On the right track?

A brief review of monitoring and evaluation in the humanitarian sector

Saul Guerrero
Sophie Woodhead
Marieke Hounjet

On the right track?

A brief review of monitoring and evaluation in the humanitarian sector

About the authors

Saul Guerrero is the Senior Evaluations, Learning and Accountability Advisor at Action Against Hunger UK (ACF-UK). Over the last ten years he has supported humanitarian organisations, UN agencies and government institutions in the design, planning, implementation and evaluation of nutrition programmes. He has worked in over 17 countries in Africa and Asia.

Sophie Woodhead has a background in languages and joined Action Against Hunger's Evaluation, Learning and Accountability (ELA) Unit in 2011. Since then, she has co-edited learning papers for the Global WASH Cluster and more recently joined the inter-agency Coverage Monitoring Network (CMN) project, a collaboration between ACF, Save the Children, Concern Worldwide, International Medical Corps, Helen Keller International and Valid International.

Marieke Hounjet is the Advisor for Knowledge, Impact and Collaboration for the Consortium of Humanitarian Agencies, a collaboration between humanitarian agencies to strengthen the third sector for more effective and higher quality humanitarian aid. She has previously worked with War Child, the Overseas Development Institute and the Institute for Development Studies and holds an MSc from the London School of Economics and an MA from Leipzig University.

Acknowledgements

Rosie McAllister and Jessica Graham helped pull together the initial data set used in this review. The timely support of the Consortium of British of Humanitarian Agencies (CBHA) proved essential in taking this review to its current form. ALNAP, and in particular Francesca Bonino, infused a great deal of enthusiasm into the process, and allowed for increased collaboration through the activation of their community of practice. Tony Beck's experience and expertise ensured a robust quantitative classification and clear qualitative analysis of the indicators in Section 3. The contribution of Professor Richard Garfield is also gratefully acknowledged; this paper would not have been the same without his pragmatism and passion. The feedback provided by John Borton, Richard Cobb, Nan Dale, Maureen Gallagher, Alistair Hallam, Michael Jensen, Mariana Merelo Lobo, Jake Phelan, Elias Sagmeister and Hillary Williams proved invaluable in reshaping this review. Finally, the authors would like to thank all the agencies (ACF, ActionAid, CAFOD, CARE, Christian Aid, Concern Worldwide, Merlin, Oxfam, HelpAge, International Rescue Committee, Islamic Relief, Plan, Save the Children, Tearfund, and World Vision) whose openness and support made this review possible.

Photos: Saul Guerrero and Samuel Hauenstein Swan

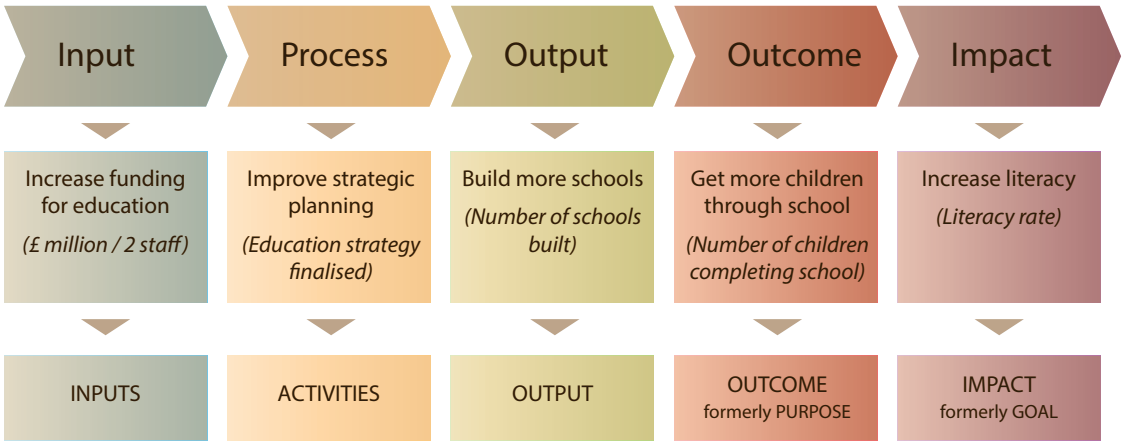
Design: Wingfinger

1 Introduction

In mid-2011, ACF International began its first systematic global reporting process, using commonly agreed objectives and indicators first developed as part of the organisation's 2010–2015 International Strategy. One challenge rapidly emerged; how to collect the data needed from over 40 field missions (under multiple headquarters) in an efficient manner. The original objective of this review was rather simple: to understand how other humanitarian organisations undertook such processes of reporting and data management. As discussions with different organisations began to take shape, however, a number of more profound questions about the past, present and future of monitoring and evaluation (M&E) in the humanitarian sector began to emerge. As such, we came to understand that the practical and the more theoretical questions were closely intertwined.

These discussions did not take place in isolation; they occurred at a time when the debate around the concept of "value for money" (promoted mainly by the UK's Department for International Development, DFID) was gathering pace. Although the notion of "value for money" was not new (the public sector has for many years used this term to define *"whether or not an organisation has obtained the maximum benefit from the goods and services it both acquires and provides, within the resources available to it"*[1]), its introduction into the humanitarian debate about quality

Box 1 Definitions of components in the result chain



Source: DFID (2011) *How to Note: Guidance on using the revised Logical Framework* (London)

The following OECD-DAC definitions were used:

INPUT: The financial, human, and material resources used for the development intervention.

