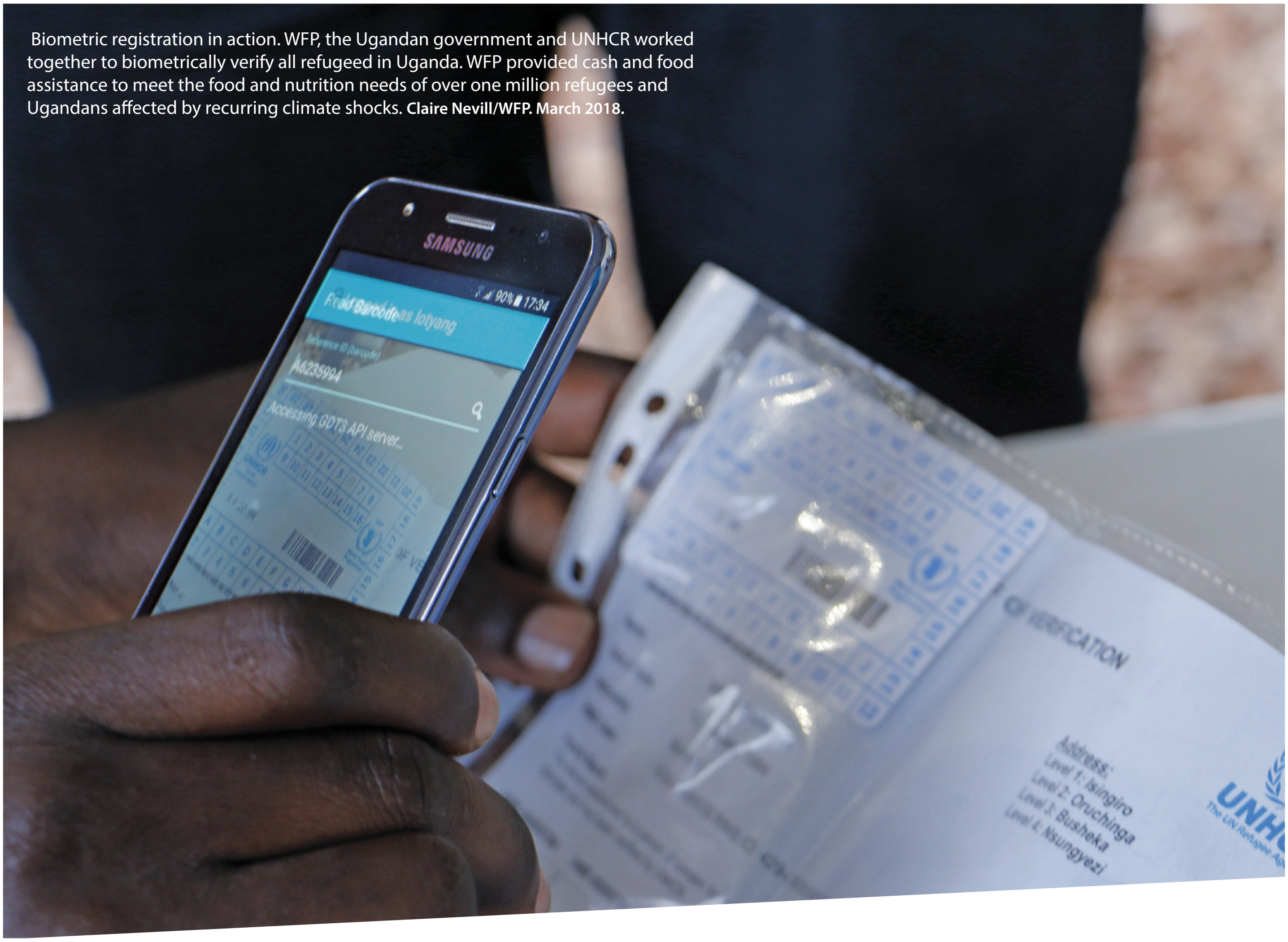
The Cash Learning Partnership

CASE STUDY:

DATA RESPONSIBILITY AND DIGITAL REMOTE TARGETING DURING COVID-19

GiveDirectly.



Biometric registration in action. WFP, the Ugandan government and UNHCR worked together to biometrically verify all refugeed in Uganda. WFP provided cash and food assistance to meet the food and nutrition needs of over one million refugees and Ugandans affected by recurring climate shocks. **Claire Nevill/WFP. March 2018.**

ACKNOWLEDGEMENTS

This case study was commissioned by CaLP with funding from the German Federal Foreign Office (GFFO).

The research for this publication was conducted between June and September 2020 by Linda Raftree, an independent consultant.

The publication was developed by Linda Raftree (@meowtree), with editorial and content supervision and support from Anna Kondakhchyan (@akondakhchyan). This case study benefited from invaluable input from GiveDirectly and many organisations and CVA practitioners, who have generously contributed their lived experience to the research process and who, because of the sensitive nature of this research, shall remain anonymous.

The views expressed in this publication are the authors' alone and are not necessarily the views of the donors or of CaLP member agencies.

CaLP is a dynamic global network of over 90 organisations engaged in the critical areas of policy, practice and research in humanitarian cash and voucher assistance (CVA) and financial assistance more broadly.

For more information, please visit CaLP website at www.calpnetwork.org

Follow CaLP on Twitter: @calpnetwork

Front cover image: A participant involved in GiveDirectly's cash programming in Kenya.

In early 2020, the COVID-19 pandemic arrived and turned the world on its head. By the end of 2020, some 75 million worldwide confirmed cases had been reported to the WHO, including 1.7 million deaths.[1] As opposed to a 'normal' crisis, the pandemic has been global in nature, affecting every country in the world to some degree. The pandemic overwhelmed health systems, morgues, supply chains and families. It shut down economies and businesses, both large and small, with a few exceptions, and led to growing food insecurity. The pandemic cancelled social gatherings, shuttered vital social services, and left many without a social or financial safety net.

