

r-tm

Measuring welfare when it matters most



Contents

Acknowledgments	V
Introduction	VI
Chapter 1: Decentralized data collection in rural Malawi for rapid welfare monitoring	1
Introduction	1
Description of the Malawi RFMS	2
Sampling Design for the RFMS	2
Questionnaire	3
Hiring and Training of Local Enumerators	5
Data Entry	5
Management and Quality Control	6
Coordinating with Multiple Stakeholders	6
Survey Implementation Costs	7
The value of the RFMS data	8
Discussion	10
Comparing RFMS and Rapid Phone Surveys	11
Choosing a Monitoring Strategy	13
References	14
Chapter 2: High-Frequency Phone Surveys: Monitoring the Effects of COVID-19 on Households and Firms in Ethiopia	15
Setting Up a Phone Survey System to Monitor the Effects of COVID-19 in Ethiopia	16
Data Ecosystem before the Pandemic	17

Survey Design	18
The High-Frequency Phone Survey of Households	19
The High-Frequency Phone Survey of Refugees	24
The High-Frequency Phone Survey of Firms	27
Implementation Arrangements	31
Main Findings	31
Household Survey	32
Refugee Survey	40
Firm Survey	44
Lessons Learned	48
Conclusion	50
References	53
Chapter 3: Representative monthly phone panel surveys: Listening surveys	57
Introduction	57
Listening Survey Design	58
Structure and content of the survey – face to face baseline and phone-based panel follow-ups	58
Alternative respondents and target topics	64
Frequency of the data collection	65
Interview duration	67
Sampling design, attrition, and weights	68
Quality control and supervision	75
Implementing partners	76
Costing and resources for Listening surveys	77
How were these surveys used?	78
Understanding and responding to shocks	78
Linking well-being and public opinion	81
Informing Policy Reform	83
Survey Experiments and Impact Assessments	85
Conclusion	86
References	88

Chapter 4: Using Geospatial Data and Modeling to Assess the Impacts of a Flood: An Application for Pakistan	91
Introduction	91
The Modeling Approach: Nowcasting with Geospatial Data after the Shock	93
The Main Channels of Impact: Loss of Income, Assets, and Purchasing Power	95
Implementing the Model in Pakistan: A Step-by-Step Guide	96
Identifying the Input Data	96
Calibrating the Damage Parameters	97
Adding Information on Household Exposure	99
Estimating Impacts across Various Groups	104
Robustness Checks	105
Cross-Validation Using Administrative Data	105
Varying Damage Parameters	106
Buffering Effects from Assets	107
Caveats and Lessons Learned	108
Annex. Geospatial Data Sources	110
References	112
Chapter 5: Frontier Approaches for Real-Time Poverty Measurement	115
Introduction	115
Novel data sources relevant to real-time	117
Measuring static welfare with novel data	123
Real-time monitoring with novel data	127
Nascent work on real-time monitoring of welfare with novel data	131
Looking forward / Conclusions	137
References	140

Acknowledgments

This edited volume was prepared by a team from the World Bank Poverty Global Department consisting of Kimberly Bolch, Maria Eugenia Genoni, and Henry Stemmler. The work was conducted under the supervision of Luis Felipe López-Calva (Global Director, Poverty Department) and Gabriela Inchauste (Practice Manager, Poverty Department). The contributed chapters were authored by Emily Aiken, Joshua Blumenstock, Erwin Knippenberg, Walker Kosmidou-Bradley, Moritz Meyer, William Seitz, Christina Wieser, Nobuo Yoshida, and Kazusa Yoshimura.

This document benefitted from consultations with many members of the Poverty Global Department as well as other World Bank teams who led the development and implementation of many of the initiatives referenced here. The team is particularly grateful to Alemayehu Ambel, Aziz Atamanov, Oscar Barriga, Paul Corral, Yeon Soo Kim, Erwin Knippenberg, Walker Kosmidou-Bradley, William Seitz, Tara Vishwanath, Christina Wieser, and Nobuo Yoshida for serving as chapter reviewers throughout the drafting and revision process.

Additionally, the team would also like to thank Federico Haslop for his research support and Juliana Soares for her support in organizing the production process. This volume received editing support from Robert Zimmerman and design services from Carlos Reyes and Gabriel Lora. This volume benefitted from financial support provided by the Umbrella Facility for Poverty and Equity.

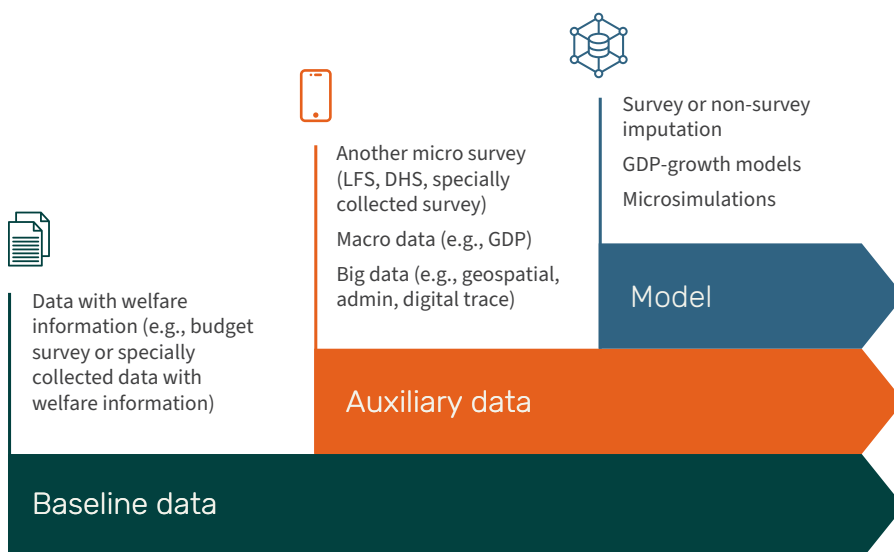
Introduction

In a global context marked by heightened uncertainty, the ability to act on reliable, up-to-date information is more essential than ever. As emphasized by the [World Development Report 2021: Data for Better Lives](#), data is a foundational input for development, enabling governments to tailor policies to people’s needs and respond effectively to shocks. Yet in many countries, the information required to guide decisions on poverty and vulnerability is not available when it is needed most.

Traditional household surveys, which underpin official poverty estimates, remain indispensable—but are often conducted too infrequently to inform timely policy action. This is particularly true in low-income countries and in fragile and conflict affected situations, where surveys may be implemented with even larger lags due to financial and operational constraints.

In response to this challenge, the World Bank’s Poverty and Equity Global Practice (GP) has been deploying innovative approaches for more timely welfare monitoring. These approaches typically work by integrating traditional surveys (“baseline data”) with alternative high-frequency data sources (“auxiliary data”) and applying a range of modelling approaches (Figure I.1). While most approaches rely on a strategic combination of these three elements, others focus solely on monitoring welfare through the direct collection or use of more high-frequency data (for example, by implementing rapid surveys or using administrative data). While the development and testing of these approaches has been ongoing for around a decade, it was accelerated by recent crises, such as the COVID-19 pandemic and climate-related disasters. In recent years, these methods have also increasingly been able to leverage more frontier data sources and methodologies, such as big data and machine learning.

FIGURE I.1 The ingredients for real-time welfare monitoring



In 2023, the Poverty Global Department launched an initiative to take stock of this growing body of knowledge. What did we know about which real-time monitoring (RTM) approaches worked best in different settings? A key milestone in this agenda was the publication of [Measuring Welfare When It Matters Most: A typology of approaches for real-time monitoring](#). That publication mapped out the broader landscape of existing RTM approaches, reflecting on relevant use cases and caveats and providing a summary of key methodological resources. The aim was to guide practitioners in choosing the most context-appropriate tools to answer their questions.

This edited volume was prepared as a complement to that publication, aimed at those readers interested in learning more about how RTM approaches have been practically applied on the ground. *Measuring Welfare When it Matters Most: Learning from Country Applications* delves deeper into selected examples, offering a more detailed look at how to design and implement high-frequency monitoring systems in different types of country settings and in response to different types of policy questions. The chapters walk the reader through these case studies and reflect on methodological best practices, practical challenges, and lessons learned.

This volume includes five chapters featuring country examples from across multiple regions—including Eastern & Southern Africa, Central Asia, and South Asia. Each application focuses on a different type of RTM approach. **The first three chapters showcase how different types of “rapid surveys” can be used to collect new high-frequency data**, exploring examples of how a decentralized model can enable more frequent collection of face-to-face survey data, the rapid deployment of high-frequency phone surveys to reach households and firms during a crisis, and the creation of longer-term monitoring systems based on representative phone panel surveys. **The last two chapters showcase how different types of existing high-frequency data sources can be better leveraged**, exploring examples of how “novel” data sources such as remote sensing data and digital trace data can be used to help update welfare estimates.

Chapter 1 presents the **decentralized face-to-face data collection approach** used in Malawi’s Rapid and Frequent Monitoring System (RFMS). By relying on locally hired enumerators and narrowing the survey scope, RFMS enables the collection of high-frequency, low-cost data that complements traditional surveys. The system has been especially valuable in tracking the impacts of cyclones and informing resilience programming in rural districts.

Chapter 2 explores the **high-frequency phone surveys (HFPS)** used in Ethiopia to monitor the impacts of COVID-19 on households and firms. These surveys, which can be implemented quickly and at a fraction of the cost of traditional surveys, provided near real-time insights to inform the government’s response. The chapter reflects on methodological trade-offs, data quality considerations, and lessons for future crisis monitoring.

Chapter 3 introduces **Listening Surveys**, nationally representative phone panel surveys conducted monthly in countries such as Tajikistan, Uzbekistan, and Kazakhstan. These surveys combine core welfare indicators with timely data on public perceptions, helping policymakers understand not only economic conditions but also how the public perceives and reacts to reforms. By capturing short-term fluctuations and sentiment, Listening Surveys fill a critical gap in policymaking, particularly where traditional data systems are too infrequent or rigid.

Chapter 4 highlights the use of **geospatial data in a vulnerability model** to rapidly assess the poverty impacts of the 2022 floods in Pakistan. By integrating flood exposure maps with household survey data and damage functions, the analysis produced estimates on poverty impacts within just two weeks of the

disaster—supporting emergency response, resource allocation, and advocacy efforts. The chapter demonstrates how such methods can deliver timely, policy-relevant insights when conventional data are lacking.

Chapter 5 reviews the use of **novel data sources** for measuring poverty and welfare in developing countries. It takes stock of the latest research on how remote sensing data, mobile phone data, and web data can be used to measure welfare—highlighting examples from nine country case studies. While innovations over the past decade have made it possible to use these approaches to reliably produce estimates of welfare levels in a cross-section, future research is needed on how to produce better estimates of welfare changes over time.

Collectively, these chapters illustrate how real-time monitoring is not a single tool, but a growing suite of approaches that can be tailored to different questions, constraints, and country contexts. Whether through phone interviews, geospatial models, or community-based data collection, these efforts reflect a shared goal: bringing timely, actionable welfare data into the hands of decision-makers when it matters most.

As countries seek to build more resilient and adaptive social and economic systems, the value of such information will only grow. This volume is offered as a practical resource for teams across the World Bank and beyond working to bridge the gap between data and action—to ensure that the evidence used to inform policies reflects the current reality of the circumstances it is trying to change.

1.

Decentralized Data Collection in Rural Malawi for Rapid Welfare Monitoring

— Kazusa Yoshimura¹ and Nobuo Yoshida²

Introduction

Face-to-face surveys are a crucial instrument for collecting reliable and representative information about households. However, large-scale implementation of these surveys typically requires a significant investment of time—in some cases, it can take several years to fully collect and process the data. Often, there is a need for more up-to-date monitoring of welfare conditions than traditional face-to-face survey methods allow. In some contexts, however, face-to-face surveys can be conducted more frequently and in a timelier manner by narrowing their scope and streamlining the data collection process. This type of rapid face-to-face survey can help collect the information needed to answer targeted and time-sensitive questions, such as how household dynamics have evolved in the wake of a natural hazard.

Like many other countries, Malawi conducts a large national survey every three years—the Integrated Household Survey (IHS)—to collect data on a wide range of socio-economic variables. The IHS data is rich in content and highly useful, but the survey is not designed for monthly or quarterly monitoring of living conditions. When sudden shocks occur, more rapid and frequent data collection becomes essential to effectively monitor impacts and inform policy responses.

This chapter describes a decentralized face-to-face data collection approach, based on local enumerators, to monitor welfare conditions and inform resilience programming in 10 districts of southern rural Malawi. The Rapid and Frequent Monitoring System (RFMS) leverages local enumerators to conduct low-cost,

¹ World Bank.

² World Bank.

high-frequency household surveys, enabling more “real-time” tracking of food security, economic resilience, and the impacts of climate shocks. This approach has complemented official household surveys by providing high-frequency data between survey rounds, thereby strengthening evidence-based policymaking.

The second section describes the design of the Malawi RFMS. The third section summarizes the survey implementation. The fourth section concludes with a discussion of how this approach compares with phone surveys and reflects on the sustainability of the RFMS going forward.

Description of the Malawi RFMS

The first attempt to build a monitoring system in rural Malawi began in 2016, when Catholic Relief Services (CRS), in collaboration with Cornell University, initiated monthly data collection in one district in southern Malawi. The goal was to monitor the resilience of households affected by flooding, using a protocol known as Measuring Indicators for Resilience Analysis (MIRA).³

Following this successful proof of concept, the MIRA protocol was expanded to 2,100 households across three districts. In 2019, it was further scaled up to become the RFMS, with the World Bank and the Malawi National Statistics Office joining the effort.

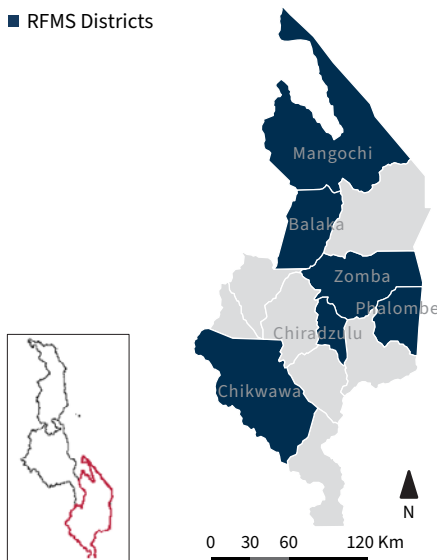
The RFMS is designed to be representative of rural southern Malawi and conducts face-to-face household-level data collection on a monthly basis. It tracks key indicators such as food security, coping strategies, and vulnerability to shocks. The system also responds to ad hoc data requests from the government and development partners. The overarching goal is to enable government agencies, development partners, and local communities to monitor and better understand household well-being and resilience, thereby informing more effective resilience programming.

Sampling Design for the RFMS

The RFMS covers 10 districts in Malawi’s Southern Region, targeting 4,500 households. The selected districts align with Malawi’s Resilience Focus Zone, as the survey is intended to support resilience-related programming.

³ The data collection continued until January 2025.

FIGURE 1.1 RFMS districts



The sampling frame is based on listing information and cartography from the 2018 Population and Housing Census. The RFMS was first rolled out in six districts—Balaka, Chiradzulu, Chikwawa, Mangochi, Phalombe, and Zomba—and later expanded in July 2021 to include four additional districts: Nsanje, Mulanje, Machinga, and Thyolo (Figure 1.1). In each district, 400 to 450 households were selected from randomly sampled Enumeration Areas (EAs).

An additional 1,600 households were oversampled in Balaka, Chikwawa, Mangochi, and Phalombe to ensure representativeness in areas where USAID and FCDO were implementing

resilience-building projects. As a result, the initial survey covered 4,200 households, expanding to a total of 6,000 households with the inclusion of the new districts.

Questionnaire

Since August 2020, the RFMS has implemented a baseline survey followed by monthly surveys for 4,200 households across six districts in southern rural Malawi (Balaka, Chiradzulu, Chikwawa, Mangochi, Phalombe, and Zomba).

Because the questionnaire is comprehensive, some modules are administered bimonthly or quarterly to keep the overall length manageable at any given time. The questions asked every month are referred to as the “Core Modules” and include topics such as shocks experienced in the past month, coping strategies, food consumption, and health status. Other modules—covering livelihoods, WASH and nutrition, and project-specific topics such as the adoption of targeted technologies—are asked either twice a year or once annually. A few additional questions designed to enable monetary poverty estimation through survey-to-survey imputation techniques are administered quarterly. The frequency and sequencing of the different modules through August 2021 are shown in Table 1.1.

The RFMS is designed to track indicators that are likely to change quickly in response to shifting climatic and economic conditions. The system is flexible, allowing modules to be added or removed as needed to respond to emerging shocks and needs. For example, when the COVID-19 pandemic began, a module was added to assess its effects on households. Similarly, when cyclones Ana and Freddy struck southern Malawi, new questions were introduced to capture flood-related damages, and data collection was conducted immediately following the events.

TABLE 1.1 RFMS questionnaire module schedule

Module	Content	Aug 2020	Sept	Oct	Nov	Dec	Jan 2021	Feb	Mar	Apr	May	June	July	
Baseline	Fixed infrastructure, complete household roster													
Monthly “Main”	Food Security, Shocks, EW Indicators, Migration													
Monthly Individual Health	Health, shock follow-up													
Covid-19	Knowledge, experience, impacts of shutdown													
SWIFT +	Poverty ranking, assets													
Livelihoods	Livelihoods, agriculture													
WASH & Nutrition	Water sources, sanitation													
Project-Specific	Exposure to technologies, program participation													

Hiring and Training of Local Enumerators

A key feature of the RFMS is its decentralized data collection approach. Enumerators are hired and based in the villages where they collect data. This model enables frequent, shock-responsive data collection while significantly reducing implementation costs—particularly for transportation and lodging. In addition to lowering marginal data collection costs, engaging locally embedded enumerators (who are residents of the areas they cover) fosters community trust and provides employment opportunities for local youth.

Another feature of the system is its joint funding and shared management. The RFMS is supported by multiple donors and managed as a common data collection infrastructure—reducing duplication, lowering overall costs, and ensuring consistent monitoring across programs.

Because enumerator quality is critical to RFMS success, the hiring process followed strict and consistent standards. Unlike a one-off survey, RFMS requires a longer-term commitment, making the relationship between candidate enumerators and villagers an important consideration. In Malawi, CRS leveraged its strong ties with local churches to help identify and introduce candidates. The National Statistics Office (NSO) participated in the hiring process to ensure enumerators met national quality standards.

While most enumerators were recruited from the target villages, some came from neighboring villages. In such cases, relocation was discussed with local leaders. Most selected enumerators were recent high school graduates.

Once the questionnaire was finalized and programmed using Computer-Assisted Personal Interviewing (CAPI), decentralized training sessions were conducted across various districts. Unlike traditional surveys, where training is centralized and conducted once, RFMS requires ongoing training. The questionnaire evolves over time, with rotating modules added or adjusted. Enumerators must therefore understand the structure and purpose of core and rotational modules. Whenever new modules are introduced, refresher training is conducted.

Data Entry

RFMS uses CommCare software for CAPI-based data collection. The system includes case management features that prompt enumerators about previous shocks

reported by households, allowing for seamless tracking and longitudinal analysis. CommCare is well-suited for managing repeated visits to the same households.

Management and Quality Control

RFMS implementation requires strong planning, real-time data quality checks, and timely data analysis and dissemination. In Malawi, these functions are managed by a central team composed of CRS staff based in-country and partners from Cornell University and the World Bank operating remotely.

Monthly surveys are typically conducted during the same week each month and take only a few days to complete. Enumerators use smartphones to capture responses, which are uploaded automatically to the cloud—enabling near real-time data visualization and quality checks.

Once data is uploaded, the central team reviews it via CommCare. If any issues are identified, they inform local CRS staff, who then coordinate with supervisors and enumerators. After cleaning, key indicators are published on a user-friendly dashboard. A summary of findings is also shared with target households and communities through the enumerators.

Supervisors play a central role in ensuring quality. They serve as liaisons between the enumerators and central team, provide regular feedback, and ensure timely, high-quality data collection. They also mediate if enumerators face challenges with communities—especially in cases where enumerators are not from the village they are assigned to. Supervisors visit enumerators at least once per month.

Coordinating with Multiple Stakeholders

RFMS was established to consolidate data collection efforts across donors and development organizations in Malawi and provide a unified, coordinated platform. It is a joint project, co-owned by multiple donors and involving a range of counterparts.

In the first year, coordination meetings were held every two weeks to align project design with partner interests, including oversampling locations and module content. Larger stakeholder meetings, including line ministries, were held quarterly to review findings and discuss how to integrate RFMS into government data systems.

Survey Implementation Costs

RFMS supports the monitoring of multiple programs while remaining flexible in content and geographic coverage. To meet these goals, it requires centralized management and significant investment in enumerator and supervisor recruitment and training.

While the initial setup involves substantial fixed costs, the marginal cost per household is very low. Fixed costs include the operation of a central office, coordination with government and partners, design and revision of survey instruments, and establishment of data systems. Setting up this infrastructure required \$2–3 million.

Marginal costs are minimized through the use of local enumerators, as pioneered under the MIRA methodology. Hiring enumerators from the villages they serve eliminates the need for travel and lodging. As a result, the per-household cost of survey implementation is just \$2–3—significantly lower than both traditional face-to-face and telephone surveys. Table 1.2 summarizes the RFMS features and costs.

TABLE 1.2. Malawi RFMS features and costs

Topics	RFMS
the level of representativeness	Rural South Malawi
# of projects	Multiple projects
Areas for hiring of local enumerators	Village
Location of data collection	House of sample household
Fixed costs	\$2 - 3 million
Marginal costs	\$2 - 3 per household
Frequency	Monthly
Survey period	August 2020 onward
Data collection agency	CRS/Cornell University

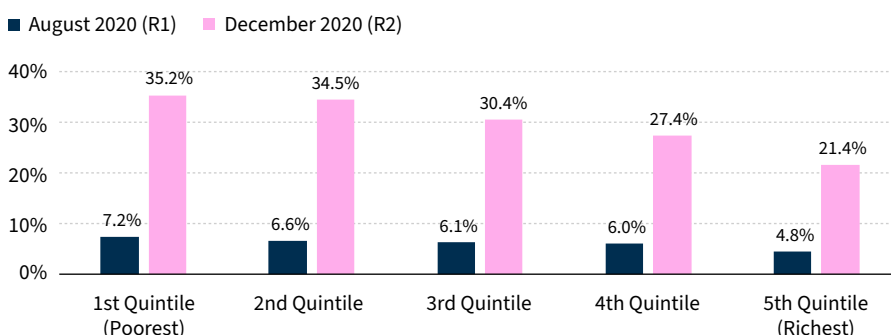
The value of the RFMS data

Malawi has frequently been exposed to various weather-related shocks, including dry spells and flooding. Because the majority of the population still relies heavily on maize farming—and typically cultivates it only once per year—they are particularly vulnerable to these natural disasters. In addition, the COVID-19 pandemic hit Malawi hard in 2020, as it did in many other countries. RFMS data helped assess the impacts of these events and distinguish them from seasonal changes (for more details, see Yoshimura et al., 2023).

As part of the RFMS, households are asked each month whether they have been affected by any natural and/or socioeconomic shocks. For example, Figure 1.2 shows the percentage of households across consumption quintiles who reported experiencing a dry spell in the month preceding the survey. In both August and December 2020—particularly in December—it is evident that poorer households were more likely to report being affected by drought.

Although droughts are covariate shocks, wealthier households tend to be less vulnerable to their effects. This is because they own more assets, practice crop diversification, and use improved or drought-resistant seeds. Moreover, better-off households often have additional sources of income beyond farming.

FIGURE 1.2 Percent of households vulnerable to drought

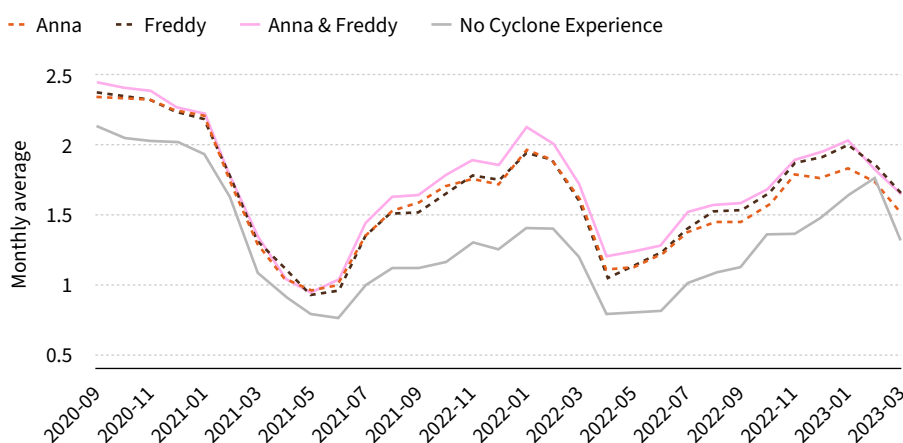


Source: Authors' estimation using data from RFMS in August and December 2020.

Another example of recurrent weather events in Malawi is flooding. In January 2022, Tropical Storm Ana struck southern Malawi, followed by Cyclone Freddy in March 2023. Shortly after each of these events, the RFMS added questions to capture the impacts of the cyclones and immediately began interviewing affected households.

Figure 1.3 illustrates the relationship between food security and vulnerability to shocks, using the Household Hunger Scale (HHS) to compare households affected by Cyclone Ana and/or Cyclone Freddy with those that were not affected. The results clearly show that households impacted by the cyclones had higher food insecurity scores—even before the events occurred. This indicates that those vulnerable to shocks are structurally worse off, underscoring the importance of tracking the same households over time to distinguish the effects of shocks from pre-existing structural disadvantages.

FIGURE 1.3 Vulnerability to cyclone and food security



Source: Estimation by CRS using data from RFMS during 2020 – 2023.

Beyond specific findings, RFMS data has proven to be a valuable tool for donors, program implementers, policymakers, and communities. A key feature of the RFMS is its engagement with existing community and district structures to help them understand and use the data. CRS has developed simple, visual dashboards at both the community and district levels to present key results from the monthly data. In addition, the data has been used to identify households that are currently vulnerable, as well as those most likely to be affected by future seasonal or unexpected shocks.

RFMS data has enabled timely analysis of which interventions contribute most to improved resilience, allowing program teams to make quicker adjustments than would be possible with traditional monitoring systems. The system has also been used to compare the frequency and duration of shocks experienced by households receiving support versus those not covered by programs, helping assess

the effectiveness of interventions. The survey tool includes a case management feature that prompts respondents about previously reported shocks, enabling continuous tracking and analysis of household conditions over time.

The data has contributed to the estimation of poverty rates, inequality measures, and income data, supporting the design of more informed pro-poor policies. Furthermore, machine learning models based on RFMS data have been used to forecast food insecurity and support anticipatory action in response to emerging risks.

Discussion

The implementation of the Rapid Feedback Monitoring System (RFMS) offers important insights into the potential of decentralized, face-to-face data collection models. RFMS, with its innovative approach of hiring local enumerators and employing rapid, real-time data collection methods, has proven effective in providing timely information on key socioeconomic indicators such as poverty and food security. This model enables fast data collection and adaptability to changing conditions, making it well-suited for continuous monitoring of vulnerable populations.

However, while RFMS has very low marginal costs per round of data collection, it requires a significant initial investment, resulting in high fixed costs. When there is a need for large-scale, frequent data collection across multiple projects over a sustained period (e.g., 3 to 5 years), RFMS can be a cost-effective option due to its low per-round costs.

The choice of data collection method depends on several factors, including survey objectives, questionnaire complexity, and the characteristics of the target population. It is also important to recognize that data collection methods can be complementary. Traditional household surveys remain foundational but can be augmented by phone or decentralized face-to-face surveys for more frequent monitoring. When surveys require extensive and complex questionnaires, the population is geographically concentrated, and primary sampling units are not widely dispersed, traditional approaches may still be preferable.

Comparing RFMS and Rapid Phone Surveys

Phone interviews offer another option for rapid monitoring, as discussed in the following chapters of this volume. Phone and decentralized face-to-face data collection methods each have distinct advantages and limitations. The choice between them depends on context, objectives, and available resources. Each involves trade-offs; below we summarize key considerations based on the RFMS experience.

Coverage and Representativeness

Face-to-face surveys using local enumerators offer clear advantages in reaching remote, marginalized, and phone-inaccessible populations. This is especially relevant in rural or low-income settings like Malawi, where mobile phone access is limited. Locally hired enumerators are particularly valuable in fragile settings or small, sparsely populated areas (e.g., small islands), where phone coverage is poor or unreliable. Shared language and cultural familiarity with respondents can reduce security risks and enhance trust. Local enumerators also tend to have better awareness of safety conditions in their areas (see, for example, Hoogeveen and Taptué, 2019). In countries with high travel costs—such as those with many islands—local hiring can reduce logistical expenses and offer a viable alternative to traditional surveys.

By contrast, phone surveys tend to be more efficient in urban or middle-income areas with high phone penetration, but they may systematically exclude poorer households without reliable phone access, introducing bias.

Data Quality

RFMS-type surveys enable direct interaction, allowing enumerators to build rapport and observe non-verbal cues. However, a key challenge is the limited availability of qualified local enumerators. In rural Malawi, the number of individuals with secondary education is low, making it difficult for enumerators to operate CAPI tools and tablets effectively, even with comprehensive training. Additionally, training large numbers of local enumerators can be costly and logistically complex. While error detection features in CAPI and a mix of in-person and remote supervision help address these challenges, basic capacity constraints remain.

Decentralized face-to-face methods like RFMS may also face social desirability bias or local political pressures due to familiarity between enumerators and respondents.

Conversely, phone surveys can reduce interpersonal bias by using standardized scripts and offering respondents greater anonymity. They also allow for the recruitment of highly qualified interviewers from across the country, bypassing geographic limitations. However, phone surveys are limited in their ability to probe responses or clarify complex questions, making them less suitable for long or multi-topic questionnaires. Moreover, in many low-income contexts, phone ownership is not widespread, which can introduce sampling bias.

Cost and Operations

Phone surveys are typically cheaper, faster to deploy, and particularly useful for rapid assessments or crisis monitoring. They require no travel, minimal field planning, and shorter training periods. A phone interview typically costs \$15–\$20 per respondent. Given the low startup costs, phone surveys can be more cost-effective than RFMS when the sample size is small or data collection is infrequent. However, they often face high staff turnover and declining respondent engagement over time.

In contrast, surveys using local enumerators—such as RFMS—benefit from contextual knowledge and achieve higher response rates. Although RFMS requires significant upfront investment to establish a central management unit, it has lower marginal costs per round compared to phone surveys. If implemented by a National Statistical Office (NSO), the upfront costs may be much lower, given their established infrastructure and experience in decentralized data collection (e.g., censuses involving locally hired and trained staff).

Cost considerations are highly context-specific. In Malawi, for example, the World Bank and the Japan International Cooperation Agency (JICA) partnered with the Malawi NSO to collect frequent evaluation data for JICA's agriculture commercialization project using decentralized data collection. To mitigate capacity constraints, the NSO hired enumerators at the sub-district rather than village level, which increased per-round costs. Still, the combined fixed and marginal costs averaged about \$15 per household—equivalent to phone survey costs.

Sustainability and Enumerator Management

Phone surveys are easier to scale using centralized call centers but often suffer from high staff turnover and declining respondent engagement. Conversely, surveys using local enumerators benefit from contextual knowledge and achieve better response rates. While turnover can also affect locally hired teams, in rural Malawi it has remained low due to high salaries compared to local income opportunities. However, in areas with more employment options, enumerator turnover may increase, leading to higher training and recruitment costs.

Choosing a Monitoring Strategy

In summary, the choice of monitoring strategy should be guided by the context, nature of the questions, and available funding. Phone surveys are well-suited for rapid deployment and are more cost-effective than traditional surveys. However, they often suffer from sampling biases, particularly in developing countries, and are constrained by unstable phone connections and limits on questionnaire complexity.

RFMS and similar face-to-face approaches involve higher initial setup costs but much lower per-round costs. If data collection is expected to continue over several years, RFMS may be more cost-effective than phone surveys. Moreover, RFMS can help reduce sampling bias significantly.

While it is easier to recruit highly qualified enumerators for phone surveys, RFMS faces challenges in sourcing qualified local personnel. Both methods face difficulties collecting complex data, though for different reasons: phone surveys are constrained by communication limitations, while RFMS may be constrained by enumerator capacity.

References

- Ballard, T. J., Coates, J., Swindale, A., & Deitchler, M. (2011). *Household Hunger Scale: Indicator Definition and Measurement Guide*. Washington, DC: FANTA-2 Bridge, FHI 360.
- Kilic, T., Serajuddin, U., Uematsu, H., & Yoshida, N. (2017). *Costing Household Surveys for Monitoring Progress toward Ending Extreme Poverty and Boosting Shared Prosperity* (Policy Research Working Paper No. 7951). Washington, DC: World Bank.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434), 473–489.
- Hoogeveen, J., & Taptué, A.-M. (2019). Iterative beneficiary monitoring of donor projects. In J. Hoogeveen & U. Pape (Eds.), *Data Collection in Fragile States: Innovations from Africa and Beyond* (pp. 215–234). Cham: Palgrave Macmillan.
- Yoshida, N., Yoshimura, K., Cardona, L., & Li, X. (2023). *Monitoring the Smallholder Horticulture Empowerment Promotion (SHEP) Project in Malawi Using SWIFT: Project Completion Report*. Tokyo: JICA.
- Yoshida, N., & Yoshimura, K. (2025). *Guidelines for High-Frequency Data Collection for Fragile Populations* (Unpublished manuscript).
- Yoshimura, K., Aron, D., Campbell, J., Li, X., Upton, J., Yoshida, N., & Zhang, K. (2022). Rapid Feedback Monitoring System (RFMS) – Real-time, cost-effective, shock-resilient monitoring of living conditions and food security. Paper presented at the *UNECE Expert Meeting*.

2.

High-Frequency Phone Surveys: *Monitoring the Effects of COVID-19 on Households and Firms in Ethiopia*

– Christina Wieser⁴

High-frequency phone surveys (HFPS) have become an increasingly popular data collection tool, especially in contexts where in-person surveys are difficult to conduct. Their main advantage lies in speed, frequency, and cost: they can be deployed rapidly, repeated often, and conducted at a fraction of the cost of traditional face-to-face surveys. In Africa, for example, HFPS were estimated to cost up to 30 times less than in-person surveys (Zezza et al. 2022). However, this efficiency comes with limitations. HFPS typically have a more limited scope and often exclude detailed modules on consumption and expenditure. Like other survey methods, HFPS can be structured as either cross-sectional surveys or short- to medium-term panels, where the same households or individuals are re-interviewed over time.

This chapter explores the design and implementation of HFPS in Ethiopia to monitor the impact of the COVID-19 pandemic on households and businesses. It discusses key methodological choices, implementation challenges, and lessons learned, and highlights the value of phone surveys in supporting crisis response and policymaking. The analysis underscores the importance of aligning survey design with context and objectives, and of carefully managing trade-offs between rapid data collection and ensuring data quality and representativeness.

