



1

Fragility and Innovations in Data Collection

Johannes Hoogeveen and Utz Pape

1 Introduction

Fragility, conflict, and violence (FCV) represent a critical development challenge that threatens efforts to end extreme poverty and promote shared prosperity. Two billion people live in countries where development outcomes are affected by FCV, including many countries in Africa. Of the 38 countries on the World Bank's official 2018 FCV list, 20 can be found in Africa. Moreover, while the global share of the extreme poor living in conflict-affected situations is about 20%, this number is much higher in Africa, around 32%. In fact, nearly 80% of all poor people living in conflict-affected situations reside in Africa (Fig. 1).

J. Hoogeveen (✉) · U. Pape
World Bank, Washington, DC, USA
e-mail: jhoogeveen@worldbank.org

U. Pape
e-mail: upape@worldbank.org

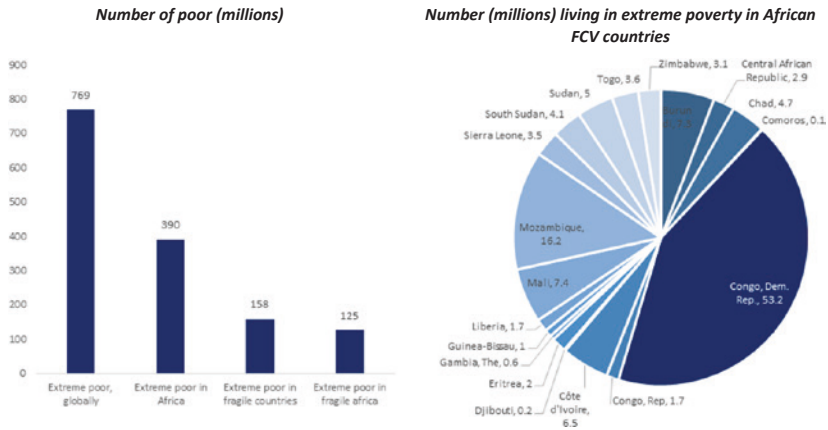


Fig. 1 Extreme poverty (2017 or latest available number) (Source World Bank, Poverty and Equity Data Portal, accessed November 2017)

Particularly worrisome is that between now and 2030, the share of extremely poor people living in FCV countries is expected to rise from 20 to 50%. Given that most of these people are likely to be in Africa, it is unsurprising that at the 2015 Annual Bank Conference on Africa, Makhtar Diop, the then World Bank’s vice president for the region, emphasized not only the importance of fragility, but also the need for a much more profound inquiry into its drivers and consequences: “Conflict and fragility exact a costly toll on the economies of Africa. As we scale up our operational work in fragile states, a better understanding of the causes and impacts of conflict and fragility can help to prevent some of the deadly conflicts at the community level.”

A better understanding of socio-economic well-being of citizens in such countries as well as measuring the impacts of shocks and conflicts start with better data. Data deprivation is a pressing problem in FCV settings for both decision makers and its citizens, and in particular, for the poor, who often lack voice and agency, and who may remain invisible unless data identify their existence and state of being. The need for reliable data on living conditions in fragile situations is even greater, and yet data deprivation tends to be worse in such contexts. Data can provide evidence on the plight of some of the most vulnerable populations,

such as the displaced, or those affected by natural disasters, violence, famine, or epidemics, and can facilitate the formulation of policy responses by decision makers. As such, there is an urgent need for data in fragile situations.

This book attempts to address this data challenge. It reflects work carried out by World Bank staff from the Poverty and Equity Global Practice and by others covering our experiences in fragile situations, facing challenges around data collection, mostly in Africa in the Central African Republic, the Democratic Republic of Congo, Liberia, Madagascar, Mali, Malawi, Nigeria, Senegal, Sierra Leone, Somalia, South Sudan but also in Iraq, Jordan, Lebanon, and Yemen.¹ Typical welfare surveys such as the Living Standard Measurement Surveys (LSMSs) and Household Budget Surveys (HBSs) that are implemented in a large number of countries are not always appropriate for these situations. Because of the pressing demand for data, there has been significant support for experimentation and innovation around data collection methods. This has allowed us to develop solutions suitable for these contexts, which are often equally relevant for non-fragile settings.

