nature, was getting people to commit to and then attend interviews. People within the humanitarian sector have high workloads and unpredictable travel, and senior leaders are regularly responding to similar interview requests. As a result, the interview team became overly reliant on personal connexions and relationships, influencing respondent selection. Some groups were under-represented because, despite numerous invitations, very few individuals agreed to be interviewed. As with the case studies, a further constraint with the interview approach was that the results were almost entirely perceptual.

Aid recipient survey

Selection of countries and participants

For this *State of the Humanitarian System* report, ALNAP again commissioned GeoPoll to carry out telephone surveys in DRC, the Horn of Africa (Kenya and Ethiopia), Iraq and Afghanistan. The SOHS 2012 was one of the first major surveys of aid recipients in humanitarian action, reaching 1,104 aid recipients in DRC. This iteration surveyed 5,000 aid recipients across the five countries. These countries were chosen to represent humanitarian responses in a variety of geographical areas and contexts. The selection was partially influenced by the choice of case study countries, and aimed to include more conflict contexts to make up for their under-representation in the case studies.

The survey used a probability sample of mobile phone respondents, with a two-step sampling methodology:

1. **Pre-stratification:** respondents were selected based on key factors such as age, gender or location, to allow for comparisons across different demographic groups and to ensure representation. Respondents in the GeoPoll database are incrementally indexed for pre-stratification purposes. With respect to location, as the aim of the survey was to elicit responses from people who had been in receipt of aid, certain geographical regions (where a higher proportion of the population had been in receipt of humanitarian aid) were chosen for the survey in DRC, Iraq and Kenya.

Table 7 / Geographical regions chosen for the aid recipient survey

DRC	Iraq	Kenya
Majority of the sample from the following provinces (though not limited to these provinces):	Limited to the following provinces:	Limited to the following locations:
North Kivu	Anbar	Turkana
South Kivu	Ninewa	Wajir
Katanga	Erbil	Marsabit
Maniem	Kirkuk	Isiolo
	Salahal-Din	Mandera
	Baghdad	Garissa
	Dahuk	Nairobi
	Diyala	Tana River
	Sulaymaniyah	
	Babylon	

2. **Simple Random Sample (SRS):** respondents were randomly selected from the GeoPoll database (a database of all mobile phone subscribers or mobile subscribers with the largest service providers in the country) to participate in a survey. GeoPoll asked eligibility questions at the beginning of the survey to determine if the respondent had been an aid recipient within the past two years. Only those who were eligible continued on to complete the survey. Recipients were all aged 15 years and over.

In total, 693,795 surveys were sent out, 31,987 people responded and 5,000 completed responses were received (including 331 responses to the CATI voice questionnaire (see below) in Iraq). The relatively low number of completions among those who responded is largely a function of the eligibility question: many respondents had not received aid in the past two years.

Questionnaire structure

ALNAP provided GeoPoll with the content of questions for the survey, which used the same or slightly modified questions from the 2012 and 2015 editions to provide consistent comparisons over time. The 2018 survey used text-based (SMS) survey instruments (with the exception of Iraq) to collect data, which was the same method used in the 2012 and 2015 editions. In Iraq, GeoPoll collected a portion of the responses with computer-aided telephone interviewing (CATI). CATI was used as a supplemental method in order to increase response rates.

Constraints and limitations

The methodology used for the aid recipient survey suffers from a number of potential biases.

Selection bias: as there is no overall, country-level list of aid recipients, it is not possible to conduct a probability sample specially targeting all aid recipients. Rather, GeoPoll targeted the whole population, and then screened out those who were not aid recipients. In some cases, it may be theoretically possible to conduct a probability sample of the whole population, using census data as the sampling frame. However, it is not clear that aid recipients mirror the entire population in composition, and so data on the overall population would not represent the subset of the population who had received aid.

The aid recipient survey uses a database of mobile phone subscribers as the sampling frame. The degree to which this reflects the entire population will differ from country to country, depending on the proportion of the population who are mobile subscribers (see Table 8).

Table 8 / Rates of mobile phone ownership

Country	Mobile phone penetration	Unique subscribers (m)
Kenya	59%	28.3
Ethiopia	34%	34.7
DRC	26%	21.0
Iraq	50%	Not applicable
Afghanistan	41%*	Not applicable

Source: GSMA The Mobile Economy 2017, <https://www.gsma.com/mobileeconomy/>

Notes: m: million. As Afghanistan was not included in the GSMA report, we have inserted the following calculation as an estimate: the GeoPoll database in Afghanistan as a percentage of the national population, multiplied by the reverse proportion of mobile network operators (MNOs) on which GeoPoll is active.