While countries adopted various approaches to dealing with COVID-19, the short- and long-term need for Cash and Voucher Assistance (CVA) at unprecedented scale for previous cash and social protection beneficiaries and for newly vulnerable, unstable, or insecure populations quickly became clear. At the same time, quarantines, lockdowns, travel restrictions, and bans on group assemblies forced a shift in standard humanitarian and social protection playbooks. The COVID-19 response drove a rapid move to remote and digital channels for targeting, registration, delivery, and monitoring of CVA. Some predict that these shifts will lead to a 'new normal' that brings opportunities for improved scale and efficiency of CVA through digitization. However, there are ethical and data responsibility concerns that need to be considered and mitigated for immediate implementation and for future programme design.[2]

This case study explores digital remote targeting approaches that GiveDirectly is using for cash assistance and ways that the organization is addressing data responsibility.

THE CHALLENGE: TARGETING REMOTELY, QUICKLY, INCLUSIVELY, ACCURATELY, AT SCALE

The World Bank predicts that COVID-19 could push up to 150 million people into poverty by the end of 2021, depending on how severely the pandemic affects the world economy. This would be the first time that global poverty has risen in the past 20 years.[3] This unprecedented situation has brought new challenges: **how to deliver CVA at huge scale; as quickly, inclusively and accurately as possible; to previous recipients and to a rapidly increasing number of people who are now eligible due to COVID-19, and all without physical contact?** The pandemic has made standard CVA processes impossible in some contexts. The scale of the crisis has led to discussions, collaboration and cross-learning between humanitarian organizations and those working on social protection programming. While social protection programming is normally done in partnership with governments, humanitarian CVA tends to be directed towards populations that government social protection programmes do not cover. COVID-19 has blurred some of these lines, and many in the CVA sector are calling for closer collaboration and learning.[4,5]

In the case study below, we discuss how one of GiveDirectly's COVID-19 response programmes leverages machine-learning technology to complement an existing government run social safety net.

Due to COVID-19 lockdowns and rules which limit travel and mobility in order to stop the spread of the virus, organizations have limited their door-to-door vulnerability and needs assessments (VNA). Many have shifted to conducting CVA targeting over the phone, yet this has proven time consuming. It is also particularly challenging for an enumerator to assess the reliability of the information the applicant is providing over the phone, increasing the susceptibility to fraud and gaming. Phone registration also makes it difficult to include the most vulnerable, who may not have mobile phone access. It is also common for people in many countries to share and borrow phones and to change their SIM cards regularly, which complicates mobile phone–based remote CVA processes. Some organizations are working through community focal points; however, there are concerns that this can put community members at risk of contracting the virus. It can also lead to favouritism or manipulation of lists, or put community members in a position of having to manage demands for cash. During COVID-19, it is difficult to establish direct monitoring by a third party or provide back-up of an outside party for conflict resolution. While Know Your Customer (KYC) regulations have been waived or reduced in some cases, for example, with flexible KYC and onboarding and reduction of KYC for small transfers, KYC is difficult for organizations to conduct when registering recipients by phone, as it can be more difficult to prove identity remotely.[6,7,8]

1 World Health Organization, Coronavirus Disease Dashboard.
2 CaLP (2020) State of the World's Cash.
3 The World Bank, 'COVID-19 to Add as Many as 150 Million Extreme Poor by 2021', 7 October 2020.
4 SPACE Social Protection Approaches to COVID-19 Expert advice helpline, July 2020.
5 Social Protection in Crisis Contexts Sub Working Group.
6 Save the Children, 'Tip-sheet: Adaptations in how we identify, register, and verify CVA beneficiaries in the time of COVID-19', April 2020.
7 Mercy Corps, 'COVID-19 Tipsheet: Evidence-based Participant Selection and Targeting', April 2020.
8 GSMA 'Tracking mobile money regulatory responses to COVID-19', July 2020.

COVID-19 has made it dangerous for people to assemble in groups, so even if remote targeting can be done, delivery of cash and vouchers at scale is a challenge. When populations are experiencing a government-mandated lockdown, their options for physically withdrawing cash are limited. There are fears that the virus can spread on surfaces, meaning that keypads, point-of-sale device screens, ATMs, and even pens used to sign for cash-in-hand payments might not be possible to use. In some cases, aid workers are also on lockdown which further limits options. Where mobile kiosks are closed, it is difficult for people to cash out, and where they are open, kiosk owners and CVA recipients could be at risk of spreading the virus or of being punished for violating government lockdowns if they are supposed to be closed.[9,10,11]

Challenges due to the pandemic have led to all kinds of innovations and adaptations focused on remote targeting, enrolment, verification and delivery. While required for the ongoing COVID-19 crisis, these new approaches might be useful in future responses that require remote targeting and delivery, such as in fragile contexts. If they prove to be more cost-effective in the long term and/or able to reach scale more quickly and efficiently, they might become a common complement to existing in-person methods. **While there are fears of exclusion with digital methods, in some cases, new methods might actually be more inclusive compared to traditional means of humanitarian targeting, which still often miss some of the most vulnerable.** It remains to be seen how well these approaches achieve timeliness, adequacy and coverage, and how well agencies can evolve them to address multiple risks including fraud, mistakes related to inclusion or exclusion, misinformation and unrest. **New approaches bring new data protection and ethics considerations that organizations are exploring as they develop data-heavy, non-traditional methods.**

REMOTE TARGETING USING NON-TRADITIONAL DATA SOURCES

GiveDirectly, together with Innovations for Poverty Action (IPA) and the Center for Effective Global Action (CEGA) are exploring an emerging approach to remote targeting that uses non-traditional data sources to pinpoint geographic sub-areas with the poorest individuals and to subsequently identify those who might qualify for an emergency cash transfer due to COVID-19. Early results indicate that this approach has been effective at quickly delivering COVID-19-related cash transfers to a large number of individuals living in extreme poverty in record time. The method has elicited debate, however, in that it uses mobile phone call detail records (CDRs) and other forms of unconventional data sources.[12] **There are privacy concerns related to the use of this data as well as concerns about the potential that the most vulnerable who may not have access to mobile phones are excluded. At the same time, there is some evidence showing that big data methods can actually lead to greater inclusion,[13] especially when the alternatives for in-person enrolment remain scant.** It remains to be seen whether this type of non-traditional approach to remote targeting could be attempted in a highly fragile context, or if it would be suitable and safe to use for targeting in contexts with a large number of transient migrant, internally displaced or refugee populations with limited identity verification.