Through our experiences in identifying innovative ways to collect data, we have learned three lessons. First, it is possible to collect high-quality data in fragile settings. Doing so may require adaptations to the data collection process but situations in which no information can be collected are rare. Second, data collection in fragile contexts does not need to be more expensive than in other settings. In fact, the costs associated with many of the innovations discussed in this book compare favorably to more traditional data collection methods. Third, a careful assessment of the data needs of decision makers is essential. Often relatively easy-to-collect information goes a long way toward meeting their demands, as long as it is provided in a timely fashion. This holds

¹Not all these countries are on the fragile country list maintained by the World Bank and downloadable from: <http://www.worldbank.org/en/topic/fragilityconflictviolence/brief/harmonized-list-of-fragile-situations>. When countries are not on the fragile country list, the discussed approaches were typically applied during an emergency, as was the case during the Ebola crisis in Sierra Leone and the 2016 floods in Malawi.

particularly in volatile situations. Hence it may be sufficient to demonstrate whether respondents can engage in certain income-generating activities, without measuring how much income is actually earned. Perception questions, eliciting information about trust, security, or development priorities, tend to be very informative for decision makers in unstable settings where rumors spread quickly and where opinion polls and (objective) media reporting are absent. In other instances, simple to collect information does not suffice. We present such a case in Chapter 9 for Somalia where estimates of poverty had to be produced even though interviews could not be lengthy for security reasons, precluding asking detailed consumption questions.

There was also a fourth lesson: technology is not a panacea for all data collection issues and not everything works. We considered machine learning and big data, but these approaches were not successful. Cloud computing and improvement of statistical learning algorithms enable the use of satellite images and other sources of big data, but satellite images can be expensive, the methodologies can be complex, and external validity is at times difficult to ensure. Some data collection exercises were discontinued because of a lack of funding (and by implication, a lack of demand). Tablets facilitated electronic data collection and reduced field supervision, but in some situations, its use complicated data collection as it raised suspicion from respondents or unwanted attention from thieves. Improved mobile phone coverage also created the opportunity to use mobile phone interviews for data collection in insecure areas, but the resulting information may not be representative of the population.

It has been immensely rewarding to find ways to produce reliable data in the face of significant challenges: absent sampling frames, high levels of insecurity, and limited budgets. We feel privileged to have been given the opportunity to collect data that has helped inform decision makers at critical junctures of the development process. However, we also realize that our work is far from complete. With adaptations, many of the innovations presented in this book are scalable. This holds for district censuses, which are highly suited to inform decentralization processes, or Iterative Beneficiary Monitoring (IBM), which can be used to improve project performance in any context. Rapid consumption

surveys have the potential to significantly reduce the cost of collecting consumption data, and sampling frames derived from satellite images can be used more systematically to update sampling frames. Moreover, with cell phone coverage continuously improving, mobile phone surveys (examples presented in this book are monitoring the Ebola crisis and people displaced by the crisis in Mali, and to inform a famine response in Nigeria, Somalia, South Sudan, and Yemen) that can be scaled up rapidly during a crisis deserve to become part of the regular tool-box of disaster planning, as they can offer timely data when a crisis is imminent.

2 Data Collection in Fragile Situations

Fragility, conflicts and violence affect data collection in multiple ways. The capacity to implement and analyze complex surveys tends to be limited and resources to pay for data collection are scarce as the revenue generating capacity in FCV settings tends to be constrained and because funding for data collection competes with other urgent needs. For these reasons, few household surveys are implemented in fragile situations, or if they are, are not implemented regularly or without covering the entire territory. In addition, risks in FCV countries are oftentimes elevated, because of violence but also because of other dangers, such as disease. In Somalia, for instance, a traditional household consumption survey with interview lengths exceeding several hours was not possible given the level of insecurity and danger imposed to enumerators if spending more than one hour with a household. During the Ebola crisis, enumerators could not travel and collect information from respondents using face-to-face interviews because of the risk of infection.

Data collection during conflict is also affected by poor road quality, inadequate telecommunications infrastructure and, at times, populations that are hostile to representatives of the central government offering little in terms of key public services. The reason for these challenges is because conflicts tend to occur in locations that are physically distant from administrative centers, isolated, have low population density and few key public services, and which bear the brunt of weak state capacity.

Collecting data in such situations is not only logistically challenging, but people living in these areas often feel little loyalty to the distant capitals that have historically ignored them and may be hostile to anyone seen to represent the state.

Mobile target populations are a further complication often associated with data collection in fragile situations. Mobility is a challenge not only because pastoralists tend to live in distant, low-density areas that are often the theaters of conflict, but also because displacement is a major issue during times of insecurity. During the crisis in northern Mali, for example, 36% of the population fled the area, and in the Central African Republic, 25% of the population was displaced. The United Nations High Commissioner for Refugees (UNHCR) estimated that by the end of 2016, there were 5.1 million refugees in Africa, with the Central African Republic, the DRC, Somalia, South Sudan, and Sudan being the major sources of refugees. The number of internally displaced people (IDPs) is even higher, with almost 9 million displaced people between these five countries alone.

