



HAL
open science

Quality and relevance of the evidence on the situation of refugees in Jordan in an emerging context of humanitarian evidence-based policy-making

Gwendoline Promsopha, Ingrid Tucci

► To cite this version:

Gwendoline Promsopha, Ingrid Tucci. Quality and relevance of the evidence on the situation of refugees in Jordan in an emerging context of humanitarian evidence-based policy-making. 2019. hal-02417426

HAL Id: hal-02417426

<https://hal.science/hal-02417426v1>

Preprint submitted on 18 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Quality and relevance of the evidence on the situation of refugees in Jordan in an emerging context of humanitarian evidence-based policy-making

Gwendoline Promsopha*

Ingrid Tucci*

*Aix Marseille Univ, CNRS, LEST, Aix-en-Provence, France

DOCUMENT de TRAVAIL

Introduction.....	Erreur ! Signet non défini.	2
Counting refugees and assessing their situation: a debate for policy, polity and politics		5
What counts as evidence	Erreur ! Signet non défini.	
Emergence of evidence-based decision-making in the humanitarian sector		6
<i>Evidence and decision-making: evidence for what?</i>		8
<i>Decision making in the humanitarian sector: on the basis of what kind of evidence?</i>		8
What kind of evidence? Analysis framework and related questions		10
Data and methodology		11
Results.....		12
<i>Actors involved in the production of data and diffusion of results</i>		12
<i>Sources of evidence quality variability</i>		14
<i>Relevance of evidence on refugees in Jordan</i>		21
Discussion and avenues for future research		25
References		27
Appendices		31

Introduction

The Syrian crisis led to a massive displacement of Syrians within Syria, in neighbouring and more distant countries, creating the conditions for one of the biggest and most visible humanitarian crisis of the last decade. Since the turn of the “new aid paradigm” in the 2000’s the humanitarian sector experiences a growing focus on accountability and effectiveness of aid (Dijkstra, 2015)¹ and a global trend toward an evidence-based humanitarian policy-making approach. In 2003, “The good humanitarian donorship” initiated by the Swedish government formulated 23 principles that were agreed upon by the major donors (Olin & Schreeb 2014), which stresses that humanitarian funding is allocated “in proportion to needs and on the basis of needs assessments.”² The World humanitarian Summit (2016), the creation of an international Expert Group on Refugee and Internally Displaced Persons Statistics (EGRIS) by the United Nations Statistical Commission (UNSC) (2016) or the first UN World Data Forum in 2017. The growing importance of evidence-making in humanitarian action resulted in an acceleration of data collection to inform humanitarian policy: various types of data are being increasingly collected in humanitarian crisis such as Syria, such as monitoring and counting data, micro-data from surveys, qualitative data from key informants, interviews or focus group discussions, data from experiments, big data. These data are expected to help decision-makers to shape the nature and scale of the humanitarian response in crisis areas.

The issue of evidence as a basis for humanitarian action (evidence-based approach or policy-making, i.e. EBPM) has been addressed extensively in the scientific and in particular in the grey literature (see Bradt 2009, Knox Clarke & Darcy 2014; Christoplos et al. 2017). The EBPM rests on the gap hypothesis between science and policy, which considers that decision-makers do not sufficiently use available (scientific) knowledge. EBPM assumes that high quality data are a prerequisite to adequate policy and effectively provides right responses to needs. This assumption has been contested since (scientific) knowledge also have a legitimating function, e.g. for delaying decisions or for achieving public acceptance of decisions that were already adopted by decision-makers (Faist 2011).

While we learn a lot from this literature on the use of evidence in decision making, we still know little about the nature, quality and production process of the evidence itself. This issue is at least equally important: indeed, the use of evidence is only valuable if the evidence itself has a certain quality, if it addresses properly the population of concern, and is produced with transparency. This issue is all the more important in a sector which deals with vulnerable populations such as refugees and is frequently entangled with heated debates on migration and hosting policies. Ironically enough, there is little evidence on the empirical bases and methodologies used to produce so-called evidence in the humanitarian field. The first objective of this paper is therefore to bring a new perspective on EBPM in the humanitarian sector, through the analysis of the quality and relevance of evidence produced about

¹ We thank Thomas Faist for his helpful comments on an earlier draft.

² Good Humanitarian Donorship. Principles and good practice of good humanitarian donorship. Stockholm; 2003. Available here: <https://reliefweb.int/sites/reliefweb.int/files/resources/EN-23-Principles-and-Good-Practice-of-Humanitarian-Donorship.pdf>

refugees in Jordan in the context of the Syrian crisis. Jordan is one of the main host countries for Syrian asylum seekers in the Middle Eastern region - the number of registered Syrian refugees increased by 58% reaching 671350 persons/ the number of registered Syrian went as high as 671350 according to - with one of the highest refugee density of the region after Lebanon. Amman was also chosen to host the sub-regional response facility for the Syria crisis, leading to a significant presence of humanitarian actors in the country. Jordan became member of the IOM in 1999, and is also the first country among Arab states to become a Delivery as One country (DaO) within the UNHCR framework, i.e. a pilot country adopting standard operating procedures in order to match better UNHCR funding programmes and national priorities.³

In this paper, the quality of evidence is considered as multidimensional and defined along data quality, transparency, accuracy of the data analysis and triangulation. We also study the relevance of the evidence⁴ as the degree of adequacy between the nature of produced evidence and the actual needs and situation of potential beneficiaries. We also investigate the nature and evolution of actors involved in the funding, production and diffusion of evidence. A comprehensive set of 243 documents dealing with the situation of refugees in Jordan and published between 2012 and 2018 were downloaded from three platforms (Reliefweb⁵, data.unhcr and REACH Initiative⁶) that collect, centralise and diffuse information to humanitarian actors. On the basis of those documents, more than 150 variables were created, coded and analysed quantitatively. We use the Multiple Correspondence Analysis approach to produce a typology of documents based on various indicators of evidence quality, and analyse variability in geographical coverage, issues and sectors analysed.

The objective of this paper is not to inform about the use of evidence in decision making in the humanitarian field (as done by von Schreeb et al. 2008, Olin & von Schreeb, 2014, Clarke & Darcy 2014), it is rather concerned with the issue of the nature, methodological quality, relevance and function of evidence through a thorough investigation of what is supposed to be the first “input” of EBPM, i.e. data, reports, needs assessment, evaluations, and so on.

The first section addressed the complex issue of counting and assessing needs of refugees, section 2 shows how difficult it is to define what an “evidence” is. The third section provides a review of the critical literature on the evidence-based approach focused on its implementation in the humanitarian sector; questions the specificities of the sector that may matter when assessing humanitarian evidence; and presents our framework for the rest of paper/how we will use the concept in the paper. Section 4 describes our empirical methodology and database. Section 5 present our results: an overview of the reports and of the actors involved in their production, the variability of the evidence quality according to a typology of reports constructed using the method of multiple correspondence analysis (Le Roux & Rouanet 2010) and the relevance of the data used for refugees in Jordan. We end this paper with a discussion of the results on the legitimizing function of evidence production

³ <http://jo.one.un.org/en/what-we-do/>

⁴ Relevance could be understood as the adequacy between the evidence produced and the actual needs of vulnerable population but this would necessitate another research design.

⁵ Reliefweb is the world’s largest humanitarian information platform.

⁶ REACH Initiative is a consortium founded jointly by the NGO Acted, the Think-Thank Impact, and UNOSTAT.

(Boswell 2008; 2009) for organizations in the field and against the background of the quantification processes that can be observed more generally (Porter, 1996; Desrosières, 2010).

What counts as evidence

Generally speaking the idea of “evidence-based policy” has been developed from evidence-based medicine in 1971 (with Archie Cochrane’s work). It has then been extended to all sectors of public and private decision-making, with the explicit purpose of breaching the gap between scientific evidence and decision-making and to improve the effectiveness and impact of policies (Oliver et al, 2014). More recently, it has become prevalent in economic development policy under Duflo’s and Banerjee’s work (2011) through the set of evaluations conducted by the Abdul Lateef Jameel Poverty Action Lab (J-PAL), with an increasing emphasis on good governance, accountability and the production of evidence through Randomized Control Trials (RCTs) before implementing large-scale development programs.

Yet, there has been extended debates on what is meant by “evidence”, and on its real added-value for policy-making. The nature of evidence differs according to the question asked (Bowen et al. 2005). Evidence loosely defined as “scientific research”, i.e. RCTs, cohort or time series analyses, to observations, experiences and qualitative studies (Bowen et al 2005), is generally considered “hard evidence”, but other types of information such as expert knowledge or consultations are also used for policy-making. In a frequently criticized “hierarchy of evidence”, RCT has become “king in the castle” (Bedecarrats et al. 2018), and a “gold standard” (Pritchett & Sandefur, 2013) for both international donors and researchers. In all fields dealing with evidence-based policy making, a same emphasis can be found on the issue of attribution and causality so as to find sound methodologies and insure that objectively measured impacts can be attributed to a specific project, in the perspective of generalization and replicability. Yet, “evidence” is a complex notion since, as knowledge, it is socially constructed and never fully objective: the interpretation of an evidence “is subject to the way in which an individual understands the social world and what they consider to be important” (Clarence 2002, p. 5). What counts as evidence for decision making can even be contentious (Martson & Watts 2003). Some authors argue that evidence derived from RCTs or other methodologies in one context might be irrelevant for a similar social problem in another context (external applicability, see Davies et al. 2000)⁷, and other authors propose to move away from the hierarchy of evidence and to adopt the idea of “appropriateness”⁸ for decision making (Parkhurst & Abeysinghe 2016). Beside the mere challenge

⁷ With regards to RCTs, Bruno (2015, quoted in Abdelghafour 2017) see this type of evidence-based policy making as “an art of emphasizing objectivity as a guarantee of realism and efficiency”, a desire for robust evidence that faces many challenges due to political and financial constraints as well as to the various practices in the private and public sector in the production of this kind of evidence.

⁸ Decision making processes imply considering complex situations and combining different types of evidence so that a single method aiming at tracing causality is important but might not be enough to grasp the whole social problem to be solved.

in producing timely research matching the temporality of policy (Oliver et al. 2014), many critical arguments about evidence-based policy making have spread in the field of medicine, health and social science, regarding the positivist framing of most of evidence based approaches, high valuing clear measurement from a (larger) sample, the omission of the role of “meanings” for individuals (Greenhalgh & Russell 2009).

Scientific knowledge can be also be used by politics and organizations for alternative purposes. Weiss (1979) proposes six models of research utilization: in the political model, research is used by politicians or organizations to reinforce predetermined opinions and positions, and is ignored when contradicting these opinions. Another risk in EBPM is to set the politics aside at the expense of democratic control on decisions (Hawkin & Prakhurst, 2016). Finally, evidence also has a (symbolic) legitimizing function for political and social action, and more generally for organizations, as shown by Boswell (2008; 2009) in her study on the role of expert knowledge in the European migration policy.⁹ According to Boswell, organizations need to secure internal but also external legitimacy. The methods for gaining and securing legitimacy differ according to the type of organization: political organizations reinforce their legitimacy through rhetoric and the decisions they take, while, action organizations, derive their legitimacy from their societal interventions (Boswell 2008, p.5).