Remote targeting can include a variety of data sources, including satellite imagery, call detail records (CDRs), financial service provider data, and other datasets, including traditional sources such as existing lists from community-based organizations or government social protection programmes. Software can be programmed to identify the poorest regions of a country based on certain aspects revealed by satellite images, for example. The images capture features such as types of housing and roads and sizes of farming plots, and these provide a rough estimate of socio-economic status and quality of life. Machine learning can identify patterns and construct poverty maps that estimate levels of poverty down to a one-kilometre grid cell level at higher speeds and accuracies than approaches requiring field staff. Rather than relying on existing administrative-level aggregates of poverty estimates which are based on small samples of household-level data, this method produces high-resolution maps that combine satellite data with household-level survey data and highlights specific sub-areas where there is a high degree of poverty. In urban areas, for instance, clusters of poorer households that sit adjacent to a wealthy district can be pinpointed, whereas with traditional targeting methods, these households might be missed if the median wealth of the area is relatively high.[14] This is one way that those testing these new approaches hope that targeting could be made more inclusive.

Governments and other organizations normally do not have the resources to blanket an entire subregion with a cash programme. They might also not have updated, comprehensive data on where people live, household-level economic status or the effects of a particular crisis on the household, especially during COVID-19, when going door to door is not possible in most countries. So, a

9 Save the Children, April 2020.
10 Mercy Corps, April 2020.
11 Key Informant Interview with Mike McCaffrey, Ulana Insights, September 2020.
12 Innovations in Poverty Action, RECOVR Webinar, 14 July 2020.
13 Blumenstock, J. 'Using Mobile Phone and Satellite Data to Target Emergency Cash Transfers', Center for Effective Global Action, January 11, 2021.
14 Joshua Blumenstock, Recovr Webinar, 14 July 2020.

second step is needed to pinpoint individuals for CVA once a specific sub-geographical area is identified as one that meets general criteria for assistance.[15,16,17] This is where other databases such as tax returns, occupation records, voter registration lists and/or CDRs from mobile network operators (MNOs) come in. While there are a wide variety of unconventional data sources to analyse, MNOs in particular are a ubiquitous source of individual data records, and they have up-to-date information on the cell phone usage patterns of their subscribers. If strong privacy protections are incorporated to protect individuals, these usage patterns can be analysed to give an indication of changing economic status.[18,19,20]

For example, work by Blumenstock and team at the Center for Effective Global Action (CEGA) has shown that a mobile phone subscriber's behaviour (as recorded in CDRs) can accurately predict their socio-economic status. The intuition for this is simple: wealthy people use their phones differently than poor people. They spend more on airtime, they have different social networks, they make different types of calls (e.g., international, long-distance), they use more data, etc. Blumenstock's insight was to use machine learning to discover these patterns and from there to identify the people with the greatest need within a wider population. For a CVA programme, this kind of information could allow implementers to reach out to potentially eligible individuals by SMS to make them aware of the availability of a CVA and encourage them to register to undergo a closer vetting of eligibility.[21,22] In some cases, a subscriber's approximate location can be inferred at a fairly high level (Administrative level 1 and 2) through CDR to broadly verify if the individual lives in the target area. Many MNOs already do this – they assign users to their Most Used Cell Tower (MUCT) – to gain a better understanding of geographical usage across the country. This, of course, requires strong privacy protections and robust data-sharing agreements between operators and research institutions, or CVA implementing agencies, whether government or humanitarian.

In some cases, these emerging targeting methods are designed to provide a secondary social safety net to capture individuals who are not on existing CVA recipient lists yet whose vulnerability has increased due to the shock of COVID-19. **If carefully integrated into social protection programmes, the idea is that non-traditional targeting methods could identify and target individuals and sub-groups of households that are generally missed on official lists, for example individuals who are not a part of formal systems** or marginal communities located within wealthier areas.[23,24] In other cases, these methods are aimed at targeting a specific sub-area that has been locked down due to COVID-19 and identifying those most in need to provide supplementary short-term CVA to weather the crisis.[25]

In the humanitarian CVA space, these methods could play a role in delivering CVA during COVID-19 and beyond, however, using this type of data to identify or to reach extremely vulnerable groups such as refugees and migrants could be difficult or risky, as we will explore throughout this case study.

BOX 1: MOBILE NETWORK OPERATORS (MNOS) COLLECT MULTIPLE TYPES OF DATA AS PART OF THEIR OPERATIONS?

MNOs collect metadata, including:

- cell phone tower-level usage and location data;
- mobile money transactions;
- data about the browser and type of phone;
- mobile usage.

15 Joshua Blumenstock, Recovr Webinar, 14 July 2020.
16 Innovations in Poverty Action, RECOVR Webinar, 14 July 2020.
17 Key Informant Interview with Han Sheng Chia, GiveDirectly, September 2020.
18 Joshua Blumenstock, Recovr Webinar, 14 July 2020.
19 Blumenstock, J., Cadamuro, J., On, R. 'Predicting poverty and wealth from mobile phone metadata', Science, 27 November 2015.
20 Naef, E. et al. (2014) 'Using Mobile Data for Development', The Bill and Melinda Gates Foundation.
21 Key Informant Interview with Han Sheng Chia, GiveDirectly, September 2020.
22 Cina Lawson, Minister of Posts, Digital Economy and Technical Innovation, Togo and Shegun Bakari, Senior Advisor to the President of Togo, RECOVR Webinar, 14 July 2020.
23 Joshua Blumenstock during the Recovr Webinar, 14 July 2020. July 14, 2020
24 Mark Laichena, Africa Operations Director, GiveDirectly, during the Recovr Webinar, 14 July 2020.
25 Cina Lawson, Minister of Posts, Digital Economy and Technical Innovation, Togo and Shegun Bakari, Senior Advisor to the President of Togo, Recovr Webinar, 14 July 2020.

BOX 1: MOBILE NETWORK OPERATORS (MNOS) COLLECT MULTIPLE TYPES OF DATA AS PART OF THEIR OPERATIONS? (CONT)

Domestic MNO service providers may also have access to:

- unique identifiers for the SIM card and device (IMSI and IMEI numbers);
- time and location of transactions, such as calls and messages;
- billing data;
- data obtained during SIM-card registration, including national ID number and date of birth and, in some countries, biometric ID such as fingerprints or photographs.