ACTIVITY: Actions taken or work performed through which inputs, such as funds, technical assistance and other types of resources are mobilized to produce specific outputs.

OUTPUT: The products, capital goods and services which result from a development intervention; may also include changes resulting from the intervention which are relevant to the achievement of outcomes.

OUTCOME: The likely or achieved short-term and medium-term effects of an intervention's outputs.

IMPACT: Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.

Source: OECD-DAC *Glossary of Key Terms in Evaluation and Results Based Management*

1 University of Cambridge, *A Brief Guide to Value for Money*

2 As a DFID-funded document on "value for money" clearly states ,"At a time of reduced public expenditure and rationalization of resources, this term has gained wide currency in the formulation of economic policy imperatives. It is part of the Structural Reform agenda for DFID, and forms part of wider organizational changes initiated in DFID following the election of the Coalition Government in May 2010." (Barnett, Chris et al (2010) *Measuring the Impact and Value for Money of Governance & Conflict Programmes* (Final Report, ITAD, December 2010; p4)

1 Introduction

and performance was more novel.[2] As we continued to engage with both debates, a common question repeatedly surfaced; do we maybe need a new quality framework that promotes improved M&E in humanitarianism?

For over two decades, there has been a clear recognition (amongst humanitarian agencies and donors alike) that measuring the effect of humanitarian interventions must go beyond counting beneficiaries. When the SPHERE Project was created by a group of NGOs and the Red Cross and Red Crescent Movement in the late 1990s, their aim was precisely to *"…help improve the quality of assistance to people affected by disaster or conflict as well as the accountability of humanitarian agencies and states towards their constituents, donors and affected populations"*.[3] Since then, other initiatives and guidelines, including *The Good Enough Guide*,[4] the Ombudsman Project, the HAP Standard in Accountability and Quality[5] and the more recent Joint Standards Initiative[6] (to name but a few) have developed evolving approaches for monitoring and improving the performance of humanitarian action. These collaborative initiatives are the cause (and the effect) of the growing importance of understanding the changes brought about by humanitarian action. But to what extent have humanitarian organisations operationalised this interest in quality and introduced and mainstreamed it as part of their M&E systems? Do humanitarian organisations successfully measure the effect of their work, and what defines the extent to which this occurs? If additional efforts are needed to better capture the effect of humanitarian action, what specific gaps should these address? In short, defining how to improve our M&E systems should start by reviewing how effectively we currently do so, and the process that defines and shapes them. These questions became the focus of this review.

To answer them, the review required a tangible set of data capable of elucidating how different organisations understand and implement monitoring and evaluation. A review of all indicators used by member organisations of the Consortium of British Humanitarian Agencies (CBHA) was therefore commissioned. The analysis used a data set of more than 1,680 indicators provided by eleven (11) agencies.[7] These were classified according to the Logical Framework results chain (baseline, input, process, target, output, outcome and impact), using definitions from a DFID guidance note[8] (see Box 1). "Indicators" which did not meet basic requirements for a sound indicator (e.g. "year by year reduction of number of children excluded from mainstream school education is observed") were removed from the analysis.[9] The indicators were reviewed for quality and the overall set was reviewed for gaps. Basic statistical analysis was carried out related to indicator and sectoral focus. The indicators, however, offered little in terms of the process that led to their adoption. To complement this, an online questionnaire was shared with all 15 members of the CBHA.[10] The questionnaire contained eight basic questions about the motivation and rationale behind the adoption of specific M&E models and indicators (see Annex 1).

With the data in hand, we set out to answer two fundamental questions; how does the humanitarian sector currently use monitoring and evaluation to measure the quality of its work and how did it get here? Rather than providing a descriptive analysis of different practices in the sector, the review focuses on three major forces shaping the answers to these questions; the intentions of humanitarian organisations, the choices made along the way, and the way in which monitoring and evaluation data is collected and used to measure quality.

3 The SPHERE Project in Brief

4 Oxfam GB/ECB *The Good Enough Guide: Impact Measurement and Accountability in Emergencies* (Oxfam Publications, p6)

5 Established in 2003, HAP International is the humanitarian sector's first international self-regulatory body. The latest version of the HAP Standard in Accountability and Quality Management was published in 2010.

6 The Joint Standard Initiative integrates three of the largest quality assurance projects in the humanitarian sector; the Humanitarian Accountability Partnership (HAP), People in Aid and the SPHERE Project.

7 The CBHA members that were not included in this review either lacked strategic indicators or were in the process of developing/reviewing them at the time of writing.

8 DFID (2011) *How to Note: Guidance on using the revised Logical Framework* (London)

9 Approximately 50 indicators

10 At the time of writing, the CBHA was composed of fifteen member agencies: ACF, ActionAid, CAFOD, CARE, Christian Aid, Concern Worldwide, Merlin, Oxfam, HelpAge, International Rescue Committee, Islamic Relief, Plan, Save the Children, Tearfund and World Vision.

2 Intentions

What we want to measure

Monitoring and evaluation frameworks, more than any other component, reflect the ambitions and expectations of humanitarian organisations. They are the language chosen to communicate their aims, the progress made and the changes that are needed to achieve these aims. But what defines the type of framework that is ultimately chosen by organisations?