⁴ World Bank. We would like to acknowledge the contributions of the following individuals: Federico Haslop, Consultant in the EPVGE department, for his research support; Yeon Soo Kim, Senior Economist in the EPVGE department, as the author of Box 1; Jeffery Tanner, Senior Economist in the OPSSR department, as the author of Box 2; and Alemayehu A. Ambel, Senior Economist in the DECLS department, who was a co-researcher on the Ethiopia high-frequency phone survey and reviewed this draft.

Setting Up a Phone Survey System to Monitor the Effects of COVID-19 in Ethiopia

The pandemic had severe, though short-lived effects on the Ethiopian economy. External demand collapsed, particularly in the garment, textile, and agricultural sectors. Remittances fell by 10 percent in fiscal year 2019/20, though they began to recover in the first half of fiscal year 2020/21. Foreign direct investment was severely affected. Inflows dropped by 20 percent in 2019/20, leading to weaker reserves (Sánchez-Martín et al. 2021). Firm revenues and household incomes were badly depressed.

Aspects of the pandemic response, such as school closures, were long-lived and may leave scarring effects on future economic performance. The majority of students suffered deficits in access to remote learning, and these deficiencies were more apparent in already deprived households. The pandemic also had a large impact on health. The country experienced a pandemic-related cumulative excess mortality rate that reached 337 deaths per 100,000 people, which exceeded mortality rates in African countries of comparable wealth, such as Burkina Faso (195), Guinea (169), and Rwanda (250).⁵

COVID-19 related restrictions were short-lived and not as stringent in Ethiopia as in other countries. The government of Ethiopia introduced containment policies quickly following the report of the first case on March 13, 2020. It declared a five-month state of emergency in April 2020. Land borders were closed; schools were locked; interregional public transport and public gatherings were banned; and nightclubs and entertainment venues were shut down. By April 2020, over 42 percent of firms in Addis Ababa were completely closed, and none were fully operational, primarily because of government-imposed movement and business activity restrictions under the state of emergency. The government accompanied this policy with tax forgiveness policies for firms, prohibitions on laying off workers, and fiscal spending packages for food distribution, health care, shelter, agriculture, and refugee support.⁶

⁵ Refer to COVID-19 Data Explorer (dashboard), Our World in Data, Global Change Data Lab and Oxford Martin Program on Global Development, University of Oxford, Oxford, UK (accessed August 26, 2024) <https://ourworldindata.org/coronavirus#explore-our-data-on-covid-19>.

⁶ Policy Responses to COVID-19: Policy Tracker (dashboard), International Monetary Fund, Washington, DC (accessed August 28, 2024), <https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19#E>.

To detect the effects of the pandemic, the World Bank Ethiopia team, in collaboration with the government, designed and implemented three high-frequency phone surveys, one with firms, one with households, and one with refugees. These phone surveys were established to monitor the effects of the COVID-19 pandemic on Ethiopia's economy and people and inform interventions and policy responses. The next section describes how the high-frequency phone surveys were used to monitor the impacts of COVID-19 on households, firms, and refugees in Ethiopia and highlights the main findings of each survey.

Data Ecosystem before the Pandemic

Ethiopia collects a wide range of survey data through various national and sector-specific surveys. These are crucial in informing policy decisions, guiding development programs, and understanding socioeconomic conditions. Ethiopia's survey data collection encompasses a broad spectrum of topics, from demographics and health to agriculture and economic activities. The surveys are instrumental in shaping public policy, monitoring development progress, and addressing key challenges in the country.

Traditional data collection efforts include census and household surveys, but at infrequent intervals. Ethiopia's last census was collected in 2007 and is outdated. The Central Statistical Agency is responsible for the collection of national household surveys. The national household surveys used to monitor poverty and other socioeconomic outcomes—the Household Consumption Expenditure Survey and the Welfare Monitoring Survey or their previous iterations—are collected infrequently (2005, 2011, and 2016).⁷ The last survey collected in 2021 was never made public. Other household survey data collection efforts include the infrequent National Labor Force Surveys (1999, 2005, 2013, and 2021), the Urban Employment Unemployment Survey (12 rounds collected between 2003 and 2020), the Ethiopia Demographic and Health Surveys (2000, 2005, 2011, 2014, and 2016), and the Agricultural Sample Surveys.⁸ In addition, the World

⁷ HCE (Ethiopia Household Consumption Expenditure Survey) (dashboard), Global Health Data Exchange, Institute for Health Metrics and Evaluation, University of Washington, Seattle, <https://ghdx.healthdata.org/series/ethiopia-household-consumption-expenditure-survey-hce>; WMS (Ethiopia Welfare Monitoring Survey) (dashboard), Global Health Data Exchange, Institute for Health Metrics and Evaluation, University of Washington, Seattle, <https://ghdx.healthdata.org/series/ethiopia-welfare-monitoring-survey>.

⁸ AgSS (Agricultural Sample Survey, Ethiopia) (dashboard), Institute for Health Metrics and Evaluation, University of Washington, Seattle, <https://datacatalog.ihsn.org/catalog/?page=1&sk=Agricultural%20Sample%20Survey%20Ethiopia&country%5B%5D=66&ps=15>; Ethiopia DHS (Ethiopia Demographic

Bank, in collaboration with the Central Statistical Agency, collects the Ethiopia Socioeconomic Survey (2012, 2014, 2016, 2019, and 2022).⁹

The national statistical system currently does not collect high-frequency data. The COVID-19 monitoring surveys represented the first time the government undertook a high-frequency data collection effort. For this reason, the World Bank funded the data collection. Because of the numerous large-scale research initiatives in Ethiopia, many private sector research and data collection firms are active and contribute to the country's data ecosystem. They range from firms specializing in market research to entities focused on public health, social research, and technology-driven solutions. The World Bank was thus able to draw from a strong pool of potential firms for the surveys.

Survey Design

COVID-19 monitoring surveys were a critical need in Ethiopia because of the widespread impacts of the pandemic on households and firms. The spread of the virus exacerbated existing vulnerabilities, particularly among poorer households and among refugees, who faced heightened economic and social challenges. High-frequency phone surveys were an appropriate tool for capturing the real-time effects of the pandemic on these populations, providing valuable data on employment, income, and access to essential services. By including refugees and poorer households in the surveys, policy makers could gain a comprehensive understanding of the impact of the virus across various segments of society. Because of the possibility of infection, face-to-face surveys were not feasible, and data were collected by phone. The first case of COVID-19 in Ethiopia was confirmed on March 13, 2020, and survey enumerators and interviewers were sent to the field the following month, in mid-April. The next subsection provides detailed information on the creation of the sample frame and the data collection methods.

and Health Surveys) (dashboard), DHS Program, ICF International, Rockville, MD, https://dhsprogram.com/Countries/Country-Main.cfm?ctry_id=65&c=Ethiopia&Country=Ethiopia&cn=&r=1; NLFS (Ethiopia National Labor Force Survey) (dashboard), ILOSTAT, International Labour Organization, Geneva, <https://webapps.ilo.org/surveyLib/index.php/catalog/LFS/?page=1&sk=Ethiopia%2F&ps=15&repo=LFS>; UEUS (Urban Employment Unemployment Survey, Ethiopia) (dashboard), Institute for Health Metrics and Evaluation, University of Washington, Seattle, <https://datacatalog.ihsn.org/catalog/?page=1&sk=Urban%20Employment%20Unemployment%20Survey%20Ethiopia&ps=15>.

⁹ ESS (Ethiopia Socioeconomic Survey) (dashboard), Microdata Library, World Bank, Washington, DC, <https://microdata.worldbank.org/index.php/catalog/6161>.

The High-Frequency Phone Survey of Households

The phone survey of households monitored the economic and social impacts of and responses to the pandemic among households. The household survey interviewers called a sample of households every three to four weeks over a 12-month period for a total of 11 survey rounds. The initial panel consisted of approximately 3,000 households that had access to mobile phones and that were located in urban or rural areas nationwide. High attrition, partly associated with an armed conflict in Tigray, one of the regions in the north of the country, and partly because of difficulties in maintaining the engagement of households in the repeated cycles of data collection, meant that only around 2,000 households completed all rounds of the survey. This subsection briefly describes the methodological aspects of the survey by summarizing the more thorough discussion offered by Ambel, Bundervoet, et al. (2020).

The selection of adequate survey frequency was crucial because of the rapidly evolving situation in infections, policy measures to curb the spread of the virus, and the substantial impact on lives and livelihoods. A high-frequency approach was considered important to capture timely and relevant data, which was essential for informed decision-making in such dynamic circumstances. However, the high frequency posed feasibility challenges, given the constraints on resources and budget. Additionally, there were trade-offs to weigh. A pace of survey rounds that was too rapid risked overwhelming the ability to process and utilize the information in policy making, but a pace that was too slow might not support timely and effective decision-making. To balance these factors, the team opted to interview households every three or four weeks for the first six months, providing critical insights during the most volatile period. The frequency might subsequently be reduced to maintain sustainability while capturing sufficient data.

Sampling Frame and Weights

The household phone survey sample is a subsample of the Ethiopia Socioeconomic Survey 2018–19. This approach followed the first-best sampling strategy for high-frequency phone surveys, which involves basing the survey sample on an existing representative (face-to-face) survey, such as the Living Standards Measurement Study or the Demographic and Health Surveys (Himelein et al. 2020). The Ethiopia Socioeconomic Survey is a regionally and nationally representative sample of households in Ethiopia. Of the 6,770 households in urban and rural areas responding to the survey, 5,374 had at least one valid phone number. To account for nonresponse and attrition, all 5,374

households were called during round 1 of the survey. Of these, 3,249 responded to the survey in round 1 and were called repeatedly during every survey round (refer to table 2.1).

TABLE 2.1 Survey respondents, rural and urban areas, by round, number

Location	Round 1 Apr 22–May 13, 2020	Round 2 May 14–June 3, 2020	Round 3 June 4–26, 2020	Round 4 July 27–Aug 14, 2020	Round 5 Aug 24–Sep 17, 2020	Round 6 Sep 21–Oct 13, 2020	Round 7 Oct 19–Nov 10, 2020	Round 8 Dec 1–21, 2020	Round 9 Dec 28, 2020–Jan 22, 2021	Round 10 Feb 1–23, 2021	Round 11 Apr 12–May 8, 2021
Rural	978	940	934	838	775	760	716	576	553	537	442
Urban	2,271	2,167	2,124	2,040	1,995	1,944	1,821	1,646	1,524	1,641	1,540
National	3,249	3,107	3,058	2,878	2,770	2,704	2,537	2,222	2,074	2,178	1,982

Source: Ambel et al. 2022.

Linking phone survey households with data from the Ethiopia Socioeconomic Survey allowed the construction of a socioeconomic profile for the entire population. The sample is evenly distributed across consumption quintiles nationally: 21 percent in the poorest quintile and 18 percent in the richest quintile were represented in the survey. In rural areas, 87 percent of household heads were engaged in agriculture, and 60 percent had no formal education. About 62 percent of rural households had a modern roof; 68 percent had access to improved water; 94 percent owned their dwellings; and only 4 percent owned televisions. In urban areas, 96 percent had a modern roof; 98 percent had access to improved water; 42 percent owned their dwellings; and 52 percent owned televisions.

Substantial effort was made to avoid attrition. Phone surveys are prone to high attrition. For this reason, strict survey protocols were implemented to reach households. Each household in the sample was called up to three times a day over a three-day period (nine calls total) before being flagged as a nonresponse. Moreover, all 3,249 households that responded in the first survey round were called in each subsequent round. The survey objective was explained to each

household contacted, and consent to participate was obtained. Interviews were conducted in Amharic, Afan Oromo, Afar, Somali, Tigrigna, and Wolayita. Yet, response rates were low, and attrition was a concern, especially in rural areas. In survey round 9, the rural sample only contained 553 households, compared with 978 households that responded in round 1.

Sampling weights were applied to make the sample representative of the whole population and to adjust for round-on-round attrition. As discussed by Ambel, Bundervoet, et al. (2020), to obtain unbiased estimates, the reported information was adjusted using sampling weights according to the methodology outlined by Himelein (2014). The steps included the following:

- Begin with base weights from the National Statistical Office 2018/19 for each household.
- Incorporate the probability of subselection of round 1 unit for each of the phone survey households, calculating the probability of selection for each of the 20 strata of the National Statistical Office by creating the numerators as the number of completed phone interviews and the denominator as the number of households identified by the National Statistical Office in each stratum.
- Pool the weights in the first two steps above.
- Derive attrition-adjusted weights using a logistic response propensity model based on household head characteristics, household characteristics, and dwelling characteristics.
- Trim weights by replacing the top 2 percent of observations with the 98th percentile cutoff point.
- Post-stratify weights to known population totals to correct for the imbalances across the urban and rural sample thereby ensuring that the survey distribution matches the distribution in the Ethiopia Socioeconomic Survey.

Content of Modules

The questionnaire was designed to be short because of the nature of a phone survey and included fixed and rotating modules. The questionnaire modules were tailored to collect individual- and household-level information on pandemic knowledge, behavior (such as sanitation and socialization), fulfillment of basic needs, employment, income, coping strategies, food security, and aid and assistance (refer to table 2.2). Typically, the respondent was the household head, but, if the head could not be reached after multiple attempts, another knowledgeable household member was selected.

TABLE 2.2 Content modules of the household survey and the survey rounds

Module	Survey round
Knowledge of actions to reduce exposure and change behavior	R1, R3, R6
Willingness to take COVID tests	R6
Willingness to be vaccinated	R6, R10
Access to essential medicines and staples	R1–R7, R11
Education	R1–R5, R8, R11
Early childhood development	R10
Access to health services	R1–R6, R8–R11
Household income dynamics and coping strategies	R1–R6
Household debt	R8
Food security	R1–R6, R11
Aid and assistance	R1–R7, R9
Employment and business	R1–R9, R11
Water and sanitation	R4, R9
Agriculture	R3–R6, R9
Locust invasion	R4, R7
Migration	R8

Source: Ambel et al. 2022.

Challenges

Low phone penetration rates in rural Ethiopia raised a concern about representativeness. A major concern was the low phone penetration rate in rural areas, where only about 40 percent of households had access to phones while access was at more than 90 percent in urban areas. There were also systematic differences between households with phones and households without phones. Households with phones exhibited higher total consumption, educational attainment, and access to improved water and sanitation. The household survey sample is thus representative only of households with phone access.

In phone surveys, the ownership of the phone that is called is crucial. Typically, phone surveys reach the household head or another senior household member, especially in households in which few members have phones. Consequently, many questions rely on proxy respondents, that is, for example, the household head might answer the labor market questions for other members. The approach is often necessary to maintain the short duration of the survey call, which is important in phone surveys. However, it is critical to consider the potential imprecision and bias this may introduce, especially relative to household members who are less likely to own a phone, such as women. Some surveys address this issue by actively calling other household members and even providing phones to household members who do not have them, but this raises ethical considerations. In an emergency, these approaches are, in any case, not practical. This means the survey must rely on respondents who are available despite the potential bias.

Updating the survey roster was a crucial part of the household phone survey in Ethiopia. While the full roster of the Ethiopia Socioeconomic Survey 2018–19 was available, it was two years old when the household phone survey was launched and required an update. There were several possible solutions. One option was to create a new roster, but this posed challenges during phone interviews as not always all household members are accounted for. Instead, the option adopted in the household phone survey was to update the existing roster by filling the survey form with the roster from 2018–19 and asking whether each of these household members was still part of the household. If a member was still part of the household, all subsequent questions were asked. If not, the survey inquired about the reason for the departure of the absent member and removed such individuals from the household roster. After reviewing the names on the roster, the interviewer also asked about any new household members that might have been missed, such as children born in the previous two years or new members added through marriage. This ensured an accurate update of the household roster and facilitated more effective data collection.

Linking the household phone survey with the Ethiopia Socioeconomic Survey enabled the comparison of statistics before and during COVID-19, providing valuable insights into the pandemic's impact. This link was particularly useful in the analysis of labor market outcomes. However, a potential bias may be introduced because of the differences in the data collection methods of the two surveys: face-to-face surveys and phone surveys may yield different results because of differences in the dynamics, particularly the visible cues possible in a face-to-face interview.

The High-Frequency Phone Survey of Refugees

Pandemics and measures taken to curb the spread of disease can have more detrimental effects on population groups already in precarious circumstances because the capacity of such groups to cope with shocks tends to be more limited. To determine more accurately the socioeconomic effects of the pandemic on a particularly vulnerable group in the population, the World Bank, the Ethiopia Refugees and Returnees Service, and the United Nations High Commissioner for Refugees (UNHCR) collaborated to integrate refugees in the household phone survey.¹⁰ The high-frequency phone survey of refugees was therefore undertaken as a booster sample of the national household phone survey to monitor the impact of COVID-19 on camp-based and out-of-camp refugees (The refugee phone survey methodology is discussed in detail by Wieser, Dampha, Beltramo, et al. 2020).

Two rounds of the refugee phone survey were conducted. Round 1 coincided with round 6 of the household phone survey, and round 2 coincided with round 7 of that survey. The survey among refugees was thus implemented after the start of the phone survey among households, as well as after the phone survey among firms. By the time the interviews among refugees commenced, the effects of COVID-19 had subsided. The refugee survey sample consisted of 1,650 refugee households per round in three locations near UNHCR suboffices. The sample included only households with mobile phone access.

The questionnaire was adapted slightly to the refugee context, but the main parts of the refugee survey were similar to the corresponding parts of the national household survey to facilitate comparability. The refugee questionnaire included material relevant to refugees and a new module on social relations and the views of refugees on Ethiopian society and the government. The modules covered knowledge about the pandemic, behavior (such as handwashing and avoiding gatherings), access to basic needs, employment and nonfarm business, income and coping strategies, access to water, sanitation, and hygiene, and access to aid and assistance.

Sampling Frame and Weights

The sampling frame of the refugee phone survey was drawn from the refugee registration database of the Ethiopia Refugees and Returnees Service and UNHCR.

¹⁰ This effort was funded through the World Bank–UNHCR Joint Data Center on Forced Displacement (refer to <https://www.jointdatacenter.org/>).

The UNHCR proGRES database includes phone numbers of registered refugees by location (UNHCR 2018). To manage and support refugees, UNHCR has six suboffices, in Addis Ababa, Assosa, Gambella, Jijiga, Melkadida, and Shire, along with settlements around Addis Ababa. The sample was drawn from lists assigned to each of the UNHCR suboffices.

The challenges, coping strategies, and types of assistance available for refugees, even from the same country of origin, vary by geographic region. In Ethiopia, the geographic division of UNHCR suboffices and phone penetration rates were used to determine the stratification that would yield the most robust and representative results. Seven possible strata were identified: Addis Ababa town, settlements around Addis Ababa, and the five suboffices outside Addis Ababa. The refugee phone survey included three of these seven survey domains to represent specific refugee groups based on phone penetration rates, as follows:

- *Addis Ababa refugees*: Out-of-camp refugees living in Addis Ababa face differing protection and humanitarian needs. They include various nationalities, but 87.0 percent are Eritreans.
- *Eritrean refugees*: Camp-based refugees are primarily served by the UNHCR sub-office in Shire, where 99.9 percent of the refugees are of Eritrean origin.
- *Somali refugees*: Camp-based refugees are primarily served by the UNHCR sub-office in Jijiga, where 97.7 percent of the refugees are of Somali origin.

To monitor the impacts of the pandemic on distinct groups, the refugee phone survey representatively sampled both camp-based refugees (Eritreans and Somalis) and refugees living under the out-of-camp policy. Out-of-camp refugees are those who have permits to live outside refugee camps or settlements. They must cover all living costs themselves or with the support of relatives, friends, or other sponsors. The survey design allowed for a comparative analysis of the pandemic's impacts on camp-based versus out-of-camp refugees. The findings revealed significant differences between these groups. For instance, total household income had declined among a quarter of the refugees. The out-of-camp refugees were particularly affected by income loss. Despite the higher share of income loss, out-of-camp refugees received the least assistance due to their refugee status because they did not qualify for aid from national authorities or UNHCR. This highlighted the unique vulnerabilities faced by out-of-camp refugees during the pandemic.

The three survey domains were used as explicit sampling strata. The sample size was determined using power calculations, considering design effects and intra-cluster correlation coefficients from the 2017 Refugee Skills Survey (Pape 2019).

The selected sample sizes were 480 households in Addis Ababa, 580 households among Eritreans in Shire, and 590 households among Somali refugees in Jijiga (refer to table 2.3). Overall, 858 camp-based refugee households were surveyed, representing 51 percent of the sample.

TABLE 2.3 Refugee camps included in the refugee survey sample

Sample stratum	Targeted sample size	Actual sample size, R1	Actual sample size, R2
Addis Ababa	480	526	484
Eritrean refugees	580	561	422
Somali refugees	590	589	523

Source: Wieser, Abebe, and Asfaw 2021.

The sample was drawn using a simple random sample method without replacement. Anticipating a high nonresponse rate based on the household phone survey experience, a stratified sample of 3,300 refugee households was selected for the first round. To obtain unbiased estimates, survey weights were constructed. Even though information was missing at some steps, the same procedure was followed that was applied in the household phone survey (Himelein 2014):

- Begin with base weights. Base weights will equal 1 for all intents and purposes.
- Derive attrition-adjusted weights for all individuals by running a logistic response propensity model based on characteristics of the household head (that is, education, labor force status, demographic characteristics), characteristics of the household (consumption, assets, financial characteristics), and characteristics of the dwelling (house ownership, overcrowding). While the UNHCR proGRES database is limited in socioeconomic variables, it includes key characteristics of the refugee household head and the refugee household.
- Trim weights by replacing the top 2 percent of observations with the 98th percentile cutoff point.
- Post-stratify weights to known population totals to correct imbalances across the sample, ensuring that the survey distribution matches the distribution in the UNHCR proGRES database.

The same replacement and nonresponse procedures were applied in the refugee phone survey and the national household phone survey. Each sampled household was called up to three times a day at different hours, with at least three hours between each call, over a minimum of three consecutive days, totaling nine attempts. The respondent was typically the household head. If the household head could not be reached despite multiple attempts, another knowledgeable household member was selected. If neither of these options yielded a positive response, the refugee household was replaced.

Challenges

Phone penetration rates are low among refugees, which affected the representativeness of the sample. Information from UNHCR, including phone penetration rates, was crucial in determining the best stratification to yield robust, representative results on refugee populations. However, phone penetration was extremely low among some refugee groups. Because of the feasibility constraints of the survey costs, only locations with a phone penetration rate higher than 30 percent were included in the sample: Addis Ababa, among Somali refugees in Jijiga, and among Eritrean refugees in Shire. To identify potential bias in the survey, the team compared socioeconomic outcomes among refugees with and without phone access. Refugees with phones were generally characterized by higher educational attainment, smaller households, and longer stays in Ethiopia. This disparity suggests that the survey may not fully capture the experiences of the most vulnerable refugees, highlighting a significant challenge in ensuring comprehensive and equitable data collection.

As with the household phone survey, the conflict that erupted in Ethiopia in November 2020 caused a significant decline in the refugee survey response rate. Among refugee respondents, Eritrean refugees were particularly affected by the outbreak of conflict because many had to flee conflict areas and move to different refugee camps after the clashes. It was not possible to contact them during the second round of the refugee survey.

The High-Frequency Phone Survey of Firms

The pandemic resulted in an unprecedented shock to the productive sectors of economies. To monitor the effects of COVID-19 on firm operations, the World Bank, in collaboration with the Job Creation Commission, implemented the

high-frequency phone survey of firms.¹¹ The data collected were based on a sample of 645 firms in industry and services using a list of registered firms provided by the Ministry of Trade and Industry. The firm phone survey revealed how businesses were affected by the pandemic and how they responded, including hiring and firing. Expectations of future operations and future labor demand were analyzed to tailor interventions and policy responses and monitor the impacts of the pandemic and of the interventions and policies more effectively.

The firm phone survey monitored responses to the COVID-19 crisis, focusing on economic activities and the effects of the pandemic on firm operations, revenues, and employment. The survey methodology is discussed by Abebe, Bundervoet, and Wieser (2020). A sample of firms in Addis Ababa was contacted every three weeks between April and October 2020, spanning eight survey rounds (refer to table 2.4).¹² The survey began with a sample of 645 firms in Addis Ababa and ended with 344 firms that had responded to all survey rounds. Survey implementation was challenging because of the large rates of nonresponse and the number of firms going out of business. In round 1, only 46 percent of the firms responded.

TABLE 2.4 Rounds and corresponding data collection periods

Round	Data collection period
1	April 15 and May 5, 2020
2	May 6 and May 27, 2020
3	May 29 and June 18, 2020
4	June 22 and July 14, 2020
5	July 23 and August 15, 2020

¹¹ Though the term firm is used for ease of understanding, the sample actually consists of establishments. An establishment is an economic unit that produces one predominant activity, typically at a single physical location. A firm, meanwhile, may consist of one or more than one establishment.

¹² The original sample design accounted for firms in other cities of Ethiopia, but those firms were ignored after the challenges observed with respondents in Addis Ababa.

Round	Data collection period
6	August 17 and September 8, 2020
7	September 13 and October 4, 2020
8	October 6 and October 26, 2020

Source: Wieser, Abebe, and Asfaw 2021.

Sampling Frame and Weights

The sampling frame was based on a registration database of firms. The list of registered establishments was obtained from the Ministry of Trade and Industry. It included 403,039 firms in Addis Ababa. The list was not frequently updated and had to be cleaned before use. Firms with missing or invalid phone numbers were removed, resulting in 389,927 firms in Addis Ababa. This cleaned list was validated by Ethio telecom, which retained only active phone numbers, creating a final sampling frame of 288,660 firms in Addis Ababa.¹³

Because of budget constraints, a panel of roughly 500 firms was selected to be contacted during eight survey rounds. The sample was stratified by establishment size (proxied by available capital) and industry classification.¹⁴ The stratification process considered two industry classifications, industry and services, and three firm size groupings, micro (below the 25th percentile in capital), small and medium (25th to 75th percentile in capital), and large (above the 75th percentile in capital). The sample thus consisted of six strata:

- Micro establishments in industry
- Micro establishments in services
- Small and medium establishments in industry
- Small and medium establishments in services
- Large establishments in industry
- Large establishments in services

¹³ At the time of survey design, Ethio telecom was the only telecommunication provider in Ethiopia, and every phone number was an Ethio telecom number.

¹⁴ The use of capital as a proxy for firm size is imperfect because capital is only weakly related to employment size in Ethiopia.

In Addis Ababa, industry firms were oversampled to ensure representativeness because the industry sector was too small, and there was a concern about representativeness if attrition proved high. The sample was drawn using simple random sampling without replacement. Anticipating a high nonresponse rate, 1,450 establishments were sampled in Addis Ababa. Additional firms were sampled to replace nonrespondents. Each firm was called at least three times over three consecutive days (nine attempts total) before replacement.

To obtain unbiased estimates, a sampling weight was applied. Because of limited information on capital, sector, and location, a weighting class adjustment was performed. Cells (region, by sector, by size) were constructed to obtain accurate counts, and weights were applied to ensure that the survey distribution matched the sampling frame distribution.

Challenges

The challenges involved in the high-frequency phone surveys among firms were severe. Initially, the survey was based on lists from the Ministry of Trade and Industry, but these lists were outdated, particularly regarding phone numbers. The team had to clean the list of registered establishments in Ethiopia by removing those with missing or invalid phone numbers, which resulted in the removal of roughly 20,000 establishments of 400,000 in Addis Ababa alone. Subsequently, all phone numbers from the cleaned list were shared with Ethio telecom, and only active phone numbers were retained. This step led to the removal of approximately 100,000 phone numbers, underscoring the challenges related to the quality of the sampling frame.

Despite these efforts, reaching firms remained challenging. Business owners are often busy and reluctant to participate in surveys, and the stressful period for businesses during COVID-19 further reduced the availability of firms for the survey. Moreover, many firms had ceased operations during this time, leading to unanswered calls. Another significant challenge was the lack of clarity on who the respondent should be, as this information was not provided in the sampling frame. Consequently, the survey interviewers had to identify the appropriate respondents, obtain their contact information, and ensure these individuals were knowledgeable and willing to participate. This process was time-consuming and required multiple callbacks to secure an appropriate respondent for each firm. The survey became expensive because of the follow-ups, and the firm conducting the survey struggled to cover the costs.

The selection of appropriate respondents is a critical step that can significantly influence the quality and reliability of the results of a survey among firms, especially surveys carried out by phone. The collection and interpretation of some statistics may be more complex in this format. Among large firms, the identity of the best possible respondent is often unclear. Should a finance officer be chosen to respond to questions on profitability and viability? Should an operations officer be selected for questions related to the impact of COVID-19 on firm operations? Since only one respondent can be selected, this choice may introduce bias. If the firm has complex operations, there is a risk of obtaining only rough estimates if there is no ready measure of revenue or turnover. Conversely, smaller and less formal firms may not track these metrics as rigorously, which could also affect the accuracy of the data collected.

Implementation Arrangements

Because of the quickly changing reality of COVID-19 and its impacts, rigor and speed needed to be prioritized in survey implementation. The World Bank team undertook the sampling, questionnaire design, and weighting. In addition, an experienced survey firm was hired to implement the phone surveys.

The pandemic imposed certain restrictions on the normal operation of the survey. For example, the existence of COVID protocols meant that the training of enumerators and interviewers had to be accomplished remotely rather than in person. Furthermore, the data collection process through phone calls was not carried out at a call center, as would normally occur, but at the homes of the enumerators and interviewers.

Several steps were taken to avoid attrition. Each sample household was called up to three times a day at varying hours, with a minimum of three hours between each call. This was done for a minimum of three consecutive days, meaning there was a total of nine attempts per household. Only then would the enumerators and interviewers be given access to another household on the respondent replacement list.

Main Findings

The main findings are presented by survey type. Box 2.1 describes how high-frequency phone surveys provided timely information on household welfare globally during the pandemic.

Household Survey

The high-frequency phone survey of households ran for 11 rounds between April 2020 and May 2021. The survey responses helped in understanding household responses to the pandemic, track the impacts on employment and income, and measure how access to education suffered during lockdowns. Special reports were prepared to analyze sex-biased outcomes and changes in poverty. This subsection generally summarizes the results presented by Ambel et al. (2022).¹⁵

Perceptions of the Pandemic and Behavioral Responses

At the onset of the pandemic, nearly all households (99.7 percent) had heard of COVID-19 and were well-informed about preventive measures. In the early days of the pandemic, the government introduced mobility restrictions and market closures to prevent the spread of the virus. To determine if people were aware of the need to change their behavior, the survey asked questions about perceptions and behavioral responses to the pandemic. Initially, 98 percent of respondents washed their hands more often; 96 percent avoided handshakes; and 83 percent avoided gatherings. However, by September (round 6), adherence had declined, and only 43 percent avoided gatherings, while 73 percent avoided handshakes (refer to figure 2.1). Concern about contracting COVID-19 also declined, from 71 percent in June (round 3) to 56 percent in September (round 6). Furthermore, the number of people who saw the pandemic as a threat to their household finances also fell, from 60 percent to 44 percent of households from June to September 2020. In September 2020 (round 6), 86 percent of respondents said they would definitely get tested for COVID-19 if the tests were free, and 98 percent were willing to be vaccinated at no cost. By March 2021 (round 10), 97.0 percent were still willing to be vaccinated, but, by December 2021, only 1.4 percent of Ethiopia's population had been fully vaccinated.¹⁶

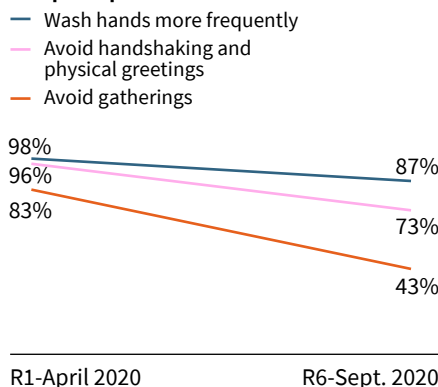
Access to Necessities

Access to basic food items was not a big challenge during the pandemic. From April to November 2020 and again in April 2021, respondents were asked whether their households were able to buy sufficient medicine and enough of the most

¹⁵ Reports, methodological documentation, and results tables related to the survey can be found in World Bank (2022a).

¹⁶ Refer to CRC (Coronavirus Resource Center) (dashboard), Center for Systems Science and Engineering, Johns Hopkins University, Baltimore, <https://coronavirus.jhu.edu/>.

FIGURE 2.1 Share of respondents who adopted preventive behavior



Source: Adapted from Ambel, Cardona-Sosa, et al. 2020.

important food items. When they were not, they were asked for the main reasons. Most households were able to buy necessity items. Teff was the scarcest.¹⁷ A third of households were unable to buy sufficient quantities. While 71 percent could buy enough medicine at baseline, this increased to 94 percent in July 2020 and stayed above 90 percent for 11 months. Reduced income and higher prices were the main reasons for the inability to afford basics. The price increases became more significant in May 2021 because of inflation.

Access to Health Services

Less than a third of households needed medical attention between April 2020 and May 2021, and, for most of them, access was not a problem. Indeed, the share of households able to access health care services rose from 86 percent in April 2020 to 96 percent in May 2021. The main barriers to access cited among survey respondents were lack of money, facility closures, and supply shortages in the facilities. Although the share of households needing medical care fluctuated, peaking at 40 percent in May 2021, it maintained a slight upward trend over time.

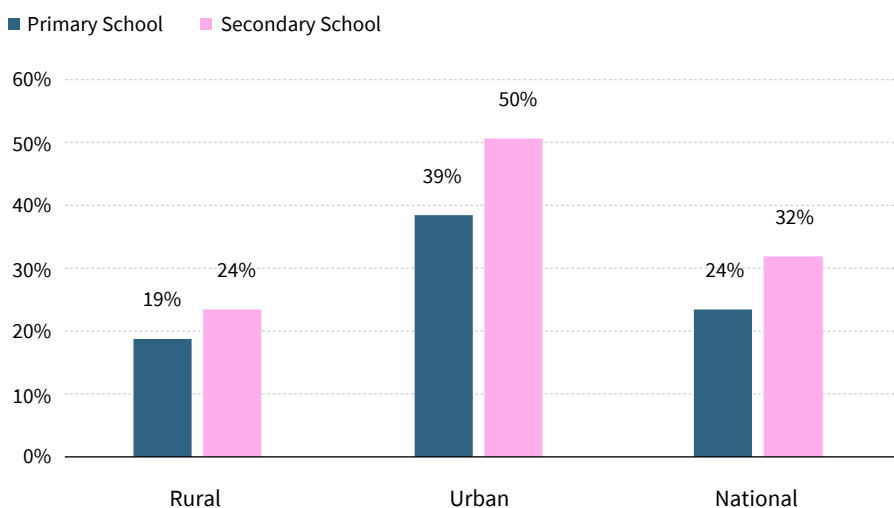
Access to Education

The survey allowed the impact of school closures on education access to be tracked among urban and rural children. The government closed all primary and secondary schools for more than seven months beginning on March 16, 2020. This affected more than 26 million students and 700,000 teachers. The policy raised concerns about learning loss among children in poorer households and permanent dropout among rural households, among which the pre-pandemic early drop-out rates were already high. Nationwide, only 24 percent of primary-school students and 32 percent of secondary-school students were engaged in distance learning activities during round 4 of the survey (July-August 2020, four months

¹⁷ Teff (*Eragrostis tef*) is an annual grass, a species of lovegrass, native to Ethiopia. It is a staple, one of the most important cereals in Ethiopia and Eritrea, cultivated for its edible seeds and also for its straw to feed cattle.

after schools had closed) (refer to figure 2.2). Deprivation in access was severe particularly in rural areas, where only 19 percent of children who had previously attended primary school and 24 percent of children who had previously attended secondary school were engaged in learning activities.