Data collection in FCV settings is also affected by the absence of adequate sampling frames, which may have been lost or are simply out of date. In the case of the Central African Republic, for instance, during the civil war, much of the data infrastructure (buildings, books, maps, servers, and computers) was lost to looting. However, even without the looting, sampling frames would no longer have been valid as a large proportion of the population had become displaced. Finally, there is often time pressure, as decision makers require accurate information with a quick turnaround. In the Central African Republic, following the signing of the Peace Accord, the team had 90 days to prepare, field, and analyze a survey to yield representative data on the development priorities of citizens. The pressure to inform decision makers during or directly after a disaster can be even higher, for example, in the Ebola-affected countries, or for the drought response in Nigeria, Somalia, South Sudan, and Yemen.

Because traditional data collection methods are not always suited to fragile situations, this book presents innovations developed to deal with some of these challenges. Some, though not all, were also motivated by the fact that data needs in fragile situations are different. There is much

more emphasis on timely data that can monitor a given situation than on in-depth analyses to inform policy decisions. For example, policy-makers in insecure settings often prefer knowing where schools are and whether they are still functioning, rather than seeing a detailed analysis of whether the rate of return to education is higher at the primary or tertiary level. This reality has shaped some of the data collection processes presented in this book, as questionnaires in these contexts can be less comprehensive. This in turn can be effectively combined with mobile phone interviews as a data collection method, which typically should not last longer than 20–30 minutes, and interviews by locally resident enumerators who cannot be retrained for every new questionnaire. District surveys introduced in the Central African Republic and Mali capitalized on the realization that an index reflecting the degree of public service provision (health, water, education, and infrastructure) at the lowest administrative level was a pragmatic alternative to a more detailed poverty map, which would take a long time to create. The IBM approach introduced in Mali which offers feedback to project staff drawn from light data collection exercises, was developed to complement project supervision missions, which had become difficult to conduct due to security concerns. The approach relies on highly simplified data collection tools, which ensure focus, speed and allow to keep cost down.

Simplifications, are not always possible. In Somalia, for instance, up-to-date poverty estimates were needed to inform the Heavily Indebted Poor Countries (HIPC) process. Under normal circumstances, estimating poverty requires administering a lengthy consumption module that takes several hours to complete. However, due to security concerns, it was advised that the maximum duration of a household interview should not exceed 60 minutes. This time restriction meant that a lengthy consumption module was not possible, even if questions about education, health, and perceptions were dropped. Using a new questionnaire design with smart sampling techniques at the level of questions solved this challenge.

To structure the book, we organized it into three parts. Part I: “Innovations in Data Collection” presents ways to collect data that are cognizant of security and other risks, as well as the specific data needs

of decision makers in FCV countries. The first three chapters in this section discuss data collection using mobile phone interviews. Chapter 2 provides an example of this method during the Ebola crisis in Sierra Leone. Chapter 3 describes how mobile phone interviews were used to inform a response to the drought in Nigeria, Somalia, South Sudan, and Yemen. Chapter 4 reports an exercise to track people displaced by the crisis in northern Mali. Chapter 5 discusses how, in situations where travel by outsiders is too dangerous, data collection may still be feasible by relying on locally recruited, resident enumerators who are trusted by their community. Chapter 6 discusses the district survey and Local Development Index introduced in the Central African Republic. It informed the Recovery and Peace Building Assessment and collects much of the data that feeds into the national monitoring system.

Part II: “Methodological Innovations” presents innovations with respect to collecting data and sampling. To deal with the absence of sampling frames in the DRC and Somalia, satellite images and sophisticated machine learning algorithms were used to estimate population density and demarcate enumeration areas (Chapter 7). The same chapter also showcases a novel sampling approach implemented in the Afar region of Ethiopia to ensure that pastoralists were adequately included. This approach was also used in Somalia to avoid listing exercises that were viewed with suspicion by community and authorities. Chapter 8 discusses sampling for representative surveys of displaced populations, using the example of Syrian refugees and host communities in Jordan, Lebanon and Kurdistan, Iraq. Chapter 9 offers a solution for those interested in collecting poverty estimates for insecure locations in which the time available for face-to-face interviews is too limited to implement lengthy household consumption expenditure surveys that are generally used for measuring poverty. Chapters 10 and 11 discuss how to elicit truthful information from respondents. Chapter 10 focuses on asking questions about sensitive issues such as e.g. loyalty to controversial groups while Chapter 11 deals with how to avoid strategic responses when respondents might expect benefits to be associated with certain answers.