The questionnaire carried out with CBHA agencies focused on this question. The results showed that humanitarian organisations, like their donors, are primarily concerned with measuring their impact, and understanding who they are reaching (and therefore also excluding) through their work (Figure 1). It is the desire to measure the effects of their intervention (positive or negative, directly or indirectly, intended or untended) that guides them.

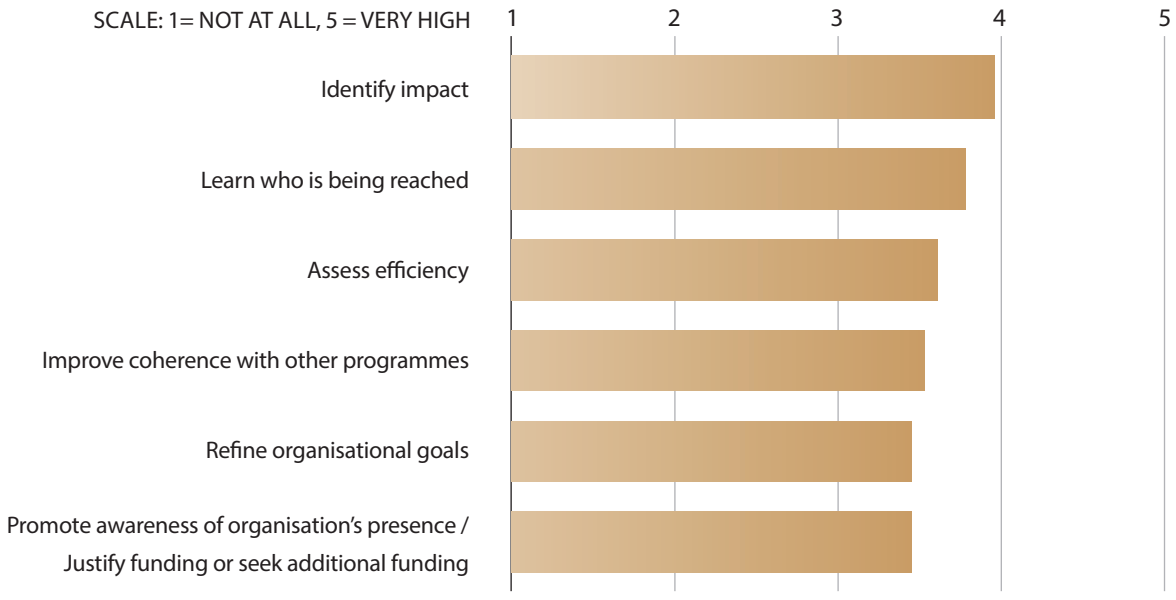
But whereas there is consensus on impact as the primary objective of humanitarian M&E systems, the questionnaire also showed that these systems and processes are seen as potentially serving a wider range of uses (see Table 1).

Four key purposes were identified:

- determining whether interventions meet the needs in relation to needs assessment, theory of change and, primarily, beneficiary perspectives
- determining the effectiveness of interventions in relation to their original design objectives and plans
- establishing levels of inclusivity and coverage
- facilitating learning.

The organisations surveyed see impact as a key objective of M&E and articulate their aspirations in ways that are consistent with this end of the result chain spectrum. Yet, organisations must choose concrete indicators to construct their M&E systems. Do the indicators ultimately selected reflect these aspirations, views and prioritisations?

Figure 1 What are your objectives in doing M&E?



2 Intentions

What we want to measure

Table 1 Question 5: If you could learn only three (3) things by doing monitoring and evaluation, what might they be?

	1	2	3
NGO 1	Testing and validating our theory of change	Developing robust value for money case	Demonstrating and communicating impact of campaigning
NGO 2	Are we changing people's lives for the better?	Is the quality consistent, and if not, where are the most significant gaps?	Are we reaching the highest possible proportion of the affected population?
NGO 3	Which interventions are rated highest by beneficaries	Which interventions are valued most by partners and staff	Which types of interventions have most enduring positive impact
NGO 4	Status of implementation	Weakness of implementation	Contribution to change
NGO 6	How effective we are	What we need to do differently	The ability of our partners to deliver on their commitments
NGO 8	How to develop a monitoring plan?	How to ensure beneficiary accountability?	How to measure impact?
NGO 10	That we have delivered against our commitments	That our assumptions on how best to facilitate positive change are correct	That the end users (patients, participants and communities) as well as partners are satisfied with our support
NGO 11	That we are responding to people's needs	That we are making a difference	We have done what we said we would do
NGO 12	What we have achieved	What our beneficiaries think of what has been achieved	How to do it better
NGO 14	How well we are doing with the resources we have and through the approaches we have chosen	How we could do better and where we can be sharing our lessons to improve programme quality across all that we do	Staff, partners and other key stakeholders are passionate about understanding their performance, the change they contribute to and identify ways for improvement
NGO 15	Has the situation for people improved?	Were the most vulnerable included?	Has the intervention led to a sustainable change?