FIGURE 2.2 Children closed out of school who were engaged in learning activities, by type of education, round 4, August 2020



Source: Adapted from Wieser, Ambel, et al. 2020a.

The school closures in Ethiopia affected public and private institutions and all income segments. Nonetheless, more well-off segments—such as parents who are well educated or who are able to pay for private tutors—were more likely to be able to provide learning opportunities for their children. Children in wealthier households thus enjoyed greater exposure to distance learning activities. These differences were especially stark in urban areas during round 2 (May 14–June 3, 2020), when only 26 percent of children in the poorest quintile were engaged in distance learning, in contrast to 65 percent in the richest quintile (refer to table 2.5).¹⁸ The already wide learning gaps between children in poor versus more well-off households and between rural versus urban households likely widened during the pandemic.

¹⁸ Household income data were obtained from the original dataset used for the sampling frame (Ethiopia Socioeconomic Survey), which contains a full consumption module. Household income data are thus representative of the situation in 2019, when the survey took place.

TABLE 2.5 Engagement in distance learning among children, by location and income quantile, round 2

Location	Q1 (poorest)	Q2	Q3	Q4	Q5 (richest)
Rural	21.7	23.3	26.4	30.5	36.7
Urban	26.3	49.2	43.3	56.1	64.5
National	22.2	27.5	30.6	42.7	53.9

Source: Adapted from Wieser, Ambel, et al. 2020b.

The pandemic has had lasting effects on the many children who have permanently dropped out of school. Schools were progressively reopened. By December 2020, schools had reopened for 69 percent of school-age children, and, by May 2021, schools had reopened for 98 percent. The share of children who did not attend school even though their schools had reopened fell from 6 percent in December 2020 to 2 percent in May 2021, with no differences by residence or sex, although more girls than boys started to attend in December 2020. Despite the encouraging return rates, whether there are long-term consequences in learning remains to be seen.

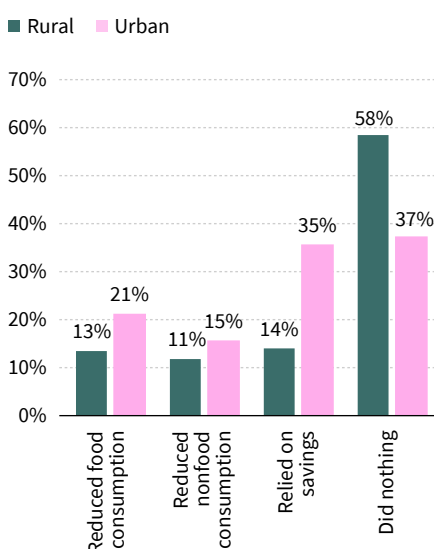
Household Income and Employment

The pandemic had detrimental effects on employment and incomes at the outset, but labor markets recovered relatively quickly, though the impacts were lasting. Although the state of emergency declaration—in effect between April and September 2020—prohibited firms from laying off workers, layoffs occurred in the sizable informal sector. At the onset of the pandemic, 80 percent of respondents reported reductions or losses in nonfarm business income, and 40 percent reported losses in farm income. Over time, the share of respondents citing loss of income declined. Households initiated various coping strategies, including selling assets, borrowing money, and reducing consumption (refer to figure 2.3). The pandemic generated job losses. By April 2020 (round 1), 8 percent of respondents had lost their jobs since the outbreak, and around 63 percent of this group indicated COVID-19 as the cause. The job losses were more severe in urban areas (20 percent) than rural areas (3 percent) (refer to figure 2.4). While employment rates had recovered by December 2020, many people had shifted to more vulnerable types of employment, such as self-employment, casual employment, and family work.

Non-farm family businesses suffered as a result of the pandemic, especially in urban areas. Before the pandemic, about 25 percent of households owned non-farm family businesses. By April 2021, ownership had dropped by 3 percentage points. This trend was particularly dramatic among urban households, among which ownership fell from 38 percent pre-pandemic to 31 percent in April 2020 and to 25 percent in April 2021.

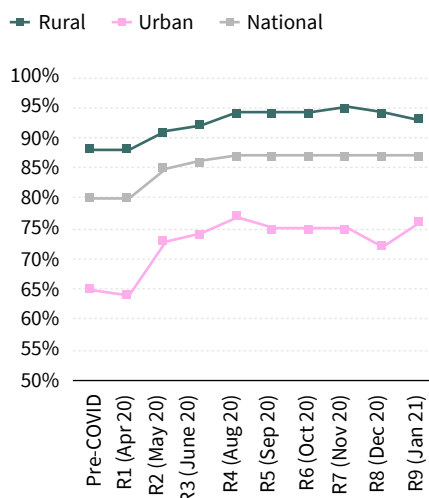
Social assistance by the government and other actors played a minor role in Ethiopia during COVID-19. Despite significant income losses, only 8 percent of households reported receiving assistance in April 2020. With time, the government's role in providing assistance decreased as non-governmental organizations (NGOs) became more prominent.

FIGURE 2.3 Coping strategies, by location, average, April–October 2020



Source: Adapted from Ambel, Cardona-Sosa, et al. 2020; Ambel et al. 2021.

FIGURE 2.4 Respondents who worked in the seven days preceding the interview, by location and round



Source: Ambel et al. 2022.

Rural Economy

The pandemic affected urban areas significantly more than rural areas. About 80 percent of Ethiopians reside in rural areas. Rural Ethiopia proved resilient, with a quick recovery in activities and income despite initial losses associated

with mobility restrictions. Among rural households, the largest adverse impact occurred in the early months of the pandemic. For example, by April 2020, half of rural households had experienced income losses. Nonetheless, assistance was sparse. Only 10 percent of households reported they had received assistance from any source during the pandemic.¹⁹

Poverty

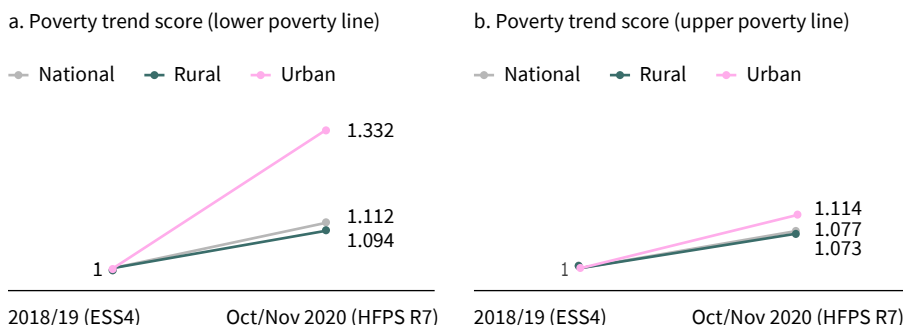
Estimates indicate that income and employment losses during the pandemic led to significant increases in poverty in urban areas. While phone surveys offered the crucial ability to track outcomes even during lockdowns, they needed to be brief and could not include the comprehensive modules typically used to estimate household welfare through income or consumption data. To address this issue, the methodology of the Survey of Well-Being via Instant and Frequent Tracking (SWIFT) was used in the high-frequency phone survey among households, which permitted the estimation of poverty rate projections from phone data (Wieser et al. 2022).²⁰ Nationally, the share of people living below the lower poverty line rose by 11.2 percent between the Ethiopia Socioeconomic Survey 2018–19 (wave 4) and the household phone survey in October–November 2020 (round 7) (refer to figure 2.5).²¹ The pandemic’s impact was particularly severe in urban areas, where poverty rates surged by 33.2 percent compared with 9.4 percent in rural areas. Conversely, inequality showed the most substantial rise in rural areas, reversing the trend in 2018–19, when urban areas were more unequal. In urban areas, the Gini coefficient increased slightly, from 0.375 to 0.383, while, in rural areas, it rose more sharply, from 0.366 to 0.393.

¹⁹ The Productive Safety Net Program was the main source of government support, especially in the early months.

²⁰ SWIFT is a survey-to-survey–based methodology that applies machine learning and multiple imputation techniques to estimate household expenditure and income by collecting data on poverty correlates through simple questions (Yoshida et al. 2015).

²¹ Because the Ethiopia Socioeconomic Survey is not used to estimate official poverty statistics, it does not refer to poverty lines. Instead, the poverty line was set here at the 23.5th percentile of the Ethiopia Socioeconomic Survey data because the then most recent official poverty rate in Ethiopia was 23.5 percent (in 2016). The upper poverty line was set at the 40th percentile to align with the World Bank’s goal of shared prosperity, which tracks the income growth of the poorest 40 percent of the population.

FIGURE 2.5 Poverty trends from the Ethiopia Socioeconomic Survey 2018–19 to the high-frequency phone survey, round 7



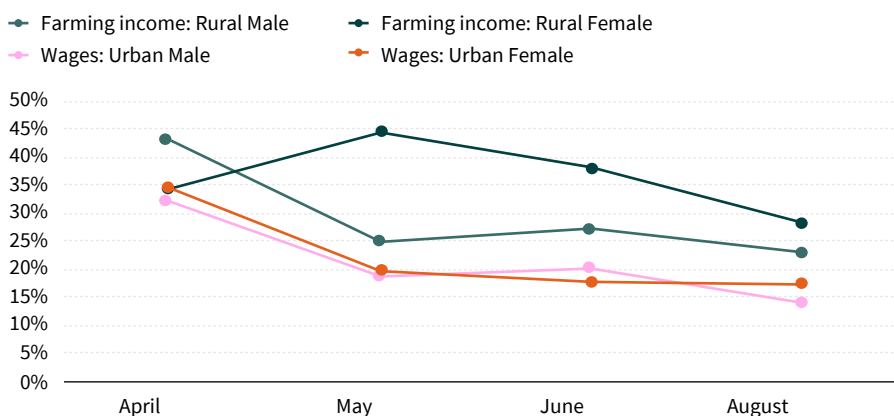
Source: Wieser et al. 2022.

Note: The figures display poverty rates relative to the baseline value (set to 1). They should be interpreted as proportional changes, where values above or below 1 indicate an increase or decrease in poverty relative to the baseline.

Gender

Female-headed households in rural areas were particularly affected during the early months of the pandemic. Ebrahim et al. (2020) explore the gendered impacts of COVID-19, focusing on outcomes from rounds 1–4 (April to August 2020). The results revealed that, four months into the pandemic, women, especially women who headed households, faced unique challenges, while remaining committed to COVID-19 preventive measures. The gender disparities were especially pronounced in rural areas. For instance, while women generally adhered more strictly to COVID-19 preventive measures and expressed greater concern about the disease than men, rural women reported they had less access than men to medical treatment. Rural women also experienced the highest income losses, surpassing both urban women and men overall (refer to figure 2.6). While urban female- and male-headed households reported similar levels of concern about food security, rural female-headed households expressed significantly more worry than male-headed ones. Employment rates were lower among women, but it is difficult to disentangle the effects of the pandemic from preexisting structural issues in the job market. Nonetheless, as primary caregivers, women likely faced additional challenges in maintaining stable employment during the pandemic.

FIGURE 2.6 Respondents reporting diminished farm and wage income, by sex and location, April–August 2020



Source: Ebrahim et al. 2020.

Box 2.1 COVID-19 Monitoring Survey Harmonization

The COVID-19 pandemic had a global impact, leading to an increasing need for timely information on livelihoods and health. However, governments faced challenges in their ability to collect the information needed because of contagion risks. To address this, the World Bank launched high-frequency phone surveys in April 2020 to monitor the welfare impacts of COVID-19. Data collection continued into 2022 in some countries. By the end of the data collection initiative, nearly 500 rounds of phone survey data had been collected across 93 countries located in all regions, representing more than 4.5 billion people. More than half were low- or lower-middle-income countries, and a quarter were in fragile and conflict-affected settings. In 14 countries, the high-frequency phone surveys included samples of forcibly displaced populations, such as refugees, internally displaced persons, and Venezuelans displaced abroad.

The high-frequency phone surveys provided one of only a few sources of timely and broadly representative household-level data during the pandemic, particularly in low- and middle-income countries. Core survey modules usually covered employment, food security, education, health services, safety nets, and coping mechanisms and were often tailored to the evolving needs and priorities of individual countries. The overall survey design was maintained such that many indicators could be harmonized ex post to enable cross-country comparisons. This vast harmonization effort resulted in the COVID-19 Household Monitoring Dashboard.^a

The data were invaluable in tracking the welfare impact as the pandemic unfolded, tracing out trends in employment and income across and within countries as restrictions on economic activity were relaxed, as documented in a number of country, regional, and global reports. They also included ground-level evidence that, although many governments offered assistance to cushion the impact of the shock, the support was largely inadequate to meet the scale of the crisis. Analysis has also warned of potential long-term scarring effects on human and physical capital. Thus, vulnerable households were more likely to resort to negative coping strategies, such as the sale of assets or reduced food consumption, or to experience learning losses. The resulting erosion of human and physical capital could lead to heightened inequality in the longer term.

Source: Adapted based on Brunckhorst, Kim, and Cojocar 2023.

a. COVID-19 Household Monitoring Dashboard, World Bank, Washington, DC, <https://www.worldbank.org/en/data/interactive/2020/11/11/covid-19-high-frequency-monitoring-dashboard>.

Refugee Survey

Two rounds of data collection among refugees were completed: The first round of the joint national and refugee high-frequency phone surveys was implemented between September 24 and October 17, 2020, and the second round between October 20 and November 20, 2020. Results were reported in two different briefs (Wieser, Dampha, Ambel, et al. 2020; Wieser et al. 2021).²² Box 2.2 describes the global effort of the World Bank to collect high-frequency phone surveys among refugees.

Access to Necessities

Because of the assistance they received in camps, refugees were able to obtain basic food items to a greater extent than the rest of the population. Thus, 96 percent of refugees interviewed in late September and early October 2020 reported that they were able to buy oil, while only 77 percent of non-refugee population did so. This was not true of medicines. A slightly higher share of non-refugee households (95 percent) than refugee households (89 percent) reported they could purchase medicines.

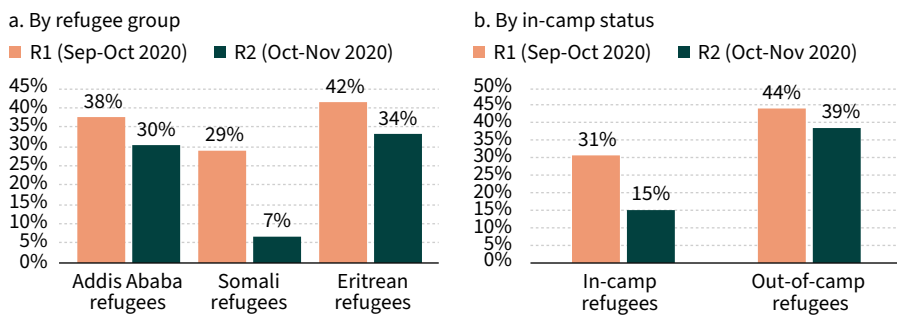
²² Reports, methodological documentation, and result tables related to the survey can be found in World Bank (2022a).

Education

The low school attendance among refugee children before the pandemic was exacerbated by COVID-19. When the schools were closed, there was a concern about the impact on learning and drop-out rates. Schools provide additional services to refugees by aiding them in achieving the transition to normalcy and protecting them from exposure to risks, such as gender-based violence, pregnancy, or early marriage. Among the total population of primary- and secondary-school-age refugees, only 20 percent and 5 percent of children were attending primary and secondary schools, respectively, before the pandemic. In contrast, among non-refugee households, almost 70 percent of primary-school-age children and 20 percent of secondary-school-age children attended school. Not all refugee groups show low attendance rates. The results of round 2 of the survey indicate that, while 65 percent of children in Somali refugee households attended primary school, this was true of only 18 percent and 16 percent of refugees in Addis Ababa and Eritrean refugees.

School closures rendered access to education access more challenging during the pandemic. Among refugee children who attended school before the closures, only 38 percent of primary-school-age children and 51 percent of households with secondary-school-age children were participating in distance learning activities when the first round of the survey was implemented after schools had closed. Four weeks later, during the second survey round, engagement in distance learning among children whose schools had not yet opened had dropped by another 13 percentage points among children of primary-school age and 30 percentage points among children of secondary-school age. Participation in learning activities was not only lower among in-camp refugees, but the participation of this group also exhibited a larger drop relative to out-of-camp refugees between survey rounds (refer to figure 2.7).

FIGURE 2.7 Primary-school-age children in school before COVID who participated in learning activities during school closures

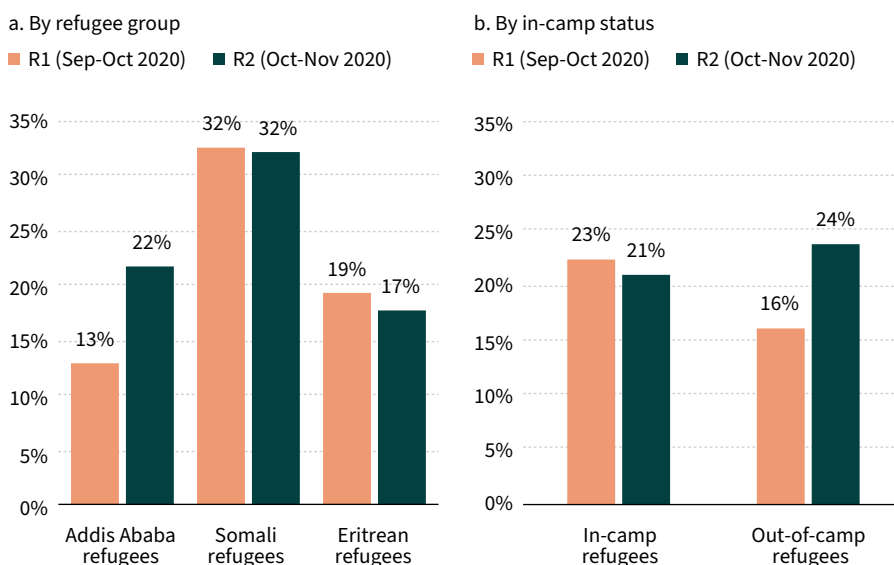


Source: Wieser, Dampha, Ambel, et al. 2020; Wieser et al. 2021.

Household Income and Employment

Employment rates were significantly lower among refugee respondents than among nonrefugees. Before COVID-19, 28 percent of refugee respondents had jobs. At the time of round 1 and round 2, employment among refugee respondents had dropped by 10 percentage points and were below 22 percent overall. Not all groups fared equally. Employment rates were highest among Somali refugees (32 percent), followed by refugees in Addis Ababa (22 percent), and Eritrean refugees (17 percent) (refer to figure 2.8). Differences between in-camp and out-of-camp refugees were not as stark.

FIGURE 2.8 Share of respondents who reported they were currently working



Source: Wieser, Dampha, Ambel, et al. 2020; Wieser et al. 2021.

Income sources among refugees before round 1 were diverse and depended largely on whether the refugees were living in or outside of camps. Among refugees living in Addis Ababa and Eritrean refugees (among whom about 50 percent were living outside camps), wage employment was the main source of livelihood. A large share of refugees relied on remittances as the main source of income. Among Somali refugees, almost all of whom are based in camps, assistance from the government, the international community, or NGOs was an important source of income.

Many refugees suffered income losses because of the effects of the pandemic on nonfarm businesses. Around 57 percent of refugee households that cited nonfarm business as a livelihood reported less income from that source, and 55 percent of households with income from wages had lost some or all of the income from that source within six months of the outbreak of the pandemic. A third of the refugee households that indicated remittances from abroad as a source of income (almost half of all refugees), reported a decline in the remittances. Indeed, 27 percent of the refugee respondents and 45 percent of the refugee respondents located in Addis Ababa reported a reduction in income since the onset of the pandemic. It seems that, among the non-refugee population, household income losses were more prevalent at the onset of the pandemic, but these households seem to be farther along in recovery compared with refugees. Non-refugees were more than twice as likely to report an increase in total household income in September 2020 (11 percent) compared with refugees (5 percent).

Assistance

Because of the assistance provided in the camps, refugees fared well during the pandemic. Around 30 percent of refugees received assistance from the government, NGOs, or faith-based institutions between March and September 2020. Somali refugees, all of whom live in camps, were more than twice as likely to receive assistance relative to Eritrean refugees (among whom only around half live in camps). The refugees in Addis Ababa received the least assistance because of their out-of-camp status. The most important types of assistance were free food and direct cash transfers.

Box 2.2 Including Forcibly Displaced People in High-Frequency Phone Surveys during the Pandemic

As the COVID-19 pandemic and mitigation measures began to affect local and national economies, concerns arose about marginalized groups, particularly forcibly displaced populations, who might be especially vulnerable to isolation from social safety nets. In Spring 2020, the World Bank–UNHCR Joint Data Center on Forced Displacement collaborated with UNHCR and World Bank country teams to expand high-frequency phone surveys, initially conducted among nondisplaced populations, to include internally displaced persons and refugees.

Between 2020 and 2021, the Joint Data Center supported multiround surveys among displaced people in eight countries using a similar methodology. Through the efforts of the center, the World Bank, and UNHCR, high-frequency phone survey data among displaced people were also collected in 14 countries during this period, including Bangladesh, Burkina Faso, Chad, the Democratic Republic of Congo, Costa Rica, Djibouti, Ecuador, Ethiopia, Iraq, Jordan, Kenya, Mexico, Somalia, Uganda, and Yemen. All survey instruments and anonymized data were made publicly available.

In addition to country-level briefs, aggregated results from subsets of these surveys were published in two reports covering each of the first two years of the global pandemic (Tanner et al. 2021; World Bank 2023). These reports revealed that displaced populations were almost always more likely than nondisplaced populations to experience welfare shocks. This was partly because of labor market challenges. Displaced populations were more likely to be employed in sectors vulnerable to shocks, leading to higher job losses. They were also less able to relocate geographically to find new employment. Displaced children were less likely to visit health clinics if ill, a significant public health concern during the pandemic, although displaced people were more likely than the nondisplaced to engage in COVID-19 protective measures. Displaced children were also more likely to drop out of school and face delays in reenrollment or not reenroll at all. Displaced populations often faced extreme shocks reflected in basic welfare measures, such as food security.

Firm Survey

This section summarizes the results of the high-frequency phone survey among firms presented in Wieser, Abebe, and Asfaw (2021).²³ The insights from these surveys were crucial for tailoring interventions, implementing effective policies, and monitoring the outcomes of the interventions and policies.

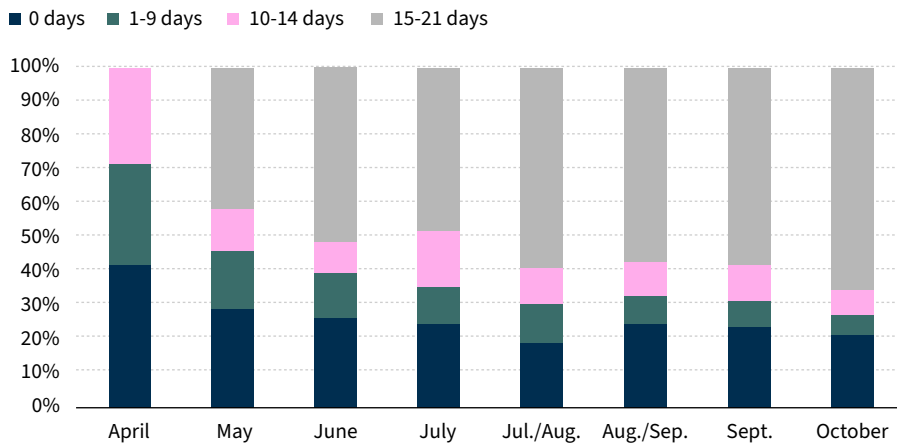
Overall Impacts

The pandemic affected more than 90 percent of firms in Addis Ababa. In April 2020, at the time of the survey round 1, more than 42 percent of firms were closed, and no firms were fully operational, primarily because of government-imposed movement and business activity restrictions under the state of emergency

²³ Reports, methodological documentation, and results tables related to the survey can be found in World Bank (2022a).

(refer to figure 2.9).²⁴ By September, after the state of emergency had been lifted, three-quarters of firms were fully operational. Thus, most COVID-19-induced closures proved temporary.

FIGURE 2.9 Firms, by number of days in operation during the previous 21 days in each survey round



Source: Wieser, Abebe, and Asfaw 2021.

Loss of Revenue

Firms faced significant financial stress and were struggling to pay rent, invoices, staff wages, and social security. Early in the pandemic, sales revenue dropped sharply. Many firms were operating at zero revenue, particularly own-account firms, microfirms, and firms in the industry sector.²⁵ In April 2020, firms were earning only 43 percent of the average monthly revenue earned during the previous year, which had declined to 12 percent by round 4 of the survey (June–July 2020). Revenues gradually recovered as restrictions eased, but the performance varied across sectors. The service sector experienced increased revenues at full operation, but the revenues in industry remained low.

²⁴ The government declared a five-month state of emergency in April 2020 to curb COVID-19, imposing restrictions on gatherings, alcohol sales, restaurant hours, transportation services, and sporting activities.

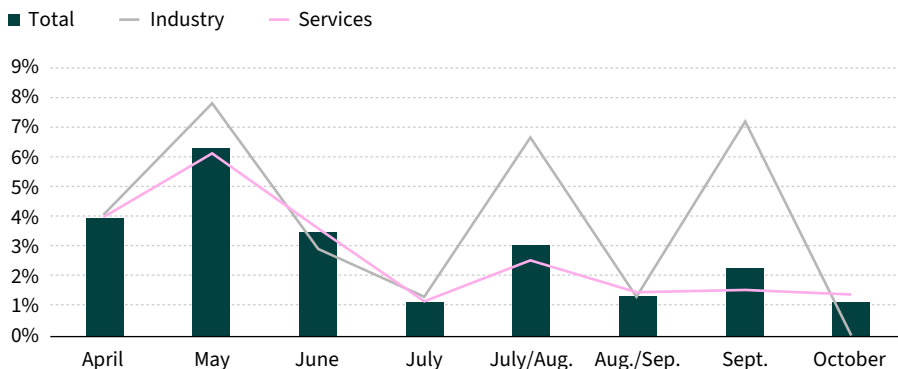
²⁵ Three firm size groupings—micro firms (below the 25th percentile in capital), small and medium establishments (25th–75th percentile in capital), and large establishments (above the 75th percentile in capital)—were used in the stratification process. Own-account firms are defined as firms in which only the owner is working in the firm, and there are no payroll employees.

A collapse in consumer demand represented the main impact of the pandemic on firm operations. About 62 percent of firms were reporting that the lack of consumer demand was the most significant shock in April 2020. By October 2020 (round 8), nearly 90 percent of firms still identified the demand shock as the primary issue. Disruptions to supply chains also became significant. In April 2020, 8 percent of firms reported concerns about supply chains. The share had risen to 26 percent by the final round.

Impact on Employees: Layoffs and Hiring Behavior

In the early phase of the pandemic, firms responded by granting paid or unpaid leave, laying off workers, and adjusting wages. Relatively few firms reduced the working hours of one employee or more (12 percent), cut wages (8 percent), or granted leaves of absence with or without pay (5 percent and 11 percent, respectively). About 6 percent of firms laid off employees between survey rounds 1 and 2 (refer to figure 2.10). The limited number of laid-off employees may have been linked to the state of emergency, which ran from April to September 2020 and prohibited firms from laying off employees. When layoffs occurred, though, they affected temporary workers more frequently.

FIGURE 2.10 Firms implementing layoffs since the previous round



Source: Wieser, Abebe, and Asfaw 2021.

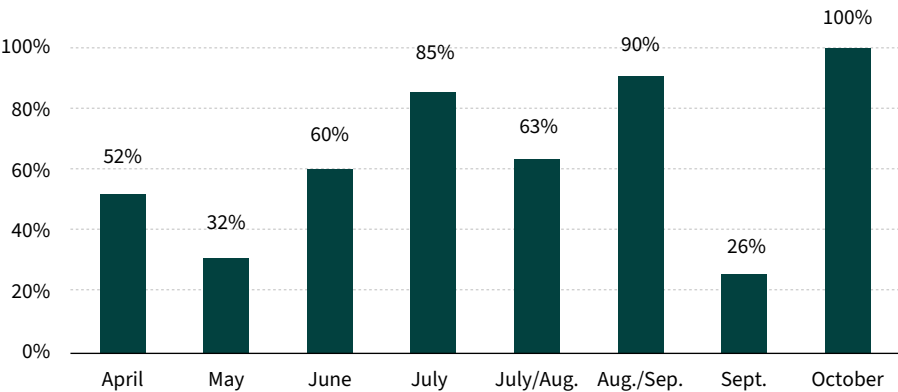
Over time, firms shifted from granting leaves and implementing layoffs to reducing salaries. Eight months into the pandemic, salaries had fallen by more than 15 percent, on average, relative to the same month of the previous year. While firms recovered relatively quickly in hiring behavior, particularly after survey round 4, hiring expectations remained subdued. Only 2 percent of firms reported in round 8 that they planned to hire, down from 4 percent in round 1. However, firms in

Addis Ababa did not expect additional COVID-19 challenges in the medium term, and layoff expectations declined significantly between rounds 1 and 8.

Gender-Biased Impacts

The pandemic disproportionately affected woman-owned businesses and women employees, exacerbating existing inequalities. Women were disproportionately affected by the increased demand for child and family care because of the confinement measures, and woman-owned firms were more likely to close. In April 2020, 35 percent of woman-owned businesses (compared with 23 percent of man-owned businesses) reported forced closures. While layoffs were limited, when the pandemic broke out in April 2020, about 70 percent of the workers laid off were women, although women only accounted for 42 percent of the workforce (refer to figure 2.11). Layoffs among women rose throughout the pandemic.²⁶

FIGURE 2.11 Laid-off employees who were women



Source: Wieser, Abebe, and Asfaw 2021.

Government Response

The government implemented various support measures, including covering the costs of operations, reducing utility costs, deferring payroll taxes, providing wage subsidies, and offering zero-interest loans. However, by October 2020, only 9 percent of firms had received any government support. The policies were more

²⁶ Across the three survey rounds, a total of 244 workers in 39 firms were laid off. The sample size is thus small and represents an important caveat.

relevant to larger firms, while most Ethiopian firms are own-account or micro-firms. Firms indicated that waiving tax payments, covering operational costs, freezing loan repayments, and extending loan terms or partial debt relief were the most relevant support measures.

Lessons Learned

Experience with high-frequency phone monitoring surveys in response to an emergency, such as a pandemic, was limited in Sub-Saharan Africa. There have been two notable efforts to collect high-frequency phone monitoring data in Africa: (1) Listening to Africa and (2) Ebola monitoring in Sierra Leone. Listening to Africa is a collaborative initiative involving national statistical institutes and NGOs across Sub-Saharan Africa in piloting the regular collection of information on living conditions using mobile phones (Hoogeveen and Etang Ndip 2017). The approach integrates face-to-face surveys with follow-up mobile phone interviews, enabling welfare monitoring. The Ebola monitoring project in Sierra Leone was conducted by the government of Sierra Leone with support from the World Bank and in partnership with Innovations for Poverty Action. The objective of the mobile phone interviews was to monitor and measure the key socioeconomic effects of the Ebola virus (Etang Ndip and Himelein 2020).

Experience has otherwise been limited in carrying out phone surveys in the context of an emergency, such as the global pandemic. Because of this and the urgent need to implement a sound high-frequency monitoring system quickly, Ethiopia has been at the forefront in the implementation of the survey series. Within only a month of the outbreak of the pandemic in Ethiopia, the survey enumerators and interviewers were in the field. The rapid implementation of the survey meant that no experience was available on appropriate COVID-related questions, and the team had to apply novel methods. It also ensured that high-quality data were available quickly, providing important insights to policy makers. This could only have been achieved through an experienced and dedicated team—the World Bank, the government, and the survey firm—that put in a significant amount of time and effort.

Two key factors in the success of the phone surveys in getting information to policy makers quickly were (1) the streamlined data collection and analysis workflow and (2) a straightforward internal approval process. The team-defined data collection and analysis workflow involved clear responsibilities and an established, closely monitored timeline. The identification and use of standardized data protocols and

programming files ensured easy data processing and analysis at every round. This allowed the time between the end of each round of data collection and the final analysis to be short. An abbreviated clearance process before the first survey round was implemented to ensure that the results could be reported and made publicly available quickly. All of this ensured that the briefs and data tables were ready within three or four weeks of the end of the survey round data collection.

There were nonetheless significant challenges involved in using phone survey findings for actual policy making in Ethiopia. Although the COVID-19 pandemic erupted in Ethiopia in March 2020 and the high-frequency phone survey was launched promptly in April 2020, providing quick results to policy makers, the take-up of the findings in policy making was limited. Despite the Jobs Creation Commission's efforts to design policies supporting private businesses and providing cash assistance to informal workers through the Urban Productive Safety Net Project, these initiatives were never fully realized. This was not because of issues with communication or the quality of the findings, but rather the challenging environment for policy making marked by constant shocks and competing demands. In November 2020, conflict broke out in northern Ethiopia, causing devastation in lives and livelihoods. Consequently, the government's attention shifted to addressing the conflict and preventing the spread of violence, rather than mitigating the impact of COVID-19. This left little space for policy makers to focus on introducing support measures for households and firms based on the phone survey findings.

From the technical side, the biggest challenge to phone survey monitoring systems in settings such as Ethiopia is gaining access to accurate information on active phone numbers. The high-frequency phone survey in Ethiopia could only be carried out because the Ethiopia Socioeconomic Survey had been implemented a short time previously. This ensured the team had access to updated lists of phone numbers that could be called. The experience of COVID-19 showed the need to invest in high-quality, updated, representative phone number databases that can be utilized in such initiatives.

If high-frequency monitoring is to be successful, it must be implemented quickly, but the analysis and assessments must be transparent. In the case of Ethiopia, the team opted to put all outputs—survey analysis, sampling documentation, questionnaires, data tables—on a stand-alone survey website to facilitate access to the information by anyone. In addition, access to anonymized microdata was provided. This helped to publicize the results and to ensure that everyone had easy access to the relevant information.