Counting refugees and assessing their situation: a debate for policy, polity and politics

The idea that migration policy should follow evidence and economic or social rationales has been widely discussed yet, faces a number of political and methodological challenges.

First, collecting data to build evidence in the forced migration sector faces a number of issues that require specific ethical and methodological cornerstones: the political and sensitive nature of the debate, the vulnerability of the population involved, technical and methodological difficulties such as access to population and sampling. One of the most important type of data produced in the field of migration deals with the counting of flows and stocks of migrants and foreigners. Despite their lack of accuracy, such data are regularly used in media or political discourses leading to frequent debates and controversies (Singleton 2016; Baldwin-Edwards et al 2018, Steinhilper and Gruijters 2018) and making the counting a political count of the uncountable, especially when estimating the number of so-called irregular migrants (Vogel et al. 2011). Although less visible and commented, large and representative survey data are also produced by research agencies or the scientific community, in an attempt to fill the gap between scientific evidence a policy making. Yet, large representative sample are mainly found in developed countries and have their own flaws: specific populations remain hard to capture in large surveys, i.e temporary mobile persons such as seasonal workers, posted workers, and refugees. Qualitative studies or smaller-scale surveys on refugees have been conducted in various European countries to fill the gap (see, e.g., Phillimore 2011, Korac 2003; Campbell 2006,

⁹ Knowledge is of course broader than evidence but the latest is part of it.

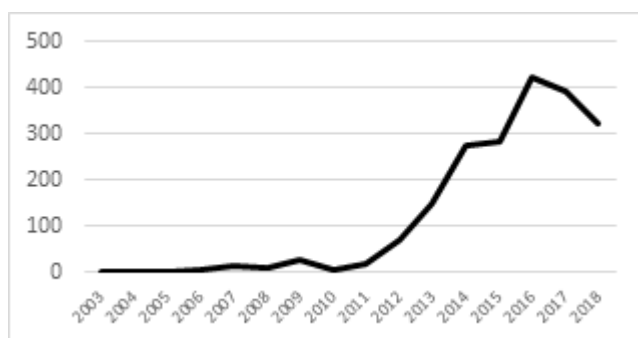
Colic-Peisker & Tilbury 2007) however we still know little about the situation of refugee, even in the developed countries.¹⁰

In low income countries on the other hand, Harrell et al. (1992) underline how refugees have progressively become both the means and ends of humanitarian actors, as these progressively took charge of managing refugee crises and took responsibility in raising funds. Producing counts of refugees and monitoring data then became a strategic tool to achieve both functions, i.e. raising funds and providing aid, but led to a number of scientific and operating controversies such as: is it possible to accurately count refugees? Are count data the best way to measure needs? Is counting without consequence for the refugee population, or does it produce another type of oppression and an “ideology of control” (Harrell et al. 1992)?

In countries experiencing large refugees’ inflows such as Lebanon or Jordan, large surveys conducted by research agencies on refugee populations are lacking. Data is mainly produced by international organizations - governmental or not – following recent incentives to adopt “evidence-based policy making” (EBPM) , but with important obstacles in the diffusion and sharing of the data, which are circumscribed to the circle of humanitarian actors and invisible to the scientific community (Levine 2016).

As a consequence, we observe an increase in the production of documents and reports including data aiming at evidence-making on refugees, as can be seen in the figure 1 below, along with a general objective for humanitarian actors of providing evidence for decision making and publicizing the outcomes and impact of the aid delivered.

Figure 1: Number of documents posted on Reliefweb about refugees (World)



Notes: In this graph only documents classified in reliefweb as assessments, analysis, evaluations and lessons learns and UN documents are taken into account. .

Emergence of evidence-based decision-making in the humanitarian sector

¹⁰ A larger survey now exists for example in Germany, a country that experienced a large inflow of asylum seekers from the summer 2015 (Kühne et al. 2019).

The evidence-based approach progressively entered the humanitarian field during the 1990s (Sutcliff & Court 2005; Knox Clarke & Darcy 2014). Evidence-based practice in the humanitarian field aims at better prioritizing needs and objectifying humanitarian action, optimizing funding to reach more efficiency, and legitimating the selection of project for funding. Yet, the humanitarian sector is singular in its potential for evidence production for a number of reasons:

- Besides the lack of historical relations between humanitarian actors and research, countries in need of humanitarian assistance have little resources, offer complex political and economic contexts, and strong public statistical service are not likely leading to high entry costs for drawing random samples and sound analysis. Since monitoring and rationalization efforts are rather recent for humanitarian actors, relevant skills are likely to be lacking, especially for statistical analysis.¹¹ The extensive professionalization of the humanitarian sector (Kukkonen 2018) is still lagging behind for evidence production, justifying the implementation of collaborations between data-specialist agencies, donors and scientists (Banatvala & Zwi 2000).
- In crisis contexts, there is a need to produce short-term and fast results. According to Levine (2016) the timing and priorities of humanitarian actors and academics widely differ, explaining the lack of connexion between the two spheres. Ethical considerations - even if raised in the guidelines on data collection of international organizations (Banatvala, & Zwi 2000) - as well as logistics funding issues are other establish stronger relations with the academic community (Levine 2016). Finally approaches in the two communities differ: because they depend on public attention and funding, humanitarian actors sometimes dramatize situations (see, e.g., De Chaine 2002), a strategy deemed inappropriate in the academic sphere.
- The main focus of the humanitarian sector in its “efficiency for action” is on matching the “needs” of beneficiaries. Cost-effectiveness or long-term impacts are more marginal issues in the humanitarian sector, despite recent efforts for “resilience-building” in the Syria crisis (Gonzalez, 2016).
- There is a potential conflict between the rationalization of action implied by EBP and the “good Samaritan” and strong ethical considerations such as those in place in the humanitarian sector (Dijkzeul et al. 2013).
- As in other sectors related to social policy, the humanitarian sector undergoes a paradox between the common good issue of facilitating transparency and sharing of evidence; and on the other hand the competitiveness within the funding system might produce incentives to win calls for funding, and not to diffuse the data.

¹¹ A study about professionalization conducted among a sample of ca. 1300 humanitarian workers, indicates that statistical analysis and demographics is the core knowledge humanitarian workers estimate to have the least understanding of, while needs assessments are in the top position and monitoring and evaluation methods comes in the fourth position of the skills humanitarian workers have at least some understanding of (Walker & Russ 2010, p. 31).

Several questions arise from the use of evidence for decision making in the humanitarian sector: The first question is whether EBPM allows to allocate aid to those who need it the most. Even if essential, this question has already been widely analysed in the literature and will only marginally be referred to in this paper. The consensus in this extensive research is that, despite all efforts made to strengthen the quality of humanitarian action (the SPHERE movement, see Buchanan-Smith 2003), aid is not necessarily allocated where it is the most needed, but rather respond to political or geopolitical issues at the macro level¹² or interrelations at the micro level (Clarke & Darcy 2014). The second set of questions can be summarized as “*Evidence for what?*” understood as “which function for evidence”. Even if not the main focal of this paper, this issue is important to contextualize our main research question guiding our empirical analysis: *Which evidence?*; How are the evidence produced, based on which types of data and analysis, and with which quality?

Evidence and decision-making: evidence for what?

Beside the self-evident function of guide for action, better efficiency and accountability, the production of evidence also carries a symbolic role: by producing data on international migrants, including refugees, international organizations obtain an external legitimacy as organizations that produce data, assess needs and (seem to) act according to individuals’ needs. As research on the linkage between migration research and policy in Europe suggests, producing evidence might be a strategy of gaining legitimacy for organizations in the sector (see Boswell 2009), especially given the nature of the humanitarian environment, its widespread financial instability and the multitude of organizations. As proposed by Hilhorst and Jansen (2010), the humanitarian arena only exists because it is discursively created by the various stakeholders, reinforcing the need of humanitarian actors to show that they are “doing good work” (p.1122). The necessity of gaining legitimacy in humanitarian action has contributed since the beginning of 2000 to an increase in impact assessments in crisis regions (Watson 2008). In a way, the emergence of a “humanitarian government” as described by Agier (2010) also leads to the development of government policy instruments for the control and governance of populations, in which the production of data and indicators will probably play an important part both to guide policy making and for ex-post policy legitimation.

Decision making in the humanitarian sector: on the basis of what kind of evidence?

The second issue raised previously – *which kind of evidence?* – is as important as the first one but far much less research and only in specialised literature. Yet, without precise knowledge on the type of data and information produced by the humanitarian sector and about the kind of evidence it represents, it seems difficult to understand the role of evidence in decision-making processes. The kind of evidence produced shapes the robustness and validity of the recommendations found in reports produced by international organizations, but also informs on the system of legitimation built by international organizations. If international organizations were compelled to provide robust evidence to legitimate their action, they would certainly do it. Otherwise, producing evidence of any kind could

¹² This view is in line with the tradition of the Science, Technology and Society (STS) studies pointing at the societal and social embeddedness of scientific production and technological developments, and vice versa.

be considered as “window dressing”, and this even if a needs assessment might of course be important for the vulnerable populations concerned by the assessment.

A thin research on the quality of evidence can be found in humanitarian medicine and health policy, and in several papers in the grey literature -yet based solely on literature reviews or non-transparent methodologies (“some” interviews with key informants, expert knowledge). In humanitarian medicine and health, the criteria for “sound” evidence are mainly experimental or related to quantitative science and not always transferable to other fields. Traditionally, the quality of evidence in EBPM is associated with scientific criteria borrowed from experimental and quantitative methodologies and focusing on attribution, causality and statistical robustness (Sacket et al. 1996). As already developed in a previous section, RCT is the golden standard in evidence-based policy but its prevalence is criticized by some for confining knowledge (Marston and Watts 2003). These last authors differentiate between hard - quantitative and/or experimental data - and soft evidence - qualitative data. Yet, the specificities of the humanitarian sector require an adaptation of the traditional EBPM standard to crisis environment and vulnerable populations. For instance, quantitative surveys do not allow respondents to describe their suffering (Roberston et al. 2002). However, despite the recent emphasis on evidence and several attempts to set up standards in the production of evidence, the humanitarian sector lacks a clear definition of what an “evidence” is (Kukkonen 2018).