Know Your Customer (KYC) requirements when using mobile money transfer for a cash transfer programme might include:

- the sender's and recipient's phone numbers;
- the date and time of the financial transaction;
- the transaction ID;
- the location and size of the transaction;
- the store where it was conducted;
- any agents involved at either end.[26]

Apps and SMS collect transaction details as unencrypted SMS messages, including:

- account balance;
- date of transaction;
- agent ID;
- transaction ID;
- transaction type (deposit, withdrawal, etc.);
- transaction amount and recipient's phone number, name and national ID.

These data points can allow inferences about behaviours, for example:

- whether a person belongs to a particular social group;
- if a person or group was singled out for humanitarian assistance during a particular period of time;
- the approximate movement of individuals after a crisis, based on location records of where they conducted transactions;[27]
- who is part of a person's network, based on subsequent transfers that don't involve a humanitarian organization – information can be inferred about these persons, even though they were not directly involved in the cash transfer programme;[28]
- whether people in a general geographical area are following quarantine or lockdown mandates;[29]
- whether an applicant should be considered creditworthy, and the probability of loan repayment.[30,31]

CASE EXAMPLE: GIVEDIRECTLY'S WORK IN EAST AND WEST AFRICA

When COVID-19 arrived in Togo in early April of 2020, the government put in place curfew and restriction of movement orders and closed schools to prevent the spread of the disease. The Government of Togo had concerns about how COVID-19 was causing increased poverty and food insecurity, however, and wanted to support the informal sector through cash transfers.

The emergency was announced on 1 April 2020, and a government-to-person cash transfer programme for the informal sector using mobile money was launched the following week in the geographic areas with the highest numbers of COVID-19 cases. This scheme, named NOVISSI, was launched prior to involvement with GiveDirectly. NOVISSI was entirely digital, with USSD used for registration and onboarding. Targeting was carried out using two eligibility requirements: geographical location and occupation

26 Mas, I. and Morawczynski, O. 'Designing Mobile Money Services – Lessons from M-PESA', Innovations: Technology, Governance, Globalization 4, no. 2 (2009): 77–91.
27 Hamilton, I. A. 'Compulsory selfies and contact-tracing: Authorities everywhere are using smartphones to track the coronavirus, and it's part of a massive increase in global surveillance.' Business Insider, April 24, 2020.
28 Privacy International and the International Committee of the Red Cross (2017) 'The Humanitarian Metadata Problem: Doing No Harm in the Digital Age'.
29 Cooper-Smith is doing this in Malawi. Flowminder is using this type of analysis for work in the Democratic Republic of the Congo, Ghana, Haiti, Namibia, Mozambique and Curacao.
30 Shema, A. (2019) 'Effective credit scoring using limited mobile phone data'. Proceedings of the Tenth International Conference on Information and Communication Technologies and Development.
31 It should be noted that there are widespread privacy and ethics concerns about the use of MNO data for these purposes. In April of 2020, the GSMA issues privacy guidelines on using MNO data for COVID-19. These offer useful guidance for other initiatives as well. See also Privacy International and the ICRC's work on the implications of metadata on privacy and harm prevention in humanitarian work.

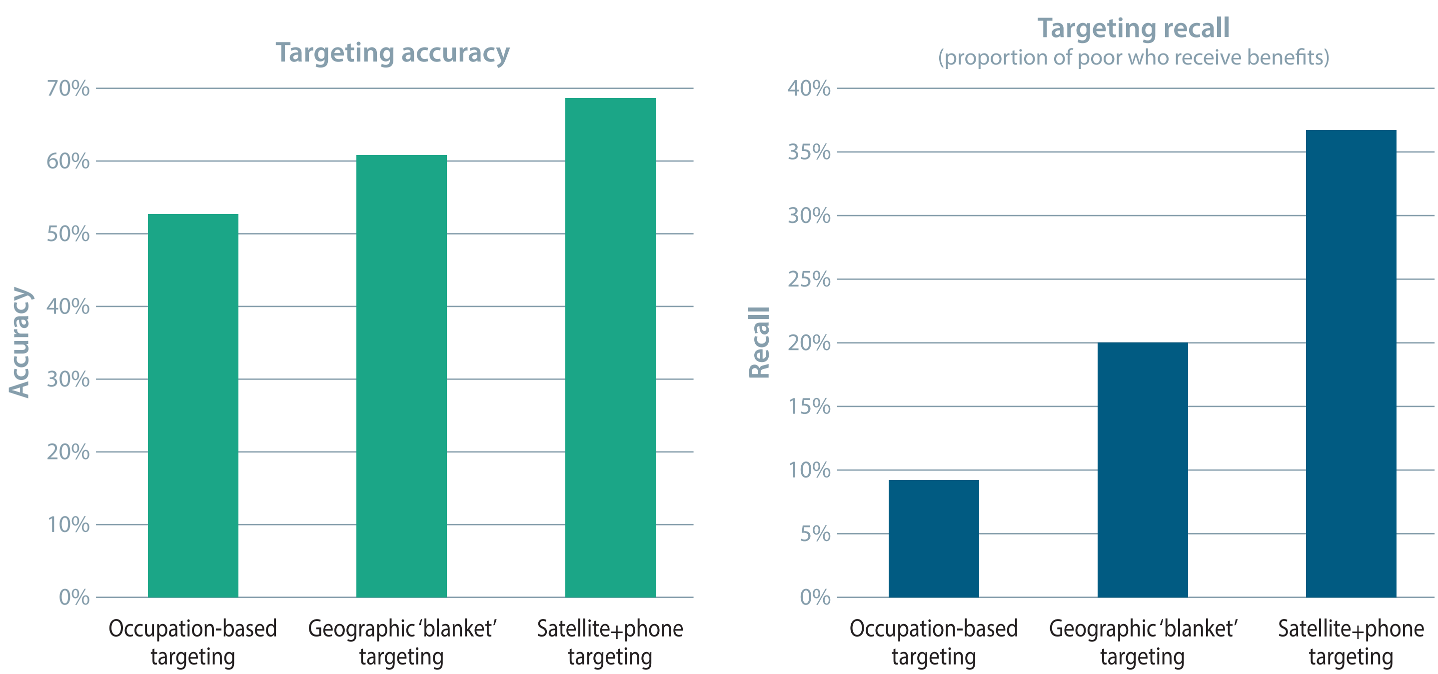
in the informal sector. Applicants to the USSD were checked against an existing voter database that held up-to-date information from a recent election to verify this. A real-time statistical dashboard was set up, as well as a logging and audit process. For verification, the voter card from recent elections, which covers 93% of the adult population, was used. In the first week, 35% of the adult population (1.38 million) registered to receive a transfer, and ultimately 15% (567 thousand) received the transfer within the three-week period following the lock-down order.[32,33,34]