The results serve to emphasize the value of the ready availability of a high-frequency monitoring system to address unexpected events. For example, the phone survey among firms was conducted during a period of protests in the country. The protests had far-reaching consequences, including significant violence and a two-week internet blackout. Setting up the phone survey structure rapidly allowed the team to assess the impact of this event on firms immediately. About 49 percent of the firms in Addis Ababa were affected by the instability, and 47 percent were directly affected by the unrest. The internet shutdown exerted an impact on 6 percent of firms. Among firms that reported that they had been affected by the riots, around 71 percent reported that the unrest had an impact on sales; 29 percent experienced notable supply chain disruptions; and about 9 percent suffered direct physical damage to property. These results could not have been reported without the phone survey system.

High-frequency phone surveys can also be useful in noncrisis situations, in monitoring welfare, and in gauging perceptions and responses to policy decisions. By continuously monitoring a population's well-being and the public's reaction to policy changes, governments can make more well-informed decisions and adjust strategies to meet the needs of the people. In Ethiopia, there was an initiative to monitor the responses to and impact of the government's homegrown economic reform agenda on the population using high-frequency phone surveys through the Urban High-Frequency Survey. This initiative aimed to provide timely insights into how the reforms were affecting household welfare and economic conditions. However, despite the potential value of this data, the initiative received limited attention from policymakers and was discontinued after just four rounds. This discontinuation highlights the challenges of sustaining such monitoring efforts and ensuring that the collected data is effectively utilized in the policy-making process.

Conclusion

The experience with high-frequency phone surveys in Ethiopia highlights key lessons for gathering rapid, policy-relevant data in challenging contexts. There may be risks, but flexibility and adaptability are essential for success. Striking a balance between speed and rigor ensures that data are both timely and reliable. Early planning on team composition, workflow bottlenecks, and management clearance decisions can help streamline operations and prevent delays.

Engagement with governments is crucial to impact. While this can sometimes create barriers to rapid implementation, the buy-in of policy makers is necessary for

the take-up of findings and the sustainability of monitoring systems. Investing in a well-maintained and accessible database of phone numbers enhances efficiency and sustainability. Transparency in data collection, from sampling methods to openly sharing findings through accessible platforms such as websites, strengthens credibility and facilitates the broader use of the insights.

Ultimately, the high-frequency phone surveys require significant effort and commitment. Their usefulness depends on dedicated investment in design, execution, and continuous improvement. Beyond data collection, the real value lies in ensuring that the evidence generated informs policy decisions and operational responses, contributing to more beneficial and responsive development interventions.

Finally, while high-frequency phone surveys have expanded rapidly in recent years, several important areas still require further research and testing.

This document notes potential bias arising from who owns or answers the phone. Evidence from other World Bank phone survey programs highlights that this bias can be particularly consequential for certain types of indicators. For example, labor market outcomes are especially sensitive to respondent selection, as shown in studies where surveys targeting household heads (often through phone lists) reported higher employment rates than those using random digit dialing (RDD), which reached a broader mix of household members (Brubaker et al., 2021; Gourlay et al., 2021). In contrast, when comparing differences between population groups within the same country—such as rural versus urban households or gender gaps—relative patterns tend to remain consistent across phone and face-to-face modalities, even if levels differ (Ambel et al., 2021). This suggests that while phone ownership and respondent bias can affect the accuracy of absolute estimates, phone surveys may still provide valuable insights on disparities between groups. Nonetheless, more detailed analysis is needed to guide appropriate use and interpretation of phone survey data across different sectors and contexts.

Another key challenge is minimizing attrition and maintaining respondent engagement across repeated rounds, particularly in low-resource settings. The effectiveness of different design choices—such as rotating panels, shorter instruments, or incentive schemes—remains underexplored.

There are also persistent questions about the comparability of phone-based responses to those collected in face-to-face interviews, especially for complex or sensitive topics such as consumption, food security, and psychosocial well-being.

In addition, strategies to reach populations without reliable phone access—such as shared phones, mobile kiosks, or hybrid models—need further piloting.

Beyond these operational concerns, there are unresolved methodological questions that warrant deeper investigation. These include the extent to which post-stratification weights can adequately adjust for selection biases in phone survey samples, and how best to design sampling frames in contexts where phone ownership is uneven. Importantly, these questions are not limited to the context of the COVID-19 pandemic; answering them could contribute to the broader development of reliable, cost-effective remote data systems for future use. Continued research in these areas is essential to improve the credibility, inclusiveness, and long-term utility of high-frequency phone surveys.

References

- Abebe, Girum, Tom Bundervoet, and Christina Wieser. 2020. "Sampling Design." *Monitoring COVID-19 Impacts on Firms in Ethiopia*, Report 2 (May 15), World Bank, Washington, DC.
- Ambel, A. A., et al. 2021. *Lessons from High-Frequency Phone Surveys on COVID-19: The World Bank's Experience in Sub-Saharan Africa*. World Bank.
- Ambel, Alemayehu Azeze, Lina Marcela Cardona Sosa, Asmelash Haile Tsegay, and Christina Wieser. 2022. "Results from 11 Rounds of High-Frequency Phone Surveys of Households from April 2020 through May 2021." *Monitoring COVID-19 Impacts on Households in Ethiopia*, Report 10 (January 18), World Bank, Washington, DC.
- Ambel, Alemayehu Azeze, Lina Marcela Cardona-Sosa, Asmelash Haile Tsegay, and Christina Wieser. 2020. "Results from Six Rounds of High-Frequency Household Phone Surveys." *Monitoring COVID-19 Impacts on Households in Ethiopia*, Report 7 (December 21), World Bank, Washington, DC.
- Ambel, Alemayehu Azeze, Lina Marcela Cardona-Sosa, Asmelash Haile Tsegay, and Christina Wieser. 2021. "Ethiopia High Frequency Phone Monitoring Household Survey Round 1 to Round 9: Tables." *Monitoring COVID-19 Impacts on Households in Ethiopia*, World Bank, Washington, DC. <https://documents1.worldbank.org/curated/en/256341617317644165/text/List-of-Tables.txt>.
- Ambel, Alemayehu Azeze, Tom Bundervoet, Asmelash Haile Tsegay, and Christina Wieser. 2020. "Monitoring COVID-19 Impacts on Households in Ethiopia." *Survey Methodology Document* (May 28), World Bank, Washington, DC.
- Betthäuser, Bastian A., Anders M. Bach-Mortensen, and Per Engzell. 2023. "A Systematic Review and Meta-Analysis of the Evidence on Learning during the COVID-19 Pandemic." *Nature Human Behaviour* 7 (3): 375–385.
- Brubaker, J., et al. 2021. *Measuring Employment in a Pandemic: Do Phone Surveys Understate Employment Rates?* Policy Research Working Paper No. 9705, World Bank.

- Brunckhorst, Ben James, Yeon Soo Kim, and Alexandru Cojocaru. 2023. “Tracing Pandemic Impacts in the Absence of Regular Survey Data: What Have We Learned from the World Bank’s High-Frequency Phone Surveys?” *Policy Research Working Paper* 10585, World Bank, Washington, DC.
- Dai, Ruochen, Hao Feng, Junpeng Hu, Quan Jin, Huiwen Li, Ranran Wang, Ruixin Wang, Lihe Xu, and Xiaobo Zhang. 2021. “The Impact of COVID-19 on Small and Medium-Sized Enterprises (SMEs): Evidence from Two-Wave Phone Surveys in China.” *China Economic Review* 67 (June): 101607.
- Ebrahim, Menaal Fatima, Alemayehu Azeze Ambel, Niklas Buehren, Tom Bundervoet, Adiam Hagos Hailemicheal, Girum Abebe Tefera, and Christina Wieser. 2020. “Gendered Impacts of the COVID-19 Pandemic in Ethiopia: Results from a High-Frequency Phone Survey of Households.” *Monitoring COVID-19 Impacts on Households in Ethiopia*, Report 5 (October 12), World Bank, Washington, DC.
- Etang Ndip, Alvin, and Kristen Himelein. 2020. “Monitoring the Ebola Crisis Using Mobile Phone Surveys.” In *Data Collection in Fragile States: Innovations from Africa and Beyond*, edited by Johannes G. Hoogeveen and Utz Johann Pape, 15–31. Washington, DC: World Bank; Cham, Switzerland: Palgrave Macmillan.
- Gourlay, S., et al. 2021. *High-Frequency Phone Surveys on COVID-19: Good Practices and Lessons Learned*. World Bank.
- Himelein, K., S. Eckman, J. G. Kastelic, K. R. McGee, M. Wild, N. Yoshida, and J. G. Hoogeveen. 2020. *High Frequency Mobile Phone Surveys of Households to Assess the Impacts of COVID-19 (Vol. 2): Guidelines on Sampling Design*. Washington, DC: World Bank Group.
- Himelein, Kristen. 2014. “Weight Calculations for Panel Surveys with Subsampling and Split-off Tracking.” *Statistics and Public Policy* 1 (1): 40–45.
- Hoogeveen, Johannes G., and Alvin Etang Ndip. 2017. “Let’s Invest in Mobile Phone Surveys to Monitor Crises.” *Africa Can End Poverty* (blog), June 27. <https://blogs.worldbank.org/en/africacan/lets-invest-in-mobile-phone-surveys-to-monitor-crises>.
- Ohnsorge, Franziska L., and Shu Yu, eds. 2022. *The Long Shadow of Informality: Challenges and Policies*. Washington, DC: World Bank.

- Pape, Utz Johann. 2019. "Ethiopia, Skills Profile Survey 2017: A Refugee and Host Community Survey." April 18, World Bank, Washington, DC.
- Sánchez-Martín, Miguel Eduardo, Samuel Mulugeta, Zerihun Getachew Kelbore, and Christina Wieser. 2021. "Ethiopia Economic Update: Ensuring Resilient Recovery from COVID-19." March, World Bank, Washington, DC.
- Shafi, Mohsin, Junrong Liu, and Wenju Ren. 2020. "Impact of COVID-19 Pandemic on Micro, Small, and Medium-Sized Enterprises Operating in Pakistan." *Research in Globalization* 2 (December): 100018.
- Tanner, Jeffery, Harriet Kasidi Mugera, Domenico Tabasso, Maja Lazić, and Björn Gillsäter. 2021. "Answering the Call: Forcibly Displaced during the Pandemic." *JDC Paper Series on Forced Displacement* 2 (August 5), World Bank–UNHCR Joint Data Center on Forced Displacement, Copenhagen.
- UNHCR (United Nations High Commissioner for Refugees). 2018. "From proGres to PRIMES." March, UNHCR, Geneva. <https://www.unhcr.org/blogs/wp-content/uploads/sites/48/2018/03/2018-03-16-PRIMES-Flyer.pdf>.
- Wieser, Christina, Girum Abebe, and Adamsu Asfaw. 2021. "How Have Firms Fared in Times of COVID-19 in Addis Ababa? Evidence from Eight Rounds of High-Frequency Phone Surveys." World Bank, Washington, DC. <https://doi.org/10.1596/36664>.
- Wieser, Christina, Alemayehu Azeze Ambel, Tom Bundervoet, and Asmelash Haile Tsegay. 2020a. "Ethiopia High Frequency Phone Monitoring Household Survey, Round 4: List of Tables." *Monitoring COVID-19 Impact on Households*, World Bank, Washington, DC. <https://documents1.worldbank.org/curated/en/175151601574426469/pdf/List-of-Tables.pdf>.
- Wieser, Christina, Alemayehu Azeze Ambel, Tom Bundervoet, and Asmelash Haile Tsegay. 2020b. "Results from a High-Frequency Phone Survey of Households." *Monitoring COVID-19 Impacts on Households in Ethiopia*, Report 2 (June 26), World Bank, Washington, DC.
- Wieser, Christina, Nfamara K. Dampha, Alemayehu Azeze Ambel, Asmelash Haile Tsegay, Harriet Kasidi Mugera, and Jeffery Tanner. 2020. "Results from the High-Frequency Phone Surveys of Refugees." *Monitoring COVID-19 Impact on Refugees in Ethiopia*, Report 1 (December 18), World Bank–UNHCR Joint Data Center on Forced Displacement, Copenhagen.

Wieser, Christina, Nfamara K. Dampha, Alemayehu Azeze Ambel, Asmelash Haile Tsegay, Harriet Kasidi Mugeru, and Jeffery Tanner. 2021. “Results from the High-Frequency Phone Surveys of Refugees.” *Monitoring COVID-19 Impact on Refugees in Ethiopia*, Report 2 (March 3), World Bank–UNHCR Joint Data Center on Forced Displacement, Copenhagen.

Wieser, Christina, Nfamara K. Dampha, Theresa Parrish Beltramo, and Ibrahima Sarr. 2020. “Monitoring COVID-19 Impacts on Refugees in Ethiopia.” *Survey Methodology Document* (December 6), World Bank–UNHCR Joint Data Center on Forced Displacement, Copenhagen.

Wieser, Christina, Nobuo Yoshida, Shinya Takamatsu, Kexin Zhan, and Danielle Aron. 2022. “Poverty Projections and Profiling Based on Ethiopia’s High-Frequency Phone Surveys of Households Using a SWIFT-COVID-19 Package.” Paper presented at the *International Association for Research in Income and Wealth–Tanzania National Bureau of Statistics Conference*, “Measuring Income, Wealth, and Well-Being in Africa,” Arusha, Tanzania, November 11–13.

World Bank. 2022a. “Phone Survey Data: Monitoring COVID-19 Impact on Firms and Households in Ethiopia.” *Ethiopia Brief*, February 10. <https://www.world-bank.org/en/country/ethiopia/brief/phone-survey-data-monitoring-covid-19-impact-on-firms-and-households-in-ethiopia>.

World Bank. 2022b. *World Development Report 2022: Finance for an Equitable Recovery*. Washington, DC: World Bank.

World Bank. 2023. *Displaced during Crisis: Lessons Learned from High-Frequency Phone Surveys and How to Protect the Most Vulnerable*. Washington, DC: World Bank.

World Bank. 2024. *Poverty, Prosperity, and Planet Report 2024: Pathways Out of the Polycrisis*. Washington, DC: World Bank.

World Bank, UNESCO (United Nations Educational, Scientific and Cultural Organization), and UNICEF (United Nations Children’s Fund). 2021. “The State of the Global Education Crisis: A Path to Recovery.” UNESCO, Paris; UNICEF, New York; World Bank, Washington, DC.

Yoshida, Nobuo, Ricardo Munoz, Alexander Skinner, Catherine Kyung-Eun Lee, Mario Brataj, Spencer William Durbin, and D. Sharma. 2015. “SWIFT Data Collection Guidelines Version 2.” World Bank, Washington, DC.

3.

Representative Monthly Phone Panel Surveys: *Listening surveys*

— William Hutchins Seitz²⁷

Introduction

Phone surveys with a panel design—repeatedly interviewing the same households over time—offer a powerful way to track changes in welfare, behavior, and other outcomes. They are especially valuable during periods of rapid change, such as pandemics, conflicts, or natural disasters. Compared to one-time surveys, panel surveys allow for richer analysis of dynamics, such as shifts in employment or coping strategies. While they pose challenges like attrition and respondent fatigue, they also provide deeper insights into household dynamics and potential causal relationships.

This chapter introduces the *Listening surveys*, a series of nationally representative, high-frequency phone panels that combine traditional welfare indicators with timely questions on public perceptions. The surveys are currently active in Tajikistan (since 2015), Uzbekistan (2018), Kazakhstan (2020), the Kyrgyz Republic (2021 and 2022), Ukraine (2022), and Indonesia (2024).

Listening surveys fill a critical gap in policymaking. They capture public views on economic and social issues—something policymakers commonly demand—but that are usually absent from traditional household surveys. Official data systems tend to be infrequent and rigid, and while political polling may collect information on perceptions, it is rarely combined with economic details or links to welfare outcomes. As a result, decision-makers often lack the evidence they need to understand public sentiment and manage reform risks.

²⁷ World Bank.

Without timely data, reform can stall. Uncertainty over how the public will respond—particularly during disruptive or complex changes—can cause policy-makers to delay or avoid needed action. Even well-designed policies may face backlash if the social and political context is poorly understood. Listening surveys help reduce this uncertainty by providing a fast, flexible platform to capture public opinion alongside detailed socioeconomic data.

By collecting monthly data from the same households, these surveys shed light on short-term fluctuations in welfare—measuring volatility and transitions often missed in standard surveys. Topics such as income, employment, food security, access to services, and subjective well-being can be tracked with a level of detail and timeliness that supports more responsive and informed policymaking.

The remainder of this chapter will outline the design and cost of Listening surveys, highlight practical applications from several countries, and offer reflections on when and how this approach may be most useful.

Listening Survey Design

High-frequency phone survey designs have been tested and refined over more than a decade. The *Listening survey* model presented here is the result of both formal evaluations and iterative learning, combining effectiveness in data collection with a relatively low burden on respondents and survey teams.

This design was first fully implemented in *Listening to Tajikistan* in 2015, then expanded in *Listening to the Citizens of Uzbekistan* two years later. It has since been applied in Kazakhstan, the Kyrgyz Republic, Ukraine, and Indonesia, with additional surveys currently under preparation in several other countries.

Beyond practical implementation, the design is shaped by the need for affordability. This is especially important because the value of Listening surveys grows over time, making it essential to sustain them within the modest budgets typically available for data collection.

Structure and content of the survey: face to face baseline and phone-based panel follow-ups

The information collected in Listening surveys can be grouped into two categories: (i) information that is relatively or fully time-invariant, and (ii) information—such as

welfare or public perceptions—that may vary in the short term. Examples of category (i) include the birthdates of household members, completed education levels of adults, sex, and other demographic characteristics. Housing conditions are also typically treated as relatively time-invariant. Category (ii) encompasses a wide range of time-sensitive topics. Examples from Listening surveys include income and its volatility, job changes, health status, access to services, perceptions of policies, assessments of local or national economic conditions, electricity outages, and more.

The standard Listening survey design breaks apart these two types of information into separate phases of data collection—the “baseline” survey, and the “panel” survey. This sequence integrates the strengths of in-person sampling and long-duration interviews for the baseline, with the flexibility and cost-effectiveness of phone-based surveys for the household panel. Examples of questionnaires of both types are available online.²⁸

An important feature of the Listening surveys is that they build on a nationally representative face-to-face baseline survey. This initial in-person interview leverages the advantages of in-person survey designs to ensure representativeness, and collects time-invariant information, contact details, data that cannot be gathered by phone (such as anthropometrics or cognitive tests), and time-intensive modules like consumption expenditure or detailed employment histories. These baseline interviews are typically comprehensive, often following the Living Standards Measurement Study (LSMS) design, and last between 1.5 to 2.5 hours per household. Sample sizes generally range from 3,000 to 5,000 households, depending on the country and the needs of the project.

Standard modules in the face-to-face baseline survey include: (i) a household roster with contact information, education levels, migration history, and demographics; (ii) employment of all household members; (iii) household consumption and expenditure; (iv) health and disability; (v) household income; (vi) housing conditions and access to basic services; and (vii) community-level issues and local amenities.

While only the first module (demographics and contact details) is strictly necessary for launching the second phase, the baseline provides a critical opportunity to collect data that is difficult or impossible to gather over the phone. In some countries, additional modules have been included for environmental concerns (e.g., air quality, lead exposure), time use, financial conditions, and individual cognitive skills.

²⁸ For more information, visit the World Bank's webpage on Listening Surveys (<https://www.world-bank.org/en/topic/poverty/brief/world-bank-listening-surveys>).

The second phase is a continuous phone survey with a panel of households randomly drawn from the baseline sample. Respondents are contacted approximately every 30 days. This monthly panel enables real-time tracking of short-term changes in both traditional and non-traditional welfare indicators. Data collection ideally begins shortly after the baseline, as delays tend to increase the risk of attrition. For efficiency, the phone survey is pre-populated with baseline data and typically lasts 20 to 30 minutes. Monthly samples include 1,000 to 2,000 households, resulting in 12,000 to 24,000 interviews annually.

The standard panel questionnaire includes seven core modules:

1. **Household Updates:** Captures changes to the household roster (e.g., births, deaths, or moves).
2. **Shocks and Events:** Covers disruptions to services (water, electricity, sanitation, etc.) and major events like crises or natural disasters.
3. **Well-being:** Includes measures of financial status, mental health, food security, coping strategies, illness, and subjective poverty.
4. **Migration:** Tracks migration history, intentions, and remittances—distinguishing between household and non-household migrants.
5. **Activities:** Focuses on employment and labor force participation, as well as non-market activities.
6. **Perceptions and Views:** Gathers opinions on government performance, corruption, social assistance, inflation, tax policy, and other relevant policy issues.
7. **Flexible Modules:** Additional thematic modules are added periodically based on evolving needs.

Most questions in the core modules are asked every month, enabling time-series and panel analyses. Other questions and standalone modules are added periodically or on-demand—for example, on climate change, internet use, or policy experiments related to gender norms or subsidy reform. In these cases, listening surveys function as a flexible and cost-effective data platform, leveraging existing infrastructure to reduce marginal costs. The typical lead time to implement a new module is one to two weeks, enabling rapid response to emerging issues. This flexibility has proven especially valuable in crisis settings—for example, during the early stages of COVID-19 when Listening surveys were already active and provided critical real-time insights.

Finding the right balance between consistently repeated questions and periodic modules has been a key learning area in the development of high-frequency

surveys. While both types are useful, some of the most important insights have come from repeated measures, which allow for robust time-series analysis—an essential strength of the Listening design.

Finally, the distinction between time-varying and time-invariant questions is not always straightforward. Experience has shown the value of taking an expansive view of what might change meaningfully over time. For instance, self-assessed income class or subjective well-being—often assumed to be stable—can in fact vary significantly in the short term. This is illustrated, for example, in the panel transition matrix for life satisfaction in Indonesia (see Table 3.1), which captures meaningful fluctuations over time.

TABLE 3.1 Monthly transition matrix for life satisfaction in the Listening to Indonesia survey in 2024

	Very dissatisfied	Dissatisfied	Neither	Satisfied	Very satisfied
Very dissatisfied	27.62	33.15	16.57	14.36	8.29
Dissatisfied	4.55	46.4	19.98	25.19	3.88
Neither	0.53	5.34	54.55	32.04	7.54
Satisfied	0.45	3.83	22.05	65.69	7.97
Very satisfied	0.89	2.02	21.17	30.26	45.66

Source: Listening to Indonesia Survey.

Note: The diagonal includes the share of respondents providing the same response in the previous and following month. All others – the “off diagonal” responses – signify a change in reported life satisfaction.

A note on measuring poverty using Listening surveys

Many countries face challenges in using gold-standard household surveys to frequently measure monetary poverty. When standard sources are unavailable or infrequent, the question often arises: can Listening surveys serve as an alternative source for poverty measurement? The answer depends on the context and specific objectives.

Several Listening baseline surveys have included full consumption, expenditure, and income modules, providing in principle all the components needed

to measure monetary poverty using traditional methods. However, none of the countries currently implementing Listening surveys lack a functioning national poverty measurement system with reasonable frequency. For this reason, listening baseline surveys have not typically been used to generate official headline poverty estimates published by governments.

Because Listening baseline surveys are usually one-off efforts, they are a poor substitute for a routine, in-person household survey system. Where such systems already exist, relying on Listening surveys for poverty estimates adds limited value and can create confusion due to inevitable differences in fieldwork protocols, measurement methods, and data processing decisions. One notable exception is when the baseline provides an opportunity to experiment with alternative welfare measurement approaches—something that can be difficult to do within national statistical systems. This has been the case in countries like Tajikistan, Uzbekistan, and Indonesia, where Listening baselines were used to test innovations in survey design rather than to publish poverty statistics.

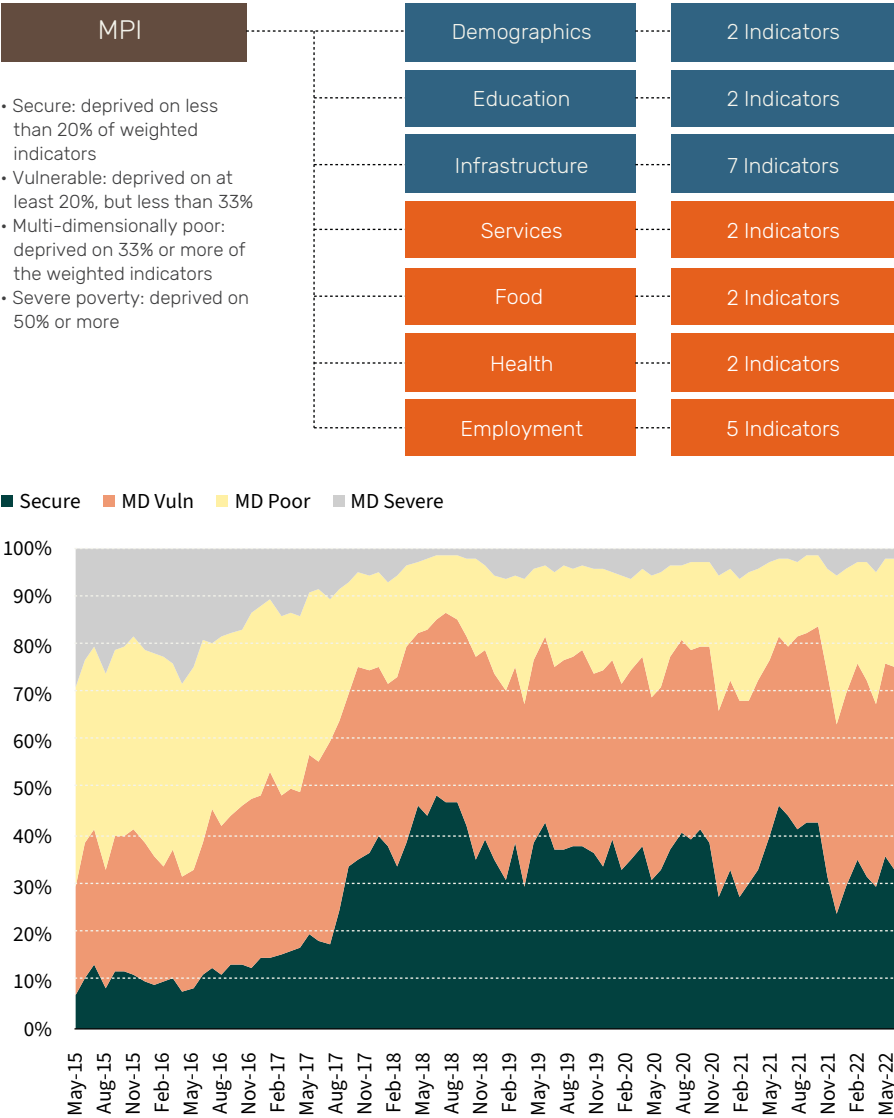
The limitations are more pronounced when it comes to measuring monetary poverty in the monthly Listening panel surveys. Accurate measurement of household consumption or expenditure requires detailed interviews that often last several hours—clearly incompatible with the shorter 20–30-minute format of high-frequency phone surveys. Imputation methods have shown mixed success in this context, and out-of-sample predictions are especially prone to bias.

While income is somewhat easier to collect than consumption, it still presents challenges. Listening surveys ask respondents to report total household income from key sources—wages, self-employment, pensions, remittances, social assistance, agriculture, and others. Although this measure has proven consistent and useful for tracking trends over time, it may miss certain income sources (especially those unknown to the respondent) or contain bias (such as incomplete reporting of agricultural net income, due to difficulties accounting for input costs). Thus, income data in Listening surveys are valuable for understanding changes over time within the panel but are not fully comparable to standard income or consumption-based poverty measures used in national statistics.

In contrast to these limitations in capturing monetary poverty, listening surveys have proven especially useful for tracking subjective and non-monetary poverty. Both the levels and short-term changes in these indicators tend to correlate with income, while also capturing drivers of welfare volatility that traditional surveys often miss. For example, Figure 3.1 illustrates the use of a high-frequency

Multidimensional Poverty Index (MPI) in Tajikistan. Figure 3.2 shows evidence from Indonesia, where financial stress significantly varied month to month—so much so that 40 percent of respondents who self-identified as poor in one round had not said they were poor in the previous month.

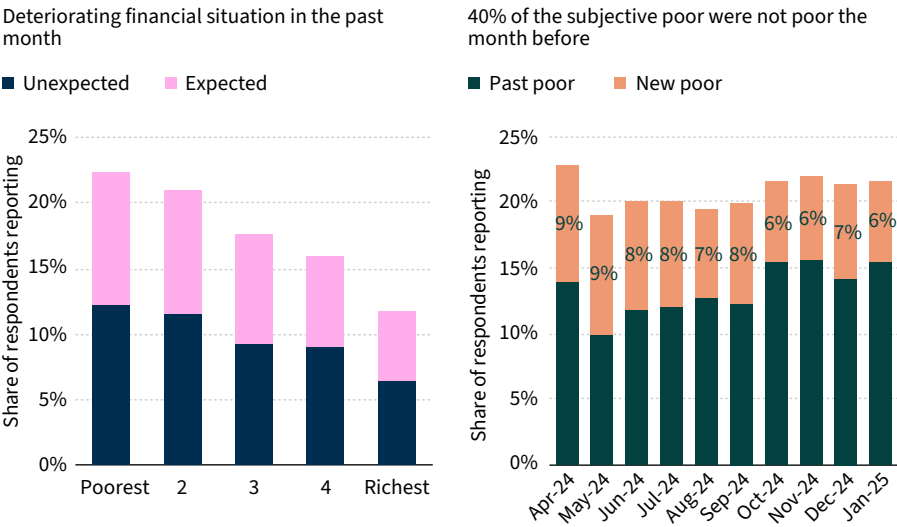
FIGURE 3.1 Listening to Tajikistan’s time-varying multi-dimensional well-being index



Source: Listening to Tajikistan Survey.

These findings underscore a core strength of the Listening survey approach: the ability to capture dynamic, often-overlooked aspects of welfare, particularly in subjective or multidimensional dimensions. While not a replacement for traditional poverty measures, listening surveys provide a valuable complement—highlighting welfare trends that matter for policy but are difficult to detect using conventional survey tools.

FIGURE 3.2 Financial uncertainty and subjective poverty in Indonesia



Source: Listening to Indonesia Survey.

Alternative respondents and target topics

Including targeted questionnaires for specific respondents beyond households—such as local community leaders—can yield valuable insights. In the case of Uzbekistan, *mahalla* leaders, with their close ties to the community, offer an in-depth understanding of local needs, challenges, and dynamics. As traditional neighborhood organizations, *mahallas* are central to both local governance and social cohesion. Engaging their leaders directly in surveys allows for a more nuanced view of how policies are experienced and executed on the ground—insights that household surveys alone may not capture.

This approach not only enhances the richness of the data but also helps identify gaps in service delivery and areas where government programs may require adjustment. For example, *mahalla* leaders in Uzbekistan often serve as the first point of contact for residents seeking financial support from government, making their input crucial for understanding regional variation in how national policies are implemented. Their perspectives can also provide culturally grounded context, adding depth to the interpretation of household data.

This model can be applied in other countries as well. Involving local leadership in survey design and data collection can support more responsive and effective policy development, helping tailor interventions to the specific needs of diverse communities.

Frequency of the data collection

One of the key strengths that distinguishes Listening surveys is their frequency. The current monthly cadence was reached through trial and error. Initial rounds of *Listening to Tajikistan* were conducted weekly, then shifted to every 10 days, then biweekly, before finally settling on a monthly schedule. Other efforts, such as the *Listening to Africa* surveys, experimented with lower frequencies. Experience shows that monthly frequency offers several advantages, falling into two main categories: (i) it aligns naturally with how households experience and report economic conditions, and (ii) it enables a stable, manageable workflow for survey teams.

A major reason for choosing a monthly frequency is that many economic activities—such as wage payments, rent or mortgage obligations, utility billing, and food purchases—follow a monthly cycle. Respondents are also more likely to recall events and expenditures accurately over this period, compared to shorter or longer intervals that may not match their natural budgeting rhythm. Surveys conducted weekly or biweekly risk overburdening respondents and introducing fatigue, while quarterly or semiannual surveys suffer from greater recall bias. Monthly surveys strike a practical balance, ensuring frequent, high-quality data while keeping respondent burden manageable.

Monthly data collection is also especially useful for assessing the effects of shocks. In a typical difference-in-differences research design, analysts need multiple rounds both before and after a treatment to establish trends and interpret impacts. This makes continuous data collection more effective than intermittent surveys, which may miss key transitions or require longer timelines to gather sufficient observations.

Furthermore, many household activities—particularly in agriculture and the informal sector—exhibit strong seasonality. A single annual survey conducted at the same time each year may miss this variation, leading to seasonal bias. Monthly surveys capture seasonal fluctuations as they happen, such as post-harvest abundance versus pre-harvest scarcity, or holiday-related changes in employment and consumption. This allows policymakers to anticipate predictable pressures—such as food insecurity during the lean season—and design timely, targeted interventions.

Monthly frequency also minimizes recall error and reduces reliance on retrospective reporting. Long recall periods are known to distort responses; studies show that extending recall windows leads to underreporting of consumption and participation. While cross-sectional surveys attempt to compensate with retrospective questions, the accuracy is often poor. The Listening survey minimizes these issues by using short recall periods—sometimes as brief as one or two days for questions on electricity or water outages (Seitz, Kudo, & Azevedo, 2023), and one week for employment questions (Heath et al., 2021). By asking about only the last month’s activities, responses are more accurate and sensitive to real-time changes. Repeated monthly interviews also allow researchers to track transitions (e.g., job loss, income recovery) and event sequences (e.g., income decline followed by reduced spending), offering a more detailed view of household dynamics and cause-effect relationships.

Monthly frequency also appears to reduce attrition. Although not rigorously tested, implementers report that regular, predictable contact helps maintain participation and builds respondent habits. Lower-frequency surveys make it harder to establish this routine, increasing the risk of dropouts over time.

From an operational perspective, monthly surveys create a steady workflow that supports staff retention and improves data quality. Infrequent or project-based surveys force firms to hire on short-term contracts, increasing turnover as experienced enumerators leave for more stable employment. By contrast, monthly surveys allow firms to offer continuous work, creating stable, professional survey teams. This “call center” model improves efficiency, lowers per-round costs, and enhances quality through better supervision and staff development.

A consistent team also leads to better data: enumerators become more skilled with each round, familiar with the questionnaire, and capable of spotting inconsistencies or probing effectively. The regular cadence of monthly surveys also allows firms to maintain infrastructure, such as call centers, and to plan logistics more effectively.

Over time, high-frequency survey systems like Listening can strengthen national data capacity. In the Kyrgyz Republic, for example, the Listening survey is used by the National Bank for routine inflation forecasting. In Tajikistan, it serves as the official source for Sustainable Development Goal (SDG) monitoring in areas such as electricity and water access.

In sum, monthly frequency offers a practical and cost-effective balance—enhancing recall accuracy, capturing seasonal and shock-driven fluctuations, supporting strong implementation teams, and delivering timely insights for decision-makers.