In their exhaustive study of evidence produced on nutrition during the Ethiopian famine, Spiegel et al (2004) finds that evidence is generally provided in the form of needs assessments with cross-sectional surveys, includes frequent methodological errors, and in 95% of the cases the evidence were neither valid, i.e. they did not measure what they intended to, nor precise, i.e. they did not involve probabilistic sampling or had small sample size. Given sample sizes, the methodology and data treatment is also of concern with an expected variability of estimates across reports. Although Spiegel et al (2004) recognize the importance of rapid needs assessments for humanitarian action¹³, there is a significant risk of “convenience” samples and biased results, questioning the generalization of results or relevance of recommendations. In another study, Thompson (2017) shows how short-term action is often privileged in refugee crises¹⁴. This might focus the attention on refugee camps or certain emergency sectors at the expense of others. Blanchet et al. (2017) also identify research gaps in analysing 345 studies published between 1980 and 2014 on health issues, in armed conflicts, natural disaster and cute crisis. They show that evidence are relatively weak: experimental designs are not systematically applied, changes in outcomes are not tested statistically, control groups are lacking and statistics are not disaggregated by age or gender). In reaction, the humanitarian sector has enacted standards and processes such as the HNTS by WHO, the SMART project, SENAC project, IPC tools and so on. Yet Darcy and Hoffman (2003) conclude to an even greater confusion produced by the growth and uncoordinated multiplication of standards. More recently, collaborative

¹³ “It is not our intention to demean humanitarian NGOs’ motives or quality of work, but rather to document the inadequacies and suggest practical means to rectify the situation in the future.” Spiegel et al. 2004 p. 617.

¹⁴ Given the fact that most refugees populations have been settled in some countries for decades (the Palestinian populations is a key example), one question arises about the need to shift from humanitarian aid to development cooperation, a longstanding debate.

attempts such inter-agency assessment, the IASC¹⁵ needs assessment task force and what Garfield (2011) call the Common Needs Assessment (CAN) have been put forward but with limited success: collaborative attempts follow too many objectives to provide clarity; are excessively focused on quantitative and survey-based data, create rivalries between participating. Listing eight criteria for the quality of evidence in the humanitarian sector - accuracy, representativeness, relevance, generalisability, attribution (causal linkage) and clarity around contexts and methods, Knox-Clarke and Darcy (2014) concludes to weak evidence in the sector.

What kind of evidence? Analysis framework and related questions

In this paper, we propose a framework to assess the quality of evidence and analyse sectorial evolutions. It is built on the general literature on evidence-based policy assessments and the quality of evidence, but also takes into account the specificity of the humanitarian context and its limitations for the implementation of research designs. “Quality” is not an easy to define and depends on the chosen definition of evidence. For Boaz & Ashby (2003), methodological quality is only one dimension of the research quality in the context of evidence-based policy making, and must be complemented by the usefulness for policy and practice, the appropriateness of the methods to the aims of the study, and the quality of reporting so that results can be traced by others. But research quality is a much broader topic than what concerns us here: evidence quality. Shaxson (2005) states that the quality of evidence is related but must be distinguished from the broader idea of usefulness, i.e. what is useful and relevant for practitioners. For him, usefulness of evidence has many dimension including “quality”, credibility, relevance and practicability. Rather than distinguishing between usefulness and quality, we propose a broader definition with two dimensions of evidence quality: *internal quality* which assesses the quality of evidence production and the general coherence of information upon which it is supposed to describe or analyse, and *external quality*¹⁶ which investigates the relevance of the information regarding to actual situations.

Based on the literature (Clarke & Darcy 2014; Christoplos et al. 2017), we include as **internal validity**: *data quality* (sampling method, sample size¹⁷ in case of a quantitative study, type of population included, and as a consequence, potential for generalization), *appropriateness of the data analysis* (producing appropriate and sound data analysis relative to the nature of the data, causality making), implementation of *triangulation* (for instance through mixed methods, different sets of data, reference to a larger literature and other operational reports) and *transparency* (quality and completeness of information on sampling, methodologies and limitations of the study). **External validity** relates to the relevance of issues and populations analysed relative to humanitarian priorities,

¹⁵ International Accounting Standards Committee (IASC)

¹⁶ The two dimension makes refer to the idea of external and internal validity in experiments (see Mc Dermott 2011).

¹⁷ We do not address here extensively the question of evidence quality in research on which a lot was published by medical scientists and public health researchers, setting for example criteria of evidence quality (see for example the GRADE guidelines, Balshem et al. 2011) or showing that

and accounting for vulnerable population. Disaggregated data for example, by sex, age or location, might be indispensable for the assessments of situations and for aid delivering in an efficient and effective manner (Mazurane et al. 2013).

Drawing of this framework of analysis and the review of the literature, we develop a series of hypothesis. First, given observation from the literature we expect an “actor-level” effect on quality: given entry costs in producing data; the potential role of legitimacy issues in producing evidence; the heterogeneity between actors in terms of operational objectives; financial capacities or statistical skill, there might be interesting reconfigurations of actors in the production of data and evidence, through time and with consequences on the nature and quality of evidence. Moreover, as proposed by the literature, we expect a number of bias in the different steps of data collection and analysis, which could affect the external quality of evidence and carry potential consequences for the shaping of humanitarian responses.

Data and methodology

We apply this framework to a systematic review of documents published on several platforms on Syrian refugees in Jordan. Syria was officially proclaimed in a civil war at the mid of the year 2012. Then on, the number of internally displaced Syrians and asylum seekers rose dramatically in neighbouring countries like Lebanon, Turkey and Jordan. Jordan is one of the most important receiving countries of Syrian asylum seekers in the Middle Eastern region and host the sub-regional response facility to the Syria crisis in Amman: by the end of 2012, 115 558 Syrian refugees were registered by the HCR in Jordan, 500 000 in July 2013 and it has remained over 600 000 persons after July 2014.¹⁸ While camps were soon built to take charge of the refugees (the main being Zataari and Azraq camps), 79% the Syrian registered population actually lives in (peri)-urban and rural areas in Jordan (UNHCR, “External Statistical Report on UNHCR Registered Syrians as of 15 September 2017). Adding unregistered refugees and using the 2015 Jordan census brings much higher figures: 1 265 514 Syrian citizens lived in Jordan in 2015, among which 95% in urban areas¹⁹ and including 953289 Syrian registered refugees. Those different figures show that counting individuals, in particular in crisis situations, is not an easy task and lead to important variations in estimations. Several international and non-governmental organizations are present since the beginning delivering humanitarian aid, some of them were instructed by the Jordan government to deliver this aid only to those who have the full documentation.²⁰

The data used in this paper was collected on different platforms that diffuse information on the situation in regions of the world where humanitarian work is deployed. We focused on four platforms in particular that gather the most part of the documents published online: REACH Initiative Resource Centre (130), UNHCR Data portal Assessment Registry (69), Relief Web (41) and ALNAP (3). We

¹⁸ <https://data2.unhcr.org/en/situations/syria/location/36>

¹⁹ http://dosweb.dos.gov.jo/censuses/population_housing/census2015/census2015_tables/

²⁰ see http://cadmus.eui.eu/bitstream/handle/1814/34904/mpc_2015-02_pb.pdf?sequence=1 page 7

choose reports classified on those platforms as: Assessments, Analysis, Evaluation and lessons learned, Reports, Reports and policy papers, UN document. After having identified the documents and excluded duplicates as well as entries without downloadable documents, the analysis draw on a total of 243 reports that claim to produce evidence on refugees in Jordan and were published between 2012 and 2017.

We listed 7 topics to be explored: source of the document and involved actors, platform and publication, topics addressed, geographical zone, data used, analysis, references and lexical issues. A set of variables was defined for each topic and the variables were coded for each report meaning that each coder went through a set of reports. Most of the variables, e.g. topics addressed in the report, organizations involved, geographical zones, have been objectively measured, while a small part of the variables are more subjective, for instance the adequacy between results and conclusions, or the clarity of assessment/evaluation objectives. To ensure inter-coder reliability, a subset of the reports was simultaneously coded and compared ex-post.²¹ In this phase but also throughout the coding process, in-depth discussions were conducted to understand potential differences and enabled a same understanding of what was measured. In some cases, the difference in understanding and coding was recurrent and strong so that we decided to exclude the variable. The data set produced comprises circa 160 variables.

The sections below present the actors involved in the production of the reports and, secondly, the results from a Multiple Correspondence analysis (MCA). MCA is a form of principal component analysis applied to categorical data (Le Roux & Rouanet 2010; Husson & Josse 2014). Its purpose is to explore and summarize the data in order to identify similar or opposed characteristics of the reports and the dimensions structuring the most the reports. It is followed by the hierarchical clustering method in order to build a typology of reports based on a set of variables aiming at defining the quality of the evidence produced.²²

Results

Actors involved in the production of data and diffusion of results

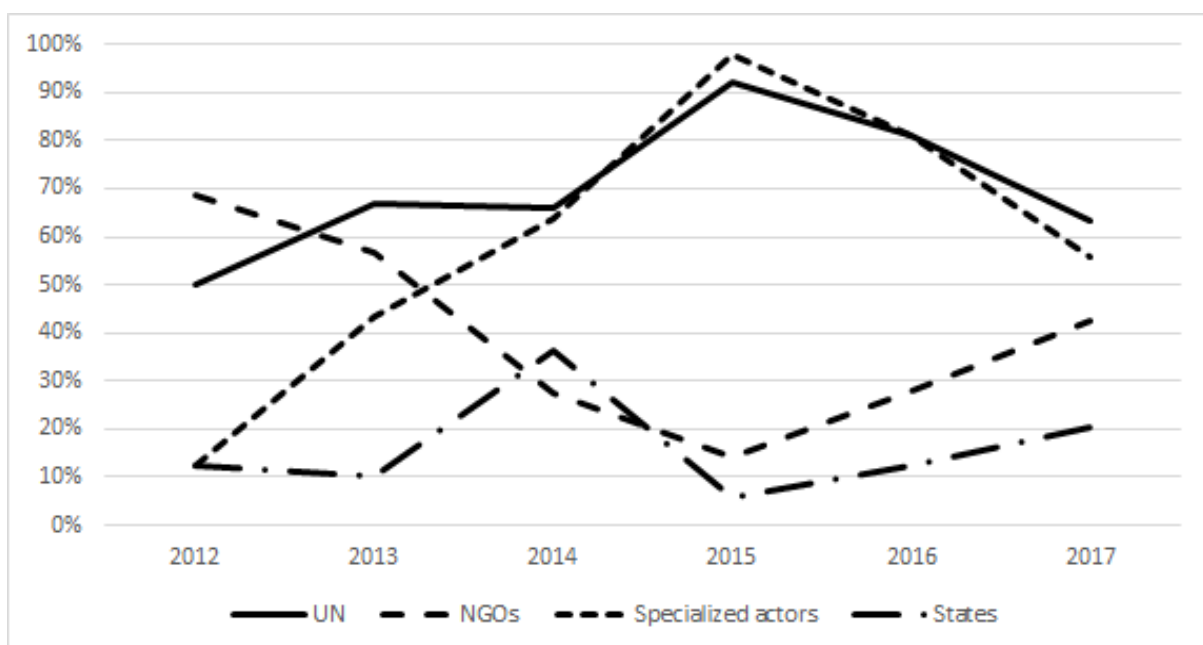
Four types of actors were identified in the documents: the first three are classical actors in humanitarian crisis: the various agencies of the United Nations with a prevalent role of the UNHCR but also interventions by the UNICEF, WFP, UN Women and UNFPA, and sporadically other international organizations such as WHO, the World Bank, the IOM, or the ILO; international and more scarcely local NGOs; and governments, both Jordanian and foreign. The fourth type is a new actor in the humanitarian sector: a set of organizations - either private or non-profit - specialized in data collection, production and dissemination of information. An good example of this category of

²¹ We thank Léa Macias for her collaboration at the beginning of this research. The paper was written in the framework of a research project funded by the French Research Agency (ANR) and titled "Time of conflicts / time of migration: Reflections on the categories and genealogy of migration in the Middle East" (coordinated by Kamel Dorai).