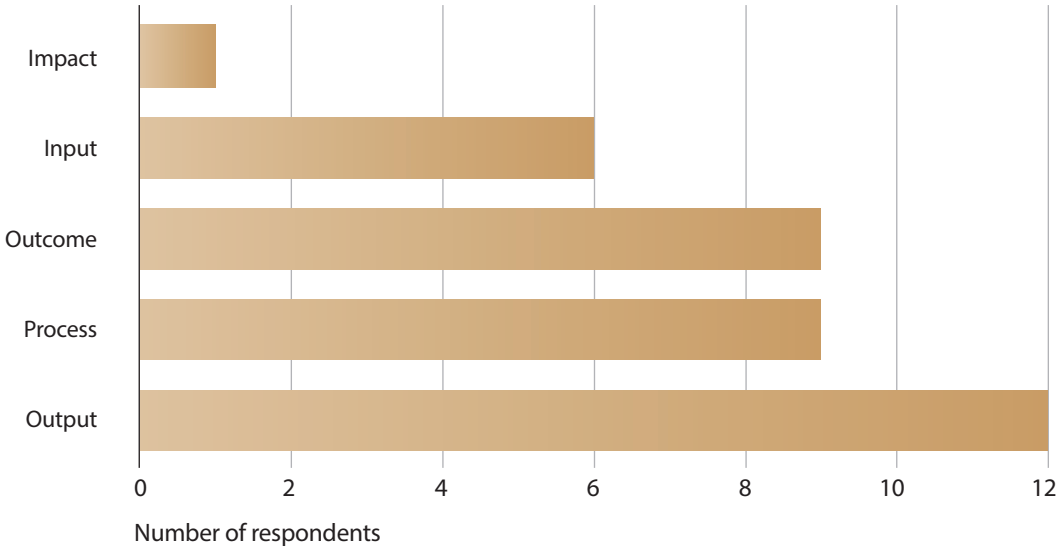
3 Choices

What we actually measure

The questionnaire measured the perception amongst organisations about what their M&E systems commonly measure. The results showed a high degree of consensus, namely that current systems are insufficient and/or incapable of measuring what they see as the main objective: impact (Figure 2). Where does this discrepancy originate from?

Part of the answer lies in the fact that monitoring and evaluation systems are defined by more than just the desires of individuals within organisations. At a global level, most organisations aim to align their monitoring and evaluations systems to their organisation's strategic priorities.[11] One of the organisations surveyed, for example, monitors poverty levels in its programme areas by evaluating changes in three different dimensions (inequality, assets and returns, and risks and vulnerabilities). Another organisation systematically evaluates women's empowerment horizontally (as a cross-cutting issue) and vertically within all programme areas. This is positive insofar as it allows organisations to link their global aims with the programmes and activities that they implement. The problem, however, is that strategic priorities are often defined by outcomes (e.g. reduction in the prevalence of under-five severe acute malnutrition) that organisations can only contribute to rather than fully achieve independently. In many organisations the selection of indicators is the result of internal pull and push forces. In the words of one of the respondents, an organisation's *"fundraising and communication teams are always keen to be able to talk about our impact. Technicians tell them that this cannot be done, and after debating this we settle on the middle ground which is around outputs and outcomes."*[12] These debates play a part, but there are more deliberate steps being taken to move M&E away from impact measurement. *"[E]ach NGO normally only has a small influence on long-*

Figure 2 What do you think your organisation's M&E indicators most commonly measure?



11 Online Questionnaire with all 15 members of the CBHA (August 21st – September 4th, 2012)

12 Ibid.

3 Choices

What we actually measure

term social change" writes Alex Jacobs, *"…[o]ther factors are normally more important, like government, politics, economic opportunities, friends and family, culture or even the weather. Ultimately, people make their own choices about their lives."* In this context, Jacobs goes on to argue, *"…NGOs can achieve most by managing and measuring their own performance, rather than poor people's long-term social change."*[13] Whether NGOs' work can lead to an attributable level of social change is open to question, but what seems clearer is that such change cannot be attributed singularly to one organisation. John Borton, for example, argues that

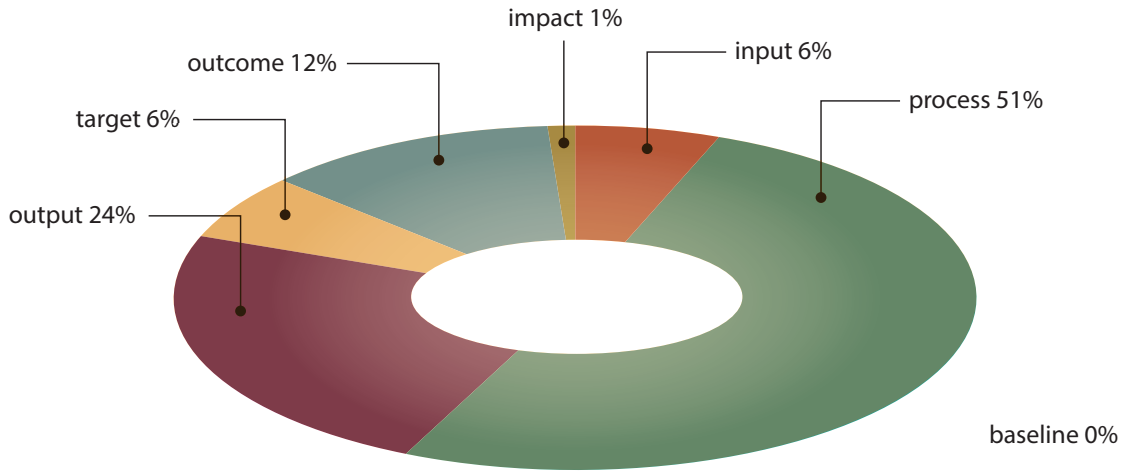
"…the impact indicators that agencies want to be measuring (such as Child Mortality Rate and Under-Five Mortality Rate) cannot be measured by agencies at programme level. They need robust studies at the meso-level (or 'operation level' to put it in plainer language). The 'problem' stems from trying to answer the questions from data at too low a level."