Part III: “Other Innovations” presents a project that used video testimonials (Chapter 12) as a unique and cost-effective way to give external

Table 1 Topical guide to this book

Chapter topic	Collecting data unsafe environments	Monitoring of crises, projects and country programs	Asking sensitive questions	Dealing with inadequate sampling frames	Following mobile populations	Collecting data rapidly	Promoting development outcomes	Repeated data collection and panel data
Mobile phone surveys	✓	✓				✓		✓
Rapid response survey	✓	✓				✓		
Tracking displaced people					✓			✓
Resident enumerators	✓					✓		✓
Local development index		✓				✓		✓
Geo spatial sampling				✓				
Sampling displaced populations				✓				
Rapid consumption surveys	✓					✓		
Studying sensitive topics			✓					
Eliciting accurate responses			✓					
Video testimonials							✓	
Iterative beneficiary monitoring		✓				✓		

audiences a perspective on the lives of survey respondents. In South Sudan, a web portal was created where one can watch short video testimonials of respondents describing their situation in their own words, which not only provided the necessary context for the quantitative results, but also gave a voice to the poor. In Chapter 13, IBM is discussed, which relies on light-touch, repeated data collection exercises to create dynamic feedback loops for project staff. IBM has been found to enhance the efficiency of projects and is, because of its minimalist data demands, highly suited for fragile contexts.

We have aimed to keep this book practical and accessible, focusing on illustrations and applications, as our objective is to provide the reader with examples of what is feasible. Every chapter presents the data challenge, how it was addressed, and lessons learned. For readers interested in specific topics, we present in Table 1 an overview of which chapters might be of interest. For example, if the concern is that respondents might give biased answers, because questions touch upon sensitive issues or because the respondent may believe that the right responses can result in certain benefits, then Chapters 10 and 11, which discuss methodological solutions and behavioral nudges respectively would be worth reading.

Box 1 Using tablets for data collection allows for a rich array of innovations

Using tablets or mobile phones to collect data, or more specifically, Computer-Assisted Personal Interviews (CAPI), led to more changes than making data entry obsolete. Enumerator error can be reduced with dynamic validity checks and complex skipping patterns, opening up new possibilities. The randomization of questions can now be automated, for instance, a feature that has been part of rapid consumption surveys (Chapter 9) and list experiments (Chapter 10). Complex survey skipping patterns, not possible in paper questionnaires, become an additional option.

To improve accuracy, CAPI can identify implausible responses and request enumerators to verify or correct their responses before proceeding. This has proved useful in consumption modules, where responses can be assessed against caloric needs, or where unit values can be checked

against plausible price ranges. Photos can also be used to obtain more reliable estimates of otherwise hard to quantify, and seasonably variable units such as a “heap” or “bunch.”

The use of tablets also improves supervision. GPS locations can be collected in the background, allowing supervisors to assure that enumerators are where they are expected to be, and also assess the spatial distribution of a sample. Tablets can monitor the time it takes to record answers, and interview snippets can be recorded randomly. These features can quickly confirm whether interviews are actually conducted, reducing the need for unannounced supervision visits.

Enumerators can also take advantage of the additional hardware included in tablets. For panel surveys, households can be given a barcode, which can be photographed or scanned with a tablet, thus reducing the frequency of mistakes. The ability to take pictures and shoot video can be used to enrich feedback in other ways as well. Chapter 12 presents an instance where enumerators were trained to use their tablets to record—after the formal interview—stories about the experiences of interviewees.

Where mobile phone networks are available, tablets can send data for aggregation and real-time analysis, significantly reducing the time it takes to produce results. As data is typically sent into the “cloud,” such analysis can be done anywhere across the globe. The Rapid Emergency Response Survey presented in Chapter 3 made use of this feature. When enumerators are in the field for a long time, or when questionnaires needed to be updated because errors need to be corrected, the use of tablets allows for remote questionnaire management, a feature used in Chapter 5 to provide resident enumerators with new survey instruments and questions.

The opinions expressed in this chapter are those of the author(s) and do not necessarily reflect the views of the International Bank for Reconstruction and Development/The World Bank, its Board of Directors, or the countries they represent.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 3.0 IGO license (<https://creativecommons.org/licenses/by/3.0/igo/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the International Bank for Reconstruction and Development/The World Bank, provide a link to the Creative Commons license and indicate if changes were made.

Any dispute related to the use of the works of the International Bank for Reconstruction and Development/The World Bank that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the International Bank for Reconstruction and Development/The World Bank's name for any purpose other than for attribution, and the use of the International Bank for Reconstruction and Development/The World Bank's logo, shall be subject to a separate written license agreement between the International Bank for Reconstruction and Development/The World Bank and the user and is not authorized as part of this CC-IGO license. Note that the link provided above includes additional terms and conditions of the license.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