²² The MCA analyses as well as the hierarchical clustering were performed in R using the package factomineR (Husson et al. 2017).

actor is REACH initiative, a consortium founded jointly by the NGO Acted, the Think-Thank Impact, and UNOSTAT, whose central mission is to “effectively respond to the needs of crisis-affected communities”²³ through the collection and production of data and information. REACH progressively became a pivot actor in the production of needs assessment and evaluation: in 2015 REACH contributes to 91% of existing report published on the three web platforms, in 2018, in 78%. Private actors such as polling organizations or consulting firms emerge in the market, even if they remain marginal (in 2017 they intervene in about one report out of 10). These results suggest both a tendency to outsource data collection, analysis and production of final reports to specialized actors, creating a market for evidence that might become profitable for a number of private or for-profit organizations²⁴; and a concentration in the hand of one dominant but not-for-profit player, REACH, who obtain a leading position in the Jordanian evidence market, a situation which might produce a form of standardization of processes and methodologies, types of data collected, in the types of biases involved in analysis, and in the presentation of the final “product”, i.e. reports. For instance, most reports produced by REACH have the same structure (executive summary, context, methodology, findings, limitation), the context section are standardized and are sometimes copy-pasted from one document to the other. Meanwhile, such concentration allows economies of scale since REACH re-uses a same dataset (especially with large samples) to produce multiple assessments on different issues and for different funders.

Figure 2 Evolution of the share of actors involved in the production of reports



²³ From REACH website <https://www.reach-initiative.org/who-we-are/>.

²⁴ In the field of international development cooperation, systematic reviews are recently increasingly used for the purpose of evaluation (Mallet et al. 2012), it seems to be less the case in the humanitarian field

From the first three years of the crisis until 2015, the role of NGOs in the production of assessments in Jordan decreases significantly (Figure 2). It increases again later on but remains at a relatively low level compared with the UN organizations whose implication is strong in the production of results over the whole period (over 50 % of the reports in all years). The UN has developed strong collaborations with REACH initiative, with almost two third of the reports involving both a UN organization and REACH. The involvement of States in the production of evidence on the situation of refugees is moderate but peaks in 2014, with mainly the UK, countries from Northern Europe, Australia, Canada, USA, Switzerland involved as donors, and Jordan. 83% of the reports involve a collaboration between different actors and when a collaboration exists, it is more often with the participation of UN agencies than without.

One question arising from this actors' perspective is whether certain types of documents are associated with a certain type of actor. How far does the quality of the evidence produced vary according to the actors involved in financing, collecting and producing of the results?

Sources of evidence quality variability

We propose to answer this question by assessing the internal quality of the evidence in these documents using several dimensions (see the section "What kind of evidence? Analysis framework and related questions") and related indicators:

- *Transparency* : presence of an information on sampling and methodology, and degree of detail
- *Quality of the methodological approach*: type of data, type of sampling method, sample size
- *Accurateness of data analysis*: type of analysis, use of statistical tests or not, multivariate, bivariate or univariate statistics, inclusion of population controls.
- *Data triangulation*: use of mixed data, references used in bibliography, cartographic analysis,

The variables for each dimension are presented in Table 1 and used to perform the Multiple Correspondence Analysis. MCA is an inductive method that enables the identify several axes structuring the sample of reports (see appendix A2) and to identify classes or types of reports using the hierarchical clustering method.

Table 1 Variables included in the Multiple Correspondence Analysis

Quality of the methodological approach	
SAMPsize	1=Quali, 2<=499, 3=500-999, 4=1000-2999, 5>=3000, 6=Unknown sample
QUANT	1 = YES, 2 = NO
QUAL	1 = YES, 2 = NO
typesample	1=Random, 2=Non random, 4=Sweep, 5=No sampling/Not informed
typedanalyse	1= Impact assessment, 2=Policy paper, 3=Needs assessments, 4=other
Accurateness of data analysis	
STATS	0 = No statistics, 1 = Descriptive statistics, 2=Tests&Econometrics
cartoanalyse	1 = YES, 2 = NO
popcontrole	1 = YES, 2 = NO
Triangulation	

refsourceshorseval	1= No references, 2 = Other references than from assessments, 3 = No other references than from assessments
refsourceshorssyr	1= No references, 2 = References not on Syria, 3 = No references not on Syria
refsourcesong	1= No references, 2 = References from NGOs, 3 = No references from NGOs
MIX	1= Quali or Quanti only, 2 = Mixed Quanti&Quali, 3 = Mixed Quali, 4=Other method
Transparency	
SAMPINFO	0 = No information on method, 1 = Simplified information, 2=Detailed information
GUIDE	0 = Quantitative, 1 = No interview guide, 2=Interview guide in the report

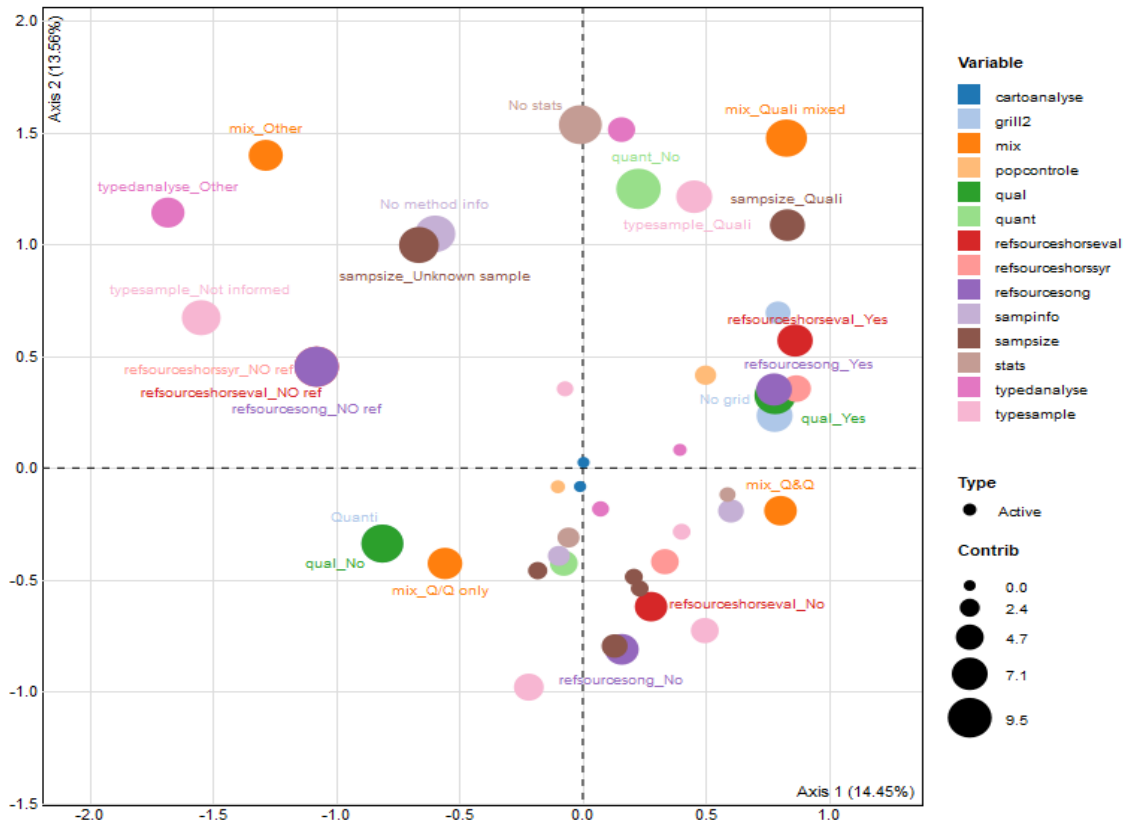
The representation of the first two axes is shown in figure 2 below, the size of the circles is proportional to the contribution of the respective modality of the variable to the construction of the dimensions. The larger the circle, the most important is the contribution.

The **first (horizontal) axis** opposes (on the positive side) reports that use references, draw on qualitative data or mix qualitative and quantitative data and include a detailed information to reports without any references, using a type of analysis that does not rely on any kind of collected data. This axis could be defined as a **transparency and validation axis**.

The **second (vertical) axis** opposes reports having no statistical data, constructed as policy papers, or using only qualitative data or unprecise sources of data (positive side) to reports that draw on quantitative data, mostly sweeps or using random sampling (negative side). This axis represents more a **methodological axis**.

The **third axis** (not displayed in the figure) opposes reports with statistical tests and econometrics, using a population of control and giving detailed sample information (positive side) to reports based on qualitative data, sometimes also mixed qualitative data, without statistics and only simplified information on the sample (negative side). This axis might be defined as a **robustness axis**.

Figure 2: The first two axes emerging from the multiple correspondence analysis (MCA)



After we ran the MCA, we used the method of hierarchical clustering on the basis of the first ten axes identified in the MCA in order to get classes of reports sharing specific characteristics. The four clusters that emerged from the clustering are presented in the following, using also some more qualitative information found in the reports.

Cluster 1 Reports with quantitative data and uninformed methodology

58 documents (24% of the sample), 16 pages on average

Cluster 1 largely involves documents with quantitative data but mostly non-random samples (39.7%, see table 2 for a description of the clusters) and a non-negligible share (15%) of documents based on data that are neither quantitative nor qualitative surveys nor sweeps, probably coming from unspecified sources. Compared with the other three, cluster 1 has the lowest share of documents informing about the methodology in a detailed manner. Only 5% of the reports include econometrics or statistical tests, not all includes descriptive statistics and only 12.1% of the documents considers a control/comparison population such as Jordanian, non-refugees or non-syrians.

Among the first five reports representative of this cluster, the most representative is a two-pages document about the public health profile of the refugee population with descriptive statistics only and almost no or only vague information about the methodology: *“Data is collected using standardised Health Information System (HIS) data collection and reporting tools and aims to be as accurate as possible.”* One document is a presentation of results in PowerPoint style. The Syrian Refugee Unit at

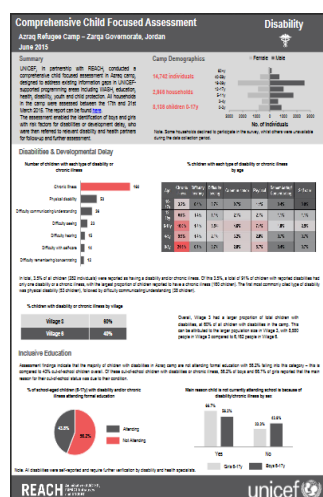
the Jordan Ministry of Labour is implicated in three documents out of these five representative ones. In the other two, REACH is implicated as a partner for data collection and analysis in the field.

