van Voorst *et al.*
Journal of International Humanitarian Action (2022) 7:20
https://doi.org/10.1186/s41018-022-00129-8

Journal of International Humanitarian Action

RESEARCH ARTICLE

Open Access

Humanitarian aid NGOs' accountability towards large donors: the case of the European Union's DG ECHO

Stijn van Voorst[1*], Sandra L. Resodihardjo[1] and Andrea Schneiker[2]

Abstract

Directorate-general ECHO of the European Commission is one of the largest humanitarian aid donors globally. Projects which it funds are often implemented by its NGO partners. This article studies how ECHO's system for assessing such projects' final results works and to what extent it provides useful information to the donor. Theoretically, it seems likely that evaluative data produced in this context is of little value, given the financial interests of the NGOs which submit the information, methodological issues related to collecting it, and the donor's limited capacity to process it. However, based on in-depth interviews and document analysis, we conclude that ECHO usually has sufficient human resources to analyse reports which NGOs submit to it. These documents are also informative about projects' direct effects, but they seem less capable of assessing long-term impacts. Furthermore, such reports seem less important to ECHO than the field visits which it conducts. These findings imply that consistent monitoring of humanitarian aid projects on the ground helps to mitigate the main weakness of a system of self-evaluation by NGOs.

Introduction

The European Union (EU) presents itself as 'one of the largest humanitarian donors' in the world (DG ECHO 2019: 2). Indeed, the OECD (2019: 70) shows that the EU's institutions are the fourth biggest donor globally, behind only the USA, the UK, and Germany. In absolute numbers, the EU spent about $16.4 billion on humanitarian aid in 2018 (OECD 2019: 70). These resources are almost entirely managed by the Directorate-General for European Civil Protection and Humanitarian Aid Operations (DG ECHO) of the European Commission.

DG ECHO does not spend this money by directly providing aid. Instead, it funds humanitarian actions performed by officially recognized partners. Some of these are governmental in nature, like international or national agencies (Broberg 2015: 254; DG ECHO 2019: 5). However, most partners belong to a group of, at the time of our research,[1] 209 non-governmental organizations (NGOs) (DG ECHO 2020; Pusterla and Pusterla 2020: 2).

To control how NGOs spend the grants that they receive, ECHO uses multiple mechanisms. Some of these provide ex-ante checks: NGOs must prove that they meet various criteria before being allowed to sign contracts with ECHO. Other mechanisms can be labelled ex-durante: the monitoring of NGOs' ongoing actions. A final group of mechanisms takes place ex-post: assessments of completed projects (DG ECHO 2019: 51).

This article focuses on the ex-post mechanisms aimed at projects' humanitarian effects (rather than their financial aspects). These encompass final narrative reports about results of completed activities which NGOs must submit to ECHO, as well as any in-depth evaluations attached to them (Broberg 2015: 263; DG

*Correspondence: Stijn.vanvoorst@ru.nl

[1] Department of Public Administration, Nijmegen School of Management, Radboud University, Heyendaalseweg 141, 6525 AJ Nijmegen, The Netherlands
Full list of author information is available at the end of the article

[1] The number of ECHO's NGO partners differed only slightly in 2014. We found no earlier figures.

ECHO n.d.). Since ECHO's subsidies to NGOs are primarily intended to help (potential) victims of disasters (Pusterla and Pusterla 2020: 2), knowing to what extent funded actions achieved that effect is particularly important for both the donor and the taxpayers which it ultimately represents.

No research has yet been conducted about how (ex-post) assessments of NGOs' activities funded by ECHO work in reality. This article seeks to fill that gap. Based on the literature on aid work, there are two reasons to expect that such ex-post assessments do not fully inform the donor. First, existing academic literature (e.g. Clements 2020; Wenar 2006) suggests that in other humanitarian aid contexts, assessing NGOs' performance is challenging. Reasons for this difficulty include actors' strategic interests in positive evaluations (Clements 2020: 2; Wenar 2006: 19) and methodological problems encountered when studying aid in developing countries (Edwards and Hulme 1996: 968; Suárez and Gugerty 2016: 2622). Second, ECHO has limited human resources. Its total staff (headquarter and field offices combined) consisted of about one thousand people in 2018 (DG ECHO 2019: 70); these professionals coordinate a broad range of policies, including civil protection within the EU, a network of aid volunteers, and projects implemented via cooperation with other international organizations or national agencies. It can be expected that this limitation makes it difficult to keep track of how projects from over 200 NGOs perform, as such work is likely to be time consuming.

Considering the importance of knowing how well public funds are spent, as well as the aforementioned potential issues with that in humanitarian aid, it is crucial to determine how informative ex-post assessments of ECHO-funded NGO activities are. This article seeks to fulfil that aim by answering two key questions. First, how does DG ECHO's system for ex-post assessments of activities implemented by NGOs function on paper and in reality? Second, to what extent do these ex-post assessments properly inform ECHO about the results of activities which it funded?

The relevance of answering these questions is twofold. First, it provides an up-to-date picture of how ECHO — one of the largest humanitarian aid donors in the world — manages one of its core activities in practice. The existing academic literature about the organization is mostly outdated (e.g. Ascroft 1999; Mowjee 1998), focused on its relation with the EU's member states (e.g. Attinà 2016; Pusterla and Pusterla 2020; Versluys 2007), related to the EU's general crisis management (Boin et al. 2013) or written from a normative perspective (Broberg 2015). Conversely, this article presents empirical data about the main channel through which ECHO currently provides aid: its partnerships with NGOs.