The fact that only those individuals with access to a mobile phone are able to participate in the survey research introduces important selection biases, when comparing respondents to the whole population. In particular, respondents will tend to be more urban, male, younger and of a higher socio-economic status.

Non-response bias: the relatively low response rates suggest that some respondents elected not to participate in the study. There could be a number of reasons for this, in particular literacy levels (as the survey was delivered by SMS, people with low levels of literacy would tend not to respond). One concrete illustration of these biases was the relatively low rate of female and older respondents. Overall, 28% of respondents were female and 72% male; 43% were aged 15–24, 33% 25–34 and 24% 35 and above.

Surveys of



1,170
aid practitioners



116
countries



28
governments



5,000
aid recipients from
Afghanistan, DRC,
Ethiopia, Iraq and
Kenya
informed the
SOHS 2018.

GeoPoll ran a test to check for any statistically significant differences between gender or age categories in each country. There were statistically significant differences in answers by gender in only 14% of the response sets,³ and by age in 42% of response sets. We considered applying a weight to the results, but in the majority of cases there was no need to apply a weight to the results (as there was no statistically significant difference based on gender or age), and on those where it would have been useful, the weights GeoPoll would need to use would be higher than typically recommended, which would have led to significant over-compensation.

These constraints are important, but it is worth remembering that it is not currently possible to create a true probability sample of aid recipients, and that any survey mechanism will suffer from selection bias. The SMS approach, while affected by the biases above, does offer the ability to collect responses from hard to reach areas (those which were not accessed by the case studies, for example). It may also avoid interview effects (such as interviewees providing responses that they think will please the interviewer) and provide better-quality data from those who respond (Link et al., 2014).

Practitioner and government surveys

The practitioner and government surveys for this iteration of the SOHS were updated to ensure that the questions asked covered all of the areas in the study matrix, but without sacrificing the comparability of the survey over time. The surveys were translated into French, Spanish and Arabic and uploaded to SurveyMonkey for dissemination.

The ALNAP team prepared a dissemination plan mapping local NGO networks and national disaster management agencies (NDMAs), liaised with the SOHS Strategic Advisory Group to spread the word and supported ALNAP Members that are operational with messaging and channels so that the surveys could reach staff on the ground. Adverts were placed on ReliefWeb and Dawns Digest, and the survey was also promoted with social media campaigns. The surveys were open for six months (from August 2017 to January 2018) and were completed by 1,170 practitioners and 38 government representatives from a wide geographical spread.

The ALNAP team cleaned and prepared all the answers collected through pivot charts in spreadsheets, to allow for cross-cutting and analysis of the data received. Datasets from past SOHS surveys were incorporated to allow for comparison of responses over time.

Literature review

The literature review was used mainly to provide information on specific areas not captured fully by other means in the Study Matrix. These related primarily to the functions of the humanitarian system beyond the provision of humanitarian assistance, including protection and resilience. The rationale for considering these elements in a separate literature review is that, while these activities should be addressed in humanitarian evaluations, there is some evidence that they do not receive consistent attention in humanitarian activities, and so will not be adequately covered in programme evaluations.

Selection of literature to review

Following a review of gaps in the SOHS evidence base in October 2017, and taking account of the limited time available, the literature review focused on the following topics:

- cash transfers and vouchers
- livelihood support and social protection
- urban humanitarian issues
- the humanitarian–development nexus
- ‘systemic’ issues – those that went beyond specific agencies or programmes
- protection and the humanitarian system.

From the search process described below, the researchers identified around 50 key sources to inform a thematic synthesis on the list of topics above. The literature search was conducted using a variety of sources, including:

- Humanitarian Policy Group and Humanitarian Practice Network
- *Disasters* journal
- ALNAP
- Groupe URD
- Feinstein International Center (Tufts)
- Refugee Studies Centre
- Chatham House
- Other academic sources
- OCHA
- IRIN
- Other relevant publications, including NGO policy papers

A JSTOR search was undertaken with the following search string: ((humanitarian) OR (disasters)). Search results were limited to works addressing the topics above, published in 2015–17 and concerning current or recent humanitarian responses. Works of a speculative nature or covering topics that are well-addressed in evaluations were excluded.