Cluster 2: Quantitative evidence from the UN/REACH collaboration cluster

65 documents (27% of the sample), 15 pages on average

Cluster 2 is similar to cluster 1 in many respect except that States are much less involved in financing or in data collection. Instead, the specialized actor REACH and the UN intervene at some stage of the process in 98% of the cases. Collaborations between these actors are also very frequent (98% of the cases against 77% in cluster 1). Those documents are mainly based on quantitative data (90.6%), especially sweep data (in almost one document on two) with a more frequent detailed information on the methodology than in cluster 1. Random samples are more significant than in cluster 1 (23.1% against 5.2%). The documents always include descriptive statistics but they all fail to have a control/comparison population. An interesting fact is that 52% of the documents have only one page and almost two third are mere “factsheets” such as the one displayed below.

Example of a factsheet: Comprehensive Child focused Assessment (disability)



Among the first five reports representative of this cluster, the most representative is a document on post-distribution monitoring using a household survey while the other four documents are one-page-factsheets concerning all the Azraq camp and using the same data, i.e. a survey among the whole camp population. The information on the methodology is more frequent than in cluster 1 but still the question of non-response relevant to estimate representativeness remains vague as illustrated by this quotation: “some households declined to participate in the survey, whilst others were unavailable during the data collection period.”

In this cluster, a toolkit provided by a web platform called “Raosoft”²⁵ is frequently mentioned to help determine sample size according to base population. Referring to Raosoft seems to guarantee the

²⁵ See <http://www.raosoft.com/samplesize.html>

robustness of the results in the eyes of humanitarian actors but also illustrates a routinization of process which ultimately obscure the importance of sampling designs for representativeness: statistically incoherent affirmations were found in several documents such as: “*Findings are statistically representative at the camp level with a 97% level of confidence and 4% margin of error*”. The standardization process can be noticed in the formal presentation of reports where UN is involved. A practical guide from the UNHCR²⁶, states: “pour faciliter l’actualisation et la comparaison de données, il est suggéré de développer un set de formats standard pour les rapports, qui peuvent inclure des tableaux, graphiques et cartes.” An example of this kind of standardized structure can be found in the appendix (A3).

Table 2: Description of the clusters

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Organizations					
% ONU	81,0	98,5	41,7	59,7	71,6
% States	15,5	4,6	27,1	25,0	17,7
% NGOs	29,3	29,2	62,5	27,8	35,4
% specialized actors	55,2	98,5	25,0	69,4	65,0
% REACH	46,6	98,5	14,6	48,6	54,7
% Collaboration	77,2	98,5	68,9	83,1	83,2
Data					
Quanti or Quali only	75,9	90,6	23,9	6,9	49,2
Quanti&Quali	8,6	7,8	2,2	93,1	32,5
Mixed quali	0,0	0,0	63,0	0,0	12,1
Other types of data	15,5	1,6	10,9	0,0	6,3
% QUANT	84,5	90,6	4,3	98,6	75,0
% QUAL	8,6	15,6	87,0	94,4	51,2
Sampling					
Random	5,2	23,1	0,0	38,9	18,9
Non random	39,7	27,7	39,6	43,1	37,5
Sweep	8,6	46,2	2,1	4,2	16,1
NI/No sampling	46,6	3,1	58,3	13,9	27,6
Statistics					
No stats	13,8	0,0	60,4	2,8	16,1
Statdescr	81,0	98,5	37,5	77,8	76,1
Test/Econometrics	5,2	1,5	2,1	19,4	7,8
popcontrol					
Yes	12,1	0,0	18,8	34,7	16,9
No	87,9	100,0	81,3	65,3	83,1
% with informed detailed methodology	12,1	21,5	20,8	59,7	30,5

Cluster 3: Qualitative data with NGOs implication

48 documents (20% of the sample), 32 pages on average

²⁶ <https://www.unhcr.org/fr/statistics/unhcrstats/4ba344089/guide-pratique-lusage-systematique-standards-indicateurs-operations-hcr.html> p. 52

Cluster 3 (as well as cluster 4) differs from the first two clusters as to the organizations involved and the type of data analysed. In Cluster 3 NGOs are present in 62.5% of the cases, followed by the UN with 42% and states with 27%. This cluster has the lowest share of collaborations between different types of actors. The data used is primarily a mix of different types of qualitative data or, in a lower proportion, a unique set of qualitative data. As a result, 60.4% of the documents have no statistics, yet surprisingly 37.5% still provide descriptive statistics. Random sampling is an absent strategy in this mainly qualitative based cluster but 18.8% of the documents include a control population.

The first representative document in this cluster is based on a qualitative study using focus groups and key informants interviews. In the Jordanian humanitarian sector, key informants interviews are used to produce general results on refugees in camps, although some reports point at limitation and a need for caution in interpreting the results: « *family and individuals figures presented in this report should be interpreted as indicative figures as these are reported by key informants within each geographic location / basic service unit.* » Other representative documents are either based on qualitative studies or have no primary data. One document involves a more “typical” qualitative study giving quotations to illustrate findings and informing on the data collection process. A third document compares the situation of refugees in Jordan, South-Sudan and Kenya, combining the counting of facilities, in-depth interviews with medical personal and focus groups with refugees. Consistent with this cluster, the methodology lacks description and percentages are given out of qualitative samples with no indications as to the basis or sources for such statistical results.

More generally the widespread use of focus group discussion in our sample can be found in its cost and time-efficiency. This practice conflicts with the usual quality standard in qualitative research. The same is true for the recurrent use of purposive sampling (Lavrakas 2008), a cheaper and cost-effective sampling method, yet leading to larger bias, potential errors of judgement by data collectors in their choice of respondents, and huge limitations in generalizing findings. Finally, the seemingly preference of NGOs for qualitative methods is consistent with their budget constrained but also their emphasis on individuals cases and stories

Cluster 4: Mixed methods data and detailed methodology

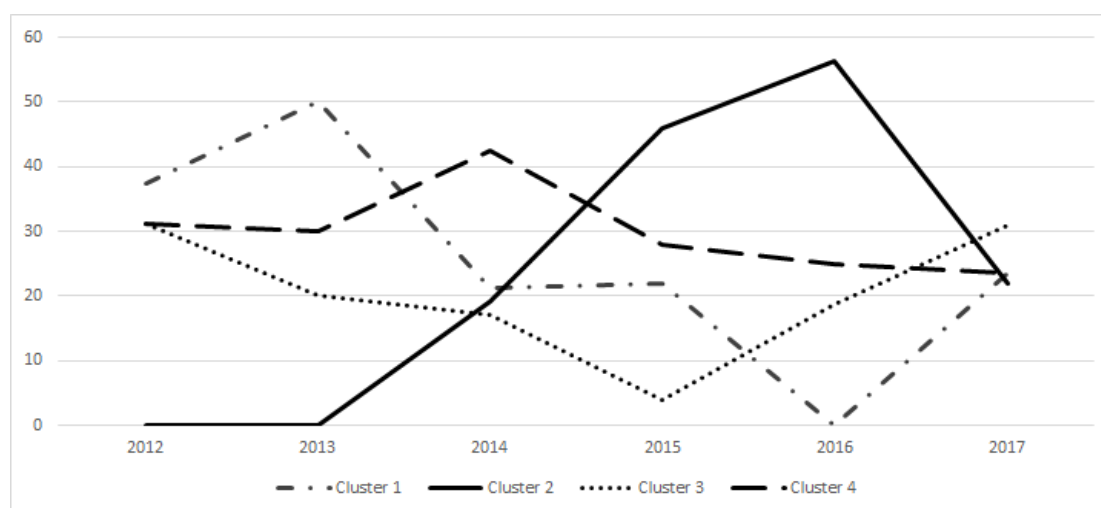
72 documents (30% of the sample), 46 pages on average

Compared to cluster 3, Cluster 4 has an important share of documents for which REACH collect the data and NGOs are not often implicated. 83.1% of the documents are the product of a collaboration between actors, although not always under the typical REACH-UN duo. This cluster is characterized by the highest share of documents mixing quantitative and qualitative data (93%) of inclusion of a control population (one third) and of random sampling (34.7%). 59.7% of the documents provide detailed information on the methodology, and 19% include more sophisticated statistical analysis such as significance tests or econometrics. The document that best represents this cluster and is the most distant from the other three clusters is a long report with a complex and well described methodology,

combining a household survey, an employer survey, a small survey with child labourers, focus groups, observations from the field as well as key informants interviews (*“The interviews helped to validate quantitative survey data, to gather experts’ views on the current policies and systems dealing with child labour, and to pinpoint areas for improvement.”*). Limitations are also mentioned in this report. Syrians are compared with Palestinians and Jordanians. Another document is based on a quantitative panel survey which is relatively rare in the sample (only 5% of the whole sample). Similar to the first report described above, this report on the welfare of refugee children also reports limitations of the data.

The evolution over time of the share of the different clusters (Figure 3) indicates that reports drawing on large data sets (sweeps mainly) take the overhead at the heart of the crisis with the installation of a large number of refugees in Jordan while reports with a higher quality of evidence, more robust and diverse methodologies (cluster 4 and 3) progressively loose in significance but return at the end of the period, signalling an increase in quality which may relate with initiatives toward resilience plans in the regional, such as the Resilience Development forum in 2015.

Figure 3: Share of the clusters over time

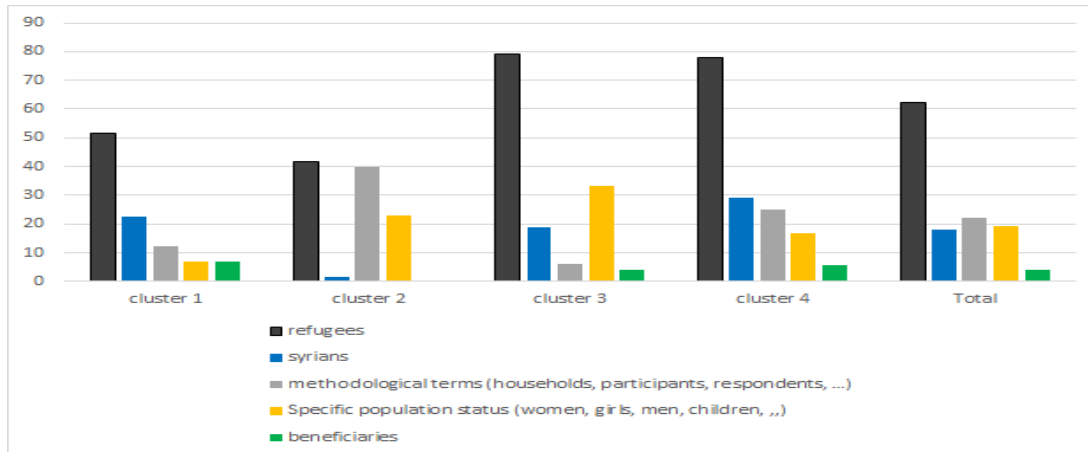


Finally, we performed a lexicological analysis to identify the terms used to designate the population studied in the reports.²⁷ Refugees are found to be largely referred to under methodological or statistical terms such as “respondents”, “population”, “key informants”, “households” (Figure 4), especially in cluster 2 characterized by the UNU/REACH collaboration and a large presence of counting of facilities such as WASH assessments: the process of standardisation and routinization of data production and reports’ writing may go along a process of objectification and, eventually, “deshumanization” of this vulnerable population. In the other three clusters refugees as rather designed as “refugees” to a large extent..