What can be achieved at this "lower" or field level is a key element highlighted by surveyed organisations to describe the choices made in selecting M&E indicators. The process of translating strategic indicators is not only influenced by the breadth of thematic areas that organisations cover, but also by the opportunities and challenges of collecting data at field level. As the agencies themselves state, *"normally indicators are not in dispute, it is more [to do with] how difficult they are to collect."*[14] This final stage of adopting indicators that can be regularly and easily collected effectively dilutes the original aims, and leads to the adoption of numerous individually-chosen proxy-indicators. The more sectors an organisation works in, the wider the range of indicators becomes, as *"…it is not possible to develop useful indicators to fill all programmes or serve information needs for all sectors."*[15] The language of monitoring and evaluation becomes more specialised, but from that stem individual dialects that make common dialogue increasingly difficult.

Adapting indicators to fit the realities of humanitarian organisations – from their strategic visions to their programmatic challenges – is a positive process, but one that (deliberately or not) refocuses monitoring and evaluation away from the stated interest in measuring impact. But what is the magnitude of this shift, and what is the resulting focus?

To determine this, the combined set of 1,680 indicators used by eleven (11) CBHA members at a global level was reviewed and classified. The classification found that differentiating

Figure 3 Classification of indicators by type (full dataset, 11 organisations)



13 Alex Jacobs, writing on http://ngoperformance.org/performance/

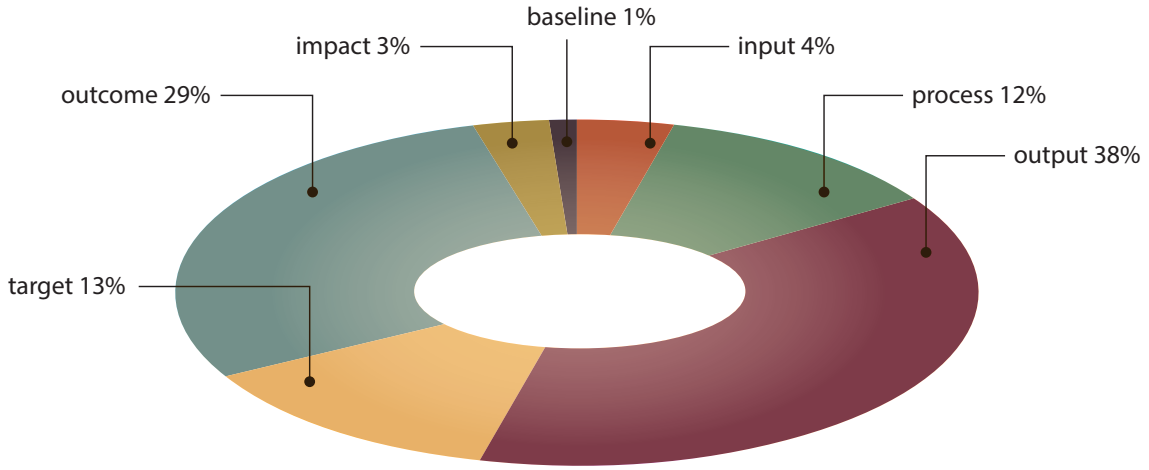
14 Op.cit.

15 Ibid.

3 Choices

What we actually measure

Figure 4 Classification of indicators by type (10 organisations)



outputs from outcomes is often difficult, and generally agency- and context-specific.[16] The biggest limitation or issue affecting the analysis came from the dataset itself; a single organisation provided 980 indicators, or 58% of the sample, mainly at the process level, whilst another organisation provided only nine indicators. This had an impact in the way the analysis was carried out. Figure 3 illustrates the breakdown by type of indicator for the full data set. As can be seen the data is distorted by the large number of process indicators from one NGO (51% of the sample).

Figure 4 therefore shows classification of indicators with the NGO which provided 980 indicators removed.

Half of the indicators[17] in Figure 4 fall in the mid-range of the results framework, either output (38%) or process (12%). The prominence of this type of indicator suggests a prioritisation of measuring what is tangible over what is meaningful and useful (see Section 2). There are relatively few impact indicators (3%), but outcome indicators make up almost one third of the sample (29%). Many of these, however, are often classified as impact; some organisations, for example, explicitly classify some indicators (e.g. net enrolment ration in primary education, or HIV prevalence in specific age-groups) as impact indicators, whereas DFID specifically classifies them as outcome indicators. Whatever the case, the shift towards the adoption of indicators at the impact end of the spectrum is an interesting trend, as outcomes are more useful for measuring actual results, but, conversely, also offer more challenges for data collection. As the ALNAP 2012 *State of the Humanitarian System* report notes:

"…outputs, while easier to measure, can be misleading as indicators. For example, seeds delivered after the end of planting season are of little value, regardless of whether they were procured at low cost. Outcomes and impact are obviously much better indicators of value, but they are also much more difficult to measure, especially in continually fluctuating humanitarian contexts. They are also subject to considerable debate, especially about whether objectives should be defined in the short or long term."