Interview duration

Interview duration is an important consideration in panel surveys, especially to ensure long-term respondent engagement. For the Listening surveys, the target is to keep interviews within 20–30 minutes, with a strict upper limit of 30 minutes for the core panel modules. Achieving this balance requires careful planning and real-time adjustments once the survey is in the field.

The first panel interview is usually the longest. It involves introducing the purpose of the survey, explaining specific questions, and collecting a comprehensive set of time-invariant information—especially if that data was not gathered in a prior face-to-face baseline. There is an inherent trade-off: while the first interview is the best opportunity to resolve missing or inaccurate information, it also sets the tone for future participation. If the initial experience is too long or repetitive, respondents may be less willing to continue in future rounds. To address this, several Listening surveys have temporarily removed selected modules from the first phone interview, reintroducing them once interview durations declined.

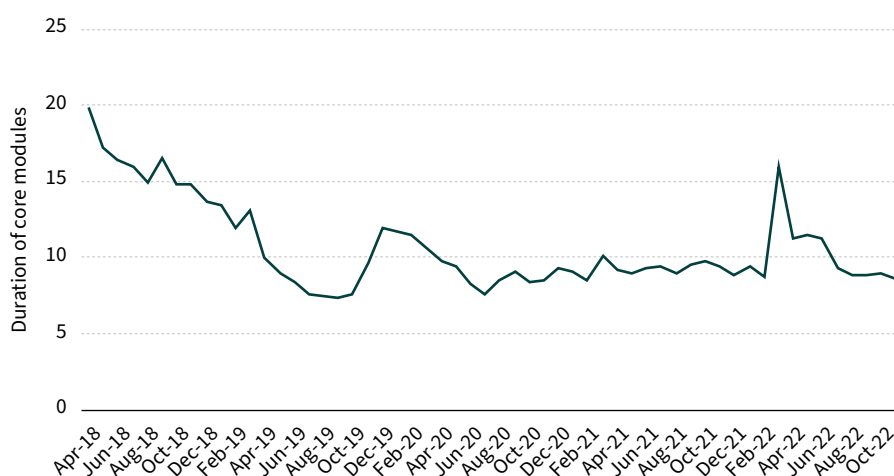
However, care must be taken not to remove all the engaging content from the first round. Modules that ask respondents for their views on current policy issues are typically the most interesting and motivating. Qualitative feedback indicates that many participants see the opportunity to “have their voice heard” on these issues as the main reason for taking part—often more so than financial incentives. Therefore, it is important to ensure that the initial interview includes not only basic demographic questions (e.g., age, education) but also topics that spark interest and encourage continued engagement.

At the same time, it is important to avoid overcorrecting in the name of brevity by removing too much content from the questionnaire. In the early stages of survey design, teams often push for shorter interviews, citing long durations observed in

piloting or pre-testing. However, these early rounds are typically not representative of routine interviews. Experience across Listening surveys shows that interview duration decreases rapidly over time. First rounds are already more efficient than pilots, and subsequent rounds often reduce interview time by half or more.

For example, in *Listening to the Citizens of Uzbekistan*, the core modules initially took around 20 minutes to complete but dropped to less than half that in later rounds (see Figure 3.3). This pattern highlights the importance of allowing space for learning and adjustment, rather than prematurely constraining the questionnaire based on early experiences.

FIGURE 3.3 Duration of the Listening to the Citizens of Uzbekistan across rounds



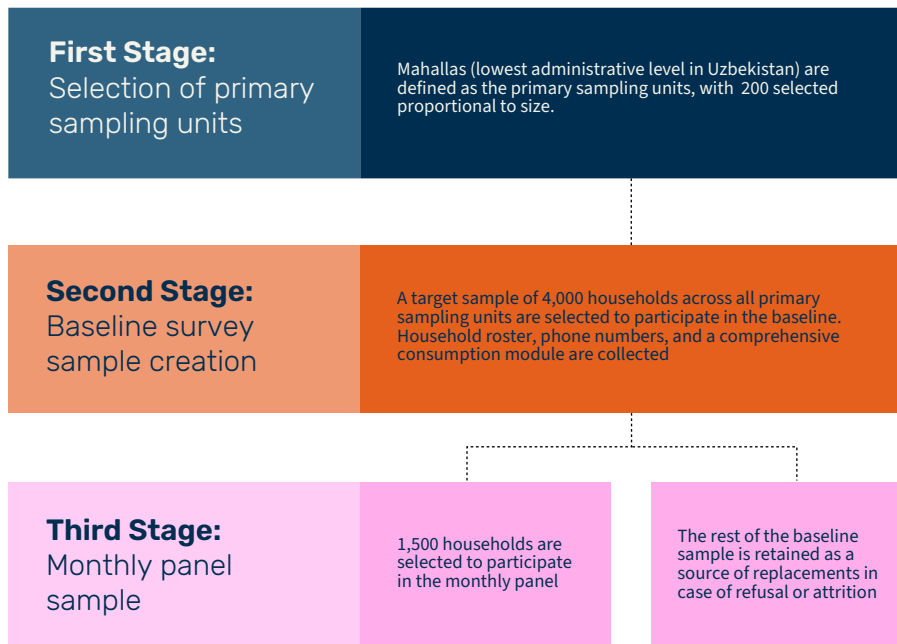
Source: Listening to the Citizens of Uzbekistan Survey.

Sampling design, attrition, and weights

Representativeness

As noted earlier, listening surveys are built upon a representative face-to-face baseline survey. As a result, they are designed to be nationally representative and follow a three-stage sampling procedure. Figure 3.4 illustrates these three stages using the *Listening to the Citizens of Uzbekistan* survey as an example.

FIGURE 3.4 Stages that compose the survey sample, with illustrative numbers from the Listening to the Citizens of Uzbekistan survey



Source: Listening to the Citizens of Uzbekistan Survey.

The first two stages use standard in-person sampling techniques. In the first stage, Primary Sampling Units (PSUs) defined here as *mahallas* (neighborhood-level administrative units)—are selected from the population without replacement. PSUs divide the population into manageable units, and the list of all PSUs with their population sizes serves as the sampling frame. Rather than sampling households directly across the country, the survey first selects a sample of *mahallas* and then, in the second stage, selects households within each *mahalla*. This clustering is essential to reduce costs; visiting randomly selected households across a wide geographic area would make fieldwork prohibitively expensive.

To ensure representativeness, each PSU is selected using a probability proportional to size (PPS) method, where the probability of selection is proportional to the population or number of households in that PSU. Stratification is often used at this stage to ensure sufficient sample allocation to relatively less populated areas, and geographies of special policy interest (see below).

In the second stage, households are selected from within each PSU using a simple random sampling method. Upon arriving in a selected PSU, the survey team first updates the estimate of the number of households—necessary because census or administrative data are often outdated or imprecise. This was particularly relevant in Uzbekistan, where the last census had been conducted long ago, and many areas had undergone significant demographic changes.

There are two common methods for conducting this second-stage sampling. The most rigorous—and labor-intensive—approach involves canvassing the entire PSU, mapping all streets, and listing all occupied dwellings. This yields a complete and updated list of households, from which a random sample (typically 8–12 households per PSU) is selected, along with a set of replacements to account for non-response. Interviews are conducted based on this list, and if a selected household cannot be interviewed (due to refusal or absence after repeated visits), a replacement household from the reserve list is used. Wherever feasible, this is the preferred method for baseline sampling. As an alternative to full listing, some surveys use an adaptative sampling design called random-route (sometimes called a “random walk”) method to pick households.²⁹

In the third stage, a stratified simple random sample is drawn from the baseline to form the monthly panel. The target panel size typically ranges from 1,000 to 2,000 households per month, balancing representativeness, statistical precision, and budget constraints. Since the panel is drawn from a known baseline sample, follow-up analyses can incorporate baseline characteristics to reweight the data and

²⁹ Although it is a relatively common practice used when either time or financial constraints are binding, it is not a fully rigorous probability sampling method. The advantage is that it does not require a complete list of households in advance and instead relies on a random starting point and systematic walking pattern. The standard procedure is to first determine a starting point in the PSU from which to begin the household selection. This is often either a prominent central location or an arbitrary point on the PSU boundary, but ideally, the starting point itself is chosen at random. Once the starting point is chosen, the enumerator walks in a specified direction, commonly chosen by either spinning a bottle or pencil on a map. This ensures the walk does not always begin or trend towards the most convenient or central area. Based on an estimate of the PSU’s size, the enumeration team sets a systematic interval to sample. For example, if the PSU is thought to have around 100 households and 10 are needed, the interval might be 10 (every 10th dwelling). Each time a dwelling is selected, the interviewer attempts an interview there, if possible, completes the interview, then continues walking in the same direction. The random route continues until the interviewer has successfully interviewed 10 households in that PSU. For cases when a selected household cannot be interviewed (either due to refusal or no contact), the interviewer immediately replaces with the next household in line. The random route method is vulnerable to bias from poor implementation. To be successful, it is important that interviewers faithfully follow the random route instructions to approximate an equal-probability selection.

adjust for any selection bias. This method contrasts with alternatives like random digit dialing (RDD), which often suffer from more pronounced selection issues.

All surveys using clustered designs are subject to intra-cluster correlation—the tendency of households within the same neighborhood to exhibit similar characteristics. This correlation reduces statistical precision, as additional observations within a cluster provide less new information. This issue is especially relevant in the third sampling stage. If the panel sample were selected using simple random sampling across the entire baseline, the distribution of households per PSU would be uneven, with some PSUs overrepresented and others underrepresented. This imbalance would reduce precision: adding more households in an already well-sampled PSU yields less marginal value than sampling in underrepresented PSUs.

To address this, the third-stage sample is stratified to ensure even distribution across PSUs. A fixed proportion of households is randomly selected from each PSU, maximizing the geographic and statistical efficiency of the panel. This approach ensures an optimal allocation of the sample from the baseline pool and supports both representativeness and analytical rigor.

Stratification and target populations

Listening surveys are typically stratified by region and urban/rural status and often include booster samples of populations of special policy interest. This approach ensures a more accurate and representative portrayal of key subpopulations within each country. Stratification reduces sampling variability within each stratum, resulting in more precise estimates, and ensures that smaller—but policy-relevant—groups are adequately captured in the survey sample.

The stratification process follows standard practices used in household survey design. It begins with clearly defining the population using a sampling frame, typically a complete list of PSUs or clusters. This frame is usually based on the most recent national census (as in Kazakhstan, Tajikistan, the Kyrgyz Republic, and Indonesia) or a recent household listing or administrative register (as in Ukraine and Uzbekistan).

Once stratification variables are determined, the population is divided into distinct strata—at a minimum by (i) urban/rural classification and (ii) administrative region, though additional criteria may also be applied. Each stratum is treated as an independent sampling domain. Random samples are then drawn

independently within each stratum using a standard multi-stage approach: PSUs are selected in the first stage, followed by simple random sampling of households within each selected PSU.

In many cases, the number of households sampled per stratum does not follow simple proportional allocation based on population size. Instead, oversampling is used to ensure sufficient representation of smaller administrative areas or groups of special interest. This results in differing selection probabilities across strata. Accordingly, sampling weights must be calculated and applied during analysis to ensure that survey results remain representative of the national population. These weights adjust for both selection probabilities and potential non-response.

Beyond geographic stratification, listening surveys often incorporate stratification by target populations. For example, in Uzbekistan, Kazakhstan, and Tajikistan, samples were designed to ensure adequate representation of social assistance beneficiaries. In Indonesia, special emphasis was placed on households operating small enterprises.

These objectives can complicate sampling procedures, particularly when reliable sampling frames for target populations are unavailable. In Uzbekistan, for instance, no data existed on the number of social assistance beneficiaries at the PSU level before fieldwork began. To address this, PSUs were first selected based on geographic stratification. Upon arrival at each PSU, enumerators worked with local officials to digitize beneficiary records. Households were then sampled using a fixed quota: a random sample of beneficiaries and a separate sample of non-beneficiaries were selected to ensure balance.

A similar approach was applied in Indonesia to oversample households with micro or small enterprises. The sample was stratified into two groups: (i) households operating a household enterprise, and (ii) those that did not. The target was for 50% of the sample to consist of enterprise-operating households. As no registry of household enterprises existed, enumerators conducted a full listing within each PSU, collecting information on household composition and whether the household operated an enterprise. The sampling protocol then required selecting five households with enterprises and five without, totaling 10 households per PSU. With 500 PSUs, the target sample was 5,000 households. Selection probabilities were adjusted based on the share of enterprise and non-enterprise households in each PSU, and final weights were scaled using estimates from Indonesia's official household survey (SUSENAS).

Attrition rates and strategies to reduce it

Panel surveys face two key challenges to representativeness over time. The first is attrition—when households are no longer willing or able to participate in the survey. Attrition in the first round is particularly difficult to address. As with any survey, non-random refusal introduces sampling bias. This bias can be partially corrected using non-response weights (i.e., assigning greater weight to respondents who closely resemble non-respondents in terms of observable characteristics). However, in the first round, the ability to adjust for non-response is limited by the lack of information about those who chose not to participate. In fact, the core issue is that these individuals refused to provide any information.

Attrition rates vary across countries—from 38 percent in Kazakhstan to just 11 percent in Uzbekistan. The Listening survey protocol allows households to skip a particular round upon request (e.g., due to the unavailability of household members). If a household does not answer the phone during its scheduled interview, enumerators are instructed to attempt follow-up at a later date. For example, if a household does not respond in April and is temporarily replaced by another from the same PSU, the protocol is to attempt to re-contact the original household in May and again in June.

Experience shows that field teams are often successful in “salvaging” these temporarily non-participating households, helping to contain overall attrition. In Indonesia, for example, about 75 percent of the sample completed at least 6 out of the first 12 monthly rounds, though only 30 percent completed all 12 rounds. In 2024, consistent participation was strong in several countries:

- Tajikistan: 1,322 out of 1,400 households participated in at least 12 months
- Uzbekistan (L2CU): 1,211 out of 1,320
- Kazakhstan: 1,173 out of 1,400
- Kyrgyz Republic: 1,302 out of 1,500
- Ukraine showed the lowest consistent participation, with 628 out of 1,358 households participating in 6 or more rounds.

These figures suggest that Listening surveys have achieved comparable or even better retention rates than many other high-frequency survey designs. For instance, the *Listening to LAC* initiative used a different protocol and faced higher attrition: in Peru, attrition reached 67 percent by the second round and climbed to 75 percent by round six (Ballivian et al. 2015). In Honduras, attrition was lower—41 percent in the first follow-up, increasing to 50 percent. In South

Sudan, attrition rose from 31 percent in the first round to 49 percent by the fourth. In Tanzania, attrition was modest, reaching 25 percent after 33 rounds (Demombynes, Gubbins, and Romeo 2013; Croke et al. 2012).

Monetary incentives have also been shown to improve response rates. All Listening surveys offer a modest financial benefit—typically \$1–2 per interview—delivered as mobile phone credit. While the impact of these incentives has not been formally studied within the Listening surveys, a literature review by Abdelazeem et al. (2023) finds that incentives can significantly increase participation: 25% for cash payments, 19% for vouchers, and 12% for lotteries.

However, financial incentives are not the only motivator. As previously mentioned, qualitative feedback from Listening survey participants often highlights intrinsic motivations, such as the desire to have their voices heard. This aligns with broader research, which shows that emphasizing the social value of participation can be just as, or even more, effective than monetary incentives (Kropf and Blair 2005).

In some cases—especially Kazakhstan and Ukraine—random-digit dialing (RDD) has been used to supplement samples affected by high attrition. RDD surveys typically involve three steps: (i) randomly generating telephone numbers, (ii) contacting them from a central call center, and (iii) administering the survey using computer-assisted telephone interviewing (CATI). Wolter, Chowdhury, and Kelly (2009) provide a comprehensive discussion of this approach.

RDD methods have both advantages and limitations. Compared to in-person surveys like the Living Standards Measurement Study (LSMS), RDD offers faster and more cost-effective data collection, especially across wide geographic areas. It eliminates the need for travel and can be deployed quickly. However, while RDD generally outperforms web- or email-based surveys in response rates, it typically achieves lower participation rates than traditional in-person surveys.

Survey weights

Attrition is often non-random, meaning that certain types of respondents are more likely to drop out of the survey over time. As a result, the sample in later rounds may no longer resemble the original population, introducing bias and potentially compromising the validity of the results. If certain groups become overrepresented or underrepresented due to attrition, survey estimates can become skewed.

To partially correct for attrition bias, listening surveys adjust sampling weights so that each survey round remains comparable to the baseline sample in terms of key observable characteristics. Reweighting modifies the contribution of each respondent such that the weighted distribution of selected variables in each round aligns with that of the baseline.

This adjustment is done using entropy balancing, a technique that recalibrates weights to ensure balance across selected covariates between rounds. Starting with the baseline weights, an adjustment factor is calculated to realign the follow-up sample's characteristics—typically using structural variables such as household size, baseline income, and demographics. In Tajikistan, for example, selected covariates include region and city indicators, as well as an urban/rural classification.

The baseline weight for each household is scaled by the inverse of its estimated retention probability (or a similar adjustment factor), resulting in a new weight that, when applied, ensures the composition of the sample in that round closely matches the baseline.

Example .do files for generating round-specific weights are available online.³⁰

Quality control and supervision

Active participation by World Bank staff in supervision, data review, and “training of trainers” has played a crucial role in ensuring the quality of Listening surveys. Data quality is further reinforced through periodic refresher sessions with enumerators review guidelines, as well as by monitoring a random subset of call audio recordings. Maintaining a high level of supervision and conducting regular spot checks is essential. Research on survey implementation has consistently shown that the certainty of being monitored is a more effective deterrent against protocol violations than the severity of penalties.

With respondents' consent and in compliance with data protection protocols, interviews can be reviewed either in real-time through live monitoring or retrospectively by analyzing recorded calls. In both cases, structured evaluation checklists are used to verify adherence to key protocols—such as accurate question phrasing, correct response recording, and sustained respondent engagement.

³⁰ For more information, visit the World Bank's webpage on Listening Surveys: <https://www.world-bank.org/en/topic/poverty/brief/world-bank-listening-surveys>.

In addition to audio reviews, regular monitoring of call and data entry metadata offers valuable insights into the consistency and reliability of data collection. Most Listening survey teams operate real-time online dashboards for ongoing supervision. Supervisors and World Bank teams track call durations to flag interviews that are unusually short or long. Breaks or pauses within calls may indicate interviewer disengagement or potential data fabrication. Similarly, the frequency and timing of call attempts provide insight into the diligence of interviewers in reaching respondents. Data entry timestamps are also analyzed, as overly rapid response entry may suggest rushed or fabricated data.

Real-time data review enhances quality control by enabling immediate detection of potential errors or inconsistencies. Automated data quality flags help identify suspicious patterns—such as identical responses across multiple interviews, excessive missing values, or logical inconsistencies. Dashboards allow supervisors to monitor incoming data, flag outliers, and track response distributions, enabling targeted spot checks when unusual trends emerge. When issues are identified, enumerators receive timely feedback, allowing them to adjust their approach and improve data quality during ongoing fieldwork.

Implementing partners

All ongoing Listening surveys are currently implemented through private firms, although earlier efforts—such as the *Listening to Africa* surveys—were conducted through national statistics offices.

Private survey firms operate under a business model that emphasizes operational flexibility. They can quickly mobilize resources, adapt methodologies, and deploy data collection teams with minimal bureaucratic delay. This agility makes them particularly well-suited for time-sensitive studies, emergency assessments, and innovative data collection approaches, such as high-frequency surveys or mobile-based enumeration.

In contrast, a key advantage of working with official statistics agencies is the opportunity for institutional capacity building. Engaging these agencies strengthens long-term data infrastructure, enhances government ownership of the data ecosystem, and supports sustainable development planning. Investments in training, methodological improvements, and technological upgrades within official agencies yield lasting benefits that extend beyond any single survey round.

The choice between using private firms and official agencies depends on the project's objectives, timelines, and the broader policy context. For rapid, targeted, and cost-sensitive surveys, private firms are often the preferred option. However, when long-term capacity building and statistical system strengthening are priorities, engaging with official agencies becomes essential.

Hybrid approaches—where private firms handle data collection under the methodological oversight of official statistics agencies—can offer a balanced solution. These models combine the efficiency and responsiveness of private firms with the institutional development benefits of public-sector engagement.

Costing and resources for Listening surveys

In some instances, listening surveys have been able to leverage other in-person surveys as baseline data sources. This approach was used in Tajikistan, Kazakhstan, and Ukraine, where baseline data collection incurred no direct cost to the Listening project. In other cases, the full cost of conducting the in-person baseline survey had to be included in the project budget.

The cost of in-person data collection varies across countries. In geographically large and diverse countries with dispersed populations—such as Indonesia—costs are significantly higher than in more compact, densely populated, and accessible countries like Uzbekistan.

Launching a typical Listening survey requires dedicated attention from core team members over several months. Preparing the baseline survey can take up to a full month of one team member's time. The launch of the panel survey usually requires an additional 3 to 6 weeks to develop call center protocols and ensure a smooth rollout of the first rounds.

Once data collection becomes routine, ongoing monitoring and administrative tasks typically require 2 to 5 person-days per month. In addition, a data management specialist is involved for approximately 5 days per month to support data processing and quality assurance.

The time required for data analysis depends on the complexity of the outputs. In recent rounds of the *Listening to Central Asia*, Ukraine, and Indonesia surveys, data analysis and presentation typically required 3 to 5 person-days per month, supported by standardized workflows and automated reporting systems.

However, additional time may be needed for custom modules or new topics that require tailored analysis not implemented in earlier rounds.

Funding for Listening surveys has varied by country and over time. In Central Asia, for instance, about one-third of survey costs were covered by the World Bank's core resources, while external partners funded the remaining two-thirds. Core survey budgets have often been supplemented with additional financing to include modules on specific policy topics or sectors.

These thematic modules have covered a wide range of issues, including energy tariff reforms and service reliability, water access and quality, vaccine availability and telemedicine, peer-to-peer lending and financial services, the performance of social assistance programs, digital access and e-government services, and local development priorities.

Support for Listening surveys has come from a diverse group of development partners, including the United Kingdom's Foreign Commonwealth and Development Office (FCDO), the Japan International Cooperation Agency (JICA), UNICEF, the Australian Department of Foreign Affairs and Trade (DFAT), and the Gates Foundation. Much of this external funding has been channeled through the World Bank's Umbrella Facility for Poverty and Equity, which coordinates support for initiatives aimed at strengthening data systems and evidence-based policymaking.

How were these surveys used?

Listening surveys have been used for a variety of purposes including understanding and responding to shocks, linking well-being and public opinion, informing policy reforms, and conducting experiments to test policy assumptions or assess policy impacts. The following sub-sections provide a selected array of examples across country settings to illustrate some of these use-cases.

Understanding and responding to shocks

COVID-19 related job destruction and impacts on income in Uzbekistan

As the COVID-19 pandemic unfolded, policymakers had to act quickly, often with limited information. In countries where they were available, listening surveys played a critical role by providing timely data on how livelihoods were affected

and who was most vulnerable. For example, in Uzbekistan, combining pre-pandemic unemployment data with real-time survey results revealed a sharp drop in employment around May 2020, followed by a gradual recovery. This information helped guide the design of social protection measures well before traditional survey data became available (Figures 3.5 and 3.6).

FIGURE 3.5 Share of households where at least one member lost their job (Uzbekistan)

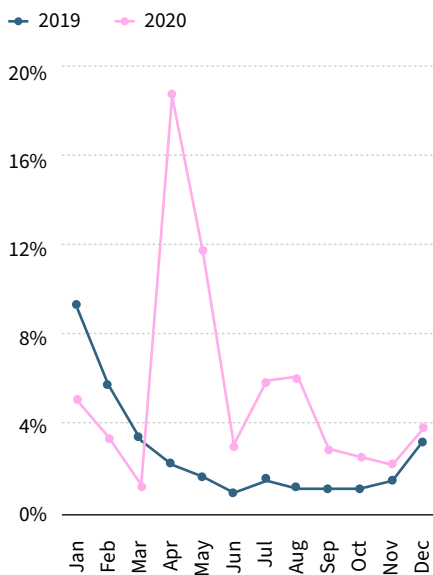
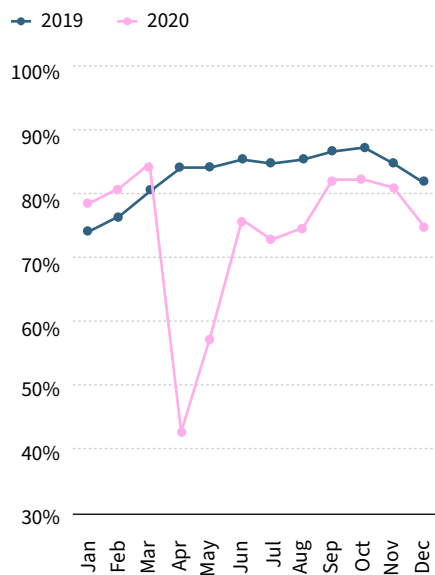


FIGURE 3.6 Share of households where at least one member is working (Uzbekistan).



Source: Listening to the Citizens of Uzbekistan.

Effects of the war in Ukraine on migration and remittances in Central Asia

Central Asian countries are among the most remittance-dependent in the world, with many families relying on income from relatives working in Russia. When the war in Ukraine began in 2022, there was considerable concern about the potential return of migrants and the economic impact on households back home. However, listening surveys quickly uncovered an unexpected trend: rather than declining, migration to Russia increased in the months following the war. A sharp appreciation of the Russian ruble boosted the purchasing power of migrants' earnings, resulting in a surge in remittance income. This led to measurable gains in food security among families with newly migrated members. Yet, as the war continued, migration slowed. The surveys captured this shift in real-time (Figures 3.7 and 3.8).

FIGURE 3.7 Uzbekistan: Share of Households with a migrating member

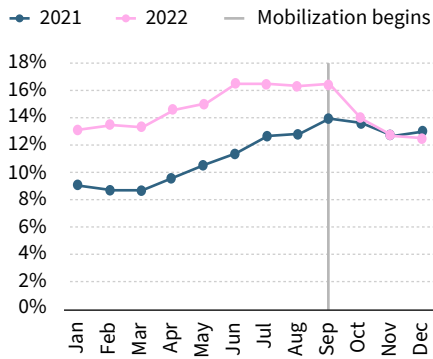
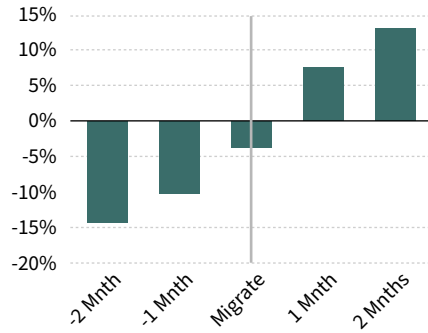


FIGURE 3.8 Uzbekistan: Share of Households reporting capacity to afford food before/after migration.

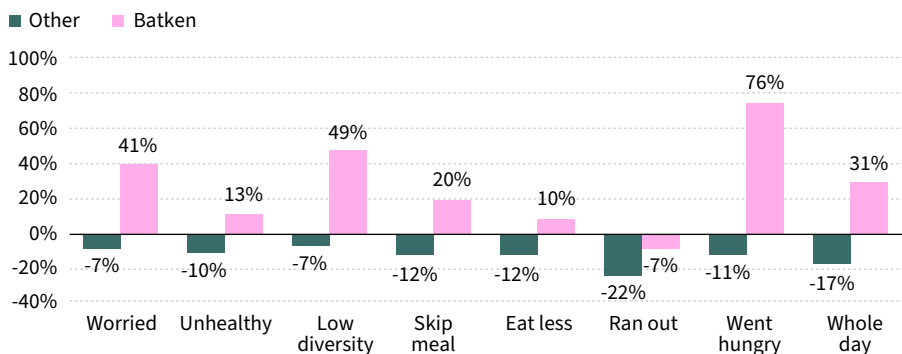


Source: Listening to the Citizens of Uzbekistan.

Food security and conflict in the Batken Region of the Kyrgyz Republic.

When the 2022 border clashes between Tajikistan and the Kyrgyz Republic occurred, the Listening surveys made it possible to conduct a rapid assessment of conditions in the Batken region, which was directly impacted by the conflict. The data revealed a sharp deterioration in food security in Batken, where nearly 80 percent of households reported experiencing hunger across one or more dimensions (Figure 3.9). These insights helped highlight the localized impact of the crisis and the need for targeted support.

FIGURE 3.9 Changes in food security over August in the region of Batken vs rest of the Kyrgyz Republic



Source: Listening to the Kyrgyz Republic.

Linking well-being and public opinion

Inflation, food insecurity and public opinion in Central Asia.

From 2022 to 2023, inflation rose across Central Asia. Surveys revealed that increased food prices significantly contributed to food insecurity, with a 1 percent rise in food inflation leading to higher food insecurity in all surveyed Central Asian countries. The surveys also showed that increases in food prices are associated with a decline in people's appreciation of the country's direction and government management (Figure 3.10 and 3.11). Conversely, decreases in food inflation are linked to lower subjective poverty and higher appreciation of government performance.

FIGURE 3.10 One percent change in food security measure vs. probability respondent satisfied with life

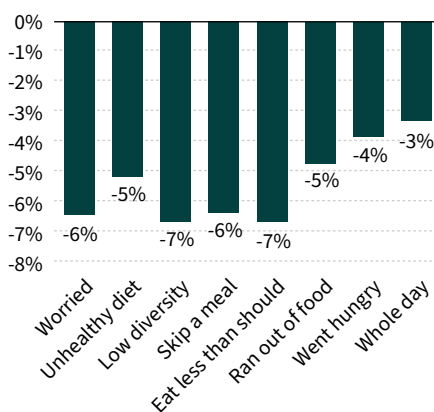
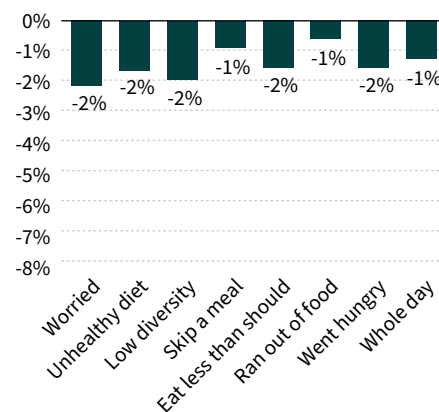


FIGURE 3.11 One percent change in food security measure vs. probability of approval of the country's policy direction



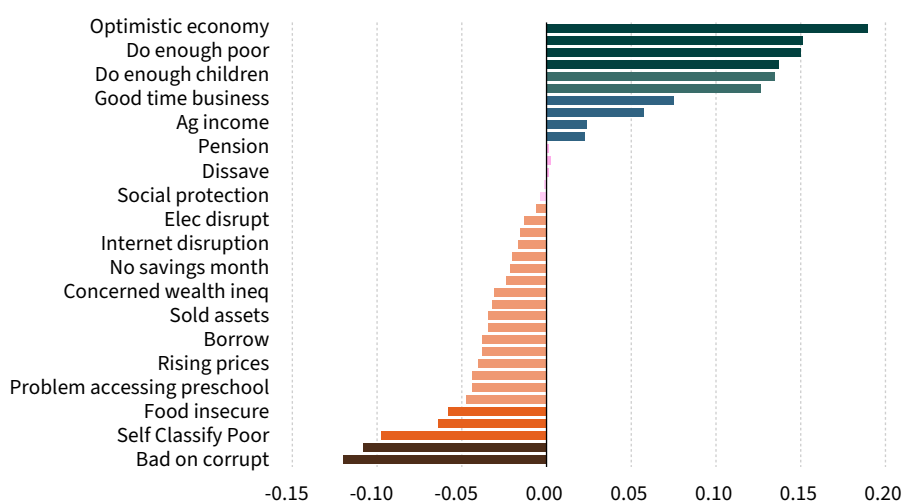
Source: Listening Surveys from Kyrgyz Republic, Uzbekistan, Tajikistan, and Kazakhstan.

Understanding the drivers of the 2022 Kazakh unrest in Kazakhstan

The survey results highlighted an accumulation of public policy concerns over the preceding year. By the time the protests commenced, less than half of the survey respondents believed that the government was engaging in an open dialogue with citizens, only 39 percent felt it was a favorable time to find employment, and merely 36 percent thought the government was adequately supporting the poor. All indicators showed significant deterioration over the preceding months (Figure 3.12). Negative views on corruption corresponded to a 12 percent decline in the

belief that the country was on the right path with reform. When respondents reported more challenging economic circumstances—particularly worsening local economic conditions, falling into poverty, and the onset of food insecurity—they were between 7 and 10 percent less likely to believe that the country was on the right track with reforms.

FIGURE 3.12 Percent change in response that the country “on right track with reform among those reporting changes in welfare “ (Kazakhstan: December 2020-January 2022)



Source: William Seitz and Metin Nebiler, 2022.

Notes: The measure of views on the country’s direction is derived from the question “Do you agree or disagree with the following statement: the country is generally on the right track on political, social, and economic reforms.” The outcome variable is coded as 0 if the respondent disagrees, and 1 if the respondent agrees. Variation is measured when a specific respondent transitions from one response to the other in consecutive months. Green represents improvement, and red represents deterioration.

The Listening to Central Asia surveys offered critical insights into vaccine hesitancy during the COVID-19 pandemic. While vaccination rates improved over time, the reasons for refusal shifted notably. In Kazakhstan (Figure 3.13), for example, distrust in vaccine producers was the top concern in April 2021 (35%), but by September, concerns about contraindications had become the leading reason (over 44%), nearly tripling in absolute terms. In Uzbekistan, the share of people who said they would “definitely not” get vaccinated fell from 20% to 14% between April and September 2021, yet concerns about contraindications rose sharply—from 18% to over 40%. Many believed vaccines were unsafe for older adults or

those with certain health conditions. The surveys revealed that default guidelines in some countries excluded high-priority groups—such as the elderly and immunocompromised—from vaccination. These findings helped prompt revisions to national vaccination guidelines in several countries.

Informing Policy Reform

Assessing changes to the Uzbekistan Propiska System

Uzbekistan has historically had one of the world’s lowest rates of internal migration, with more people moving abroad than relocating within the country. A key barrier was the Propiska system—a domestic registration requirement that restricted internal movement, particularly to the capital, Tashkent. Following a change in leadership in 2016, the government began considering reforms. However, concerns about displacement and resistance from long-term urban residents made progress uncertain. To inform the debate, the Listening to the Citizens of Uzbekistan survey included questions on public attitudes toward reform. The results were striking; 91 percent of respondents supported free movement within the country, with strong support even in Tashkent. These findings helped shape the national conversation and informed the eventual removal of Propiska restrictions. Subsequent survey rounds showed that support for the reform continued to grow, highlighting the value of real-time data in monitoring public sentiment before and after major policy changes.