Second, this article contributes to the theory about how and to what extent (international) governmental organizations can exercise control over NGOs working for them. Just like ECHO, many of these actors rely on NGOs to implement aid projects (Suárez and Gugerty 2016: 2619; Wenar 2006: 9). Thus, our work provides an interesting case study about if and how an (international) governmental organization dependant on NGOs can assess if such policies are working well. In particular, our article highlights tools which ECHO has in place to prevent overreliance on the potentially biased self-reporting by NGOs.

The next section summarizes how DG ECHO formally operates, with a particular focus on the control mechanisms aimed at NGOs related to our main questions. The 'Theoretical framework' section presents a theoretical framework about factors which may limit how useful such mechanisms are for informing donors. The methodology used in this study, which is based on a combination of document analysis and fourteen in-depth interviews with various stakeholders, is described in the 'Methods' section. The 'Results' section presents our results; the final section discusses our conclusions.

ECHO's formal control mechanisms regarding NGOs

ECHO was created in 1992 with the main purpose of responding to natural and man-made humanitarian crises (Ascroft 1999: 179). Over time, it has incorporated related tasks as well, such as managing the Emergency Response Capacity Centre and preventing disasters both within and outside the EU (decision 1313/2013 of the European Parliament and the Council). These tasks are mainly managed by ECHO's directorates A (emergency management and rescEU) and B (disaster preparedness and prevention). A deputy director-general manages the remaining directorates C (neighbourhood and the Middle East), D (sub-Saharan Africa, Asia, Latin America, and Pacific), and E (general affairs), which coordinate most relations with NGO partners in various parts of the world (DG ECHO 2021a).

DG ECHO (2019: 51) lists seven mechanisms for controlling these NGO partners. The first of those focuses on assessing entire organizations: to receive funding from ECHO, NGOs must sign its Framework Partnership Agreement (FPA), which requires going through an extensive selection process. NGOs which have signed the FPA are subjected to periodic assessments to verify if they still meet the aptitude criteria (Broberg 2015: 262).

Another three of the seven control mechanisms listed by DG ECHO (2019: 51) are (2) using needs assessments to judge what kinds of projects to fund, (3) assessing partners' proposals for individual projects leading up to

Specific Grant Agreements (SGAs), and (4) monitoring ongoing projects (Broberg 2015: 261). Typically, funded projects last for up to a year, although their duration varies.[2] In 2019, DG ECHO (2019: 3) aimed to spend 35% of its budget on direct cash transfers and 10% on education during crises; other types of funded activities include delivering food, medicine, and shelter.

The three remaining mechanisms listed by DG ECHO (2019: 51), which focus on ex-post control, are (5) checks on eligibility of expenditures, (6) financial audits, and (7) (substantive) evaluations/reviews (DG ECHO n.d.). As explained in the introduction, our study focuses on the seventh category, which can be defined as assessments of if and why actions have (not) achieved their intended results in terms of helping beneficiaries of aid.

How do these 'evaluations/reviews' formally work? When completing an action which was (partly) funded by ECHO, NGOs must submit a final report to that donor via a standardized digital 'Single Form'. This report includes a financial and a narrative part, only the latter of which is relevant for this article. The final narrative report mainly presents information about the extent to which an activity has achieved its objectives, difficulties encountered during implementation, how resources were used in relation to the objectives, and lessons for future projects (DG ECHO n.d.).

NGOs are required to substantiate final (narrative) reports via annexes. In particular, ECHO encourages them to attach an in-depth evaluation of the project produced by external consultants. If ECHO has approved the Terms of Reference beforehand, the costs of such endeavours can be entirely renumerated (DG ECHO n.d., 2014: 31). However, NGOs are not formally obliged to organize evaluations in this form (DG ECHO n.d.).

ECHO or any third party mandated by it may also evaluate an action. NGOs are legally obligated to assist in such assessments and to keep relevant documents available for that purpose (DG ECHO 2014: 32).

Theoretical framework

The need for NGOs to inform donors (like ECHO) tends to be labelled 'upward accountability' in the academic literature (e.g. Edwards and Hulme 1996: 967; Suárez and Gugerty 2016: 2620; Wenar 2006: 15). Researchers often mention evaluations as potential tools to enhance such accountability, as they provide organizations which fund aid with data about how their money is spent (e.g. Clements 2020: 4; Suárez and Gugerty 2016: 2622). However, the literature also identifies at least three issues with the use of such assessments to inform donors. These problems will be discussed in this section.

A first issue is positive bias: reports may attribute more success to aid projects than is warranted, for example by emphasizing positive results over negative findings and/or uncertainties. This phenomenon may occur when actors try to protect their interests (Aerni 2006: 29; Wenar 2006: 19). For example, organizations which deliver aid (like NGOs) can increase their fundraising if they can point to a stream of positive assessments, as such reports can strengthen donors' feelings that they are contributing to something good (Aerni 2006: 29). Organizations which fund aid may also have an interest in positive evaluations, which can be used to convince taxpayers and/or member states that the endeavour is worth their investment (Aerni 2006: 29). Empirical research shows that positive bias plays a strong role in aid projects in developing countries (e.g. Clements 2020: 6-7).