16 This is generally a challenge in Result-based Management and not specific to this data set.

17 Figure 4 sets out the classification of indicators by sector, for 10 NGOs (i.e. for the data set excluding the NGO that provided 980 indicators). The most frequently used indicator at 31% is health; followed by a general category of indicators not classifiable by sector at 19%; followed by advocacy at 11% (e.g. "Percentage of partners demonstrating increased engagement and influence in policy and advocacy work"; "Change in local people's skills, capacity and knowledge to mobilise and advocate on their own behalf"). Two other main humanitarian sectors, food/nutrition and WASH, were referred to in 9% and 6% of indicators respectively. Protection was rarely covered, and, perhaps more surprising, there were almost no shelter-related indicators included in the sample; it is difficult to come up with an explanation for the latter without further analysis of the types of emergencies being covered.

3 Choices

What we actually measure

The fact that impact indicators remain limited may reflect the intrinsic difficulties in measuring impact. But what it does serve to confirm is that monitoring and evaluation frameworks used by humanitarian organisations are incapable of providing organisations with the means by which to measure what they are most interested in.

Even if the indicators used cannot be classified as impact indicators, do they still provide sufficient information to assess quality? The analysis found three main limitations in this regard:

1. Limited disaggregation by social variables

Only 37 of the total sample of 1,680 indicators (about 2%), were disaggregated by sex, making disaggregation the single biggest gap found in the review. ALNAP syntheses of evaluations of humanitarian action have consistently demonstrated weak performance of humanitarian actors in relation to gender. The reasons for this lack of disaggregation should be analysed, as without systematic sex-disaggregation it will be difficult to determine the gender-equality effects of an intervention. The mandatory gender marker now introduced for the consolidated appeals process may lead to improvements in this regard.

There was also very limited disaggregation by age, ethnicity or disability. While clearly not every indicator should or can be disaggregated by every social variable, this and the above suggest that organisations are missing key areas of importance for humanitarian results, and there may be insufficient sense of which groups are receiving support. There is always a trade-off in M&E in humanitarian action between the need for timely and accurate information, and the time and effort needed to collect and analyse data. However, the very limited attention to social variables suggests the need for greater investment in relation to these in M&E systems.

2. Lack of qualitative indicators

A qualitative indicator is a direct perception or view of an affected person/beneficiary (e.g. satisfaction with services received). Most indicator guidance recommends a combination of quantitative and qualitative indicators; however, the NGO sample included 27 qualitative indicators (<2%). This can be interpreted as a sign that data collection is currently too top-down; the affected population needs more participatory input into agencies' M&E processes, and qualitative indicators are a useful way of fostering this.

3. Lack of attention to cross-cutting issues

(e.g. protection and gender-based violence)

These are general gaps in humanitarian action, and are reflected in the indicator set. However, as work in these areas increases, the range and utilisation of these indicators could increase.

Overall, around 70% of the sample was found to be of reasonable quality, in terms of the key elements of a good indicator: clear phrasing, being able to measure a result statement, being at the appropriate level in the results chain, and not including an activity and an indicator in the same sentence. In terms of the often used SMART metric (specific, measurable, achievable, relevant and time-phased), very few of the indicators referred to a specific time period, and about half met the other four criteria.

The review of indicators shows the extent to which the language of quality is broad and expanding beyond basic minimum standards.[18] The inclusion of many indicators for sectors which have emerged in the last decade – such as livelihoods, agriculture and DRR, which between them were mentioned in 22% of indicators – shows the willingness to expand definitions of quality to new areas of work and develop an appropriate language for it. But it is the diversity of indicators within the same sectors that is most obvious and ultimately more problematic. The use of different indicators within a common sector creates an elaborate language that makes comparative discussion more challenging, steering the sector (and donors) away from a clearly defined and agreed nomenclature to carry out a more meaningful and transparent dialogue about what works and what doesn't.