²⁷ For this, we identified the different terms used in each report, and drawing on this list we counted those terms to find the primary (occurring the most frequently) and secondary terms.

Interestingly, documents in cluster 3 - based on qualitative data and with a prevalence of NGOs - tend to give significance to sub-populations of refugees, who are referred to as “refugees” but also frequently as “children”, “girls”, “youth”, “women”, “men”.

Figure 4: Terms primarily used to refer to refugees in the report, by cluster



To summarize the results from the multidimensional analysis, Cluster 4 clearly stands out with the highest quality of evidence across most dimensions of our framework, although statistical tests or dealing with attribution or causality issues is still not widespread, even there. Reports in cluster 2 are the product of a collaboration between UN agencies and REACH providing relatively short document relying on standardised format and text structures. Clusters 3 uses qualitative data with detailed information on the methodology, strongly involves NGOs and States; while cluster 1 is mainly based on quantitative data with States playing a minor role and less frequent or detailed information on sampling methodologies. Finally, we observe that collaborations between actors can be correlated with two phenomena. Indeed cluster 2 and 4 show the highest percentage of collaboration between actors (98% for cluster 2 and 83% for cluster 4, against 77% for cluster 1 and 69% for cluster 3). In the case of cluster 4, collaboration is therefore correlated with higher quality of evidence, probably through the pooling of skills in the methodological domains and the sharing of costs. In cluster 2, what stands out in not so much the quality of evidence as the routinization of processes in data collection, analysis and results' presentation, including an objectification of populations of concern through the use of methodological lexicon for “refugees”: again, this seem to be correlated with the REACH-UN collaboration.

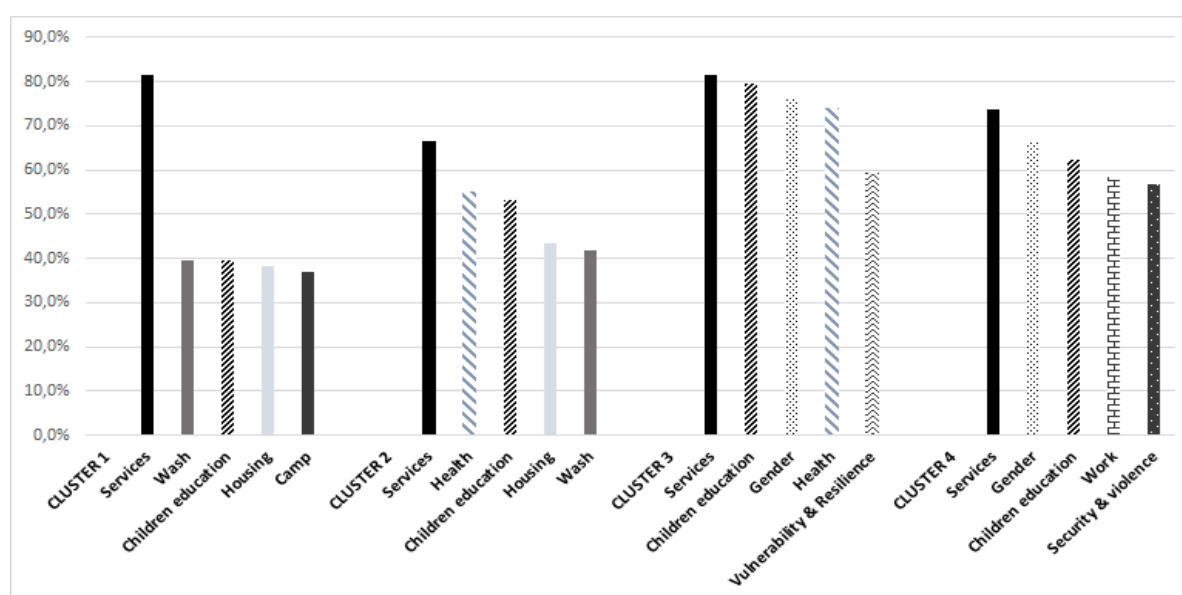
Relevance of evidence on refugees in Jordan

As mentioned in a previous section, we consider the relevance of evidence on the situation of refugees as an external dimension of the quality of the data produced in the context of humanitarian data production. The question is therefore whether the reports 1) sufficiently address different issues or life domains, 2) address without bias the population of concern, and its geographical dispersion

and 3) consider the specificities of some population groups, e.g. by disaggregating the results along gender and age. Selection bias in sampling, over or under representation of populations, issues or geographical areas might indicate a lack of relevance of the produced evidence.

As displayed in figure 5, considering the four clusters, access to services and basic needs as well as children education are the most frequently addressed. Cluster 2 – mainly UNHCR-REACH collaborations - is also especially concerned with health, housing and WASH issues - Wash and Housing concentrating important UNHCR budget effort, besides services for people with specific needs²⁸. Yet, contrarily to the UNHCR priorities²⁹, security, protection and solution orientation such as opportunities for self-reliance and later integration are much less under scrutiny.

Figure 5: The first five field addressed in each cluster



Cluster 3 and 4 are the only clusters in which gender arrives among the five most important fields. Interestingly, work and labour that are considered as central tools for long-term resilience, are only addressed in cluster 4, and left aside in cluster 2 (UNHCR) despite the priority given to resilience, autonomy and self-reliance capacities of refugees by the HCR. Our data indicates that employment is a more important issue among reports looking at non-camp refugee populations. More generally, the number of domains addressed in the different cluster differ. According to our data, reports in clusters 3 and 4 concomitantly deal with seven topics on average, while documents in cluster 1 and 2 addresses a lower number of fields topics - five on average.

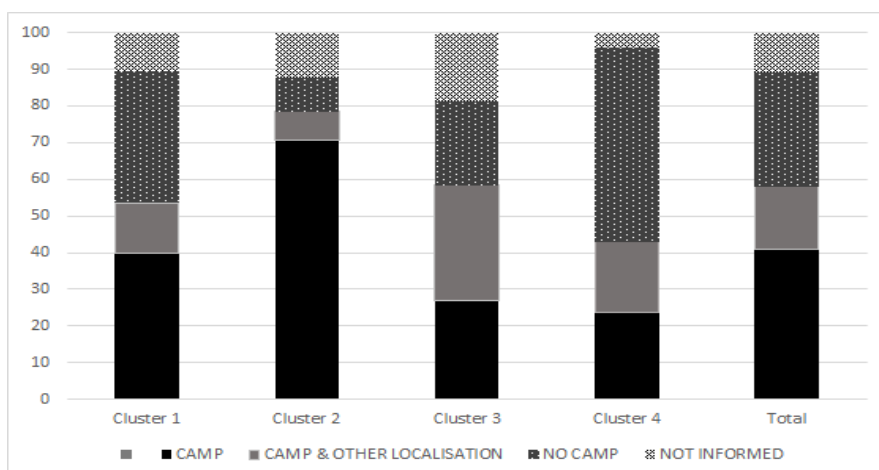
²⁸ See: <https://www.unhcr.org/528a0a2c13.pdf>

²⁹ According to the HCR websites (<https://emergency.unhcr.org/entry/258535/emergency-priorities-and-related-indicators>), the main priorities in handling refugee crises are: a favourable protection environment, documentation and registration, security from violence and exploitation (especially reducing protection risks of women, men, boys and girls), fulfilling basic needs and services (health, nutrition, water access and primary education), community-based protection and solution orientation at the onset of emergencies (find opportunities for self-reliance). For the UNHCR, the internal relevance of evidence should therefore be an evaluation of needs and policy potentials in these priorities.

Overall, 58% of the documents deal with the population in camps and 40% only analyse camp populations, while in fact 79% of the Syrian refugees do not live in camps but in urban or rural areas. Furthermore, only a very low share of document on camp populations includes a control or comparison population (1%), while a third of the reports that only deal address non-camp population include such a control population (generally Jordanian or other categories/nationalities of migrant).

As shown the figure 6, collaboration between specialised agencies and the UN as found in cluster 2 lead to a clear overrepresentation of camp population in the documents, while cluster 4 tends to focus more frequently on non-camp areas. Clusters 3 and 4 tend to have a more balanced distribution of camp and non-camp populations. Consequently, the emphasis given to camps by many documents may contribute to invisible population living in non-camp - urban or rural - areas and reduces potential for a differential assessment of the needs encountered by populations in different contexts of living.

Figure 6: Locations focused on in the reports by cluster



The Zataari camp built in 2012 is a focal point for international organizations (38.7% of documents sample Zataari), especially in cluster 2. The Azraq camp does not attract as much attention (12.8% of documents) and the city of Amman, which concentrate about one fourth of registered Syrian refugees in Jordan is largely understudied in the reports (sampled in only 13,64% documents). To summarize, there are overrepresented geographical areas, i.e. camps, at the expense of others, i.e. mainly urban areas with large share of refugees such as Amman. While the majority of Syrian refugees do not live in proper housing³⁰, the issue of housing only appear important in cluster 1 and 2, which is unsurprising since camp housing is directly managed by international organizations (see Figure 5). It therefore seems that evidence displayed by the routinized UNHCR-REACH collaboration are partly irrelevant for a large majority of Syrian refugees living in Jordan, i.e those living outside camps with distinct needs. In contrast, cluster 4 gives a higher visibility to Irbid and Mafraq - both with a large number of Syrian refugees.

³⁰ <https://reliefweb.int/sites/reliefweb.int/files/resources/67914.pdf> p. 37

Table 3: Locations focused on in the reports and share by cluster

	Syrian Refugees between 18-59 (August 2018)*	% reports on this area	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Zataari	32685	38,7	32,8	67,7	29,2	23,6
Irbib	62455	24,3	12,1	16,9	16,7	45,8
Mafraq	36651	23,9	8,6	16,9	18,8	45,8
Zarqa	23297	16,9	5,2	13,9	14,6	30,6
Amman	98408	13,6	0,0	10,8	25,0	19,4
Azraq	15628	12,8	10,3	16,9	14,6	9,7

* <https://data2.unhcr.org/en/documents/download/65827>

In a paper pointing at the more anecdote-driven rather than evidence-driven humanitarian aid in case of armed conflicts and natural disasters, Mazurana and colleagues (2013) argue that sex- and age-disaggregated data are fundamental to an efficient and effective aid delivering. In our case, gender and age are addressed respectively in 47% and 30% of the documents. This does indicate a systematic disaggregation by age and/or gender, however it shows that these dimensions are taken seriously into account. Our data (Table 4) shows that gender is less frequently accounted for in camps (and in cluster 2). Age disaggregation is more prevalent when camps are studied, qualitative data involved (cluster 3) or higher quality of evidence displayed (cluster 4).