Building on this initial programme, the Togolese government is working with Joshua Blumenstock from CEGA, IPA and the non-profit GiveDirectly to supplement NOVISSI's targeting methodologies. Blumenstock's team is building high-resolution poverty maps using a combination of satellite imagery, multiple nationally representative survey datasets and machine learning to identify areas of high poverty. They have also built a predictive algorithm that leverages MNO CDR data to identify cell phone subscribers in the country's 100 poorest geographies who are estimated to consume less than a particular threshold (e.g., $1.25 per day). When a subscriber applies on the USSD platform, their identity is verified against the voter registry and cell phone number checked against the poverty predictions developed by Blumenstock's team. If the subscriber is below the consumption cut-off, they are paid instantly via mobile money. As more funds are raised, the consumption cut-off can be raised to pay those who had initially not been eligible.

Early results are promising. In a December 2020 article, Blumenstock wrote that the Data Intensive Development Lab (DIDL) team compared this new targeting approach to alternative approaches that could have been used by the Government of Togo at the time. In particular, for the expansion of NOVISSI to rural areas, the Togolese government was considering two alternatives in addition to the phone-based approach: an expansion of the existing occupation-based targeting; and a 'geographic blanketing' approach that would provide benefits to all individuals residing in specific villages.

The DIDL team used data from the September 2020 phone survey to compare these options and compare the effectiveness of each of the different methods in reaching the poor. Preliminary results indicate that, assuming the goal is to reach the poorest 57,000 people in the 100 poorest cantons (which is what GiveDirectly had budget for), the satellite+phone approach is significantly more accurate than the alternative approaches. In particular, as shown in Figure 1, the satellite+phone approach is expected to provide benefits to nearly 2.5 times as many of the poorest citizens as a programme that targeted benefits based on occupation.[35]

FIGURE 1: COMPARISON OF THE SATELLITE+PHONE APPROACH TO EXISTING ALTERNATIVES (PRELIMINARY)



32 Cina Lawson, Minister of Posts, Digital Economy and Technical Innovation, Togo and Shegun Bakari, Senior Advisor to the President of Togo, Recovr Webinar, 14 July 2020.
33 Blumenstock, J., Karlan, D., and Udry, C. (2020) 'Using Mobile Phone and Satellite Data to Target Togo's Emergency Cash Transfer Program'.
34 Blumenstock, J. 'Machine learning can help get COVID-19 aid to those who need it most', Nature, 14 May 2020.
35 Blumenstock, J. 'Using Mobile Phone and Satellite Data to Target Emergency Cash Transfers', Center for Effective Global Action, January 11, 2021.

This partnership between Government of Togo, U.C. Berkeley and GiveDirectly is particularly significant because, while it was developed in response to COVID-19, it has post-crisis implications for the country's social protection programmes. Prior to COVID-19, the West African nation did not have a large national social protection database of extremely poor individuals. If this technology is rolled out further, it can form one large pathway that enables hundreds of thousands of vulnerable individuals to enter the social protection database in a non-crisis setting efficiently and cost-effectively. This promising technology can be a first step in helping bridge the humanitarian–development nexus.

Beyond Togo, GiveDirectly had embarked on a similar programme in Uganda that enrolled more than 40,000 individuals through a partnership with the country's two largest MNOs in mid-2020. However, instead of assigning a poverty score to individual subscribers (as was done in Togo), the programme invited and paid everyone who was associated with a particular geography of high poverty and vulnerability. GiveDirectly understood that most telecommunications companies already assigned subscribers to a 'home tower' based on a subscriber's usage. The intuition behind 'home towers' is simple – if a subscriber frequently utilizes their phone at night, or on the weekend from a particular cell tower, that individual can be inferred to be living in, or have significant ties to the region covered by that cell tower.

In addition to targeting via telecommunications companies, where possible, GiveDirectly also partnered with Community Based Organisations (CBOs) in these target geographies to enrol and pay their constituents remotely. This layered approach was aimed at enabling the organization to achieve scale and speed through the partnership with the telecommunications companies, while providing alternative pathways for enrolment through CBOs.

While the approach in Uganda does not verify an individual applicant's poverty level, it is simpler and reduces the time and expertise needed to conduct individual-level targeting by leveraging a geographical prediction model that the MNO already has. Such variations between Togo and Uganda show that, even within remote approaches that utilize MNO CDR data, adjustments can be made to account for local contexts, priorities and timeline trade-offs.

DATA RESPONSIBILITY CONSIDERATIONS

Although promising, implementers should thoroughly assess the ethical considerations that are associated with such new technologies. Ethics in humanitarian work is grounded in the principles of humanity, impartiality, neutrality and independence. A fundamental principle of humanitarian action is the principle of Do No Harm. Data ethics considers moral problems related to the use of data and algorithms, among other areas. Common ethics issues that arise when working with advanced data analytics approaches such as predictive analytics and machine learning include:

- Validity: Is the data or model representative of what is being measured?
- Bias and Fairness: Is the data skewed? Is there any prejudice or favouritism in the data or model? Has there been an over- or underestimation of what is being measured? Are some members of the population more or less represented than others?
- Ossification: Is the model (or the underlying data) codifying existing, systemic biases and thereby making it harder to change?
- Transparency and Explainability: Is there documentation of the process? Can others easily comprehend and explain how the model or algorithm(s) function?
- Privacy and Anonymity: Is the data somehow revealing the identity of individuals or groups?

These aspects come into play alongside the more concrete procedural elements of data privacy and security that must be addressed along the full data lifecycle.[36]

The abovementioned predictive models have potential benefits in terms of achieving targeting, enrolment, verification, and delivery of cash more quickly, cheaply and at greater scale. They might also achieve greater inclusiveness – paying people who otherwise would not have been reached because ground teams would not have been able to reach them in a door-to-door enrolment or would have introduced a degree of human error if they could go door to door. The models also cover individuals who, if not visited door to door, would not have been able to make the trip to a government office to apply or be screened, and thus would have been excluded.