18 The SPHERE Project (2011) *Humanitarian Charter and Minimum Standards in Humanitarian Response* (Third Edition)

4 Utilisation

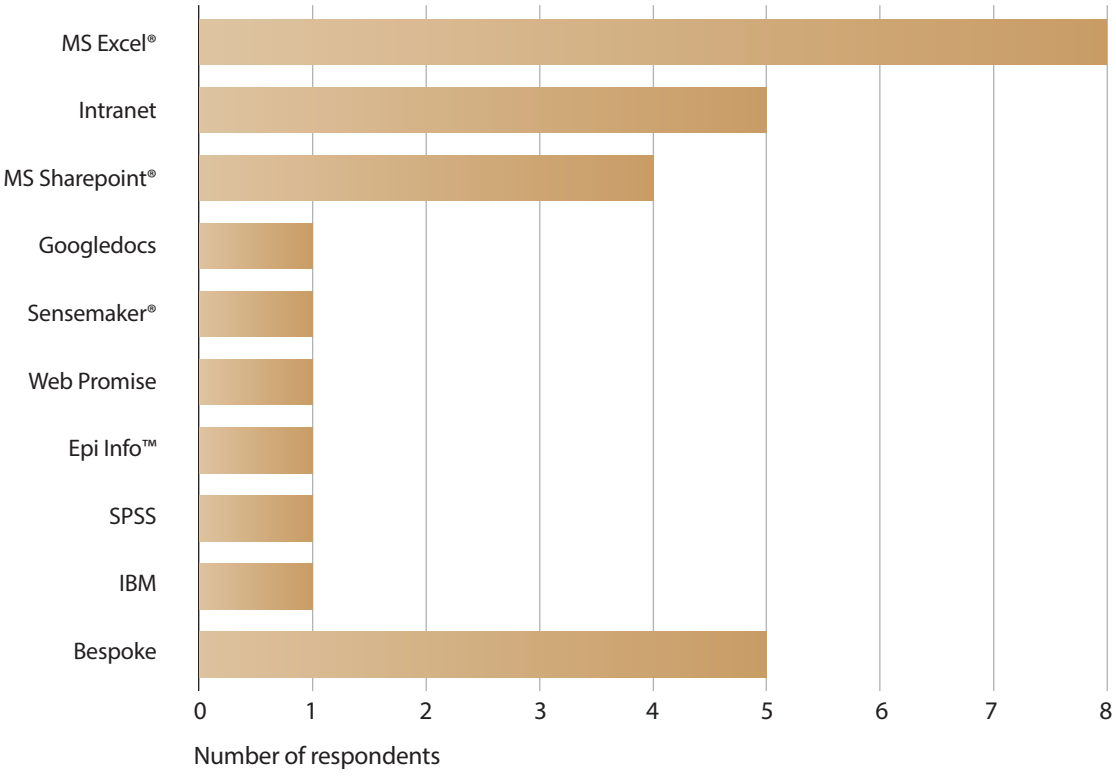
What we do with it

Humanitarian organisations want to measure the impact of their work. Yet in reconciling monitoring and evaluation systems with their own strategic priorities and the realities of what field programmes can (and should) document, an ever-expanding set of (proxy) indicators has been adopted. The third and last element that influences M&E in the humanitarian sector comes from the manner in which organisations address the act of monitoring and evaluation itself.

Though seldom openly acknowledged, upward accountability remains one of the primary driving forces behind monitoring and evaluation in the humanitarian sector. Much of the dialogue that occurs between humanitarian organisations and their donors is based on the (input, output and process-related) data collected through monitoring and evaluation systems. This creates a vicious cycle; humanitarian organisations continue to collect and feed back data that describes what they do (but says little of the quality and impact of its interventions) and donors, deprived of reliable means by which to carry out comparative analysis, are unable to make quality a more central pillar of their funding decisions. Both parties may want to change this, but so far neither has had sufficient incentives to do so.

Internally, ownership of monitoring and evaluation processes within humanitarian organisations also prevents

Figure 5 What type of IT platforms are used to store and analyse M&E data?



4 Utilisation

What we do with it

more meaningful analysis and change. For the large majority of organisations surveyed in this review, monitoring and evaluation is a standalone component that sits, often uneasily, somewhere between management and operations. About half of the organisations surveyed agree that monitoring and evaluation should be mainstreamed and made part and parcel of the way organisations operate. But all agree that this is far from the present state of affairs. Part of the problem may lie in the need to reconcile organisational aims and strategies and the (proxy) indicators that are chosen. Monitoring and evaluation advisors effectively act as translators, bridging the gap between what their indicators say (mostly about inputs, outputs and processes) and what organisations are actually interested in (outcomes and impact).

The fact that more specialised and complex monitoring and evaluation systems require specialist staff does little to facilitate the integration of monitoring and evaluation across all levels of an organisation. But the platforms to analyse monitoring and evaluation data also prevent more widespread engagement and utilisation of the information available. A high proportion of the organisations surveyed during this review continue to rely on low-tech, spreadsheet-based tools to record, collate and analyse their monitoring data (Figure 5).

Whilst other organisations have adopted mid and high-tech platforms that enhance data storage, the visualisation of this data remains limited. *"Whilst the system holds a lot of data"* one organisation stated, *"programme staff have found it difficult to access this in a meaningful way."* And this raises a second point; humanitarian staff (in particular those at field level) tasked with collecting and reporting the information needed to make these systems work, struggle to relate to the processes which they support. In the words of one of the surveyed organisation, *"…the people who really need to react to monitoring to adjust implementation do not, as yet, see the value of a global system for their purposes. As it stands, the value is seen most at national and international levels…"* This affects not only on the ability of monitoring and evaluation data to positively influence programmatic behaviour, but also the rigour and quality of the information provided.

Part of the responsibility lies in the use of (proxy) indicators (e.g. how much are we providing?) that say little about the questions that staff, at all levels, want to know (e.g. what difference are we making? How do we know? [19]). But the limited analytical capacity of reporting systems also detaches data reporting from data analysis; programme staff "talk" to the systems but the systems are not designed to talk back. Moreover, it is the absence of any real, tangible incentives that ultimately undermines the use of monitoring and evaluation systems. Humanitarian staff know that the indicators being measured say little about the quality of programmes, and that quality seldom influences the ability of organisations to continue their operations, because funding is not conditioned by what these indicators say. And this is the key point; what is needed is not only a reconsideration of how organisations collectively define quality, but a framework that helps to measure and reward it.

19 Oxfam GB/Emergency Capacity Building Project (2007) *The Good Enough Guide: Impact Measurement and Accountability in Emergencies* (Oxfam Publications, p4)

5 Conclusions

The review identified a number of areas and issues on which humanitarian organisations are in agreement. The clearest, perhaps, is the shared intention and interest in measuring the impact of their work and the simultaneous acknowledgement that current M&E systems do not facilitate impact measurement. At a time of increasing pressure for greater transparency and accountability, this is important to recognise, given its potential for positively influencing the shape of future discussions.