Given the main questions of this article, we focus on positive bias rooted in the interests of NGOs. In the context of ECHO, we expect this phenomenon to significantly affect ex-post assessments of individual projects. As was explained in the 'ECHO's formal control mechanisms regarding NGOs' section, these reports are primarily produced by NGOs. In the case of such self-evaluations, authors have both the interest and the means to selectively present results. Even when NGOs back up their reports with in-depth evaluations produced by consultants, these external parties are likely to expect that their future employment may depend on favourable reviews, which can be an incentive for them to also draw positive conclusions (Clements 2020: 4; Wenar 2006: 19).

The degree of positive bias is likely to vary between NGOs, since some of them have a culture focused on evaluation and learning which may mitigate the problem (Mitchell 2014: 610). In particular, research shows that NGOs are more likely to develop and communicate about factual evidence when they are required to do so by the countries in which they provide aid or by the governmental organizations which fund them (Mitchell 2014: 619). However, even if positive bias is not an issue in every case, it could still occur often enough to significantly affect ECHO's access to accurate information.

Hypothesis 1: positive bias significantly limits the extent to which ex-post assessments of projects funded by ECHO are informative for that donor

A second issue that could affect the extent to which ex-post assessments of activities implemented by NGOs are informative for donors is methodological problems (Mitchell 2014: 608). It can be particularly difficult to study causal links between aid projects and their (intended) effects, because many other factors, such as

[2] This information about duration and cost was received via e-mail from an anonymous ECHO official.

political stability and economic growth, can also influence these results (Clements 2020: 4; Suárez and Gugerty 2016: 2622). Such problems are even bigger when researching long-term impacts (Aerni 2006: 29; Clements 2020: 4), since the more time passes, the more other factors will be involved. Furthermore, the quality of available data (i.e. access to relevant respondents, places, and documents) can be limited in developing countries (Mitchell 2014: 608; Wenar 2006: 20), an issue which stakeholders deem particularly likely to occur in protracted humanitarian crises (Lewis and Forster 2020: 12). Considering the fact that ECHO's ex-post assessments of individual projects aim to accurately measure results (see the 'ECHO's formal control mechanisms regarding NGOs' section), we expect the methodological problems outlined above to significantly affect them.

Hypothesis 2: methodological problems significantly limit the extent to which ex-post assessments of projects funded by ECHO are informative for that donor

A third issue that could affect the extent to which ex-post assessments of activities implemented by NGOs are informative for donors is the limited resources of governmental organizations. Such actors usually lack the means to fully control the partners which they fund (Keck 2016: 4). We expect this to be a significant issue for ECHO, given its limited human resources. For example, in 2018, it employed about one thousand people across the entire world (headquarters and field offices combined) (DG ECHO 2019: 70). These professionals coordinate a broad range of policies, including civil protection within the EU, a network of aid volunteers, and projects implemented via cooperation with other international organizations or national agencies. Thus, finite resources could make it difficult to also process all data from all reports which over 200 NGOs submit.

Hypothesis 3: finite human resources significantly limit the extent to which ex-post assessments of projects funded by ECHO are informative for that donor

In the next section, we explain ho/Paraw we assessed these three hypotheses empirically.

Methods

Data collection

Our data for this study consist of in-depth interviews and coding of documents. We will first discuss the latter method. ECHO and various NGOs informed us that the final (narrative) reports described in the 'ECHO's formal control mechanisms regarding NGOs' section are confidential and thus not publicly available. However, those actors also told us that NGOs sometimes publish the in-depth evaluations which may be attached to these narrative reports on their websites. Thus, as a first step, we collected and analysed these in-depth evaluations.

From DG ECHO's (2020) list of 209 NGO partners, we selected all 137 organizations which operate in multiple EU countries or a single English-speaking country, since we found that other NGOs almost always publish reports in languages which we cannot read. Among these 137 cases are only about 63[3] different NGOs, since the list contains many national departments of the same international organizations.

Two researchers used the search engines on the selected NGOs' websites to find ex-post assessments of actions funded by ECHO, using precise keywords.[4] We also e-mailed NGOs to ask for additional reports. This resulted in a dataset of 39 in-depth evaluation reports, which are fairly evenly distributed from 2003 to 2020 and are published by 14 different NGOs. It is impossible to tell how representative this number is, since neither conducting nor publishing such reports is mandatory (see the 'ECHO's formal control mechanisms regarding NGOs' section). For periodic evaluations (e.g. annual reports about projects which were renewed for many years), we only included the most recent edition, to prevent such cases from being present in the dataset so many times that they disproportionately affect our results.

Our remaining data consist of fourteen in-depth interviews: four with officials from ECHO (two at headquarters directorates C and E; two at field offices), seven with experts from NGOs who either coordinated all contact with ECHO or some ECHO-funded projects, and three with consultants who are sometimes hired by ECHO or its NGO partners. These conversations were needed to collect any information about the final narrative reports, since these are not publicly available. The interviews were also required to gain an in-depth understanding of how ECHO's system for assessing completed projects works in reality. During the later interviews, data saturation occurred, as no new information came to the fore. Six out of the seven NGOs which we interviewed were also present in our dataset of 39 evaluations, although that was a coincidence rather than part of our selection criteria.