Table 4: Correlation between gender and age as topics addressed in the reports and share of reports accounting both topics gender and age, by cluster

	Share Gender	Share Age	Correlation between gender and age
Camp	39,4	31,3	0.5254*
No camp	51,3	23,7	0.1092
Camp & other	66,7	40,5	0.2744
Cluster 1	41,4	17,2	0.1726
Cluster 2	24,6	32,3	0.3689*
Cluster 3	64,6	35,4	0.2751
Cluster 4	58,3	33,3	0.4781*

Non camp studies show an exceptionally poor correlation of between gender and age (0,11), signalling lower use of data disaggregation. If as we propose, disaggregation is an indicator of relevance and a dimension of external quality of evidence in that it signal special attention to vulnerable populations, non-camp Syrian refugees are therefore less likely to benefit from projects and policy adapted to their needs.

Finally, a common practice in report is to provide recommendations based on evidences. In our data, half of the reports in average include some recommendations, but the practice is more frequent in cluster 3 and 4 (81% and 72%) than in cluster 1 and cluster 2 (28% and 25%). However, cluster 3 seems much less suited to make recommendations given the prevalence of qualitative, non-random data which should normally reduce potential for generalization. To better understand

recommendations in cluster 3, it is useful to recall that qualitative surveys draw the attention on subjective meaning and individuals cases, on perception of interventions and so on, which is important for aid delivering and call for funding. Nevertheless qualitative studies, contrary to quantitative surveys, draw the attention on the subjective meaning given by refugees on specific situations, on their perception of interventions etc and is more in line with the contingency approach of quality (Hilhorst, 2002) often defended by NGOs that highlights the importance of diversity, individuality and situational analysis.

Discussion and avenues for future research

Our paper contributes to a thin literature on the quality of evidence produced in crisis situations with a focus on refugees in Jordan following the war in Syria. We presented results from a systematic analysis of documents on the situation of refugees in Jordan published on the three major platforms dedicated to the diffusion of humanitarian evidence. Our paper provides first and innovative empirical results on the quality and relevance of evidence in the humanitarian sector that resonates with broader analyses of the political economy or sociology of quantification, and opens important avenues for further research.

The typology driven from the multiple correspondence analysis indicates that the type and quality of evidence is linked to the identity of the actors involved as well as implementation of collaborations. In some cases, multiple-actor cooperation involving different types of organizations increases the quality of evidence, but in other, especially when regular and routinized such as the REACH-UNHCR collaboration standardization in sampling and reporting is sometimes followed at the expense of quality. Standardization in formatting presentation and graphic design of documents sometimes gives an appearance of seriousness and robustness of the analysis, regardless of the intrinsic value in terms of transparency of the methodology, relevance of the data and analyses. As researchers, this eventually created biases in our own comprehensive understanding of the quality of evidence: our subjective evaluation of quality was sometimes influenced by the marketing of documents, the uses of graphs and maps, large samples and images, then contrasting with our objective indicators collected on the nature of data, sampling, soundness of data analysis, appropriateness. As shown here and contrarily to what has often be claimed by international donors who emphasize routine and automatized guidelines, standardization can sometimes produce worth results than pragmatism and adaptation to specific contexts, issues, and situations. Reducing standardization and routinization in favour of a promotion of skills and a reinforcement of relations with scientists and the academic sphere might reduce the reproduction of errors and inconsistencies.

Somehow, the humanitarian evidence-based movement undergoes a similar process as the “new scientific model” described by Bedecarrats et al (2018) for RCT in the development sector and the emergence of two leading specialized actors (J-Pal and the IPA) to whom are externalized many RCTs evaluations. These two NGOs with strong connection have turned into what could be called a

“family business” with a growing power on policy-making in development and in the academic world, also seen by Harrison (2011) as a “scientific monopoly”, and are sometimes criticized for a lack of critical standpoint on their methodologies and a tendency to hegemony. In the case of humanitarian research, the situation is still far from this hegemonic tendency towards experiments, but evidence-based policy making is becoming an important - inevitable sometimes - part of the legitimization process of projects and programs; it is self-sustaining in the sense that access to funding require evidence which in turn requires funding. Specialized actors such as REACH, ALNAP or others for-profit organizations have grown in the sector to fill the skill gap of other humanitarian operational agencies in an outsourcing/externalizing process. By holding the production of evidence, an important part of the decision power is transferred to these actors, which probably explains why many, such as REACH or ALNAP³¹ are held by a consortium of NGOs and international donors, but also open new market opportunities for private organizations such as polling institutes and consulting firms in the collection and analysis of data. Probably because the production of evidence in complex contexts is submitted to strong entry costs (getting to know the context, accessing the field, finding human resources for data collection), increasing returns to scale and market power are to be expected in the production of evidence, as well shown by the increasing role of REACH. This situation of quasi monopoly on evidence production also furthers the process of standardization of information mentioned earlier. Nevertheless our longitudinal results show that this quasi-monopoly is temporary and correlates with the peak of acute crisis and refugees’ settlement in Jordan that require rapid assessments and action. Later on, other actors such as NGOs re-enter data collection and analysis. This longer-term evolution could be interpreted as a progressive “division of labor”: specialised actor intervene when others are busy with providing emergency aid, but flow back later on..

As in most of the domain of policy-making, quantified data has indeed become normality also in the humanitarian field (Beerli, 2017) and large organizations such as UN agencies are the most engaged in this domain despite recommendation to use mixed methods data, i.e. including qualitative interviews, to better analyse individuals’ perceptions and strategies of self-reliance (HCR 2006³²). At the other end of the spectrum, NGOs favour mixed methods or qualitative data at the expense of large random samples, with reduced potential for generalization and making recommendations.

Numbers and quantification have a legitimation function: they “foster cooperation and control in complex systems and signal one’s legitimacy.”(Nelson Esperland & Stevens 2008, p. 411). By producing a large amount of data, mainly quantitative ones, UN organizations involved in the analysis of the refugees’ situation remain major actors. Numbers and statistics have the power of instating “facts” and give the feeling of rationalizing action. As a matter of fact, the justification of evidence-based policy making in the humanitarian sectors has directly responded to a series of crises that discredited the “good Samaritan” image in the humanitarian sector and drove doubts on the efficiency

³¹ ALNAP pools the UN, a number of NGO, the red cross and red crescent movement, research agencies. <https://www.alnap.org/help-library/alnap-governance-management-and-membership>

³² https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/documents/files/practical_guide_unhcr.pdf

of aid³³. Quantification and evidence making has been put forward as much to increase accountability as to legitimize the action and very existence of the sector. Yet, the system has become self-sustaining: for most humanitarian organizations today, evidence are necessary to obtain funds; and funds are necessary to collect and analyse data. As a consequences, the growth of evidence production is self sustained.

Traditionally, sociologist and political scientists have identified the emergence of quantification as a reflection of the emergence of modern State (Porter, 1996; Desrosières, 2010). This observation is not trivial when we know that various authors have started to describe the emergence of a “humanitarian government” (Agier, 2010) in crisis settings with weak or weakened States. Interestingly, Le Galès (2016) or other also describe the more recent reinforcement of a “government by indicators” in modern State, signalling both generalizing neo-liberal competition and State disciplining by private sector indicators, and opening the way to public policy to various types of lobby. This might be a concern to keep in mind for the humanitarian sector, especially given the emergence of various types of private actors first in the provision of humanitarian services (Börzel & Risse 2002) and more recently in the production of evidence. Moreover, as proposed by Desrosières (2014) and as suggested by the over and under representation of various categories of refugee in reports with consequences on their general visibility, the methodologies used in the production of evidence and quantification processes are far from neutral and can even create new forms of inequalities, inclusion and exclusion movement or new types of domination that are not easily spotted because they stand behind an apparent image of rationalisation and objectification. In this sense we follow the idea of Abdelghafour (2017, p. 257) “onto the ground of events, facts and evidence rather than interpretations, the belief in evidence is itself an interpretation of the world.”

Finally, this study has limitations we are aware of, but also opens avenues for future research. While this study focuses only on the Jordan case, it would be useful reproduce this kind of data collection and analysis on other contexts to have a larger database and see whether the type of reporting and of data or evidence used differs according to the context and the degree or type of implantation of humanitarian organizations in the field. The analysis of the role played by different organizations in collecting data and producing evidence as well as interactions and power relations between organizations is a blind spot of our contribution and more qualitative research is needed drawing e.g. on experts interviews and document analysis. This could allow analysing processes of legitimation through evidence production, the role played by quantification in this sector, how different actors assess this evolution while dealing with vulnerable populations and how the increasing importance of numbers might change humanitarian practices.

References

³³ Such as Bosnia in 1992-95, Rwanda in 1994, the Asian Tsunami in 2004, Haiti in 2010

- Abdelghafour, N. (2017). Randomized Controlled Experiments to End Poverty?. A Sociotechnical Analysis. *Anthropologie & développement*, (46-47), 235-262.
- Agier, M. (2010). Humanity as an Identity and Its Political Effects (A Note on Camps and Humanitarian Government). *Humanity: An International Journal of Human Rights, Humanitarianism, and Development* 1(1), 29-45. University of Pennsylvania Press. Retrieved July 10, 2019, from Project MUSE database.
- Baldwin-Edwards, M., Blitz, B. K., & Crawley, H. (2018). The politics of evidence-based policy in Europe's 'migration crisis'. *Journal of Ethnic and Migration Studies* 0:0, pages 1-17
- Balshem, Howard, Helfand, Mark, Schünemann, Holger J., et al. GRADE guidelines: 3. Rating the quality of evidence. *Journal of clinical epidemiology*, 2011, vol. 64, no 4, p. 401-406.
- Blanchet, K., Ramesh, A., Frison, S., Warren, E., Hossain, M., Smith, J., ... & Dahab, M. (2017). Evidence on public health interventions in humanitarian crises. *The Lancet*, 390(10109), 2287-2296.
- Börzel, Tanja A. and Risse, Thomas (2005). Public-private partnerships: Effective and legitimate tools of international governance. Edgar Grande/Louis W. Pauly (eds.) *Complex sovereignty: Reconstructing political authority in the twenty first century*, University of Toronto Press, p. 195-216.
- Boswell, Christina (2008). The political functions of expert knowledge: Knowledge and legitimation in European Union immigration policy. *Journal of European Public Policy*, vol. 15, no 4, p. 471-488.
- Boswell, Christina (2009). *The political uses of expert knowledge: Immigration policy and social research*. Cambridge University Press.
- Bowen, Shelley, Zwi, Anthony, and Sainsbury, Peter (2005). What evidence informs government population health policy? Lessons from early childhood intervention policy in Australia. *New South Wales public health bulletin*, vol. 16, no 12, p. 180-184.
- Bradt, D. A. (2009). Evidence-based decision-making (part II): applications in disaster relief operations. *Prehospital and disaster medicine*, 24(6), 479-492.
- Bruno, Isabelle. (2015). Défaire l'arbitraire des faits. De l'art de gouverner (et de résister) par les «données probantes». *Revue française de socio-économie*, (2), 213-227.
- Campbell, Elizabeth H. Urban refugees in Nairobi: Problems of protection, mechanisms of survival, and possibilities for integration. *Journal of refugee studies*, 2006, vol. 19, no 3, p. 396-413.
- Christoplos, I., Knox Clarke, P., Cosgrave, J., Bonino, F., & Alexander, J. (2017). Strengthening the quality of evidence in humanitarian evaluations. *ALNAP Method Note*. London: ALNAP/ODI (see www.alnap.org/system/files/content/resource/files/main/alnap-eha-method-note-5-2017.pdf).
- Clarke, P. K., & Darcy, J. (2014). Insufficient evidence. *The quality and use of evidence in humanitarian action*. ALNAP Study. London: ALNAP/ODI.
- Clarence, Emma. (2002). Technocracy reinvented: the new evidence based policy movement.
- Cochrane, Archibald Leman, et al. *Effectiveness and efficiency: random reflections on health services*. London : Nuffield Provincial Hospitals Trust, 1972.
- Colic-Peisker, Val et Tilbury, Farida. Integration into the Australian Labour Market: The Experience of Three "Visibly Different" Groups of Recently Arrived Refugees 1. *International migration*, 2007, vol. 45, no 1, p. 59-85.
- Darcy, J., & Hofmann, C. (2003). According to need?: needs assessment and decision-making in the humanitarian sector. HGP Report 15.
- Davies, H., Nutley, S., & Tilley, N. (2000). Debates on the role of experimentation. Davies, H. T., & Nutley, S. M. (Eds.). (2000). *What works?: Evidence-based policy and practice in public services*. Policy Press.251-276.
- De Chaine, D. R. (2002). Humanitarian space and the social imaginary: Médecins Sans Frontières/Doctors Without Borders and the rhetoric of global community. *Journal of Communication Inquiry*, 26(4), 354-369.