36 The Centre for Humanitarian Data (2020) 'Guidance note Series. Data Responsibility in Humanitarian Action, Note #4: Humanitarian Data Ethics'.

However, as with any novel method or approach, it is paramount to address higher level ethics and also more practical data responsibility challenges. Teams working on developing these programmes have shared how various implementation designs have been debated by the research institutions, non-profits and government agencies involved in these programmes, and that the solutions are often nuanced and complex. Any trade-offs in privacy or areas normally considered valuable in non-crisis times must be necessary and proportionate to the benefits that result. Exclusion of the poorest and most vulnerable – because they lack access to a mobile phone, because the way that an algorithm is designed excludes them, or because they are unable to manage a fully remote process for whatever reason – is a serious issue. Practical challenges abound as well, when moving from a research environment to implementation in a real-world context with real-world consequences. In normal times, as noted by teams working on these methods, these challenges would be addressed cautiously through simulations, small trials, and carefully managed and evaluated roll-outs. In an emergency, especially one at the scale of COVID-19, things move more quickly, with teams having to design new workarounds for development, verification and oversight.[37,38]

Some of the ethical and responsible data questions that arise are outlined below.

Is it ethical to quickly provide an emergency cash transfer to a large swathe of the poor who qualify, even if particular demographics of vulnerable individuals (illiterate/those with a disability/those without SIMs etc.) may find it harder to enrol? Or should organizations and governments use more traditional methods in an effort to reach the most vulnerable, knowing that it will take much longer and will cost more to undertake and will still leave out many vulnerable individuals? How can this be quantified and decided?

Alternatively, can a middle ground be found wherein such digital-remote approaches form one pathway to receiving services, but are complemented by alternative models that specifically target those that may have been left out?

Some suggest that countries should invest in better data systems and establish the foundations for more inclusive social protection systems that cater to both idiosyncratic (affecting one or few people) and covariate (affecting many) shocks.[39,40] This 'cost-effectiveness conundrum', where organizations struggle to balance doing as much good as possible with other considerations, such as prioritizing the worst off, is a common one.[41] Traditional targeting methods are not perfect. They do not always reach the most vulnerable for multiple reasons, for example, because organizations tend to respond where they are already established, median poverty levels in broad geographic areas miss out on pockets of the poorest, and data is often outdated or low quality. As discussed in the Togo case study, work by GiveDirectly CEGA and IPA is attempting to quantify the effectiveness of these remote models with regard to other available approaches.

The Sustainable Development Goals include the principle of 'Leave No One Behind', which mandates a purposeful focus on the most vulnerable and hardest to reach,[42] however, this is not a simple calculation in most cases, especially in an emergency situation. Additionally, when the short- and long-term benefits and harms have not yet been qualified or quantified because an approach is still in development or has not yet been evaluated, there is little in terms of hard evidence to help with these decisions. More evidence is needed on cost–benefit and risk–benefits of these new approaches for CVA in humanitarian assistance and in social protection.

During the COVID-19 response, for example, a large number of countries conducted mass online registration with the underlying building block being a strong ID system, and there have been examples of timeliness and effectiveness being achieved. COVID-19 has provoked a move toward individualized registries as opposed to more traditional household registries, yet there is an ongoing debate over the relative merits of both approaches. Likewise, privacy, rights and inclusion of the most vulnerable are still a challenge for digital ID systems, whether national[43] or as part of the humanitarian system.[44]

---

37 Blumenstock, 'Machine learning'.
38 Key Informant Interview with Han Sheng Chia, GiveDirectly.
39 Chirchir, R. and Barca, V. (2020) 'Building an integrated and digital social protection information system.' Deutsche-Geseltschaft fur Internationale Zusammenarbeit (GIZ) and UK Department for International Development (DFID)
40 Barca V. and Beazley R. (2019) 'Building on Government Systems for Shock Preparedness and Response: the role of social assistance data and information systems.' Australia – Department of Foreign Affairs and Trade, DFAT (2019a).
41 The Centre for Humanitarian Data, 'Guidance note Series'.
42 UN. 'Universal Values, Principle 2: Leave No One Behind'.
43 Access Now (2018) 'National Digital Identity Programmes: What's next?'
44 Goodman, R. et al. (2020) 'Review and Analysis of Identification and Registration Systems in Protracted and Recurrent Crises'. Better Assistance in Crisis (BASIC). UKAID.

How are context-specific biases or gaps in mobile data accounted for? Mobile access is unevenly distributed, and the poorest and most marginalized might not have access to a mobile phone. In some humanitarian settings, a criterion for accessing CVA includes counting the number of phones a household may own and assigning an according weight (the more phones a household has, the wealthier it is considered). Access is often in the hands of men and not women, however. People may not be able to afford phones, and they commonly share phones, change phones and lose phones. One way in which GiveDirectly, Government of Togo and DIDL are addressing this is to tie poverty scores to SIM usage, and to enable new SIM registrants to be considered for the programme as well. Tying programme eligibility to a SIM instead of a phone significantly reduces the cost of participating in the programme. Furthermore, in some countries, SIM registration is mandatory, which can help to address some of these identification challenges. However, this is not the case across the board, and does not prevent people from using someone else's SIM to communicate, meaning that cell phone data is attached to a SIM, not always to a person, and if cell phone data is used to target, this can result in greater inaccuracies in programme design.

GiveDirectly notes, however, that this factor doesn't have to lead to exclusion. The fact that a SIM card is used less can be a valuable datapoint that helps enhance prediction. Predictive targeting can also be layered with alternative methods for enrolment, meaning that there are multiple entry points into the social safety net. Failing to be classified as 'in need' by a predictive model does not have to mean that a person is excluded from a programme. Rather, it could mean they need to be routed to an alternative mechanism. Those that are identified by the predictive model are 'fast tracked' to the social programme that is most aligned with their circumstances.