Analysis of literature

The same coding system was used to organise the literature review material as was used for the evaluation synthesis, for consistency and simplicity. The synthesis process involved two elements:

1. Collating the material according to related findings on common topics.
2. Identifying findings that appeared to be broadly common across a range of evidence from the source material. Some meta-findings were also identified.

Again, the process was inductive and iterative, with provisional synthesis findings identified as the process went on, and extracts grouped under sub-headings (provisional findings or hypotheses) within each topic, according to whether they tended to confirm or contradict the finding in question. These were tested against the wider sample, with additional extracts being added as appropriate to strengthen the evidence base.



As the literature tends to take the form of ‘illustrated argument’, the potential for bias is clear. But the comparative strength of the literature review is that it allows a topic-based investigation across multiple crises and organisation types.

Constraints and limitations

Many of the same factors noted above for the evaluation synthesis also apply to the literature review. Indeed, the nature of the evidence is even more variable, being based on sometimes ill-defined questions and unclear methods of enquiry. However, a significant part of the literature reviewed itself consisted of a review of results from different studies, and so provides a useful (if not always fully reliable) meta-analysis on the topics in question. As the literature tends to take the form of ‘illustrated argument’, the potential for bias is clear. But the comparative strength of the literature review is that it allows a topic-based investigation across multiple crises and organisation types. In some areas, notably in the field of cash transfers and the comparative strengths and weaknesses of this modality compared to others (e.g. food assistance), the evidence base is relatively strong and some valid quantitative comparison appears possible. In other areas, the analysis is considerably more subjective.

One of the main constraints was available time. This was a limited and strictly topic-focused exercise, designed mainly to fill identified gaps in the evidence base for SOHS. As with the evaluation synthesis, there was considerable reliance on the judgment of those conducting the synthesis as to the relevance and strength of the evidence, as well as its interpretation.

Synthesis

Once the preliminary reports from the various elements of the research had been completed, outlining key points for each of the OECD DAC criteria and the strength and source of evidence for each, the ALNAP Secretariat team compiled nine reports, one for each of the criteria, containing all the points from the various sources. These reports highlighted where components agreed and where they disagreed. They also proposed alternative explanations for phenomena which had been observed in the components, aiming to challenge

the explanations given by interviewees, as a way of interrogating the arguments presented in the preliminary reports. The ALNAP team and the consultants in charge of each of the components met for a two-day writing conference to discuss these reports, and in particular to:

- Consider the key points and agree on those they felt were sufficiently well evidenced to be included in the final report.
- Identify any evidence that they were aware of that contradicted these points or that suggested alternative explanations.
- Weigh the balance of information where the reports pointed to disagreement: attempt to explain or resolve the disagreement on the basis of the evidence collected.
- Consider alternative explanations, and the evidence for these, and determine how viable these explanations were.
- Identify important gaps in the information, including information required to address disagreements or alternative explanations, and any research that might address these gaps.



[T]he main points included in the report were those for which there was strong or moderate evidence from a number of sources, and where there was very little evidence, or only weak evidence, to support alternative descriptions of the situation...

On the basis of this writing conference, the SOHS team agreed the main outlines of the report. The lead author then developed the report around this agreed structure. In doing so, he worked with a research assistant to identify research, other literature and additional interviewees to address important gaps, resolve contradictions in the evidence (where these had not been resolved in the writing conference), allow a more thorough assessment of alternative explanations or provide information that would enable points that were only weakly evidenced to be supported or rejected. The lead author also conducted a number of spot-checks comparing the original coded material to key points and strength of evidence, in order to validate the consultants' assessments.

In the final analysis, the main points included in the report were those for which there was strong or moderate evidence from a number of sources, and where there was very little evidence, or only weak evidence, to support alternative descriptions of the situation, or explanations as to the causes of the situation. Obviously, in some instances the balance of evidence was closer. Where this was the case, this is noted in the text. The text also includes points for which there was limited supporting evidence, but which are important because they challenge existing orthodoxies and may inspire further research. Where this is the case, the limited nature of the evidence is noted in the text.