Operationalization

For each of the evaluation reports that we found, we assessed if its main conclusions are positive. Based on the discussion about *positive bias* in the 'Theoretical framework' section, we expect that this will very often be the case, limiting the value of the reports for a donor like

[3] In a few cases, it is debatable if an NGO is a national department of an international organization, which is why we use the word 'about' here.

[4] These keywords included ECHO, European Civil Protection, EU, (European) Union and (European) Commission.

Table 1 Operationalization for document analysis

Criterion	Measurement	Examined parts of reports[a]
Positivity of conclusion	**1 = overall positive conclusion** Results of the evaluated project presented as overall effective, efficient, successful, relevant, appropriate, working, etc. **0 = overall negative or mixed assessment** Results of the evaluated project presented as overall ineffective, inefficient, irrelevant, inappropriate, not working, mixed, etc.	- Conclusion - Otherwise: summary - Otherwise: end results section - Otherwise: annexes
Triangulation	**1 = triangulation was applied** More than one method of data collection was used. Different methods include interviews, focus groups, questionnaires, observations, analysis of documents, etc. **0 = triangulation was not applied** One method of data collection was used.	- Methodology section - Otherwise: summary - Otherwise: introduction - Otherwise: annexes
Transparency of questions for respondents	**1 = questions are listed** **0 = questions are not listed**	- Methodology section - Otherwise: annexes
Transparency of respondents	**1 = respondent selection is transparent** Selection criteria and/or a list of contacted people/organizations present **0 = respondent selection not transparent** No selection criteria and no list of contacted people/organizations present	- Methodology section - Otherwise: annexes - Otherwise: introduction
Transparency of document selection	**1 = document study is transparent** Selection criteria and/or a list of studied documents present **0 = document study is not transparent** No selection criteria and no list of studied documents present	- Methodology section - Otherwise: annexes - Otherwise: footnotes (if they consistently list documents)
Transparency of other data collection methods (e.g. observations, existing data)	**1 = other methods are transparent** Selection criteria and/or lists of the cases studied with other methods present **0 = other methods are not transparent** Selection criteria and lists of the cases studied with other methods not present	- Methodology section - Otherwise: annexes
Transparency of evaluation criteria	**1 = criteria are transparent** **0 = criteria are not transparent**	- Methodology section - Otherwise: introduction - Otherwise: summary - Otherwise: tables - Otherwise: annexes

[a] We also searched the reports for these keywords: 'interview', 'questionnaire', 'survey', 'respondent', 'document', 'file', 'observ*', 'visit', 'walk', 'dataset', and 'criter*'

ECHO to distinguish between projects. We also measured the *methodological quality* of the reports via various criteria. First, we checked if they include data collected with multiple methods (triangulation), since being able to double-check information increases the validity of a study. Second, we assessed if the reports provide sufficient information for the evaluation to be replicable. Depending on the methods used, this could be lists of, for instance, questions, respondents, documents, and observation sites, or clear selection criteria for these aspects. Finally, we checked if the evaluation criteria (e.g. effectiveness, efficiency) are described. Limitations of this entire approach are addressed in the next subsection.

Table 1 summarizes the operationalization presented above. During the study, a more elaborate scorecard was used, which can be received by contacting the first author. All reports were assessed by multiple coders. First, two people completed a pilot study of ten reports, based on which the scorecard was improved. The updated version was subsequently used by two researchers, one of which had not been involved in the pilot, to assess all 39 reports. During the second phase, intercoder reliability was found to be sufficient for all criteria.[5] In the rare cases where the coders had different opinions, they discussed the matter until they reached an agreement about which score to use for the final results.

During the interviews, we first asked descriptive questions to find out if ECHO's system for ex-post assessments of projects works exactly like it is described in official documents (see the 'ECHO's formal control mechanisms regarding NGOs' section). We also asked respondents to give an overall

[5] Scores for Cohen's kappa were all above 0.550: positivity of conclusion (0.607), triangulation (0.879), transparency of questions for respondents (0.683), transparency of respondents (0.676), transparency of document selection (0.841), transparency of other data collection methods (0.800), and transparency of evaluation criteria (0.552).

assessment of how informative they think the final narrative reports, as well as external evaluations attached to them, are for ECHO to know how projects performed in reality.

Furthermore, we discussed with respondents the factors derived from academic literature (see the 'Theoretical framework' section). Regarding *positive bias*, we asked them to what extent they thought NGOs present findings in a favourable light whenever possible in the final reports and attached evaluations which they submit to ECHO. We also discussed to what extent they believe ECHO takes such reports into account when deciding about proposals for follow-up projects, to assess whether or not the system encourages bias.

Regarding *methodological problems*, we focused on the two issues explained in the 'Theoretical framework' section: access to data and the ability to establish causal links between projects and intended results. We asked respondents to what extent they think these problems affect the quality of the information which NGOs submit to ECHO in the final reports and any annexed external evaluations. Regarding *human resources*, we asked respondents to what extent they believe ECHO has the capacity to process all these documents.