In general, data quality issues in CDRs need to be calibrated for. For example, cell tower coverage might extend beyond administrative boundaries.[45] The inherent lack of mobile phone access for the most vulnerable means that MNO data may under-represent particular demographics, especially women, who are less likely to have their own phone in many parts of the world. (This can potentially be corrected for through programme design such as paying women more than men, so that in a household that has one SIM, there's an added incentive for women to enrol. This approach is being used in Togo.) Additionally, in countries with multiple MNOs, there will be variances in terms of the customer base, or a need to ensure that data from all MNOs is available, which could require long negotiating processes.

Over time, it is possible that algorithms could be developed that take into consideration data quality challenges, however, careful calibration of models will still be required. It is important that principles like fairness, transparency, accountability, trustworthiness, safety and justice are the foundations for any algorithms that are developed, and that affected populations can have a voice in defining how algorithms and predictive models are designed, according to context, and in feeding back on where they might be biased during their use.[46] Blumenstock's team at U.C. Berkeley is addressing algorithmic bias in a different way. For example, the research team, in partnership with the Institut National de la Statistique et des Etudes (INSEED), the national statistical agency of Togo, made a significant effort to ensure that the data used to train their machine learning model was representative of the target population (i.e., mobile subscribers living in the 100 poorest cantons). Through a combination of adaptive survey design and the use of survey sample weights, survey respondents were selected to ensure representation of typically hard-to-reach populations, such as the extreme poor and those living in remote villages. New respondents were drawn each morning to over-sample under-represented groups from the previous days. The research team is also ensuring that all of their code is transparently documented and can be easily understood and reviewed by a third party.

Finally, DIDL and GiveDirectly are working closely with the Government of Togo at both the national, as well as local leadership levels to ensure that community voices are heard and incorporated into programmes. GiveDirectly is in touch with more than 100 community leaders and both beneficiaries and non-beneficiaries have access to hotlines through which they can voice concerns and provide feedback as part of the programme's evaluation. The research team also plans to conduct extensive in-person surveys to quantify and document any 'targeting errors' that inadvertently occurred, with particular attention to historically vulnerable subgroups such as women, elderly, and illiterate individuals. This rigorous documentation will help ensure that the generation of the programme can better mitigate these challenges.

45 Winowatan, M., Zahuranec, A., Young, A., and Verhulst, S. (2020) 'A Data Collaborative Study: Leveraging Telecom Data to Aid Humanitarian Efforts: Lessons learned from the 2015 Earthquake in Nepal'. GovLab.
46 There are a number of artificial intelligence (AI) ethics principles and guidelines to orient the development of AI and automated decision-making, including the High Level Expert Group on Artificial Intelligence (2019) 'Ethics Guidelines for Trustworthy Artificial Intelligence' and Leslie, D. (2019) 'Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector'. The Alan Turing Institute.

How can data be accessed and handled to protect anonymity and avoid individual or group harm? MNO data access should be governed by specific data-sharing agreements between the MNO and the relevant party. These data-sharing arrangements should adhere to industry accepted principles for personal data protection such as:

1. Data minimization, where only a limited set of essential information is shared to parties that need it for specific purposes. For example, in Togo, there are multiple input datasets that various parties such as CEGA/IPA, GiveDirectly, Government of Togo and external auditors require. No party has access to every dataset, and when access is granted, only essential rows and columns with each set are given. In the case of MNO CDR, only the CEGA/IPA team has access to the data and no other party can utilize it. This data minimization structure is by design.
2. Security and encryption, where Personally Identifiable Information (PII) is securely stored, password protected and encrypted where appropriate to protect individual identity when it is not required.
3. Auditability and independent checks, which refers to the transparent auditing of data management processes, systems and outputs by third parties to ensure that data systems are performing as intended.
4. Data aggregation, referring to the practice of clustering data, when possible and not needed at the individual level, to the point where individual-level data cannot be used for unintended purposes.

There are few well-rounded global policies or laws that address non-personal or group data risks. Some data analytics firms have a policy of staying away from real-time data as a way of reducing some of these risks. Data anonymization is not always sufficient to protect individuals and groups. Advances in the field of differential privacy can be applied to obfuscating individualized movement patterns, enabling those conducting the data analysis to protect individual identities and preventing individual behaviour from being extrapolated from group behaviour. However, localization of data movement patterns could still be used to make aid conditional to communities that adhere to a lockdown law, for example.[47] In the case of the GiveDirectly-NOVISSI programme, geographical predictions are aggregated at the project level, up to the 100 canton level, covering tens and thousands of square miles. This reduces the ability for any party to monitor an individual's location with any meaningful granularity.

In order to reduce the potential for individual and group-level harms, organizations and data analytics firms need to put in place internal ethic policies and frameworks (even in in the absence of legislation) and to assess the potential for short- and long-term harms that could emerge from experimental and innovative work with data. It is critical also to ensure that all parties who are accessing and using personal, sensitive or private data have put in place clear and functioning accountability mechanisms, such as ethical review boards with consequences, data policy and procedure audits, peer review, open and transparent algorithms, and clear redress mechanisms for affected populations. GiveDirectly is working with external ethics consulting firms, as well as third party academics with expertise in machine learning and public policy to pressure-test its own positions and help improve its programmes. The CEGA/IPA team is also governed by university oversight and protocols, including an Institutional Review Board (IRB).

How are data privacy laws and legal frameworks taken into consideration? Similar to digital ID programming, a CDR-focused approach may require sharing and use of personal data that is normally not shared among international agencies, governments and MNOs for reasons of individual and corporate privacy interests, and due to legal regulations that prohibit certain data from crossing international borders without proof of adequate data protection in the recipient location. While regulations vary across countries, and many privacy regulations can be suspended to some degree during public health crises, it is important to ensure that models do not rely on emergency measures, loopholes or weaknesses in data regulation in order to function. GiveDirectly does not access MNO CDR data. It also does not have access to individual voter ID data, and is unable to individually identify a subscriber until they have applied and consented to sharing their data with the organization.

Do these approaches set precedents that cannot be reversed post COVID-19? By the end of April 2020, at least 84 countries had declared public emergencies, encouraged national lockdowns or suspended certain citizen rights, including data privacy rights, in some way because of COVID-19. While this may offer opportunities for development and humanitarian communities to utilize and leverage technology and data, it is extremely important that privacy rights and uses of data during COVID-19 do not lead to long-standing normalization of invasive privacy practices by humanitarian agencies and governments. Reasonable use of data can pave the way to 'scope creep', and it is important to think about what other uses this data could be put to, and what the potential harms at individual and societal level could be.