Informing school feeding programming in Uzbekistan

Data from the Listening survey showed that a school lunch pilot program significantly reduced food insecurity in participating areas, as illustrated in Figure 3.14. The program also received broad public support. In the final survey round for this assessment, 93 percent of respondents favored a universal free school meals program for all young children (Figure 3.15). The survey also gathered feedback from parents of children in schools with lunch programs. The most common concern was the nutritional quality of the meals. In response, government partners recognized this as a key challenge and have taken steps to improve meal standards. The Listening survey findings are now helping to shape a new national program that includes minimum nutrition requirements.

FIGURE 3.14 Count of food insecurity measures for treated region compared to a statistical counterfactual based on trends in similar regions

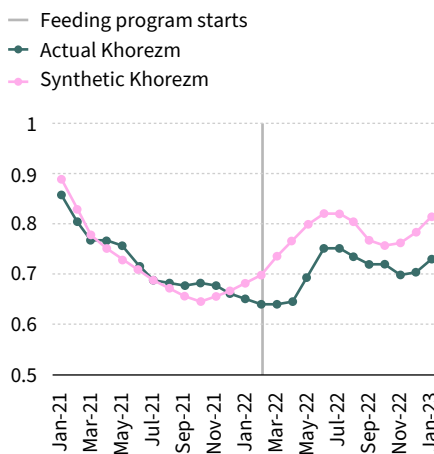
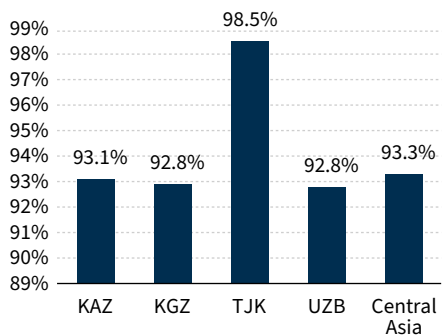


FIGURE 3.15 Share of respondents in 2023 that agree that “all schoolchildren should be given healthy school meals by the school or government at no cost to the student”

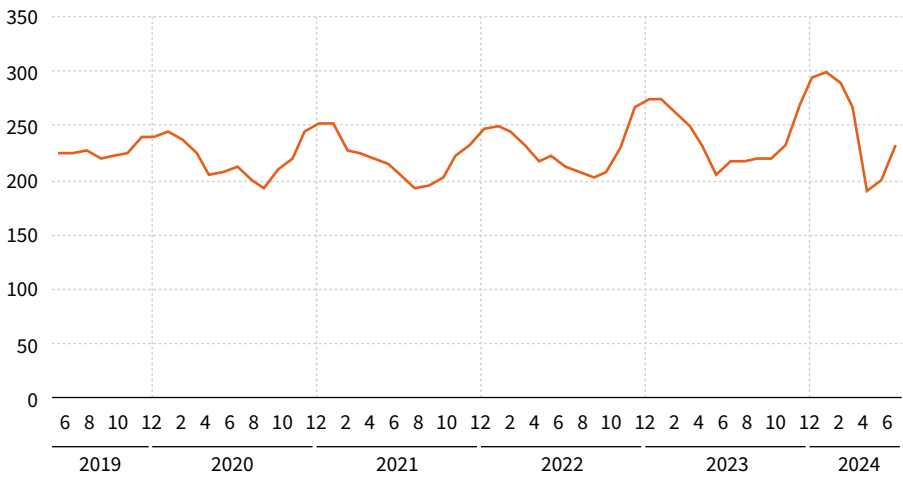


Source: Listening to the Citizens of Uzbekistan Survey.

Energy tariff reform in Uzbekistan

Energy use varies significantly with the seasons and daily routines—for example, consumption often rises during holidays when people spend more time at home. In the context of tariff reform, understanding both household energy demand and sensitivity to price changes is essential for balancing fiscal goals, climate objectives, and social impacts. Listening to Uzbekistan surveys have provided valuable data on these patterns. Figure 3.16 shows how reported residential electricity consumption has fluctuated over time. The data also enabled direct estimates of how consumers respond to price changes, helping regulators make better-informed decisions about pricing and investment. These results were essential to the design and implementation of [the Innovative Carbon Resource Application for Energy Transition Project for Uzbekistan \(iCRAFT\)](#), the World Bank’s first “policy crediting” program. Through a \$46.25 million grant, iCRAFT created incentives for energy subsidy reforms resulting in lower energy consumption and GHG emissions. The program assigns value to, and credit for, the implementation and enforcement of policies that foster emission reductions in the energy sector.

FIGURE 3.16 Dynamics of residential electricity consumption in Uzbekistan



Source: Listening to the Citizens of Uzbekistan Survey.
Note: This graph illustrates average monthly (round) consumptions since Jun of 2019. Each dot with 95% CI is average monthly consumption in UZB for a given year. The dashed grey vertical line indicates the electricity tariff change. The seasonal/monthly effects are consistent over the years. Data from the Listening to the Citizens of Uzbekistan survey.

Survey Experiments and Impact Assessments

One of the uses that the Listening surveys provide is the possibility of implementing survey experiments and impact assessments. Below are a few examples:

A recent study (Seitz 2023) used surveys to explore gender bias in how people perceive wage fairness—an issue particularly relevant in Central Asia, where traditional gender roles often shape expectations around work and caregiving. Women are typically seen as caregivers, while men are viewed as primary earners. The experiment was conducted through the Listening to Central Asia surveys in the Kyrgyz Republic, Uzbekistan, and Kazakhstan. Respondents were shown short profiles (or “vignettes”) of individuals with randomly assigned characteristics—age, gender, and salary—and asked to judge whether the person was overpaid, fairly paid, or underpaid across eight occupations, including roles like doctor, manager, and farm worker.

The results revealed clear gender biases. Across all countries and occupations, women were significantly less likely to be seen as underpaid and more likely to be viewed as fairly or even overpaid—despite having the same qualifications and

salaries as their male counterparts. These patterns were most pronounced in Uzbekistan and point to persistent perceptions that undervalue women’s labor, even when their earnings match those of men.

A randomized experiment conducted through the Listening surveys in Uzbekistan, Tajikistan, and Kazakhstan tested whether financial incentives could increase willingness to get vaccinated against COVID-19. Over 6,700 respondents were asked whether they would get vaccinated if offered varying incentive amounts—from about \$5 to \$50—or no incentive at all. Surprisingly, the results showed that financial incentives reduced vaccination intent by about 19% overall compared to the control group. The surveys also captured public opinion on the use of cash incentives. In Uzbekistan, 75% of respondents opposed them, while 58% did in Kazakhstan. Only in Tajikistan did a majority support the idea. These findings highlight the importance of context when designing behavioral policies. Cultural norms and public perceptions matter—and in some cases, incentives may backfire. To avoid unintended consequences, such programs should be carefully tested before being scaled up in new settings.

Communication modes in COVID-19. In the early months of the COVID-19 pandemic, governments worked urgently to communicate health guidance to the public. In Tajikistan, a nationwide text-messaging campaign was launched in May 2020 by the Ministry of Health and Social Protection, reaching about 5.5 million mobile subscribers—over 90% of households. Using data from the *Listening to Tajikistan* survey, a study by Seitz (2021) evaluated the impact of these messages. The results showed that people who received official text messages were more likely to report following key safety measures, such as wearing masks, limiting social visits, reducing travel, using safer greetings, and improving workplace safety. The findings suggest that text messaging was a cost-effective way to raise awareness and encourage protective behaviors, particularly in a context where reaching a dispersed population quickly was critical.

Conclusion

This chapter presents key insights from the *Listening Surveys*, a series of high-frequency phone panel surveys designed to monitor changes in welfare, behavior, and public opinion in near real-time. As the examples throughout this chapter illustrate, these surveys can be a powerful tool for understanding how households experience and respond to policy reforms and shocks—filling critical data gaps when timely information is most needed.

A central feature that underpins the quality of the Listening surveys is their foundation in a representative face-to-face survey. This approach—widely recognized as the gold standard—ensures a well-designed sample frame, allows for baseline calibration, and strengthens the overall representativeness of subsequent phone rounds. This design also facilitates integration of robust baseline welfare indicators with questions on perceptions, behaviors, and expectations, offering a richer understanding of evolving conditions and how best to respond to them through policy.

Maintaining data quality in high-frequency phone surveys requires careful attention to implementation. Strong supervision, regular enumerator training, and real-time monitoring are essential to ensure consistency, reduce errors, and build trust with respondents. Equally important is managing panel attrition. The experience from Listening surveys shows that the right frequency, light questionnaires, modest incentives, combined with transparent communication about the public value of participation, help sustain high response rates and ensure continuity in data collection.

However, these benefits are not automatic. The Listening model may not be the right fit for every context. In settings where a representative baseline is not available, where phone connections are unreliable, where attrition is too high, or where policy needs are less time-sensitive, it may be more appropriate to consider alternative approaches—such as one-off phone surveys, lower-frequency panel designs, or more targeted monitoring tools.

Experimentation is key. Adapting the survey design to local capacities, infrastructure, and policy priorities will help ensure that Listening surveys remain relevant, cost-effective, and impactful. As countries continue to confront a growing range of information needs, the value of real-time, citizen-centered data systems like Listening surveys will only increase. But their success depends on context-sensitive design, sustained investment in quality, and a clear link to decision-making.

References

- Abdelazeem, Basel, Aboalmagd Hamdallah, Marwa Abdelazim Rizk, Kirellos Said Abbas, Nahla Ahmed El-Shahat, Nouraldeem Manasrah, Mostafa Reda Mostafa, and Mostafa Eltooby. 2023. "Does Usage of Monetary Incentive Impact the Involvement in Surveys? A Systematic Review and Meta-Analysis of 46 Randomized Controlled Trials." Edited by Zhifeng Gao. *PLOS ONE* 18 (1): e0279128. <https://doi.org/10.1371/journal.pone.0279128>.
- Alesina, Alberto F., and Allan Drazen. 1989. "Why Are Stabilizations Delayed?" Unpublished manuscript.
- Ballivian, A., J. Azevedo, W. Durbin, J. Rios, J. Godoy, and C. Borisova. 2015. "Using Mobile Phones for High-Frequency Data Collection." In *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*, edited by D. Toninelli, R. Pinter, and P. de Pedraza. London: Ubiquity Press.
- Bell, Martin, and Elin Charles-Edwards. 2013. "Cross-National Comparisons of Internal Migration: An Update of Global Patterns and Trends." Unpublished manuscript.
- Conn, Katharine M., Cecilia Hyunjung Mo, and Laura M. Sellers. 2019. "When Less Is More in Boosting Survey Response Rates." *Social Science Quarterly* 100 (4): 1445–1458.
- Croke, Kevin, Andrew Dabalen, Gabriel Demombynes, Marcelo Giugale, and Johannes Hoogeveen. 2012. "Collecting High Frequency Panel Data in Africa Using Mobile Phone Interviews." *Policy Research Working Paper* No. 6097. Washington, DC: World Bank.
- Demombynes, Gabriel, Paul Gubbins, and Alessandro Romeo. 2013. "Challenges and Opportunities of Mobile Phone-Based Data Collection." *Policy Research Working Paper* No. 6321. Washington, DC: World Bank.
- Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. 1998. "The Impact of Attrition in the Panel Study of Income Dynamics on Intergenerational Analysis." *Journal of Human Resources* 33: 300–344.

- Heath, Rachel, Ghazala Mansuri, Bob Rijkers, William Seitz, and Dhiraj Sharma. 2021. "Measuring Employment: Experimental Evidence from Urban Ghana." *The World Bank Economic Review* 35 (3): 635–651. <https://doi.org/10.1093/wber/lhaa014>.
- Kropf, Martha E., and Johnny Blair. 2005. "Eliciting Survey Cooperation: Incentives, Self-Interest, and Norms of Cooperation." *Evaluation Review* 29 (6): 559–575. <https://doi.org/10.1177/0193841X05278770>.
- Seitz, William. 2023. *Preferences for Wage Discrimination against Women*. Washington, DC: World Bank. <https://doi.org/10.1596/1813-9450-10548>.
- Seitz, William Hutchins. 2020. "Free Movement and Affordable Housing: Public Preferences for Reform in Uzbekistan." *Policy Research Working Paper* No. 9107. Washington, DC: World Bank.
- Seitz, William Hutchins, Yuya Kudo, and Joao Pedro Wagner De Azevedo. 2023. *Blackout or Blanked Out? Monitoring the Quality of Electricity Service in Developing Countries*. *Policy Research Working Paper* No. 10423. Washington, DC: World Bank. <http://documents.worldbank.org/curated/en/099543304252337204>.
- Seitz, William, and Alisher Rajabov. 2021. "Crisis and Recovery: Economic and Social Monitoring from Listening to Tajikistan." Washington, DC: World Bank.
- Seitz, William, and Metin Nebiler. 2022. "Listening to Kazakhstan: Update on the Social and Economic Wellbeing for August 2022." Washington, DC: World Bank.
- Singer, Eleanor, and Cong Ye. 2013. "The Use and Effects of Incentives in Surveys." *The ANNALS of the American Academy of Political and Social Science* 645 (1): 112–141.
- Wolter, Kirk, Sadeq Chowdhury, and Jenny Kelly. 2009. "Design, Conduct, and Analysis of Random-Digit Dialing Surveys." In *Handbook of Statistics*, Vol. 29, 125–154. Elsevier. [https://doi.org/10.1016/S0169-7161\(08\)00007-2](https://doi.org/10.1016/S0169-7161(08)00007-2).
- World Bank. 2021. *Crisis and Recovery in Uzbekistan: Economic and Social Impacts of COVID-19*. Washington, DC: World Bank.

4.

Using Geospatial Data and Modeling to Assess the Impacts of a Flood: *An Application for Pakistan*

— Erwin Knippenberg,³¹ Walker Kosmidou-Bradley,³²
and Moritz Meyer³³

Introduction

A common application of real-time monitoring is to assess the impact of shocks—such as natural disasters—on poverty and welfare. As rising temperatures and more frequent natural disasters increase the urgency of understanding these effects, there is growing demand for timely estimates of how such shocks influence poverty rates. These estimates are critical to inform rapid policy responses, such as the deployment of emergency cash transfers.

The focus of this chapter is on nowcasting and real-time monitoring, which differ from much of the traditional climate-damage literature. For example, many climate-damage models aim to forecast long-term economic impacts of climate change under different scenarios (see Dell, Jones, and Olken, 2014; Auffhammer, 2018). Another common method in this literature involves using fixed-effects panel models to identify the causal impact of climate shocks on various outcomes—an approach that requires ex-post panel data and is not applicable for real-time monitoring.

³¹ World Bank, corresponding author, eknippenberg@worldbank.org.

³² World Bank.

³³ World Bank.

In contrast, real-time monitoring often relies on vulnerability functions to estimate the immediate effects of shocks on poverty and welfare. These functions assign probabilities of being affected to individuals or households based on their characteristics—a concept referred to as vulnerability (Doan et al., 2023). Vulnerability reflects a household’s ability to anticipate, absorb, adapt to, and recover from shocks (Hill and Porter, 2017), and is often derived from baseline survey data using variables such as asset ownership, employment sector, or income level. When combined with data on exposure to a specific shock—typically based on geographic location—this information can be used to simulate the impacts on household welfare.

The intensity of the impact is determined by calibrated parameters, which are based on observed effects from similar past shocks or expert judgment. These parameters are used to simulate changes in household consumption or expenditure from a pre-shock baseline.

This chapter presents a concrete example of this approach is the rapid assessment conducted after the 2022 floods in Pakistan.

The monsoon floods in Pakistan in 2022 severely disrupted the lives and livelihoods of millions, particularly in the provinces of Balochistan and Sindh. The poorest and most vulnerable were disproportionately affected because many in these populations live in flood-prone areas. In the wake of the floods, a rapid assessment was undertaken to inform the humanitarian response and the allocation of resources for recovery. However, estimating the impacts of the flood on household welfare presented a challenge because of the short timeline and the shortage of available data. Through a combination of geospatial data and modeling, the World Bank’s Poverty and Equity team in Pakistan was able to produce early estimates within only two weeks. The results indicated that the floods may have pushed up to 9.1 million people into poverty, representing an increase of 4 percentage points in the poverty rate. Not only were these estimates helpful in informing the post-disaster policy response, but they also played a key role in the government’s advocacy for loss and damages at the 2022 United Nations Climate Change Conference.

This chapter outlines the approach used in Pakistan to estimate the impacts of the 2022 floods on poverty. The analysis accounted for multiple factors affecting household welfare, including (1) household location relative to the flood exposure; (2) the extent of damage to homes, land, and livestock in districts; and (3) household characteristics that influence vulnerability to flooding.

To assess the impact, satellite flood exposure data were integrated with the Household Income and Expenditure Survey 2018–19, the official source for poverty estimation (PBS 2020). This combined dataset confirms that the most highly affected districts were poorer, with an average poverty rate of 31.4 percent compared with the national average rate of 21.9 percent.

The model also accounts for how loss is experienced by different types of households: (1) agricultural households suffered income losses because farmland was submerged, (2) households in mudbrick homes faced a higher risk of structural collapse, and (3) livestock-owning households lost animals, a critical source of income.

To quantify these losses, a damage function was applied to simulate the economic shock on affected households, reducing their estimated consumption levels. Additionally, the model incorporated supply chain disruptions and food price inflation using monthly consumer price index data of the Pakistan Bureau of Statistics. Because food represents a significant share of household consumption, these price increases also affected household welfare.

This data integration, combined with the modeling, allowed a rapid estimate of the impact of the floods on poverty, while offering crucial insights to support response and recovery efforts.

The Modeling Approach: Nowcasting with Geospatial Data after the Shock

The literature on the impact of natural disasters, especially floods, on household welfare underscores the multifaceted and enduring consequences on income, poverty, human capital, and overall well-being. As defined by the Intergovernmental Panel on Climate Change (IPCC 2022) and outlined in Doan et al. (2023), three components determine the impact (or risk) of a natural disaster on people: hazard, exposure, and vulnerability. *Hazard* is the potential occurrence of an extreme weather event ex ante. *Exposure* refers to what or who could be affected by the extreme weather event in a particular location. *Vulnerability* is the extent to which those exposed are adversely affected depending on their characteristics, including assets, livelihoods, and socioeconomic characteristics. Studies across various countries consistently highlight the severe challenges to households in the aftermath of such disasters and call for targeted policies and interventions to mitigate these adverse effects and enhance resilience among vulnerable communities.

There are four key approaches to assessing the impacts of natural disasters on household welfare depending on (1) whether the assessment takes place before or after the shock and (2) whether the model uses pre-shock cross sectional data or panel data (Table 4.1 offers a matrix summary of the different approaches). From a methodological perspective, the gold standard for quantifying post-disaster welfare changes involves comparing pre- and post-shock welfare among the same households using panel data. Because, by definition, shocks cannot be anticipated with certainty, Headey and Barrett (2015) call for a series of sentinel sites to be established in areas prone to natural disasters to monitor household welfare trends continuously and quickly establish the effect of the shock. Sentinel sites have been piloted, notably in Malawi, and are gradually being scaled up (Knippenberg, Jensen, and Constanas 2019; refer also to chapter 1, this volume). An alternative is post-shock rapid response surveys, which are necessarily short in duration, difficult to administer to a representative sample of respondents, and logistically complicated, particularly if the disaster has disrupted communications and large numbers of people have been displaced.

TABLE 4.1 Typologies in modeling the effect of natural disasters on household welfare

Data	Before the shock	After the shock
Pre-shock cross-sectional data	Homogenous simulations simulate the effect of a shock ex ante and generates probability distributions for consumption based on historical weather trends overlaid with cross-sectional data (Hill and Porter 2017)	Nowcasting with geospatial data after the event (Knippenberg, Amadio, and Meyer 2024): the case described in this chapter
Panel data	Heterogenous response functions estimate these parameters with panel data to construct response functions and simulate the effects of future climate shocks on various household types (Baquié and Foucault 2023)	Difference in differences based on sentinel data (Headey and Barrett 2015)

Hill and Porter (2017) outline an approach to simulate the effect of a shock ex ante that generates probability distributions for consumption based on historical weather trends overlaid with cross-sectional data. This creates counterfactual shock scenarios and estimates of vulnerability conditional on location and household characteristics. In an extension of this work, Baquié and Foucault (2023)

estimate these parameters with panel data to construct response functions, simulating the effects of future climate shocks among various household types.

In the case of Pakistan, there was a strong interest in estimating the impacts of the flood on household well-being and poverty. For this purpose, a novel nowcasting model was developed to estimate post-shock welfare outcomes based on pre-shock household data and remote-sensing data. Knippenberg, Amadio, and Meyer (2024) describe the methodology in more detail.

The Main Channels of Impact: Loss of Income, Assets, and Purchasing Power

The impact of the floods on household welfare was modeled to account for three distinct channels: (1) the loss of household income because of destroyed harvest, killed livestock, or inactivity of business; (2) the loss of assets, including homes, livestock, productive equipment, and household durables; (3) the shortages of real purchasing power because of rising food prices. Because the intent is to evaluate the short-term effects, the model does not factor in the potential government response through direct cash transfers or humanitarian assistance. Rather, the output of the model may inform the discussions about such a response by quantifying the extent of welfare loss.

The theoretical model draws on the canonical agricultural household model (Singh, Squire, and Strauss 1986), which posits a representative household working in agriculture. This household is both a profit-maximizing firm in that it optimizes its allocation of labor and capital to produce income and optimizes its consumption subject to income. The household's consumption is thus subject to the following:

$$p_m * X_m = p_a(Q - X_a) - p_l(L - F) - V - E, (1.1)$$

where p is price; X , consumption; Q , own production, L , labor input; F , own family labor; V , the (often implicit) cost of rent for the home; and E , outside income; m and a denote market product and the staple, respectively. Production Q is a function of labor input and capital $Q(L, K)$. Here, the assumptions are as follows:

- The flood reduces the capacity for own production (Q)
- External income is unaffected (E)
- The flood washes away the home if the home is constructed of mudbrick (V)

- In the short term, households cannot reallocate labor (L) to nonfarm activities
- Households are price takers (p_m and p_a)

In the empirical model, household welfare is proxied by Y_{it} , which is the total consumption of household i in period t (0 is the baseline, and 1 is the flood scenario), such that

$$\hat{Y}_{i,1} = Y_{i,0} \cdot \prod_s (1 - z_{i,s}) \quad (1.2)$$

with $z_{i,s} = (\beta_s | X_{i,s}, \delta_{s,d})$

Here, the welfare of household i in period 1 is a function of multiple shocks that the household experiences between periods 0 and 1, where $z_{i,s}$ is the percent decrease in household consumption of household i because of a shock through channel s . Similar to the theoretical model, the channels through which households may be affected include livelihoods, buildings, livestock, and inflation.

Specifically, $z_{i,s}$ is a function of β_s , which captures the percent decrease in household consumption because of a shock through channel s . However, a household only experiences this shock if it possesses specific household characteristics $X_{i,s}$, and conditional on exposure to the shock s within the location d of the household, such that

- $X_{i,s}$ characteristics of household i ; $X_{i,s} = 1$ if household i has characteristics that render it vulnerable to shock s (directly observed in the household survey, refer to below)
- $\delta_{s,d}$ exposure to shock s in location d , defined as a binomial distribution $p \in [0, 1]$; the probability p is defined as the percent of the population in a given district that is exposed to the flood

Implementing the Model in Pakistan: A Step-by-Step Guide

The first step is the identification of the input data.

Identifying the Input Data

To inform the microsimulation, the analysis began with an existing dataset of households in the flood-affected areas based on a nationally representative

household income and expenditure survey that had a detailed consumption module. For each individual or household, the dataset included the following:

- Welfare metrics (i.e., consumption, food security, etc.)
- Household sociodemographic variables
- Sources of income
- Ownership of assets
- Individual socioeconomic variables (educational attainment)
- Employment type (sector, formal or informal, self-employed or wage worker)

For the purposes of the model, the key indicators included household consumption Y (in period 0) and household characteristics X (in period 0) based on the representative survey. The following household characteristics $X_{i,s}$ were used to determine whether a particular household i would be affected by a shock through the channel s :

- *Livelihoods*: the extent to which households relied on agricultural income
- *Asset ownership*: whether a household lived in a mudbrick home
- *Consumption decile*: the household's ranking in consumption expenditure to determine the share of food relative to total spending and the effect of food inflation

Calibrating the Damage Parameters

The next step involves calibrating the damage parameters. The damage to households relates to the consumption loss attributable to food. Calibrating these parameters is key to informing the design of the model.

A first best approach is empirical, basing the damage on observed cross-tabulations in the datasets based on the impact channels. For example, in the case of Pakistan, the principal loss to household expenditure is assumed to occur through lost agricultural production. The damage to livelihoods was therefore benchmarked to the ratio of agricultural income to total income (refer to table 4.2).³⁴

³⁴ This is premised on the assumption that because the flood occurred immediately before harvest, households would lose 100 percent of their agricultural production, while their nonagricultural income would remain unaffected. Such assumptions can be fine-tuned to the country context to allow for only partial yield loss or reduced nonagricultural income.

TABLE 4.2 Damage to livelihood parameters, by occupation

Livelihoods	Share of households (Pakistan, rural), %	Proportion of income from agriculture, %
Self-employed: non-agriculture	16	3.0
Paid employee: construction	13	4.0
Paid employee: manufacturing	7	2.0
Paid employee: services	19	2.7
Paid employee: agriculture	10	76.0
Own cultivator	22	63.0
Sharecropper	7	72.0
Contract cultivator	3	66.0
Livestock owner	4	40.0

Note: The results are based on the ratio of income from agriculture to total income. This explicitly captures diversified households that rely on a combination of agricultural and nonagricultural income.

For asset losses, the parameter estimates are based on the estimated opportunity cost of these assets. For example, households living in a mudbrick home would have to pay the rental equivalent.³⁵

Disruptions to value chains were captured through local or national price increases, which are often reported on a monthly basis and can be made available shortly after a shock occurs. The effects of inflation were distributed to households in accordance with the share of food consumption per decile, reflecting that poorer households spend proportionally more on food.³⁶

These estimates were cross-validated against estimated damage functions in the literature to ensure external validity. For example, refer to Chen et al. (2017) on the impact of floods on yields in rural China.

³⁵ Parameters are based on the average household expenditure on rent or rent-equivalent corresponding to the value of the home using a hedonic model imputation.

³⁶ It was assumed that there was no substitution and that 50 percent of inflation could be attributed to the floods.

Consultations were conducted with sectoral experts, particularly in agriculture and disaster risk management, who could provide localized insights on damage functions. Recall that damages across channels are cumulative.

Adding Information on Household Exposure

The next step involved adding information on exposure to the 2022 floods.

Once the potential damage is defined, the household's exposure, e_d , defines the probability that the flood impact transmission channels affect a household in location d . Exposure to each transmission channel is modeled as a uniform distribution independent of other channels.

To determine the exposure of a household to a transmission channel, the location of households identified in the national household survey was combined with the estimated exposure based on the location. If household coordinates are available, these can be overlaid directly on exposure maps. If the household data do not include geospatial coordinates, shock exposure can be aggregated at the administrative unit 2 (district) or administrative unit 3 (tehsil, or subdistrict) level and merged into the household data based on reported location to determine the probability that a given household was exposed to the shock.³⁷

Geospatial data were used to estimate the extent of the exposed population, the built-up and agricultural land, and production (crops and livestock) by location to construct e_d . An advantage of geospatial data is ready availability and high-resolution coverage over affected areas. Moreover, they are often freely available to humanitarian and development actors in the aftermath of an emergency. They can then be overlaid with other geospatial data to obtain exposure estimates. The exposure estimates used in Pakistan included the following:

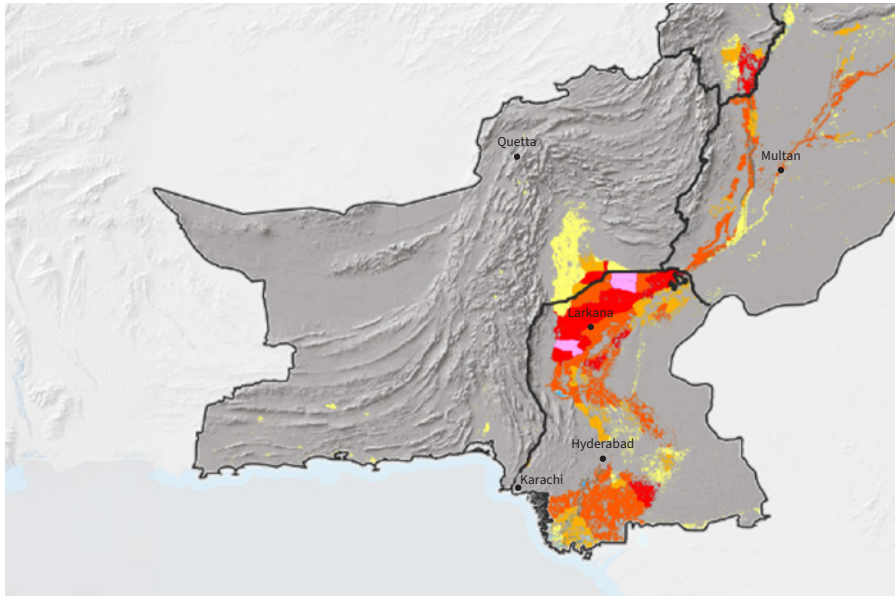
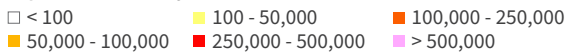
- *Population*: percentage of the population exposed (refer to map 4.1)
- *Agricultural land*: percentage of total crops lost
- *Built-up land*: percentage of houses that were kutcha houses (dwellings made of natural materials such as mud, bamboo, and grass) and that were reported destroyed or damaged

³⁷ Because exposure e_d is based on a probability function, the model was iterated 500 times to allow for variation whereby households are affected by the flood in each location. When aggregated, this produces a range of plausible welfare outcomes, allowing the construction of confidence intervals.

MAP 4.1 Establishing the probability: population exposed to flood

Population (Worldpop 2020)

exposed to flood by ADM3 unit (thesil)



Source: Knippenberg, Amadio, and Meyer 2024.

For the analysis, households were considered flood-affected if they were adjacent to or surrounded by standing water.³⁸ The flood extent was identified using remote-sensing data from multiple sources, including the Visible Infrared Imaging Radiometer Suite (VIIRS), Sentinel-1, and Sentinel-2 data.

- VIIRS offers a 300 meter resolution layer
- Sentinel-1 offers imagery at 10 meter spatial resolution using synthetic aperture radar, which delivers data day and night and despite clouds (Torres et al. 2012)
- Sentinel-2 offers multispectral imagery at 10 meter (visible and near-infrared), 20 meter (red-edge and shortwave infrared), and 60 meter (atmospheric correction) resolution³⁹

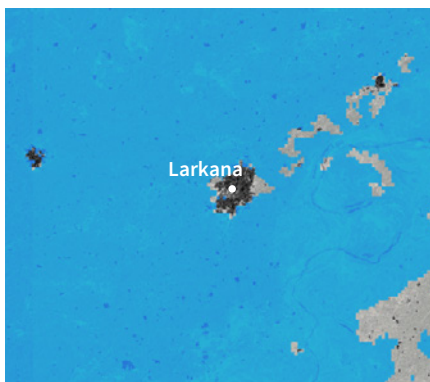
³⁸ The proposed model does not capture damage caused by flash floods, nor does it account for water depth as an input into the damage function. Because these elements are contingent on the availability of appropriate geospatial data, they are included among the proposed refinements for future iterations of the approach.

³⁹ Refer to Sentinel-2 Facts and Figures (dashboard), European Space Agency, Paris, https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Facts_and_figures.

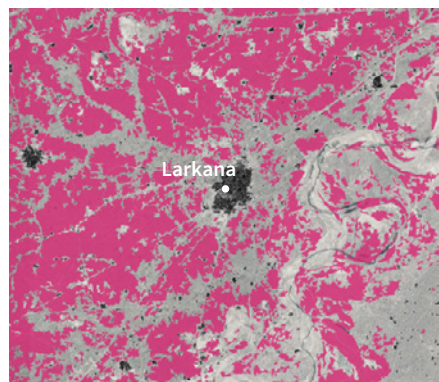
However, the representativeness of the high-resolution data suffers from limited water detection in urban areas. For the analysis, two different versions of the Sentinel data were considered, first, a version of the United Nations Institute for Training and Research (UNITAR) and, second, a version of the Ipsos Group, a multinational market research and consulting firm. The data layer provided by UNITAR has been upscaled and simplified, and urban gaps have been filled (refer to map 4.2). As a result, estimates based on the UNITAR layer show that, during the floods in Pakistan in 2022, 17.4 million people were affected, which is close to the official estimates provided by the National Disaster Management Authority. In contrast, the analysis conducted by overlaying the Ipsos layers against population density undercounts the number of people directly affected, 2.4 million in the entire country.

MAP 4.2 Differences in the extent of floods: UNITAR and Ipsos

a. UNITAR data layer estimates of flood extent



b. Ipsos data layer estimates of flood extent



Source: UNITAR and Ipsos.

The UNITAR flood extent was selected to measure exposure in the analysis. The numbers therefore refer to the total amount of value located within the flood extent for each exposed category. While a better characterization of flood intensity parameters would improve the exposure estimate and allow a calculation of direct physical impacts, the purpose of this exercise is to evaluate data that approach real-time availability after a disaster strikes. The modeling of water hazard intensity parameters (for instance, water depth, velocity, duration of submersion, sediment load) is not usually available immediately after an event. The sources of the geospatial estimates are summarized in table 4.3 and in annex 4A. Box 4.1 shows some relevant adaptations to the floods in Ghana in 2023.