We only asked questions to respondents who had knowledge about them. For instance, during the interviews with consultants, we did not discuss the final narrative reports, since these documents are only viewed by ECHO and NGOs.

Limitations of our methodology

We selected the criteria for the document analysis described above because they could be efficiently measured for several dozens of (sometimes lengthy) reports. Therefore, they probably do not fully measure the methodological quality of the evaluations, nor do they completely assess positive bias. Furthermore, we might not have found all reports, since NGOs are not obliged to make them publicly available.

Due to these limitations of the document analysis, we relied on the interviews as our main data source. That method has its own problems, since respondents might not always answer honestly (Vennix 2019: 218). Socially desirable replies are especially likely when asking about potentially fraudulent behaviour, like we did regarding positive bias. To reduce this risk, we promised all respondents full anonymity. Furthermore, we deliberately spoke to a broad range of actors: ECHO's headquarters, ECHO's field offices, larger NGOs, smaller NGOs, and consultants from various countries. Thus, we were able to triangulate data not just between the documents and interviews, but also within the latter method.

Results

Overall assessment

The interviews revealed that the final narrative reports submitted to ECHO function roughly like described in official documents: using a mostly standardized digital 'Single Form', NGOs provide information on objectives achieved, difficulties encountered during implementation, how resources were used, and lessons learned (see the 'ECHO's formal control mechanisms regarding NGOs' section for details). A few respondents criticized these reports, mainly for the reasons explored in the next subsections. However, their overall assessment of this tool was quite positive. Every respondent working for an NGO thought it made sense that they had to submit (most of) the information requested in the final narrative report, since ECHO must be able to demonstrate to the Council and the European Parliament that it has properly spent taxpayers' money.

The 'ECHO's formal control mechanisms regarding NGOs' section described that NGOs may substantiate the final narrative report with an external evaluation, which ECHO funds under certain conditions. Although none of the respondents from ECHO could tell exactly how often this happens, they consistently stated that this instrument was only used in a minority of cases. External evaluations only seem to be common for very large or unique types of projects. In most other cases, NGOs substantiate the final narrative report with an in-depth evaluation conducted internally or with no in-depth evaluation at all — although it should be stressed that in the latter case, substantiating the report with many other annexes is still required.

Consistent with the relative rarity of the in-depth external evaluations is that only about half of all respondents consider them to be useful for informing ECHO. Others argue that money could be better spent on the actual aid than on hiring consultants to write such documents, especially because of the methodological flaws discussed below.

The 'ECHO's formal control mechanisms regarding NGOs' section explained that ECHO also has the right to conduct its own evaluations. Our interviews revealed that in practice, this (almost) never happens for individual projects implemented by NGOs. Instead, such evaluations consistently focus on larger regions or themes, including how various types of humanitarian aid funded by ECHO work for specific kinds of populations and crises, as well their relation to long-term development aid. These broad reports are outside of the scope of our main question.

Table 2 Scores for individual criteria on which 39 evaluation reports were assessed

Criterion	Number of reports where the criterion is relevant (i.e. contains the relevant element/method)	Number of relevant reports with positive score	%
Positivity of conclusion	33	29	88
Triangulation	39	35	90
Transparency of questions for respondents	37	10	27
Transparency of respondents	37	13	35
Transparency of document selection	22	8	36
Transparency of other data collection methods	30	12	40
Transparency of evaluation criteria	39	34	87

Positive bias

All respondents with knowledge about the final narrative reports state that they provide some room for positive bias, in the sense that the NGOs which submit them can try to present projects' results in a way that is favourable to them. Although this is unlikely to immediately affect their chances of future funding, since decisions for a new round of grants have often already been taken when the reports are finalized, there can still be an incentive to leave a good long-term impression on ECHO.

Regarding the in-depth evaluations which may be attached to final narrative reports, Table 2 shows that out of the 33 cases studied by us which included a clear conclusion, almost 90% were mostly positive about (usually) the project's relevance, coherence, effectiveness, efficiency, impact, and sustainability. Although this does not prove bias, since it is possible that so many projects performed well, it at least suggests that the reports are of limited value for the donor to distinguish actions that worked from ones that did not.

The interviews with NGOs and consultants suggest that while many humanitarian organizations do not want to manipulate such in-depth evaluation reports, there are plenty of opportunities for those who do wish to inject a positive bias. In particular, NGOs can prescribe what questions and methods consultants should work with or ask them not to write down specific negative words. Five respondents add that the small local consultants who often evaluate humanitarian projects are particularly susceptible to such pressure, since they often heavily depend on NGOs to pay their bills.

These findings might suggest that positive bias is a large problem for ECHO. However, the interviews reveal that this risk is strongly mitigated by at least two factors. Firstly, almost all respondents stress that ECHO usually knows how projects function in reality via its regional and field offices. The technical and program officers/assistants working there visit each ECHO-funded project at least once, but usually more often. During those visits, ECHO's officials extensively speak to all stakeholders involved, including the intended beneficiaries.