47 Key Informant Interview, Xavier Vollenweider, Flowminder, August 2020.

Agencies have a role in helping to avoid this, and it is important to assess what the future might look like if certain uses of data are encouraged or permitted without mechanisms to restrict use or stop access afterwards. Data firms can push back on the idea of opening government access to MNO data as part of their internal ethics and value systems. In the case of GiveDirectly's programmes, CDR data is only shared with research institutions; it is not shared with governments. GiveDirectly ensures that aspects such as access limitation, purpose limitation, use limitation, data minimization, data sharing and data security practices are upheld, with clear accountability mechanisms put in place, which includes external expert panels reviewing data governance practices and adherence.

Are high-tech approaches feasible in fragile or unstable humanitarian contexts? Most humanitarian agencies and Cash Working Groups regularly conduct broad, multisector needs assessments that provide a sense of which geographical areas are poor and affected by disrupted livelihood or market activities. In many cases, these assessments, completed rapidly and focused on a small target geography, are sufficient for an emergency transfer. Many humanitarian agencies do blanket distributions in particular geographic areas based on this type of assessment, despite the potential risk of including some people that might not actually need a transfer.[48] During the COVID-19 response, rather than use high-tech approaches, many agencies have sought derogations, exemptions to audit requirements, and low-tech solutions. This is often for reasons of feasibility, but it also takes into account that local agencies could be excluded from humanitarian and development processes happening in their own locations if high-tech approaches that are out of their reach are favoured. These programmes, however, often cannot reach the level of scale that GiveDirectly is able to achieve, so it is important to consider the kinds of trade-offs that come with different levels of sophistication in methods.

Another consideration is that in countries experiencing conflict, such as South Sudan, mobile data collection might be suspect or forbidden by government.[49] Punitive governments might also make this type of approach too risky to consider. In other countries, network infrastructure is weak or phone access and use are extremely low, and cash is still handed out in envelopes, raising questions about the appropriateness of methods that involve CDRs.[50] In order to address these types of barriers, GiveDirectly considers liquidity, ability to distribute phones and SIMs, ability to engage the population in education activities, and independence of an MNO from government before initiating a programme such as the one highlighted in this case study. An assessment of the potential benefits and harms, as well as cost-effectiveness of options, should be undertaken before implementing any system, and the outcomes should be shared with the wider sector in order to derive learning.

How are consent, transparency and data subject rights and communication managed? Last, but certainly not least, lawful bases for data collection and use need to be established and followed.[51] Even if data protection laws have been temporarily suspended due to COVID-19, individuals should be made aware that their personal or sensitive data is being shared with third parties, for what purposes, for how long and what the potential outcomes or implications are. Even in the absence of consent and the use of another lawful basis for data collection, such as public task/public interest or legitimate interest, people have a right to know that their data is being use for additional purposes, and they should have a right to complain, to redress and to review by a human if they feel they have been adversely affected by an automated decision. They should also have the right to access an alternative way to be considered for CVA and the right to be considered without providing extraneous personal or sensitive data in exchange for CVA. Especially in the case of an automated decision, clear mechanisms for redress should be communicated widely. As part of its programme assessment, and ongoing monitoring and evaluation, CEGA, IPA and GiveDirectly are conducting community-based surveys to quantify the level of beneficiary comprehension and design new communication methodologies to explain the programme. While predominantly still a remote, contactless programme, GiveDirectly has sent small field teams to geographies that have lower than expected participation rates to collect feedback on local levels of programme comprehension. This is in addition to the availability of the call centre hotline and contact with community leaders that was previously discussed. This is ongoing work that is expected to roll out throughout 2021.

48 Key Informant Interview, Emily Savage, Consultant and Researcher on Cash and Voucher Assistance, August 2020.

49 Key Informant Interview, Thomas Byrnes, Regional Economic Recovery Coordinator, Middle East at Danish Refugee Council, July 2020.

50 Key Informant Interview, Mihai Magheru, Consultant and Researcher on Cash and Voucher Assistance, August 2020.

51 In order to process personal or sensitive data, a lawful purpose needs to be established. Lawful bases for data collection generally include some variation of the following: **Vital interests:** the processing is necessary to protect someone's life; **Public task/Public interest:** the processing is necessary for you to perform a task in the public interest or for your official functions, and the task or function has a clear basis in law; **Individual consent:** the individual has given clear consent for you to process their personal data for a specific purpose; **Legitimate interests:** the processing is necessary for your legitimate interests or the legitimate interests of a third party, unless there is a good reason to protect the individual's personal data which overrides those legitimate interests; **Performance of a contract:** the processing is necessary for a contract you have with the individual, or because they have asked you to take specific steps before entering into a contract; **Legal obligation:** the processing is necessary for you to comply with the law (not including contractual obligations).

What were short- and long-term effects of data use? It is important to put in place monitoring and evaluation processes that help determine not only whether speed, scale, accuracy and effectiveness were achieved, but also how those compare to other targeting approaches (both traditional and innovative). Likewise, it is important to identify any adjacent effects of this type of data use for targeting, including who was excluded, were there adverse effects or unintended consequences due to data use or misuse, and how this type of data processing affected important considerations such as privacy, freedom of expression, and surveillance. Similar to the point above, DIDL and GiveDirectly are assessing this throughout the performance of their programme, as well as through academic research conducted by the IPA and CEGA team.

CONCLUSION

COVID-19 has pushed forward efforts to innovate in the area of remote targeting for CVA programming. These have allowed for speed, scale and efficiency of programming. Rates of exclusion might be high in these novel approaches; however, rates of exclusion from not innovating or from 'business as usual' are also high. GiveDirectly and research partners CEGA and IPA are comparing new remote targeting approaches with traditional methods, in order to assess relative accuracy and quality, as well as speed, scale and efficiency. These approaches should also be assessed in ethical terms to evaluate trade-offs related to privacy, potential scope creep, future uses of data and targeting methodology, and normalization of the use of personal and sensitive data.

The Cash Learning Partnership

german
humanitarian
assistance
DEUTSCHE HUMANITÄRE HILFE

www.calpnetwork.org

March 2021