- Desrosières, A. (2002). *The politics of large numbers: A history of statistical reasoning*. Harvard University Press.
- Desrosières, A. (2014). *Prouver et gouverner: une analyse politique des statistiques publiques*. La découverte.
- Dijkstra Geske (2015). The new aid paradigm: A case of Policy Incoherence. *Journal of International Commerce, economics and policy*, 6(2), 1-27.
- Dijkzeul, D., Hilhorst, D., & Walker, P. (2013). Introduction: evidence-based action in humanitarian crises. *Disasters*, 37, S1-S19.
- Espeland, W. N., & Stevens, M. L. (2008). A sociology of quantification. *European Journal of Sociology/Archives Européennes de Sociologie*, 49(3), 401-436.
- Faist, T. (2011). Academic knowledge, policy and the public role of social scientists. The case of migration and development. Thomas Faist, Margit Fauser & Peter Kivisto: *The migration-development nexus. A transnational perspective*. Palgrave Macmillan, London, pp. 185-203.
- Garfield, R. (2011). *Common needs assessments and humanitarian action*. Humanitarian Practice Network, Overseas Development Institute. HPN Network Paper 69.
- Gonzalez, Gustavo (2016). New aid architecture and resilience building around the Syria crisis. *Forced migration Review*, 52, 26-28.
- Greenhalgh, Trisha and Russel, Jill. Evidence-based policymaking: a critique. *Perspectives in biology and medicine*, 2009, vol. 52, no 2, p. 304-318.
- Harrell-Bond, B., Voutira, E., & Leopold, M. (1992). Counting the refugees: gifts, givers, patrons and clients. *Journal of Refugee Studies*, 5(3-4), 205-225.
- Hilhorst, Dorothea (2002). Being good at doing good? Quality and accountability of Humanitarian NGOs. *Disasters*, 26(3), 193-212.
- Hilhorst, Dorothea and Jansen, Brom J. (2010). Humanitarian Space as Arena: A perspective on the everyday politics of Aid. *Development and Change*, 41(6), 1117-1139.
- Husson, François, LÉ, Sébastien, and PAGÈS, Jérôme. *Exploratory multivariate analysis by example using R*. Chapman and Hall/CRC, 2017.
- Korac, Maja. Integration and how we facilitate it: A comparative study of the settlement experiences of refugees in Italy and the Netherlands. *Sociology*, 2003, vol. 37, no 1, p. 51-68.
- Kühne, Simon, Jacobsen, Jannes, and Kroh, Martin. Sampling in Times of High Immigration: The Survey Process of the IAB-BAMF-SOEP Survey of Refugees. *Survey Methods: Insights from the Field (SMIF)*, 2019.
- Kukkonen, E. (2018). Inclusion of Evidence-Based Approach to Humanitarian Needs Assessment in Flash Appeals. PhD thesis, Lund University
- Le Roux, Brigitte and Rouanet, Henry. *Multiple correspondence analysis*. Sage, 2010.
- Lavrakas, P. J. (2008). *Encyclopedia of survey research methods* Thousand Oaks, CA: Sage Publications, Inc
- Le Galès, P. (2016). Performance measurement as a policy instrument. *Policy Studies*, 37(6), 508-520.
- Mallett, Richard, Hagen-Zanker, Jessica, Slater, Rachel, *et al.* The benefits and challenges of using systematic reviews in international development research. *Journal of development effectiveness*, 2012, vol. 4, no 3, p. 445-455.
- Mc Dermot, Rose. Internal and external validity. *Cambridge handbook of experimental political science*, 2011, p. 27-40.

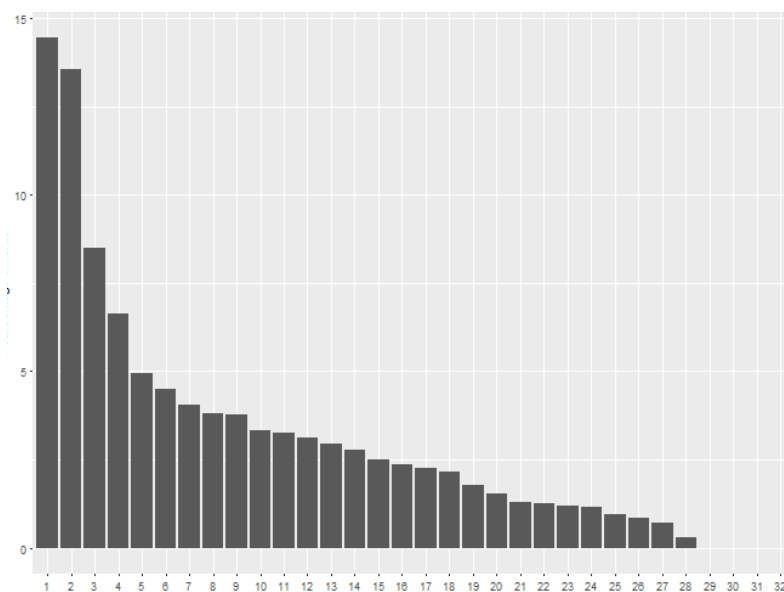
- Martson, Greg and Watts, Rob. Tampering with the evidence: a critical appraisal of evidence-based policy-making. *The drawing board: An Australian review of public affairs*, 2003, vol. 3, no 3, p. 143-163.
- Mazurana, D., Benelli, P., and Walker, P. (2013). How sex-and age-disaggregated data and gender and generational analyses can improve humanitarian response. *Disasters*, 37, S68-S82.
- Olin, E., and von Schreeb, J. (2014). Funding based on needs? A study on the use of needs assessment data by a major humanitarian health assistance donor in its decisions to allocate funds. *PLoS currents*, 6.
- Oliver, Kathryn, Innvar, Simon, Lorenc, Theo, *et al.* A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC health services research*, 2014, vol. 14, no 1, p. 2.
- Parkhurst, Justin O. and Abeysinghe, Sudeepa. What constitutes “good” evidence for public health and social policy-making? From hierarchies to appropriateness. *Social Epistemology*, 2016, vol. 30, no 5-6, p. 665-679.
- Phillimore, Jenny. Refugees, acculturation strategies, stress and integration. *Journal of Social Policy*, 2011, vol. 40, no 3, p. 575-593.
- Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press.
- Robertson, David W., Bedell, Richard, Lavery, James V., *et al.* What kind of evidence do we need to justify humanitarian medical aid?. *The Lancet*, 2002, vol. 360, no 9329, p. 330-333.
- Schreeb, J. V., Unge, C., Brittain-Long, R., and Rosling, H. (2008). Are donor allocations for humanitarian health assistance based on needs assessment data?. *Global Public Health*, 3(4), 440-447.
- Singleton, Ann. Migration and Asylum Data for Policy-making in the European Union The problem with numbers. *Liberty & Security in Europe*, 2016, vol. 89.
- Sutcliffe, S. and Court, J. (2005) Evidence-Based Policymaking: What is it? How does it work? What relevance for developing countries? London: Overseas Development Institute. Online. www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/3683.pdf
- Vogel, Dita, Kovacheva, Vesela, and Prescott, Hannah. The size of the irregular migrant population in the European Union—counting the uncountable?. *International Migration*, 2011, vol. 49, no 5, p. 78-96.
- Walker, Peter., and Russ, Catherine. (2010). Professionalising the humanitarian sector: A scoping study. Enhancing Learning and Research for Humanitarian Assistance. <http://euhap.eu/upload/2014/06/professionalising-the-humanitarian-sector.pdf>
- Watson, Cathy. Impact assessment of humanitarian response: a review of the literature. *Medford, MA: Feinstein International Center*, 2008.
- Weiss, Carol. H. (1979). The many meanings of research utilization. *Public administration review*, 39(5), 426-431.

Appendices

Appendix A1 Selected variables describing the analyzed documents (%)

<i>Year of publication</i>	
2012	7
2013	12
2014	19
2015	21
2016	13
2017	28
Share Factsheets	30
Share Assessments	25
Share Reports	28
Share other types of documents	17
Mean number of pages	28
Share of reports with one page	19
Share of reports with a bibliography	42
Share of reports with infographics	53
Share of reports with a cartographical localisation	44
Share of reports with a cartographical analysis	23

Appendix A2 Percentage of inertia by axes



Appendix A3 Example of a report summary

SUMMARY	2
Key Findings.....	3
List of Acronyms.....	6
Geographical Classifications.....	6
List of Figures and Tables.....	6
INTRODUCTION	8
METHODOLOGY	9
Phase I.....	9
Phase II.....	9
Challenges and limitations.....	10
DEMOGRAPHICS	11
KEY FINDINGS	14
Drinking Water Supply.....	14
Private WASH Infrastructure.....	16
Private WASH Infrastructure Suitability.....	19
Waste Water and Solid Waste Disposal.....	25
Communal Sewage Interceptor Tanks.....	27
Repair, Maintenance, and Reporting.....	29
CONCLUSION	35
ANNEXES	37
Annex 1: Household Questionnaire.....	37