In 2018, ECHO's field offices employed 459 people (DG ECHO 2019: 70). Most respondents from NGOs describe this staff as relatively large, especially when compared to many national donors. They also argue that the frequent field visits arranged by these program officers/assistants are not only the most crucial information source for ECHO, but also limit the room for biased information in the final narrative reports. For example, one contact person for ECHO from a large NGO stated that:

> *"[suppose] you're reporting that you're doing fantastic and you've reached all your targets and go beyond that, but the reality on the ground is that that's not likely, since other organizations that ECHO funds are not achieving what they set out for various reasons, or ECHO's own field visits are showing something different. I think organizations that would try to do that would be found out very quickly. ECHO doesn't have a huge amount of partners in countries, but they are quite close to them and quite informed on what's going on..." (interview by authors, 4-10-2021)*

Secondly, the scope for positive bias is limited by the (partly) standardized digital format in which all final reports must be submitted to ECHO. On this so-called Single Form, NGOs have a limited wordcount to describe the results of their activities based on a set of indicators defined at the start of a project. These standards are partly mandated by ECHO, although they vary to an extent based on the type of action and the partner's preferences. Examples of indicators include mortality rates, numbers of accidents, access to services (e.g. food, shelter, education), and the proportion of beneficiaries which believes that assistance was properly delivered (DG ECHO 2021b: 63).

Almost all respondents who are familiar with the 'Single Form' argue that it reduces the extent to which project results can be presented unfairly. The main reasons for this are (1) that the indicators are usually of a fairly objective and quantitative nature and (2) that the

numbers presented have to be backed up by annexed evidence. Manipulating such quantitative data is theoretically possible, but is considered too unprofessional and/or risky by NGOs. The explanations added to the numbers can present them in a favourable way, but only to a limited degree.

All in all, hypothesis 1 is not confirmed. While our results suggest that positive bias can sometimes be a problem for the reports submitted to ECHO, this issue does not seem to be significant, due to a combination of the field visits and the standardized indicators.

Methodological problems

As was explained in the 'Methods' section, we were unable to read any final narrative reports submitted to ECHO, but we could assess 39 in-depth evaluations which were probably attached to such reports. Table 2 shows that 87% of these documents clearly specify the standards used for the evaluation. Usually, these are the DAC criteria (relevance, coherence, effectiveness, efficiency, impact, sustainability), which are very common in humanitarian and development aid (Clements 2020: 1).

Table 2 also shows that in 90% of the cases, triangulation was applied. Usually, this comes down to a combination of interviews/questionnaires with site observations or document analysis. Our interviews confirm that asking questions to (intended) beneficiaries is the most common way to assess ECHO-funded projects. Usually, gaining access to such groups is no problem for evaluators from NGOs or consultants hired by them, since they already have a foothold in the places where aid is given, with the exception of a few areas where armed conflict has broken out.

As Table 2 reveals, only a minority of the reports which we assessed are transparent about all questions asked to respondents (27%), who the respondents were (35%), which documents were analysed (36%), and the case selection for other methods (40%). The interviews with NGOs and consultants confirm that the methodological clarity of such reports varies greatly. Possible reasons for this include (1) a competitive market in which evaluators sometimes promise more than they can deliver to get a job and (2) the tendency to rely on relatively small and sometimes unprofessional local consultants.

Despite their overall positive conclusions (see the previous section), 38 of the 39 external evaluations contained recommendations about how weak points of the projects could be improved. By far, the most frequent advice, which appeared in 23 reports, was to improve the indicators used for monitoring (specific forms of) impact and/or how data about these benchmarks is collected. This finding adds to the aforementioned suggestion that many evaluations are based on weak information. Other recommendations which appeared the most frequently were improving beneficiary participation (17 times), improving training for beneficiaries, especially regarding using hygiene tools (14 times), and improving gender balance (12 times). The first two of those points are substantive issues, but the third one has a methodological component as well, in the sense that it often pointed to a lack of input from men or women about project effectiveness.

Table 2 only presents a limited view of the methodological quality of those evaluations of ECHO-funded projects which we could find online. Therefore, we used the interviews to further explore this topic, by asking respondents to what extent the data which they submit to ECHO allows that organization to assess the real impact of actions. All respondents with knowledge about the final reports believe that they provide a good picture of projects' direct output, via the standardized indicators mentioned in the previous section (e.g. the number of beneficiaries reached and their satisfaction with the aid). Several experts also point out that the reports are suitable input for ECHO's strategic plans and regional/thematic evaluations, since they allow for easy comparisons between projects at the aggregate level.

Whereas most respondents from ECHO and one expert from an NGO state that the final reports and their annexes also allow for assessing projects' long-term impacts, the others do not share that view. In other words, a majority of respondents, including almost all of the ones from the implementing parties, argue that the reports cannot show the full effects of projects. A few experts state that this fact is caused by the aforementioned indicators on the 'Single Form' being too focused on the short-term, which in turn could be caused by the relatively brief duration of ECHO-funded projects (often a year at maximum). Some other respondents emphasize that areas struck by disaster are often very dynamic, which means that many variables affect local populations. Therefore, it is difficult to isolate the long-term effects of humanitarian projects, especially because control groups often cannot be used for practical and/or ethical reasons. One expert from an NGO, who also worked for ECHO in the past, illustrated this point with an example:

> *"…you have so many external elements in a crisis that you cannot do the kind of research that would just focus on, for instance, what is the impact of an increase in cash distributions? It's too specialized to be given focus." (interview by authors, 13-09-2021)*

Respondents disagree about the extent to which this limited information about long-term impacts is problematic. Some argue that humanitarian aid mainly

serves to fulfil the immediate needs of disaster-struck populations. From this perspective, short-term impacts should be ECHO's focus, while long-term effects should be aimed for by development aid, for which a different DG within the European Commission (DEVCO) is responsible. Other respondents argue that long-term effects of humanitarian projects are just as important to assess as short-term ones, for example, because ECHO also finances some multi-annual actions which require integration with DEVCO's policies.