Constraints and limitations

Two key constraints emerged in the synthesis phase, both of which relate to the nature of the information on which conclusions are drawn. The first is that the report, while aiming to discuss the system as a whole, is actually building up an assessment of the system from descriptions of its various disparate elements: from particular countries or organisations. This is problematic because there can be significant variety between one situation and another: any ‘overall’ assessment can obscure these differences, and end up describing generalities that are not true in many places. To the degree possible, the report addresses this by identifying elements that were common in the large majority of situations, and – where this was not the case – clarifying elements that were common only in one of the three contexts, or elements that appeared to be specific to a particular country or type of organisation. Nevertheless, the reader should be aware that the general performance assessments and trends identified in this report will not be true in all situations, from Afghanistan to Zimbabwe, and that any particular operation is likely to have at least some elements that differ appreciably from those outlined in the report.

By building up the sum from the parts, this approach also, arguably, misses out important elements of system performance – of those properties of the humanitarian system that are more than the sum of the parts. For example, the adequacy and added value of the collective processes (e.g. needs assessment and strategic planning) to which so much time and effort is devoted are only captured tangentially. Although the composite approach can shed light on these issues, the overall performance of the system as a system – and the ways in which that system is evolving – can only be partly captured by this approach.

The second key constraint has already been mentioned in the sections above – but holds good for the whole as well as for the parts. *The State of the Humanitarian System* report is largely based on perceptions – the perceptions of humanitarian professionals, of aid recipients and of academics and government representatives who work alongside or observe humanitarian activities. In some cases, this is exactly the right type of information on which to base an assessment. If we wish to know whether people feel that they are treated with dignity, then we are, essentially, asking about their perceptions of their lived experience. In other cases, it is much less useful. Perceptions are a poor basis on which to determine whether we are reaching all people in need, if one of the reasons we fail to reach these people is that we don’t know they are there in the first place. Perceptions are also poor guides to phenomena such as excess mortality in a population, which tends to become visible only at a scale which is not visible to the individual observer. One of the troubling elements of the state of the system in 2017 is the number of things that it doesn’t know, and doesn’t find out (see box on information gaps in the humanitarian system).

Information gaps in the humanitarian system

In researching this edition of the SOHS, it became clear that many of the most important information gaps identified in previous versions have still not been filled. In particular:

- **The number of people in need of humanitarian assistance.** Estimates for the number of people in need have improved over the past three years: rather than being based primarily on the number of people in appeals for humanitarian assistance, Humanitarian Needs Overviews now separate 'People in Need' from 'People Targeted' (generally a lower figure). However, the figures for people in need are often extrapolated from weak or outdated population data. In addition, different humanitarian programmes use different classifications and understandings of need, making it more difficult to create an accurate global total.
- **The number of people dying in humanitarian crises.** For a number of reasons, it is still extremely unusual for humanitarians to obtain and use data for excess mortality in a population affected by crisis, particularly in non-camp situations. This makes it impossible to say whether humanitarian activities are having, or have had, any effect on keeping people alive.
- **The costs of humanitarian response at an organisational level.** Different organisations use very different approaches to accounting for funds, and these accounts are generally closed (at least at any level that would allow serious examination) This prevents any serious consideration of actual or potential efficiencies.
- **The longer-term impact of humanitarian response.** Very little research is done on the longer-term, positive and negative impacts of humanitarian responses on the lives, societies and economies where they take place.

The State of the Humanitarian System report and the World Humanitarian Summit

For this edition of *The State of the Humanitarian System* report, ALNAP also developed two small research components to explore issues of performance (as defined by the DAC criteria) related to the WHS. The first, using a method modelled loosely on outcome harvesting, sought to understand how policy aims expressed in the WHS commitments are unfolding on the ground in two countries: Ethiopia and Lebanon. The second used an indicator-based approach to look at how the system might monitor progress on the WHS commitments, and how these indicators map onto the DAC criteria used to evaluate humanitarian performance. These elements are published separately, and more information on the methodology is available in these reports.

Endnotes for this chapter

1. FAO, OCHA, UNDP, UN Population Fund (UNFPA), UN Habitat, UNHCR, UNICEF, WFP and WHO.
2. Interviews in Colombia related both to ‘natural’ disaster and to conflict – but the majority were concerned with ‘natural’ disasters.
3. In Iraq: on quantity of aid, quality of aid, communication by humanitarians, ability to give feedback. In Afghanistan: on communication by humanitarians, ability to give feedback, being treated with dignity. There were no statistically significant differences by gender in CAR, Ethiopia or Kenya.