Diversity of indicators, even within individual sectors and areas of work, was another important result of the review. Organisations consistently rely on a wide range of (mostly) output indicators which are rarely disaggregated, thus limiting the depth and inter-agency comparability of the data collected. The review also found limited use of qualitative indicators and beneficiary input, as well as information management systems that intrinsically limit data visualisation and utilisation by field staff. Agencies clearly design and re-design their internal indicators within their own organisations, networks and partners. Collaboration is dispersed, and although there are some consortia and platforms focusing on these issues, comprehensive collective action is largely invisible. Changes in language will be required and this review highlights the need for a move towards a more common set of indicators. The goal is not to reinforce 'upward accountability', but to allow humanitarian agencies to benchmark their performance. The work of the BOND Effectiveness Group, and their common indicators according to themes, is a positive step in this direction. Other groups, including CDC in Atlanta, are also investing heavily in understanding the current levels of use of SPHERE standards. The sector's detailed knowledge of what can and should be measured is increasing. What is needed are more decisive moves on the part of organisations to adopt some of these indicators.

Greater adoption of common outcome and impact indicators is necessary, if not at an individual organisation's level, then certainly to evaluate collective humanitarian efforts. But these must be part of an M&E framework that also carefully assembles a range of other indicators to provide a more complete picture. As Alistair Hallam notes, *"…process and output indicators can be extremely meaningful and useful, providing one has a model of how they lead to impact. There is a time and place for impact and a time and place for process/outcome monitoring… [t]he key thing is to have an appropriate theory of change."* It is the M&E model rather than its indicators, that may ultimately define the capacity of humanitarian organisations to talk about their contributions to people's lives. Changing and strengthening these M&E models needs to happen and creating the right enabling environment for such change is key.

Incentives are a significant part of this environment. Refocusing humanitarian M&E systems to produce more meaningful data has to be linked to a system-wide rethink of the incentives for providing such data. Including the requirement for data disaggregation in the CAP appeals will likely lead to more disaggregated data, but what could be the incentive to involve beneficiaries more? Hence we need to remember that the incentives are not the real motivation for change. The real motivation must be getting closer to that desire for evaluating impact. *"If we change the incentives so that we are carrying out M&E for purposes of improving quality and learning"* writes Hallam, *"…then the indicators will, by and large, follow. Doing it the other way around will not produce the same results."*

Annexes

Annex 1: M&E review questionnaire format

The following questions relate to Monitoring & Evaluation (M&E).

1. What do you think your organisation's M&E indicators most commonly measure? (please tick as many as needed)

a. Input
b. Activity
c. Output
d. Impact
e. Outcome

2. What are your objectives in M&E? (1 not at all, 5 very high)

Response choices	1	2	3	4	5
a. learn who they are reaching					
b. identify their impact					
c. assess their efficiency					
d. improve coherence with other programs					
e. refine organizational goals					
f. promote awareness of our organization's presence justify funding / seek additional funding					

3. How were these indicators developed within your own organisation?

4. How much guidance would you like to receive from other parts of the organisation in developing/updating your indicators?

5. If you could learn only 3 things by doing M&E, what might they be?

6. In general, how much of project efforts should go into M&E during the life of a project?

7. In practice, how much do efforts mirror/differ from those described above? What elements are key in determining this?

Annex 2: Impact indicators in the review

Total sample, n=1,680

Sector	Specific Indicator	Indicator	Agency
General	Population living below $1.25 a day (PPP) – mean value for 32 countries	Impact	NGO2
General	Inequality-adjusted Human Development Index (HDI) – mean value for 32 countries	Impact	NGO2
Gender	Gender Inequality Index – mean value for 30 countries	Impact	NGO2
General	Proportion of population below $1 (PPP) per day	Impact	NGO3
General	Prevalence of underweight children under 5 years of age	Impact	NGO3
Gender	Change in women's self-efficacy	Impact	NGO3
General	Under-five mortality rate	Impact	NGO3
General	Infant mortality rate	Impact	NGO3
Gender	Maternal mortality ratio	Impact	NGO3
Health	HIV prevalence among population aged 15-24 years	Impact	NGO3
Health	Child Morbidity	Impact	NGO4
Health	HIV prevalence among young people aged 15-24	Impact	NGO4
Health	HIV prevalence among general population	Impact	NGO4
Protection	A reduction in the incidence of a specific form of GBV – to be measured only when implementing a specific GBV prevention and response programme	Impact	NGO4
Gender	Increase in women's (or other diversity group) control over important resources (land, cash, livestock, other assets...) at a household level (FIM)	Impact	NGO4
Gender	Gender Parity Ratio	Impact	NGO4
Protection	Rights fulfilled/ no longer violated	Impact	NGO4
General	Beneficiary well-being	Impact	NGO06
General	Lives saved: Number of lives saved using LiST tool at national level (from Countdown reports)	Impact	NGO10
Education	Learning outcome: % of Basic Education students in a representative sample schools/sites supported by SC, that achieve mastery of literacy in the language of instruction (P)	Impact	NGO10

www.actionagainsthunger.org.uk

www.thecbha.org