All in all, hypothesis 2 is partly confirmed. Methodological problems seem to affect the extent to which final reports about ECHO-funded projects are informative to some extent, but mainly when it comes to estimating long-term effects and not so much when it comes to short-term output.

Human resources

All ten respondents with knowledge about the matter believe that ECHO reads (almost) any final report submitted to it. In particular, NGOs state that they consistently receive detailed questions about these documents, from multiple actors within ECHO (field offices and headquarters; financial and humanitarian departments) and when needed via multiple rounds. For example, one respondent from an NGO stated that:

> *"It doesn't mean that you submit and then that's the end. They [ECHO] can come up with 100 questions if what you've reported doesn't make sense at all, if your figures do not add up or if you say something strange. So they question whatever you write in the final report." (interview by authors, 13-09-2021)*

This process of constantly questioning NGOs is enabled by the fact that ECHO's staff can spend much time on this task, given the fact that project implementation is mostly left to the partners. Furthermore, according to the respondents working for ECHO, the aforementioned standardized 'Single Form' makes reading all final reports feasible.

ECHO's officials stress that they also read in-depth evaluations attached to the final reports. However, almost all respondents from NGOs state that they rarely receive feedback on such documents, which makes them unsure if they are actually reviewed. In any case, as was explained above, such in-depth evaluations only seem to be attached to final reports for some projects from some NGOs, so workload is unlikely to be a large issue here. All in all, our expectation that limited human resources make the information submitted to ECHO less useful to it (hypothesis 3) is not confirmed by our findings.

Conclusion and discussion

This article started with two questions. First, how does DG ECHO's system for ex-post assessments of activities implemented by NGOs function on paper and in reality? Second, to what extent do these ex-post assessments properly inform ECHO about the results of activities which it funded?

Regarding the first question, both ECHO's official documents and our interviews show that at the end of each project, the implementing NGO submits a final report via a standardized digital 'Single Form'. This report includes a narrative part, which mainly describes the extent to which an activity achieved its objectives, difficulties encountered during implementation, how resources were spent, and lessons for future projects. Generally speaking, our respondents considered this to be an informative tool for ECHO.

Each final narrative report must be backed up by annexes, which may include an in-depth external evaluation. In practice, such external evaluations are only produced in a minority of cases. More often, NGOs attach an in-depth internal evaluation or no in-depth evaluation at all, because they find an (external) report to be too expensive and/or useless.

Regarding our second question, we examined three factors that can theoretically limit the extent to which assessments of humanitarian projects are informative for donors: positive bias, methodological problems, and human resources. Only the last of those three points turned out to be no issue at all for ECHO, since our results show that NGOs consistently receive feedback from multiple officials on the final reports which they submit. Since ECHO does not implement projects itself, its staff has sufficient time to read these files.

The second factor, methodological problems, somewhat limits the extent to which the reports are informative for ECHO. Our document analysis suggests that collecting data about projects via multiple methods is common practice, although even when an in-depth external evaluation is conducted the exact sources are often not transparent. Our interviews revealed that whereas the final narrative reports and their attachments usually provide an accurate picture of projects' direct effects (e.g. the number of beneficiaries reached and their satisfaction with the aid), most of the respondents from outside of ECHO are sceptical about their potential to identify long-term impacts. In theoretical terms, the system could be described as focused more on output than on outcome (Birkland 2016: 274).

A complicating factor here is that humanitarian aid is often intended for short-term relief (output). Within the European Commission, the main responsibility for aid intended for long-term development (outcome) lies not

with ECHO, but with a different DG (DEVCO). In our study, some respondents from NGOs and consultancy companies criticized the boundaries between ECHO and DEVCO. The relation between these two DGs and both types of aid in the context of the EU could be an interesting topic for future research.

The remaining factor, positive bias, somewhat limits the extent to which the reports are informative for ECHO. However, almost all respondents stress that this problem is greatly mitigated by the organization's large network of regional and field offices. The civil servants working there visit most projects multiple times to observe their implementation and talk to all stakeholders involved, so that significant problems are identified long before the final report is submitted. Thus, it would be risky for NGOs to be dishonest about results even if they would want to.

From a theoretical viewpoint, the findings discussed above imply that researchers should analyse evaluations of humanitarian aid projects in combination with associated field visits. In academic literature (e.g. Davies 1999: 152), such observations are often labelled 'monitoring' (a term which ECHO also sometimes uses): continuously assessing an ongoing policy's results. Monitoring is a common instrument for the EU's institutions to exercise control over actors who are implementing their policies (e.g. Broberg 2015: 261; Kassim and Mennon 2003: 123). The literature about this phenomenon in the context of the EU and/or humanitarian aid should be taken into account when drafting a theoretical framework for future research, instead of focusing on ex-post evaluations like this article did.