TABLE 4.3 Input data used to produce geospatial flood exposure estimates, Pakistan 2022

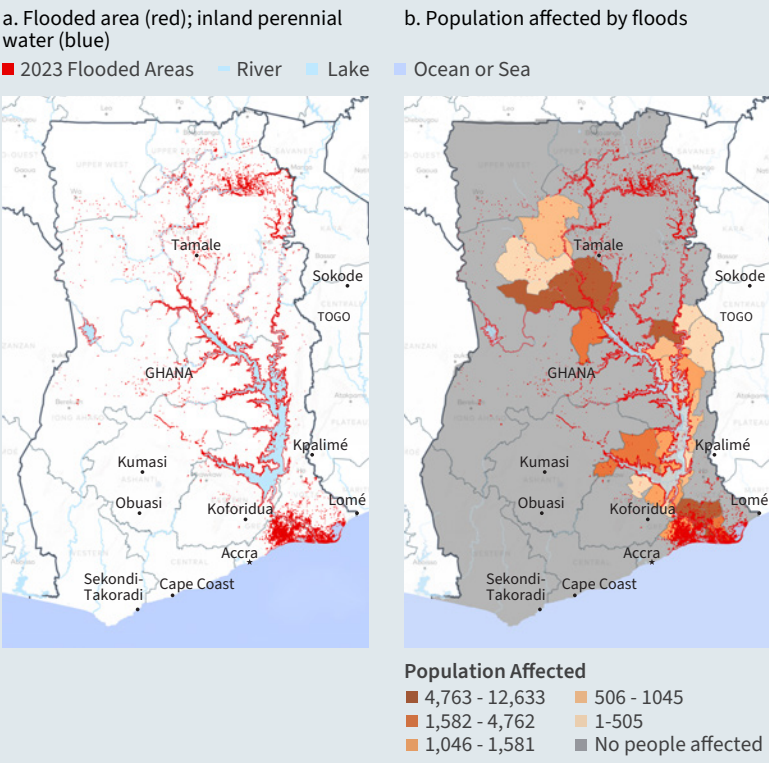
Item	Description
Flood extent July 12–August 29 (maximum)	The sum of individual extent layers obtained from UNITAR. It includes extents of individual events during the period as detected by Sentinel-1, Sentinel-2, and the Visible Infrared Imaging Radiometer Suite. This is used to estimate the total flood extent.
Population layer	Global Human Settlement Layer 2020: The distribution of the residential population is expressed as the number of people per 100 m2 cell. Residential population estimates—derived from Gridded Population of the World, version 4.11, of the Center for International Earth Science Information Network—are disaggregated from census or administrative units to grid cells, informed by the distribution, volume, and classification of built-up land as mapped in the Global Human Settlement Layer (2020) per corresponding epoch. Refer to GHSL (Global Human Settlement Layer) (dashboard), Joint Research Center, European Commission, Brussels, https://human-settlement.emergency.copernicus.eu/ .
Built-up land	World Settlement Footprint (WSF) 2019: The ratio of built-up (buildings) present in a cell is used to estimate the built-up area within the flood extent. Refer to World Settlement Footprint (WSF) 2019 (dashboard), EOC Geoservice, Earth Observation Center, German Aerospace Center, Cologne, Germany, https://geoservice.dlr.de/web/datasets/wsf_2019 .
Agricultural land	ESA WorldCover 2020 global land cover filtered for agricultural classes. Used to estimate the area of agricultural land located within the flood extent. Refer to ESA WorldCover 2020 (dashboard), European Space Agency, Paris, https://worldcover2020.esa.int/ .
Crop production	GAEZ, vC4, 2015: Total annual production (1,000 tons) for 26 crop types. Year 2015 (2014–16 average). Used to estimate agricultural production exposed to the flood. Refer to GAEZ Data Portal, Global Agro-Ecological Zones, Food and Agriculture Organization of the United Nations, Rome, https://gaez.fao.org/ .
Administrative boundaries	ADM0 (country), ADM1 (province), ADM2 (district), and ADM3 (tehsil, or subdistrict) aggregate exposure and impact estimates (OCHA 2022).

Box 4.1 Adaptations to Monitor Floods in Ghana, 2023

To estimate the flood extent in one region of Ghana, the team relied on satellite imagery in the Digital Earth Africa sandbox environment.^a The situation was complicated because fluvial (river-based) and pluvial (rain) floods occurred simultaneously. Continuous, extensive cloud cover precluded the use of Sentinel-2, Landsat, and other, traditional earth observation satellites. So, the team used Sentinel-1, a radar satellite constellation that has day-night and cloud penetration capabilities to identify water extents.^b

The exact flood times and locations were unknown in Ghana. This was a major difference between the Ghana and Pakistan experiences. The team therefore needed to devise a way to extract water coverage in large swaths of the country and, using the imagery signature of water, determine pre-flood, flood, and post-flood periods (refer to map B4.1.1).

MAP B4.1.1 Flood maps and affected populations, Ghana, 2023



While the focus had previously been on southeast Ghana, there was significant unreported flooding across the north (panel a). Overlaying the flood extent data with population data facilitated estimates of the number of people directly impacted (panel b). These flood layers could then be interacted with specific objects of interest, such as schools, health clinics, power infrastructure, roads networks, and so on. If flood depth estimates are required, the flood extent raster can be matched with Fathom3 flood predictions.^c Scaling efforts are under way to make this one-click run across countries in Africa.

a. DE Africa (Digital Earth Africa) (dashboard), Research Institute for Innovation and Sustainability, Johannesburg, <https://www.digitalearthafrika.org/>.

b. Water Detection with Sentinel-1 (dashboard), Real World Examples, Analysis Sandbox, Research Institute for Innovation and Sustainability, Johannesburg, https://docs.digitalearthafrika.org/en/latest/sandbox/notebooks/Real_world_examples/Radar_water_detection.html.

c. Global Flood Map (portal), Fathom, Bristol, UK, <https://www.fathom.global/product/global-flood-map/>.

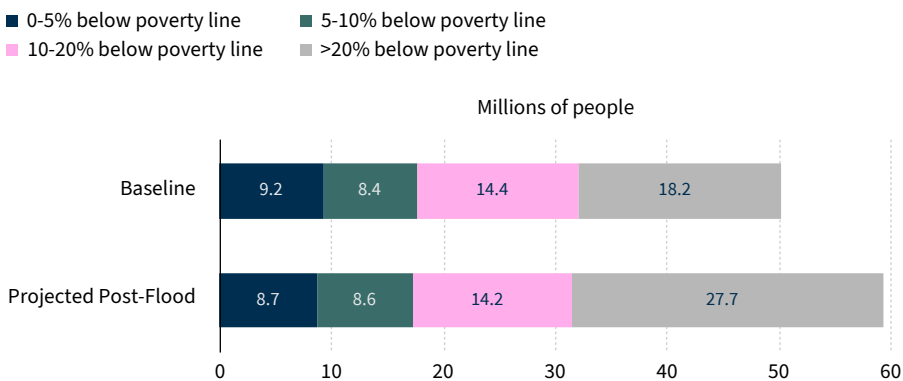
Estimating Impacts across Various Groups

Because the analysis is conducted using microdata from a national household survey, it allows for the estimation of national estimates of the welfare loss, as well as disaggregated estimates for different subgroups of the population. The results can thus be disaggregated by the sociodemographic and socioeconomic characteristics of individuals, households, and regions.⁴⁰

This microsimulation approach can capture the increase in the poverty headcount, but also the increase in poverty depth and therefore the total household welfare loss (refer to figure 4.1). These estimates can help quantify the amount of immediate relief households need to maintain consumption, thereby informing the design and magnitude of shock-responsive cash transfers; refer to the illustrative results of the study by Knippenberg, Amadio, and Meyer (2024). This does not factor in the cost of replenishing lost assets or the long-term damage to human capital.

⁴⁰ For Pakistan, weights from the microdata were updated to account for population growth. The model assumes that the population increased from 207 million to 228 million between 2018 and 2022, and accordingly, population weights are updated using historical provincial population growth rates.

FIGURE 4.1 Increase in the depth of poverty among affected households



Source: Knippenberg, Amadio, and Meyer 2024.

Robustness Checks

Cross-Validation Using Administrative Data

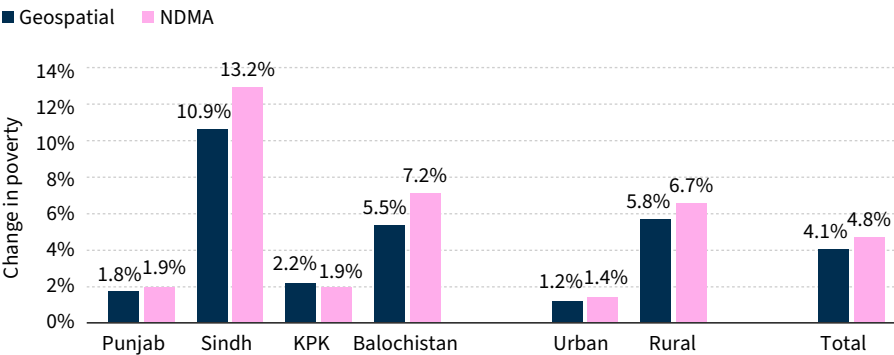
The model was validated using reports from the National Disaster Management Authority. To refine the analysis, official estimates of the affected population, buildings, land, and livestock were used in place of the geospatial estimates, and the model was rerun.

The main specification of the model relied on satellite data to calibrate the exposure parameters because this was the exposure data available immediately after the flood. As a robustness check, the same calculations were run using an alternative exposure metric, as reported by national authorities. This administrative data became gradually available in the weeks after the disaster and required extensive cleaning. Comparing exposure estimates across the two models provided an alternative estimate for the extent of exposure that can be fed into the model. This provides a cross-validation of the initial rapid geospatial estimates.

Figure 4.2 summarizes the inputs used to calculate the results, which are based on satellite data and administrative reports. These can sometimes differ substantially. Geospatial population exposure estimates are lower for Balochistan and Khyber Pakhtunkhwa because the geospatial estimates use standing water and do not account for damage from flash floods. The lower livestock estimates of the national authorities suggest that either the data are incomplete or many owners

were able relocate their livestock to higher ground. This highlights how the two data sources offer potentially complementary insights into the effect of flood exposure. While geospatial data supply a rapid post-shock estimate, administrative data provide a more nuanced ground-truth snapshot of the reality.

FIGURE 4.2 Simulated changes in poverty rates: administrative data and geospatial exposure



Source: Knippenberg, Amadio, and Meyer 2024.
Note: KPK = Khyber Pakhtunkhwa. NDMA = National Disaster Management Authority.

Varying Damage Parameters

The model combined the exposure estimates with damage parameters calibrated to the household data. While these are rooted in empirical data, these parameters were varied, as a robustness check, by + 5 and 10 percentage points. The impact on poverty varied monotonically with the magnitude of the parameter estimates. Yet, the model remains quite robust, with a lower-bound increase of 3.9 percentage points and an upper-bound increase of 4.6 percentage point (refer to table 4.4).

TABLE 4.4 Change in poverty estimates by varying shock parameters, %

Location	-10 percent	-5 percent	Original parameters	+5 percent	+10 percent
Punjab	1.7	1.8	1.8	1.8	1.9
Sindh	10.1	10.4	10.9	11.7	12.4
Khyber Pakhtunkhwa	2.0	2.1	2.2	2.2	2.3

Location	-10 percent	-5 percent	Original parameters	+5 percent	+10 percent
Balochistan	5.1	5.4	5.5	5.8	5.8
Urban	1.2	1.2	1.2	1.2	1.2
Rural	5.5	5.6	5.8	6.2	6.5
Total	3.9	4.0	4.1	4.4	4.6

Source: World Bank calculations.

Buffering Effects from Assets

Assets help a household buffer the effects of shocks. This was modeled explicitly by investigating the damage to housing, often the biggest asset of a rural household. In addition, about 10 percent of household income is drawn from agricultural and non-agricultural assets. If it is assumed that households can draw on at least as much by borrowing or drawing down liquid assets, then this may be added back in as a buffer to the consumption loss. The results suggest that cashing in income from assets would only have a small buffering effect, reducing the change in poverty by 0.3 percentage points (refer to table 4.5). In addition, many of these assets, such as land and livestock, are illiquid and may have lost their value because of the floods. Households with financial assets may also have to pay steep fees in the form of interest rates or penalties to access these assets. Privately held assets are therefore not sufficient as a buffer.

TABLE 4.5 Allowing households to front-load income from assets, %

Location	Baseline	Cashing in assets
Sindh	10.9	9.7
Khyber Pakhtunkhwa	2.2	1.9
Balochistan	5.5	5.2
Urban	1.2	1.1
Rural	5.8	5.3
Total	4.1	3.8

Source: World Bank calculations.

Caveats and Lessons Learned

The approach used in Pakistan focused on the immediate ex post impact of floods on household welfare. It offers a ready-made approach to quantifying welfare impacts in a post-disaster setting, given the limited data availability and time constraints. However, the model, the calibration, and the findings should be considered within the objective of this exercise, which is associated with certain limitations.

First, the estimates focus on the direct consequences of the flood on poverty and do not account for second-order effects, including the implications for household growth and productive capacity. They also do not account for the potential response in the form of humanitarian assistance, shock-responsive cash transfers, or insurance payouts. Instead, these first-order estimates are intended to inform the design of the response.

Second, the estimates focus on the short-term impact. The impact may vary depending on the location and design of relief and reconstruction. Even in the best case, reversing the adverse shocks to household welfare would take considerable time and some losses, such as losses in human capital and land productivity, could set in motion more durable declines in welfare. The impact of the natural disaster is also a function of the duration of the flooding. Because this analysis relies on an ex ante simulation approach, the findings do not delve into the duration of the inundation or, related to this, the duration of the limited access to services and markets.

Third, the model outcomes depend on the validity of pre-disaster household survey data. Because of natural disasters, household characteristics may change, which could undermine the reliability of preexisting household surveys. Moreover, the results of the simulation could become biased if the model relies on information from an outdated household survey. In the case of Pakistan, the data collection for the Household Integrated Economic Survey occurred in 2018/19, before the COVID-19 pandemic hit the country.

Fourth, the findings from the ex ante impact simulations are partially a function of the calibrated damages to livelihood parameters drawn from the literature and expert opinion. To the extent possible, these parameters should be validated against empirical data to ensure they align with the actual loss of livelihoods a household experiences because of a shock.

Fifth, the rapid turnaround in the case of Pakistan was only possible because many elements were already in place, most notably, preexisting household data. To be able to access and analyze geospatial data rapidly, the team invested in geospatial capacity ahead of time as part of ongoing welfare monitoring activities. This included identifying key sources of updated geospatial data, establishing protocols and scripts for accessing the data and overlaying with household data, and possessing the in-house capacity for geospatial data analytics. These basic investments in foundational data and capacity paid dividends when the disaster struck.

This work underscores the need to strengthen the ability to align poverty and other welfare measures using geospatial, administrative, and survey data. It also highlights the critical role of regularly collected household survey data—both before and after shocks—in identifying the risks that households and communities face as the climate changes. Such data can inform strategies to help households and communities adapt and build resilience. When a shock occurs, this analysis becomes even more valuable, enabling a rapid assessment of impacts and the more effective deployment of support.

Annex

Geospatial Data Sources

TABLE A1.1 Population

Name	Global Human Settlement Layer [GHS-POP]
Source	European Commission Joint Research Center
Format	Raster grid
Resolution	100 meters
Time reference	2020
Metric	Population count
License	Open
Notes	Constrained by built-up area (2020)

Source: GHSL (Global Human Settlement Layer) (dashboard), Joint Research Center, European Commission, Brussels, <https://human-settlement.emergency.copernicus.eu/>.

TABLE A1.2 Land cover, land use, and built-up area

Name	ESA WorldCover	World Settlement Footprint	Global Human Settlement Layer [GHS-BUILT-S]
Developer	European Space Agency	German Aerospace Center	European Commission Joint Research Center
Format	Raster	Raster grid	Raster grid
Resolution	10 m	10 m	100 m
Time reference	2020	2019	2020
Metric	Land cover classes	Presence of built-up (binary)	Built-up area
License	Open	Open	Open

Sources: ESA WorldCover 2020 (dashboard), European Space Agency, Paris, <https://worldcover2020.esa.int/>; GHSL (Global Human Settlement Layer) (dashboard), Joint Research Center, European Commission, Brussels, <https://human-settlement.emergency.copernicus.eu/>; World Settlement Footprint (WSF) 2019 (dashboard), EOC Geoservice, Earth Observation Center, German Aerospace Center, Cologne, Germany, https://geoservice.dlr.de/web/datasets/wsf_2019.

TABLE A1.3 Crops and livestock

Name	Gridded Livestock of the World (GLW)	Global Agro-Ecological Zones (GAEZ+)
Developer	Food and Agriculture Organization of the United Nations	Food and Agriculture Organization of the United Nations
Format	Raster	Raster grid
Resolution	10 kilometers	10 kilometers
Time reference	2015	2015
Metric	The layers contain the density of animals per pixel, with weight estimated using the random forest model. The livestock species modeled include buffalo, cattle, chickens, ducks, goats, horses, pigs, and sheep.	Crop harvest area, crop production, and crop yield maps (Fischer et al. 2012), using national-scale data on the fractional change in crop harvested area and production in 2010–15, based on statistics for 160 crops of FAOSTAT.
License	Open	Open

Sources: FAOSTAT (dashboard), Food and Agriculture Organization of the United Nations, Rome, <https://www.fao.org/faostat/en/#home>; GAEZ Data Portal, Global Agro-Ecological Zones, Food and Agriculture Organization of the United Nations, Rome, <https://gaez.fao.org/>; GLW (Gridded Livestock of the World) (dashboard), Food and Agriculture Organization of the United Nations, Rome, <https://www.fao.org/land-water/land/land-governance/land-resources-planning-toolbox/category/details/en/c/1236449/>.

TABLE A1.4 Administrative boundaries

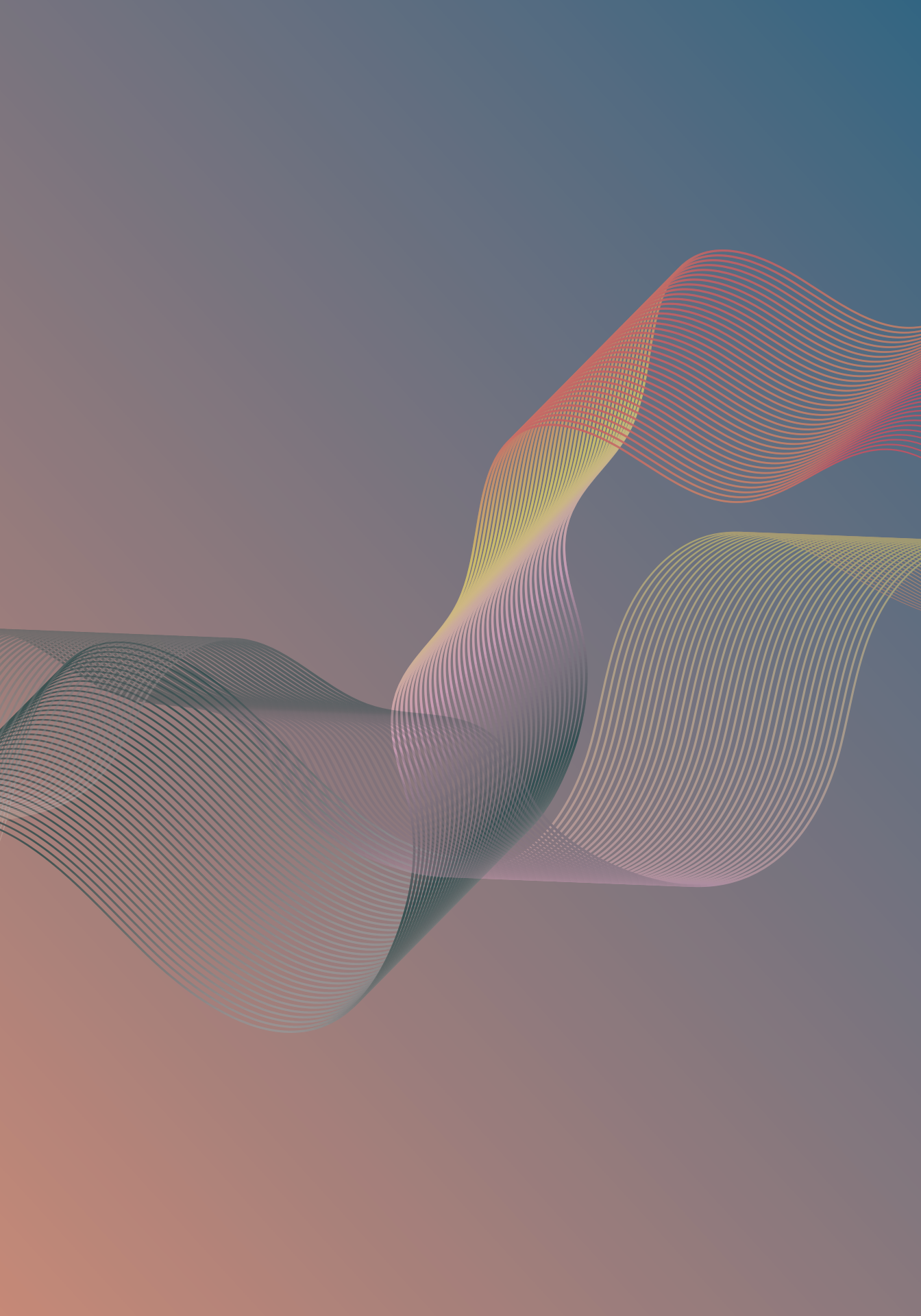
Name	Pakistan administrative boundaries
Source	United Nations Office for the Coordination of Humanitarian Affairs
Format	Nomenclature of Territorial Units for Statistics
Time reference	2022
License	Open

Source: OCHA 2022.

References

- Baquié, Sandra, and Guillem Foucault. 2023. “Background Note on Bringing Climate Change into Vulnerability Analysis.” World Bank, Washington, DC. <http://documents.worldbank.org/curated/en/099719410242336767>.
- Chen, Huili, Zhongyao Liang, Yong Liu, Qihua Liang, and Shuguang Xie. 2017. “Integrated Remote Sensing Imagery and Two-Dimensional Hydraulic Modeling Approach for Impact Evaluation of Flood on Crop Yields.” *Journal of Hydrology* 553 (October): 262–275.
- Doan, Miki Khanh, Ruth Vargas Hill, Stéphane Hallegatte, Paul Andres Corral Rodas, Ben James Brunckhorst, Minh Nguyen, Samuel Freije-Rodríguez, and Esther G. Naikal. 2023. “Counting People Exposed to, Vulnerable to, or at High Risk from Climate Shocks: A Methodology.” *Policy Research Working Paper* No. 10619. Washington, DC: World Bank.
- Fischer, Günther, Freddy O. Nachtergaele, Sylvia Prieler, Edmar Teixeira, Géza Tóth, Harrij van Velthuisen, Luc Verelst, and David Wiberg. 2012. *Global Agro-Ecological Zones (GAEZ v3.0): Model Documentation*. Laxenburg, Austria: International Institute for Applied Systems Analysis; Rome: Food and Agriculture Organization of the United Nations. https://www.gaez.iiasa.ac.at/docs/GAEZ_MD_02.02.2012.pdf.
- Headey, Derek D., and Christopher B. Barrett. 2015. “Measuring Development Resilience in the World’s Poorest Countries.” *PNAS: Proceedings of the National Academy of Sciences* 112 (37): 11423–11425.
- Hill, Ruth Vargas, and Catherine Porter. 2017. “Vulnerability to Drought and Food Price Shocks: Evidence from Ethiopia.” *World Development* 96 (August): 65–77.
- IPCC (Intergovernmental Panel on Climate Change). 2022. *Climate Change 2022: Impacts, Adaptation, and Vulnerability*. Sixth Assessment Report. Geneva: IPCC; New York: Cambridge University Press. <https://www.ipcc.ch/report/ar6/wg2/>.
- Knippenberg, Erwin, Mattia Amadio, and Moritz Meyer. 2024. “Poverty Impacts of the Pakistan Flood 2022.” *Economics of Disasters and Climate Change* 8 (3): 453–471.

- Knippenberg, Erwin, Nathaniel Jensen, and Mark Conostas. 2019. "Quantifying Household Resilience with High-Frequency Data: Temporal Dynamics and Methodological Options." *World Development* 121 (September): 1–15.
- OCHA (United Nations Office for the Coordination of Humanitarian Affairs). 2022. "Revised 2022 Floods Response Plan: Pakistan." October 4. OCHA, Geneva. <https://reliefweb.int/attachments/81b91755-8fdc-4406-8c05-75af7e7a6f73/Pakistan%20Floods%202022%20-%20Floods%20Response%20Plan%20-%20Revision%20-%202004%20Oct%202022.pdf>
- PBS (Pakistan Bureau of Statistics). 2020. "Household Integrated Economic Survey (HIES), 2018–19." June. PBS, Islamabad, Pakistan. https://www.pbs.gov.pk/sites/default/files//pslm/publications/hies2018-19/hies_2018-19_writeup.pdf.
- Singh, Inderjit, Lyn Squire, and John Strauss. 1986. "The Basic Model: Theory, Empirical Results, and Policy Conclusions." In *Agricultural Household Models: Extensions, Applications, and Policy*, edited by Inderjit Singh, Lyn Squire, and John Strauss, 17–47. Washington, DC: World Bank; Baltimore: Johns Hopkins University Press.
- Torres, Ramon, Paul Snoeij, Dirk Geudtner, David Bibby, Malcolm Davidson, Evert Attema, Pierre Potin, et al. 2012. "GMES Sentinel-1 Mission." *Remote Sensing of Environment* 120 (May): 9–24.



5.

Frontier Approaches for Real-Time Poverty Measurement

— Emily Aiken⁴¹ and Joshua Blumenstock⁴²

Introduction

Accurate and up-to-date information on the living conditions of households is indispensable for the design of effective and timely policies and interventions. Reliable data enable policymakers to design targeted social assistance programs, monitor economic progress, respond rapidly to crises, and measure the impact of policies and programs. However, many low- and middle-income countries (LMICs) lack recent data on household living conditions (Jerven 2013). For instance, Yet et al. (2020) estimate that less than half of the poorest countries completed a census between 2010 and 2020; looking further back, Serajuddin et al. (2015) estimate that 57 countries produced one or fewer national poverty estimates between 2002 and 2011.

These data gaps can have severe consequences. Barca and Hebbbar (2020) found that social registries — crucial for identifying and targeting beneficiaries of social assistance programs — are typically updated only every five to eight years in LMICs. This infrequency impairs policymakers’ ability to respond promptly to evolving economic conditions and shocks (Encinas et al. 2025). In a recent analysis of data from six LMICs, Aiken, Ohlenberg, and Blumenstock (2024) show how the gap between social registry updates can have large downstream consequences on targeted interventions: they estimate that the accuracy of social registries decreases, on average, by 9 percentage points per year, which in turn implies an increase in exclusion errors (poor households mistakenly excluded from receiving program benefits) of roughly 2 percentage points per year. For a large national program like Mexico’s Progresa, a five-year interval between PMT recalibrations would imply approximately 200,000 additional exclusion errors, leaving many vulnerable populations without much-needed assistance (Aiken, Ohlenberg, and Blumenstock 2024).

⁴¹Carnegie Mellon University Africa.

⁴²University of California, Berkeley.

Filling these data gaps with traditional surveys would be logistically challenging and extremely costly. For instance, a typical social registry requires first conducting a small household survey, to determine which household characteristics the government should use to determine program eligibility (i.e., the data needed to calibrate the targeting formula); and then conducting a census-scale survey to capture those household characteristics for every potential beneficiary of the program. Barca and Hebbar (2020) estimate that a typical household survey costs \$17 million, and the census-scale population sweep costs \$57 million. Taking a more global perspective, Kilic et al. (2017) estimate that it would cost roughly \$945M to conduct regular (3-year) household surveys in 78 International Development Association (IDA), in the period from 2016-2030. Simply put, these costs are too high for most LMICs to incur regularly.

These persistent data gaps underscore the need for alternative, more cost-effective approaches to measuring household welfare, and for monitoring the socio-economic conditions of local and national populations.

This chapter reviews recent advances in the research literature that illustrate a new paradigm for measuring living conditions. While details vary from study to study, a common feature of this new approach is the use of machine learning and related computational methods to combine traditional datasets – most often household surveys – with non-traditional sources of data, such as remotely sensed data and satellite imagery, data from mobile phone networks, and internet and social media data. The digital data sources are typically abundant, relatively inexpensive to collect, and frequently updated, creating opportunities for near-real-time insights into economic conditions. In principle, this approach has the potential to significantly reduce the cost of data collection, while simultaneously increasing the frequency and granularity with which poverty measurements are available. The hope is that cheaper and more rapid measurements can enable more timely, effective, and efficient policy decisions.

Our primary focus in this chapter is to highlight the potential for this combination of machine learning with digital data to enable extremely high-frequency – and potentially real-time – measurements of well-being.⁴³ However, as we discuss below, considerable work still needs to be done to realize this vision. In particular,

⁴³ Related work discusses the potential for near real-time monitoring using high-frequency surveys and traditional data (e.g., Dang et al, 2025, Yoshida and Aron 2024). Burke et al, 2021 and Newhouse, 2024 provide more focused reviews on the use of satellite imagery for measuring living standards and small area estimation, respectively. McBride et al, (2022) and Bolch, Genoni, and Stemmler (2024) offer broader perspectives on the strengths and limitations of newer approaches relative to traditional methods.

while a great deal of progress has been made on using non-traditional data to measure living standards in the cross-section (i.e., at a single point in time), there are far fewer instances where those data have been shown to accurately capture changes in welfare over time. Likewise, while hundreds of studies now describe innovative methodologies for measuring welfare from digital data (and how those measurements could be used *in principle*), only a handful of studies rigorously document how real-time measurements positively influence policy decisions (and how the measurements are used by policymakers *in practice*).

Thus, the chapter is organized to first provide an overview of the most relevant non-traditional data sources that are being used to measure poverty and welfare in LMICS (second section), highlighting some of the key advantages and disadvantages of each data source. In the third section, we review several of the key results from the research literature that have shown how such data have been used to measure welfare in the cross-section, at a single point in time. The fourth section then highlights several studies that show how real-time data sources can be used to monitor populations; such applications primarily illustrate how the onset of sudden events can be observed in digital traces. In the fifth section, we discuss the small handful of recent papers that explore the use of real-time data for monitoring poverty and welfare: while this is a nascent area of research, early results indicate that accurate estimates of *changes* in welfare over time are considerably harder to predict than *levels* of welfare in a cross-section. The final section concludes with a discussion of what we perceive to be the most promising and important areas for future research.

Novel data sources relevant to real-time poverty measurement

In this section we introduce four nontraditional data sources most relevant to real-time welfare monitoring. Three of these – satellite imagery, mobile phone data, and web data (and social media data) have been used in research studies and/or in practice to measure static or dynamic indices of poverty or other measures of well-being. Financial data and other sources of digitized administrative data have not yet been leveraged to the same extent for measuring and monitoring welfare, but we mention them here as they likely hold substantial potential for such applications (particularly as the quantity and availability of these types of data expands in the future). Note that this chapter does not cover administrative data typically held by governments (such as national identifier databases or social registries); we focus, rather, on nontraditional digital data sources, primarily “digital trace” data that can be repurposed for real-time monitoring of welfare.

Each data source has a unique set of advantages and disadvantages. Thus, after introducing each type of data, we briefly discuss considerations of data access and cost, data representativity, and other key strengths and weaknesses.

Remote sensing data

Data source description: Remote sensing satellites are imaging the entire earth with increasingly fine-grained spatial and temporal resolution. As of 2024, there were approximately 12,000 functioning satellites in orbit,⁴⁴ with around 1,500 focused on earth observation.⁴⁵ Much of the work leveraging remote sensing data for welfare monitoring has focused on visual imagery taken by satellites (during daytime to capture visual cues, or at nighttime to capture nighttime luminosity), but satellites often carry additional sensors, including infrared, radar, and LiDAR. Tatem et al. (2008) provide a useful history of remote sensing satellites, and Donaldson and Storygard (2016) provide an overview of applications of satellite data in the economics field.

Accessibility and cost: There are two main sources of free and publicly available satellite imagery. The Landsat satellites, operated by the US National Atmospheric and Space Agency, provide multispectral imagery at a 15 meters per pixel resolution, with an eight day reimaging cycle. Landsat imagery is available back to 1972. Sentinel satellites, operated by the European Space Agency, provide 10 meter per pixel resolution imagery with a five day repeat cycle since 2014. Landsat imagery is available from NASA,⁴⁶ Sentinel imagery is available from Sentinel Hub,⁴⁷ and both are available through Google Earth Engine,⁴⁸ which is free for nonprofit organizations, academic research, and other noncommercial purposes. Imagery from private remote sensing data providers tend to have higher spatial and temporal resolution: for example, Planet’s PlanetScope imagery has a three meter per pixel resolution, and Maxar’s SkySat imagery has sub-meter resolution. Private remote sensing data providers also provide the option for tasking, assigning satellites to image certain parts of the globe at an appointed time. However, private satellite data providers – and particular tasking – incur substantial costs.⁴⁹ One promising avenue of research is the integration of high resolution and low resolution imagery.

⁴⁴<https://sdup.esoc.esa.int/discosweb/statistics/>, last accessed 08/22/2025

⁴⁵https://www.esa.int/Enabling_Support/Space_Engineering_Technology/Earth_observation_inspires_global_inventiveness, last accessed 08/22/2025

⁴⁶<https://landsat.visibleearth.nasa.gov/>, last accessed 08/22/2025

⁴⁷<https://www.sentinel-hub.com/>, last accessed 08/22/2025

⁴⁸<https://earthengine.google.com/>, last accessed 08/22/2025

⁴⁹<https://cega.berkeley.edu/article/incorporating-remote-sensing-data-into-randomized-evaluations/>

For example, Ayush et al. (2021) propose a reinforcement learning approach to use low resolution imagery to identify areas of uncertainty where it is worth purchasing high resolution imagery for increased accuracy. Another promising option is to train machine learning models using high resolution imagery (where typically the number of “labeled” data points that are used for training are fairly few, so costs are limited), and produce out-of-sample predictions using free low resolution imagery (where typically predictions are produced for a large number of locations, so using high resolution imagery would be very expensive).

Representativity: Relative to the other data sources covered in this chapter, remote sensing data is quite representative, as it covers the entire globe. However, low-income countries tend to be re-imaged at a lower resolution and frequency than high-income countries (particularly by for profit data providers), and certain parts of the globe are disproportionately impacted by issues of cloud cover.

Advantages and disadvantages: The primary advantage of satellite imagery is its comprehensiveness and accessibility, with imagery available for the entire planet. Satellite imagery, particularly high resolution imagery, has also been shown to contain a great deal of useful signal correlated with village- and regional-level welfare: sections three, four, and five of this chapter provide a number of examples where satellite images have been used for real-time welfare monitoring and impact evaluation. While many of these examples focus on monitoring wealth or consumption, satellite imagery is also useful for measuring roof materials, electrification (particularly using nighttime lights imagery), crop types and productivity, road quality, and other measures broadly associated with welfare. Disadvantages of remote sensing data include the high cost associated with purchasing high resolution imagery, and the potentially high computational and capacity requirements to work with such imagery (particularly when the images are processed with convolutional neural networks and other computer vision techniques). Geospatial foundation models (e.g. Rusworm et al. 2023, Kerner et al. 2023) and featurization projects like MOSAICS (Rolf et al. 2021) can improve the accessibility of working with satellite-derived information from a capacity perspective, as they allow analysts to work with information derived from imagery without having to analyze the raw images themselves.

Mobile phone metadata

Data source summary: In this section, we use “mobile phone metadata” to refer to the metadata recorded by mobile network operators when calls, text messages, and other communication transactions are placed on their networks. These data are also commonly referred to as *call detail records* (CDR). Metadata on mobile phone calls

and SMS messages typically contain information on the identity of the caller and recipient (phone numbers, or pseudonymized phone numbers), the time of the call/message, the duration in the case of a call, and the cell towers through which the call or message was placed and received. Cell towers can then be geolocated to provide a rough measure of geography (the extent of geolocalisation enabled through this approach will depend on the density of cell towers in the area). Mobile phone metadata may also contain information about recharges (the time and amount of airtime top-ups) and mobile data usage (upload and download amounts per subscriber).