From a practitioner's viewpoint, the findings discussed above imply that building a strong network of regional offices can help donors to assess the effectiveness of the projects which they finance. National actors might be able to learn from ECHO in that regard — provided they have sufficient resources — as most respondents from NGOs argue that the EU has a relatively strong presence in the field, especially when compared to countries other than the UK and the USA. For future research, it would be an interesting question if donors with a less developed network of regional offices are less capable than ECHO to compensate for the shortcomings of ex-post reporting, or if they have alternative methods to do so.

Acknowledgements
No other actors contributed significantly to the manuscript.

Authors' contributions
The first author wrote most of the manuscript and collected most of the data. The other authors provided extensive feedback on the text, were present during some interviews, and contributed to the pilot study of the document analysis. The author(s) read and approved the final manuscript.

Funding
The research presented in this article received no funding.

Availability of data and materials
The data used for this article can be accessed by contacting the first author.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Public Administration, Nijmegen School of Management, Radboud University, Heyendaalseweg 141, 6525 AJ Nijmegen, The Netherlands. [2]Department of Politics & Social Sciences, Zeppelin University gemeinnützige GmbH, Am Seemooser Horn 20, 88045 Friedrichshafen, Germany.

Received: 1 June 2022 Accepted: 30 September 2022
Published online: 18 October 2022

References

Aerni P (2006) The principal-agent problem in development assistance and its negative impact on local entrepreneurship in Africa: time for new approaches. Afr Technol Dev Forum J 3(2):27–33
Ascroft E (1999) ECHO - humanitarian aid. In: Cosgrove-Sacks C (ed) The European Union and developing countries: the challenges of globalization. Macmillan, London pp 179–189
Attinà F (2016) EU's Humanitarian and Civil Protection Aid. Italy's eccentric and ECHO-consistent policy. Rom J Eur Aff 16(1):24–43
Birkland T (2016) An introduction to the policy process: theories, concepts and models of public policy making, 4th edn. Routledge, New York
Boin A, Ekengren M, Rhinard M (2013) The European Union as crisis manager: patterns and prospects. Cambridge University Press, Cambridge
Broberg M (2015) Thou shall not… misappropriate humanitarian aid: on European Union humanitarian aid and the fight against corruption. In: Zwitter A, Lamont CK, Heintze HJ, Herman J (eds) Humanitarian action: global, regional and domestic responses. Cambridge University Press, Cambridge pp 253–271
Clements P (2020) Improving learning and accountability in foreign aid. World Dev 125. https://doi.org/10.1016/j.worlddev.2019.104670
Davies IC (1999) Evaluation and performance management in government. Evaluation 5(2):150–159
DG ECHO (2014) General conditions applicable to humanitarian aid actions financed by the EU. European Commission, Brussels
DG ECHO (2019) Annual activity report 2018. European Commission, Brussels
DG ECHO (2020) List of partners. European Commission, Brussels
DG ECHO (2021a). ECHO organigramme. Brussels, Retrieved December 15 2021, from https://ec.europa.eu/echo/files/about/echo_organigramme_en.pdf
DG ECHO (2021b) Single form guidelines. European Commission, Brussels
DG ECHO (n.d.). DG ECHO partners' website. Brussels, Retrieved November 11 2021, from https://www.dgecho-partners-helpdesk.eu/ngo/ngos
Edwards M, Hulme C (1996) Too close for comfort? The impact of official aid on nongovernmental organizations. World Dev 24(6):961–973. https://doi.org/10.1016/0305-750X(96)00019-8
Kassim H, Mennon A (2003) The principal-agent approach and the study of the European Union: promise unfulfilled? J Eur Public Policy 10(3):121–139
Keck M (2016) Comparing the determinants of US-funded NGO aid versus US official development aid. Voluntas 26:1314–1336. https://doi.org/10.1007/s11266-014-9464z
Lewis H, Forster G (2020) Data collection, analysis and use in protracted humanitarian crises. Publish What You Fund
Mitchell GE (2014) Why will we ever learn? Measurement and evaluation in international development NGOs. Public Perform Manag Rev 37(4):605–631
Mowjee T (1998) The European Community Humanitarian Office (ECHO): 1992-1999 and Beyond. Disasters 22(3):250–267
OECD (2019). Development Co-operation Report 2019: a fairer, greener, safer tomorrow. Retrieved May 6 2020, from https://www.oecd-ilibrary.org/docserver/9a58c83f-en.pdf?expires=1588791723&id=id&accname=guest&checksum=E095DE1BA389A1937F3A74F66A69A981

Pusterla F, Pusterla E (2020) Humanitarian aid and the European Union. In: Thompson WR (ed) Oxford research encyclopaedia of politics. Oxford University Press, Oxford
Suárez D, Gugerty MK (2016) Funding civil society? Bilateral government support for development NGOs. Voluntas 27(6):2617–2640 https://doi-org.ru.idm.oclc.org/10.1007/s11266-016-9706-3
Vennix J (2019) Research methodology: an introduction to scientific thinking and practice. Pearson, Nijmegen
Versluys H (2007) Explaining patterns of delegation in EU humanitarian aid policy. Perspectives 28:63–84
Wenar L (2006) Accountability in international development aid. Ethics Int Aff 20(1):1–23. https://doi.org/10.1111/j.1747-7093.2006.00001

Publisher's Note