Accessibility and cost: Relative to the other data sources covered in this chapter, mobile phone metadata is typically one of the most difficult to access data types. Mobile phone metadata are recorded and held by mobile network operators (MNOs), and data sharing will depend on the interest of MNOs in collaboration and the legal frameworks for data sharing in the country in question. Mobile phone metadata have been used in a number of applications related to welfare monitoring, both in research and in practice, and a number of different data sharing models with MNOs have been explored (Milusheva et al. 2021), including direct collaboration with mobile network operators and access through government agencies (often statistical agencies or the mobile communications regulator). UNSTATS has a helpful guide to negotiating data access with MNOs (UNSTATS 2019).

Representativity: Mobile phone penetration is rapidly increasing globally: as of 2023, there were 5.4 billion mobile subscribers in the world (73% mobile phone penetration). However, there are substantial disparities by region in mobile phone ownership in use: in Africa and Asia, only 61% and 67% of individuals own a mobile phone, respectively. In low income countries globally, only 49% of individuals own a mobile phone (GSMA 2023). Moreover, within countries, wealthier, younger, and male people tend to own and use phones at higher rates than the poor, the elderly, and women (Blumenstock 2012, Wesolowski 2012). Mobile phone data are therefore not representative of entire populations, and patterns of use differ systematically across different demographics (Blumenstock et al. 2010). However, mobile phone data tend to be more representative than data sources like web and social media data that require internet access. Recent work has also begun developing statistical techniques for correcting for the representativeness of mobile phone datasets (Blanchard and Rubrichi 2025).

Advantages and disadvantages: The primary advantage of mobile phone metadata is the granular and real-time insight it enables into the behavior of populations. It is particularly useful for geospatial analysis (enabled by cell tower geolocations) and social network analysis (enabled by studying call and text message networks

among subscribers). The primary disadvantages of mobile phone metadata are its limited accessibility (requiring working directly with mobile network operators or regulators) and representativity. Limited accessibility can lead to challenges of transparency: when analysis relies on privately held data, it may not be possible for third parties or public interest groups to verify the validity of the estimates produced. Finally, more so than the other data sources covered in this chapter, estimating welfare from mobile phone data risks manipulation: in settings where people may be incentivized to appear poor (for example, when welfare estimates based on nontraditional data are used as inputs to eligibility decisions for social protection or development aid programs), people may strategically adjust their mobile phone use to attempt to “game” the eligibility threshold (Bjorkegren et al. 2021).

Web and social media data

There is a great deal of variety in types of web and social media data. Some of the most common web and social media data types used in welfare measurement and real-time monitoring are:

- Search query volumes, representing the number of users searching for different terms (such data are free and publicly available on Google Search Trends and Bing Search Trends, among others).
- Social media posts on Twitter, Reddit, Facebook, or other social media sites. While accessing social media posts are facilitated by APIs associated with each site, the cost and accessibility of scraping posts depends on the platform.
- Social media advertising data, provided by social media sites to enable targeted advertising. For example, Google and Facebook both provide information to potential advertising on the size of targeted demographics, providing useful information on the size of different types of populations using the platform, often disaggregated by country, gender, and/or age.
- Website content, derived by scraping HTML pages for text and image content. This includes information on sites like Wikipedia.
- Ground-level and crowdsourced imagery, for example from Google Street View.
- Mobility data from smartphone applications, often provided by data aggregators like SmartGraph or Cubiq.

Accessibility and cost: There is a great deal of heterogeneity in the accessibility and cost of web and social media data sources. While some data sources (like Google Search Trends or Wikipedia articles) are free and publicly available, others are only accessible by purchase or through rate-limited APIs (including most social media APIs and mobility data).

Representativity: Globally, internet access is at 68% and rising.⁵⁰ However, as with mobile phone metadata, access to the internet is lower in low-income regions (for example, 69% in Arab states and 37% in Sub-Saharan Africa.⁵¹ As with mobile phone data, the rich, people with higher education levels, younger people, and men are more likely to have access to the internet.⁵² As a result, these demographics are overrepresented in web and social media data. When it comes to language data on internet and social media sites, English is also substantially overrepresented relative to other languages, particularly languages from Africa and other poor regions (Aji et al. 2022).

Advantages and disadvantages: Web and social media data provide some of the most nuanced insights into human behavior, such as rich language data on social media for sentiment analysis and high quality ground-level imagery from Google Street View. They also have the potential for near real-time access. However, web and social media data vary substantially in their accessibility and cost, and are the least representative of the data sources covered in this chapter.

Financial data and other digitized administrative data

Data source description: With the increasing proliferation of mobile money (World Bank 2024) and digital loan products (OECD 2024), financial services data are an increasingly relevant data source for monitoring welfare. While these data types have not been used in practice to the extent that satellite imagery, mobile phone metadata, and web data have been, they are covered here because of their potential for real-time monitoring of financial well-being. Financial data include records of mobile money transactions, take-up and repayment of digital loans, and bank and credit card transactions. Beyond financial services data, other types of digitized administrative data from governments and private companies may provide unique and domain-specific opportunities for monitoring well-being.

Accessibility: Like mobile phone metadata, digitized financial records are typically privately held by the financial services operator. Access therefore depends on collaboration with such an operator. While such collaborations have been established in a research setting, there are limited examples of using such data in practice.

⁵⁰<https://www.itu.int/en/itu-d/statistics/pages/stat/default.aspx>, last accessed 08/22/2025

⁵¹<https://www.itu.int/en/itu-d/reports/statistics/2023/10/10/ff23-internet-use/>, last accessed 08/22/2025

⁵²<https://www.itu.int/en/mediacentre/Pages/PR-2024-11-27-facts-and-figures.aspx>, last accessed 08/22/2025.

Representativity: Like mobile phone metadata, take-up of financial services products are not uniform across the globe or across populations, so financial services data have limited representativity. Mobile money is perhaps the most ubiquitous and relevant such data source. Certain countries have very high rates of mobile money penetration (for example, 68% in Kenya and 64% in Rwanda according to the 2021 Findex survey, World Bank 2025). However, take-up is not uniformly high in Africa (for example, most Nigeria and Ethiopia have take-up rates between 8% and 4%, World Bank 2025), and there are limited such products outside of the African context.

Advantages and disadvantages: The primary advantage of analyzing mobile money and other financial services data is the direct insight they enable to financial well-being, including ability to save, repay loans, and transfer money to others. However, analysis of these types of data are limited by accessibility and take-up.

Measuring static welfare with novel data

While the focus of this chapter is real-time monitoring of welfare with digital data – that is, tracking changes in welfare over time with novel data sources – this is a fairly nascent topic in research and in policy. Dynamic measurement of welfare builds on a more robust literature on static poverty measurement with novel data sources. This section summarizes key work in the past decade establishing the utility of digital data sources – including remote sensing data, mobile phone data and web data – for measuring poverty and other well-being indices in static settings. We do not review financial services data in this section, as there is limited existing work using financial services data sources to measure well-being.

Satellite imagery

Work using remote sensing data to measure poverty dates back to Elvidge et al. (1997), which showed that country-level GDP measures were correlated with total nighttime luminosity, measured via nighttime satellite images (sometimes referred to as “nighttime lights” images). Later work estimated subnational GDP measures based on nighttime lights (Ebener et al. 2005, Sutton et al. 2007). The most well known work using nighttime lights to estimate poverty is that of Henderson et al. (2012), who revise GDP growth estimates for countries with low statistical capacity based on nighttime lights estimates, and propose an approach for combining traditional and nighttime-lights data for tracking GDP in these settings.

More recently, researchers have complemented or replaced nighttime lights data with daytime satellite imagery. Jean et al. (2016) develop a transfer learning approach in which nighttime lights imagery are used to “pre-train” a deep learning model which is then “fine-tuned” to predict asset-based wealth measures from daytime satellite imagery. The study finds that, in the context of Nigeria, Tanzania, Uganda, and Malawi, the approach leveraging daytime imagery substantially outperforms using nighttime lights alone when benchmarked to traditional survey-based measurements from the Demographic and Health Surveys. Later studies (Yeh et al. 2020, Chi et al. 2022) have expanded the daytime satellite-based poverty prediction paradigm globally, with Chi et al. (2022) producing publicly available wealth estimates for all low-and middle-income countries based on satellite imagery.

While satellite imagery is, to date, the nontraditional data source most used for welfare estimation, it is limited by spatial granularity and resolution. Most real-world deployments of satellite-based poverty estimation have focused on the national, admin-2, or admin-3 levels. However, recent work has shown that at these spatial scales, satellite imagery may not improve over traditional small area estimation techniques for measuring poverty rates (Mahler et al. 2022). Satellite imagery may be more valuable at higher resolutions, such as at the village level (Jean et al. 2016, Yeh et al. 2020) or household level (Watmough et al. 2019, Huang et al. 2021). Other data sources – particularly mobile phone metadata – are also well suited for prediction at the individual and household level.

Mobile phone metadata

After satellite imagery, the second most studied digital data source for poverty and welfare estimation is mobile phone metadata. While early studies showed that aspects of the way subscribers use their mobile phones correlate with socioeconomic status (Eagle et al. 2010, Frias-Martinez et al. 2012, Smith-Clarke et al. 2014), Blumenstock et al. (2015) were the first to build an end-to-end machine learning pipeline to predict individual-level poverty from mobile phone data. In the context of Rwanda, the study found high predictive accuracy for wealth indices at the individual level, community level, and district level. Later work has used mobile phone data, combined with machine learning, to estimate other measures of well-being, including consumption expenditures (Aiken et al. 2022), food security (Decuyper et al. 2014), literacy (Schmid et al. 2017), and employment (Toole et al. 2015). More recent work has also examined the use of mobile phone data for individual-level poverty targeting in the context of anti-poverty and humanitarian aid programs. In the context of Togo, Aiken et al. (2022) found that targeting

poor households for aid based on consumption measures inferred from mobile phone data was more accurate than geographic targeting or other rules-based targeting approaches, but less accurate than traditional in-person survey-based approaches to identifying poor households.

Web and social media data

While less thoroughly explored than remote sensing data and mobile phone data, a handful of studies have used web and social media data sources to produce static measures of poverty and well-being. Sheehan et al. (2019) use geolocated Wikipedia articles to predict village-level wealth indices in Ghana, Malawi, Nigeria, Tanzania, and Uganda. They experiment with a number of natural language processing techniques to extract meaningful information from the Wikipedia articles associated with GPS coordinates. They also compare their Wikipedia-based approach to using nighttime lights data to infer poverty, as well as a multimodal model leveraging both data sources; they find that Wikipedia generally slightly outperforms nighttime lights, and the multimodal model consistently performs best. In another example, Fatehkia et al. (2020) leverage social media advertising data – specifically, information on the number of people in an area accessing the Facebook platform from different device types – to predict poverty levels in the Philippines and India. The authors find fairly high predictive accuracy for identifying wealth at the subnational level based on information on device types and network access modalities in the area.

Machine learning methods

Many of the cross-sectional studies that have measured welfare with nontraditional data sources have relied on machine learning (ML) methods for producing welfare estimates (this is also often true of the real-time monitoring approaches covered in the next sections). While a full review of the machine learning methods used to construct these welfare estimates is beyond the scope of this chapter, a few key trade-offs are worth noting:

- *Supervised vs. unsupervised learning:* Most work leveraging nontraditional data for welfare measurement has relied on *supervised learning*: a small “labeled” training dataset (where ground truth measures of welfare are available at the household or village level and can be linked to the nontraditional data source) is used to train the machine learning model, and predictions are produced for all households or villages (even those where the ground truth is not available). The key challenge of supervised learning is the cost of acquiring the “labeled” training data (see comments on training data acquisition below), as well as

the challenges of linking labels to the nontraditional data source. A handful of studies therefore rely on *unsupervised learning*, using clustering methods to identify households or villages that look similar (for example, Jean et al. (2018) use unsupervised learning to produce meaningful representations of satellite images that allow for clustering of similar-looking areas).

- *Complexity of machine learning approaches:* Focusing on supervised learning approaches, ML models range from simple linear models (such as linear regression or regularized linear models like LASSO or Ridge) to more complex, non-parametric approaches (such as random forests and gradient boosting models) to multi-layer neural networks. Linear models have a number of advantages, including interpretability (coefficients in linear models can be directly interpreted as the “importance” of different variables in producing welfare estimates) and lower computational complexity. However, particularly in settings where data are complex or unstructured (such as satellite images and social media posts), nonparametric and neural networks may be a more suitable approach. Moreover, in settings with sufficient training data, the more complex machine learning models have been shown to typically be more accurate than linear approaches (McBride et al. 2017).
- *Training data:* As is often the case in supervised learning settings, additional training data is likely to improve the accuracy of welfare estimation approaches relying on machine learning (Blumenstock et al. 2015, Yeh et al. 2020, Gualavisi and Newhouse 2024, Zheng et al 2025). However, collecting “ground truth” labels for welfare measures can be expensive, as they often require time- and resource-intensive surveys. A number of studies on welfare estimation from nontraditional data have relied on labels from publicly available survey data sources, like the Demographic and Health Surveys (DHS) and Living Standards Measurement Surveys (LSMS) (Jean et al. 2016, Yeh et al. 2020). However, in some settings, the only option is to collect primary data for labels – this is often the case in settings where linking identifiers (such as phone numbers or exact household GPS coordinates) are not released in publicly available surveys. Moreover, integrating traditional “labeled” data and nontraditional data sources relies on ensuring that the data sources are compatible: most importantly, they must have at least one identifier (such as a phone number, spatial identifier, or national ID number) in common. In many settings, linking these data sources also requires ensuring common spatial and temporal resolution (for example, Aiken et al. 2022 show that it is important that nontraditional data and training “labels” come from roughly the same time period).

- *Bias and uncertainty:* Finally, and importantly, machine learning methods produce predictions of welfare that are likely to have errors. It is important to carefully validate the accuracy of machine learning methods on out-of-sample data before deploying welfare estimates based on ML models, and to ensure that predictions are not systematically biased against certain areas or groups. Moreover, while there is little work to date on uncertainty estimation for welfare estimates produced from nontraditional data (Chi et al. 2022), this is a fruitful avenue for future research. Confidence intervals for predictions could help ensure that policymakers leveraging welfare estimates based on nontraditional data have a well-calibrated sense of the accuracy of the estimates.

Real-time monitoring with novel data

Next, we review the literature on real-time monitoring with novel data sources, with a focus on low-income countries. The focus of this section is on research that has used satellite imagery, mobile phone data, internet data, or other digital data sources to monitor conditions in low-income settings. As such, the examples reviewed here cover applications well beyond welfare, including mobility, disease spread, war destruction, political sentiment, and remittances. Given that the literature on real-time monitoring of welfare specifically from digital data is fairly nascent, these examples provide useful lessons relevant to future uses of digital data for real-time monitoring of poverty and other measures of well-being.

This section is organized as five case studies, covering diverse data sources (including satellite imagery, mobile phone data, search frequency data, social media data, and mobile money data), diverse applications (from mobility to public health and natural disaster response), and diverse regions, including South and Southeast Asia, Latin America, and Africa.

Case study #1: Monitoring the impacts of violence on displacement in Afghanistan with mobile phone metadata

Mobility monitoring is one of the most robust and developed applications of real-time monitoring with digital data, with mobility monitoring feeding into decision making in public health (Wesalowski et al. 2018, Nouvellet et al. 2021), natural disaster response (Bengtsson et al. 2011), and refugee resettlement (Beine et al. 2019). Traditional mobility surveys are increasingly frequently

complemented by mobility measures derived from digital data, including mobile phone metadata (Blondel et al. 2015) and location data collected by social media applications (Antenucci et al. 2014), often accessed via a data aggregator like SmartGraph or Cubiq.

A recent research study focused on internal displacement in Afghanistan (Tai et al. 2022) provides a useful end-to-end example of how digital data – in this case, mobile phone metadata – can help monitor mobility in high-stakes settings. The authors use metadata on the locations of calls and text messages placed by over ten million mobile subscribers – geolocated via the antenna through which calls and texts are placed – to trace patterns of mobility in Afghanistan from 2013 to 2017. They pair the mobile phone data with records of fatal violent conflicts in Afghanistan from the Uppsala Conflict Data Program, which maintains a database of geocoded conflict events based on public news reports. By examining the sequence of cell towers through which subscribers place phone calls and text messages, the study identifies episodes of internal displacement at the individual level – defined in the paper as times when a person leaves their home district (as calculated based on their most frequent cell tower use) for more than a week’s time. The study then uses an event study design to identify the effect of violent events on out-migration from the district in which the event occurs. The authors find that, based on internal displacement measured from phone data, conflict in a mobile subscriber’s home district causes a 4% increase in the likelihood of their leaving that district in the subsequent week. They are also able to disaggregate the impacts of violence by perpetrator, finding that violence caused by the Islamic State leads to the highest rates of internal displacement.

Case study #2: Nowcasting disease prevalence with Internet search trends in Mexico, Thailand, Singapore, Brazil, and Taiwan

Another well-developed use case of digital data for real-time monitoring lies in the public health sector, in complementing traditional epidemiological approaches with internet search information to “nowcast” the spread of disease. Original work in this space focused on tracking spikes in seasonal Influenza in the United States using the frequency of related search terms (like “influenza”, “flu”, “cold”, “flu medicine”, and so on) (Ginseberg et al. 2009). More recently, similar approaches have been expanded to low-income settings and tropical diseases, including Dengue (Yang et al. 2017), Chikungunya (Naveca et al. 2019), and outbreaks of Ebola and plague (Aiken et al. 2020).

Yang et al. 2017, which focuses on nowcasting the spread of Dengue, demonstrates the deployment of these approaches in Mexico, Thailand, Singapore, Brazil, and Taiwan. The study obtains monthly counts of new Dengue cases in each country based on official reports from Ministries of Health (official reports of case counts are typically available with at least a month's lag), as well as weekly or monthly Google search volumes for the ten queries most correlated with the epidemiological time series in each country. The authors develop an autoregressive approach to “nowcasting” Dengue case counts, predicting the current month's case load based on case counts in the preceding months and Google search volumes in the month in the current month. The study shows that the inclusion of the Google search information improves predictive accuracy over a standard autoregressive approach (without the real-time internet search information) in four of the five countries.

Case study #3: Measuring wartime property damage with satellite imagery in Syria

With the availability of increasingly high resolution and real-time satellite imagery, there has been increasing attention in research and policy spaces on using remote sensing data for building damage detection during natural disasters and violent conflicts. Such approaches have been deployed in a number of recent conflicts to track the quantity of damage and target resources to areas with high levels of destruction, including in Syria (Mueller et al. 2021) and Ukraine (Aimati et al. 2022). Recently publicly available benchmark datasets have also been released to improve machine learning approaches to working with remote sensing data for damage detection in this setting (Gupta et al. 2019).

A recent research study using satellite imagery for automated damage detection during the Syrian civil war provides an illustrative example. Mueller et al. 2021 work with a dataset of labeled dataset of building destruction in Aleppo between 2013 and 2016, where, for parts of the city, annotators at the United Nations Institute for Training and Research hand-labeled instances of building destruction in satellite images. The authors develop a convolutional neural network – a type of deep learning model particularly useful for analyzing images – to predict whether a given building is destroyed based on the aerial imagery of the building. The study reports a high level of predictive accuracy for identifying destroyed structures – for example, their predictive model achieves a true positive rate of 80% at the cost of a false positive rate of 17%.

Case study #4: Tracking political sentiment on social media during an electoral transition in Egypt.

Social media platforms are one of the richest sources of data available for real time monitoring of public opinion. Most of this research has focused on political sentiment; while this measure is not directly related to welfare, we include examples of these studies below as they are illustrative of how social media data might be helpful for monitoring subjective well-being at a population scale. A large body of research has leveraged data from social media platforms – particularly Reddit and Twitter – for tracking political sentiment in high-income countries around elections and other key political transitions (e.g. Beers et al. 2013, Grinberg et al. 2019, Gaumont et al. 2018). Although the penetration of such social media sites tends to be lower in low-income settings, and linguistic diversity is often an additional challenge in such settings, a few research studies have used social media posts to monitor political sentiment and forecast electoral outcomes in Indonesia (Dwi Prasetyo and Hauff 2015), Venezuela (Morales et al. 2015), and India (Chakraborty and Mukherjee 2023).

The most developed example of such an approach studied political discourse on Twitter during Egypt’s 2013 coup d’état. Borge-Holthoefer et al. (2015) used Twitter’s API to collect all Arabic tweets from the summer of 2013 and filtered tweets based on keywords relating to Egypt’s political transition to around 6 million tweets related to the coup. The study then uses natural language processing methods to classify tweets as pro- vs. anti-military takeover. By classifying each user’s tweets as pro- vs. anti-military over time, the authors were able to study whether Twitter users changed camps as events unfolded. They found little evidence of users switching their stance, but observed clear changes in the volume of tweets coming from each side in response to events in the military takeover. The study also demonstrates the power of social network analysis using social media data: by examining the follower and retweet networks within and between pro- and anti-military camps, the authors evaluate the rates at which the pro- and anti-military groups recruited new members over time.

Case study #5: Measuring risk sharing after natural disasters via mobile money records in Rwanda

Mobile money data are perhaps the most exciting opportunity for real-time monitoring of financial transactions and financial services in low-income countries (particularly in the African context, where mobile money penetration is highest). In countries with high levels of mobile money penetration, the fine grained

information collected on financial transactions has the potential to provide insights into consumption patterns, savings, remittances, and uptake of lending services. However, there is limited work to date employing real-time monitoring techniques with mobile money data, perhaps due to the proprietary nature of such data (since it is privately owned by mobile money operators) and the limited number of settings in which mobile money penetration is high enough to provide a complete picture of financial systems.

Blumenstock et al. (2016) provides a relatively developed example of using mobile money records for monitoring financial transactions. The study's context is the 6.0 magnitude earthquake near Lake Kivu in Rwanda in 2008. The study first uses mobile phone metadata (records of mobile phone calls) to show that Rwandans placed calls in response to the earthquake: an estimated additional \$16,959 was spent on calls to people inferred to be living in the Lake Kivu region (based on the locations of cell towers through which they placed calls) immediately following the earthquake (relative to the amount spent on calls in normal times). The amount of airtime transfers to people living in the region also increased (by \$84) following the earthquake (the study relies on airtime transfers rather than mobile money because Rwanda's mobile money ecosystem was nascent at the time). The study is also able to analyze heterogeneity in who received such airtime transfers, finding that wealthier people and people with a history of using the airtime transfer service were most likely to be the recipient of post-earthquake transfers.

Nascent work on real-time monitoring of welfare with novel data

There are limited existing research or policy examples of using digital data for real-time monitoring of poverty, wealth, food security, or other measures of welfare. The early results in this area suggest that measuring *changes* in welfare over time is a substantially more challenging task than measuring *levels* of welfare in a given time, as in the studies on static poverty measurement reviewed in the third section of this chapter.

In this section we review five early case studies on real-time welfare monitoring with digital data sources. The first two rely on satellite imagery for measuring changes in poverty over time at the village- or household-level. The third explores whether mobile phone data can be used for impact evaluation. The fourth case study uses mobile phone data to track unemployment shocks

in an undisclosed European country, and the final case study relies on news reports posted on the Internet to forecast food security crises at a country level. Following the case studies, we briefly summarize the key current challenges to real-time monitoring of welfare, and speculate on *why* real-time monitoring seems to be substantially more challenging than the cross-sectional settings covered in the third section.

Case study #1: Measuring changes in village-level poverty over time from satellite imagery

While the seminal papers reviewed in this third section of this chapter (Henderson et al. 2012, Jean et al. 2016) showed that information in satellite images could differentiate between *levels* of poverty within countries, until recently there has been little evidence on the ability of satellite images to measure *changes* in poverty over time. Yeh et al. (2020) provided the first evidence on this question. This study used publicly available satellite imagery from Landsat to construct three-year composite images of 23 African countries. They pair these images with repeated rounds of surveys conducted by the Demographic and Health Surveys (DHS) and Living Standards Measurement Surveys (LSMS) teams. The authors train deep learning models to predict asset-based wealth from the satellite images at a village level, and assess the extent to which the predictions can measure (1) levels of wealth at a given point in time, and (2) changes in wealth over time.

The study finds that while the satellite-based wealth predictions have high predictive accuracy for differentiating between wealth levels in a single survey wave ($R^2 = 0.70$), changes in wealth prediction over time have little predictive accuracy for measuring changes in ground-truth welfare over time ($R^2 = 0.15-0.17$). The study notes that a particular challenge to measuring welfare changes with satellite imagery in this setting is the low level of wealth variation in the ground-truth survey data from the DHS and LSMS: the average wave-to-wave change in the wealth index is only 0.08 standard deviations.

Case study #2: Impact evaluation of anti-poverty and infrastructure interventions via satellite imagery in Kenya and Uganda

Building on the work of Yeh et al. (2020), two recent studies have developed remote sensing approaches to measuring welfare changes in settings where large exogenous interventions produce substantial impacts. The first study, Huang et al. (2021), focuses on a cash transfer program implemented by

GiveDirectly in 653 villages in rural Kenya between 2014 and 2017. GiveDirectly implemented a randomized controlled trial, where poor households in treatment villages received a one-time cash transfer of \$1,000. The authors extract building footprints for all households in the treatment and control groups from Google Static Maps imagery (using GPS coordinates collected during household surveys), and calculate two metrics related to building structure: (1) the size of the building footprint, and (2) the roof type (tin, thatched, or painted), based on color profiles. They find that treatment households increase building footprint by an average of eight square meters, and tin roof area by 13.6 square meters. The authors use baseline survey data from the program to calibrate Engel curves measuring the association between each of these two features and household consumption, and show that the approximated consumption impacts of the program are similar to those calculated with post-program survey data (albeit with wider confidence intervals).

Ratledge et al. (2022) use satellite imagery to evaluate the welfare impacts of a different large development intervention: the expansion of Uganda's electrical grid. The authors construct a panel dataset of grid locations in Uganda between 2010 and 2012, using publicly available data from the Uganda government and World Bank reports. They also construct a panel dataset of satellite image-level predictions of wealth, based on the standard convolutional neural network trained on DHS survey wealth indices similar to Jean et al. (2016) and Yeh et al. (2020). "Treated" areas are those that are within 2km of new distribution lines installed in 2011 or 2012, while "control" areas are those that did not receive electrification before 2016. They then use econometric techniques (matrix completion and synthetic controls with elastic net) to estimate the impacts of grid access on satellite-measured wealth. They estimate statistically significant impacts of electrification of around 0.17 standard deviations. They compare this result to a simple difference-in-differences approach implemented using just DHS survey data, and find that the impact estimate is similar, but with a wider confidence interval due to the limited number of locations included in the DHS survey relative to the satellite-based estimates.

Note that the data needs of these studies are substantial: both studies rely on precise GPS coordinates for both "treated" and "control" households (or villages); they also rely on satellite imagery from multiple time periods. However, this general approach is attractive as it can be adapted to both RCTs (as in Huang et al. 2021) and quasi-experimental settings (as in the differences-in-differences specification used by Ratledge et al. 2022).

Case study #3: Impact evaluation of anti-poverty programs via mobile phone data in Togo and Haiti

The two previous papers illustrate the potential for satellite imagery to detect the impact of large-scale development interventions. In closely related work, a pair of recent studies ask whether mobile phone data can be used to estimate the impact of smaller-scale interventions. Barriga-Cabanillas et al (2025) study a cash-transfer program in Haiti, through which the World Food Programme provided three monthly transfers of roughly USD \$50 to poor households in the south of Haiti. Aiken et al. (2025) study a cash-transfer program in Togo, which provided roughly five monthly transfers of roughly \$14 USD to poor households in rural Togo. Based on traditional impact evaluation methods using traditional data, the authors find that both programs had positive effects on welfare: In Haiti, the authors analyze phone surveys with a regression discontinuity (RD) design (since eligibility was determined based on a poverty score), and find the program had significant effects on food expenditures (0.35 standard deviations, or SD) and food consumption (0.32 SD). In Togo, where eligibility was randomly assigned, a simple RCT analysis of phone surveys indicates the program had a significant impact on food security (0.06 SD), mental health (0.07 SD), and perceived economic status (0.04 SD).

Both papers then go on to study whether the program impacts detected with traditional data can also be accurately estimated using a combination of mobile phone call detail records and machine learning. Specifically, they use methods similar to those described in Section 3 to *predict* the welfare of program participants (both beneficiaries and non-beneficiaries), and then use traditional impact evaluation methods (RD and RCT, respectively) to estimate the impact of the program on the *predicted* welfare outcomes. In both Haiti and Togo, the authors find that the impact estimates based on these predictions are imprecise and not statistically significant.

While discouraging at first glance, both papers provide helpful postmortem discussions to better understand why impact estimates based on predicted outcomes do not replicate impact estimates based on survey data. Perhaps most critically, in both cases, the specific outcomes that were impacted by the programs (e.g., food expenditures in Haiti and food security in Togo) could *not* be accurately predicted from phone data – even in the cross-section, before the programs were implemented. Other outcomes *could* be accurately predicted from phone data in the cross-section (such as a wealth index in Haiti and a proxy means test in Togo), but those outcomes were not impacted by the program

– perhaps because the interventions provided relatively small amounts of cash support. Whether mobile phone data can be used to estimate the impact of larger interventions – and particularly, the impact of interventions that affect outcomes that be accurately predicted with phone data – remains an open question and an active area of research.

Case study #4: Tracking unemployment shocks from mobile phone data in Europe

Moving away from both satellite imagery and poverty, this third case study uses mobile phone data to track unemployment shocks at the province level in an undisclosed European country. Toole et al. (2015) obtained a mobile phone dataset covers around 10 million subscribers in the country in question, with information on the identify of the caller and receiver, along with the times of calls and the location of the cell tower through which each call is placed. They pair the phone dataset with quarterly, province-level data on unemployment rates. They compute statistics on average call volume, number of incoming and outgoing calls, number of contacts, number of towers visited, and other mobility metrics at the province level (using a random sample of 3,000 mobile phone subscribers from each province). Regressing each province-month’s unemployment rate on the mobile phone variables, the authors find high predictive power for current unemployment rates ($R^2 = 0.95$) as well as for predicting the next quarter’s unemployment rate ($R^2 = 0.85$). Compared with standard autogressive approaches to forecasting unemployment, they find that using mobile phone data improves forecast accuracy (measured through route mean squared error) by 5-20%.

Case study #5: Forecasting food insecurity crises via news reports in food-insecure countries

The final case study we review leverages web data in the form of news reports related to food insecurity and famine. [Balashankar et al. \(2024\)](#) identify 37 food-insecure countries for which the Famine Early Warning Network (FEWS NET) provide classifications of food security several times per year between 2009 and 2015. They use news aggregate Factiva to identify news articles from these countries relevant to food security based on keyword filters (for articles containing words like “food prices”, “conflict”, “drought”, “flood”, and “pests” (the authors use 167 keyword filters in total). They use these articles to construct measures of the proportion of articles in a given country in a given month that contain each famine-related theme, and use machine learning to construct

a predictive model for the integrated phase classification (IPC classification) released by FEWSNET, which measures a country's level of food security. The model's output predicts whether or not a country will surpass IPC classification level three (corresponding to a food crisis) three months in advance. In addition to the news-derived features, the model also includes time-invariant risk factors like population, terrain ruggedness, and the share of agricultural land use. They compare the news model to a traditional model that relies on time-invariant risk factors and measurements of traditional drivers of food crises (including conflict fatality count, changes in food prices, an evapotranspiration index, and an inverted vegetation index), as well as three-months-in-advance expert forecasts. They find that the inclusion of news information increases predictive accuracy relative to the traditional model, reducing root mean squared error by 33%. The news model also substantially outperforms expert forecasts, reducing root mean squared error by 48%.

Challenges to real-time monitoring of welfare with nontraditional data

While this section provides a handful of examples of recent papers that use nontraditional data for real-time monitoring, this line of research is much less developed than the more robust literature on cross-sectional welfare estimation (reviewed in Section 3). This is because real-time monitoring is substantially more challenging than cross-sectional prediction: the accuracy reported in the research papers covered in this section is substantially lower than the accuracy of cross-sectional prediction; and logistically, real-time monitoring requires substantially more data than cross-sectional prediction. While to our knowledge there is no rigorous study on *why* real-time monitoring tends to have lower accuracy than cross-sectional prediction, we speculate here on a few likely reasons:

- *Variation in welfare*: The goal of welfare estimation from nontraditional data sources is typically to accurately capture differences in welfare across space and/or time. In general, this task will be easier when there is greater variation. For instance, Aiken et al. (2022) show that household-level poverty estimation from mobile phone data is easier on a national sample (where the variation in welfare is greater) than on a more homogenous rural sample (where there is not as much variation in welfare). The same logic applies when trying to estimate changes in welfare over time: if the variation in welfare over time is lower than the variation in welfare over space (which are the target of cross-sectional predictions), the time series prediction task will be more challenging. In settings where there is limited variation in welfare over time, real-time monitoring from nontraditional data will be particularly challenging.

- *Welfare measures:* Many of the studies on cross-sectional welfare measurement reviewed in this chapter focus on prediction of wealth or other slow-moving measures of well-being (Jean et al. 2016, Yeh et al. 2020, Blumenstock et al. 2015). Real-time monitoring approaches rarely focus on these measures, since they are unlikely to change much over short and medium time horizons. The measures that real-time monitoring approaches focus on – such as consumption, food security, and employment – are more likely to fluctuate over time. It could be, however, that these measures are more susceptible to measurement error (Hjelm et al. 2016, Tadesse et al. 2020) – which will make them more difficult to predict – and/or less closely related to the indicators of welfare available in nontraditional data sources like satellite imagery, mobile phone metadata, and social media data.
- *Data requirements:* While not directly related to accuracy, it is also worth noting that the data requirements for real-time monitoring are substantially greater than the data requirements for cross-sectional prediction. Studies focusing on real-time monitoring therefore typically require substantially more resources for data acquisition and computation, or must rely on lower-quality data (such as lower resolution satellite imagery or more limited mobile phone datasets), which may result in lower-accuracy predictions.

Looking forward / Conclusions

The material in this chapter is motivated by the persistent gaps in household-level economic data collection in low- and middle-income countries (LMICs). Our main objective has been to highlight the progress that has been made in recent years to address these gaps, using computational methods in conjunction with novel data sources such as satellite imagery, mobile phone records, web and social media interactions, and other administrative data. These diverse new datasets have a range of advantages – as well as limitations – that we have tried to highlight throughout the chapter. And while significant progress has been made in using these data to measure household and regional welfare – particularly at a single point in time – a great amount of work is still required to move from a single snapshot in time to the more ambitious goal of real-time poverty measurements. Before concluding, we provide a few suggestions for what must be done to make progress toward this objective.

First, the real-time measurement research agenda requires a robust and multifaceted data ecosystem. Importantly, this data ecosystem requires not just “big” data infrastructure to analyze satellite and digital trace data – though such investments

are certainly also necessary (Gelvanovska-Garcia et al. 2024). But what will be most critical is to find ways to integrate those non-traditional datasets with more *traditional* survey data, which often provide the foundation for subsequent layers of non-traditional data and analysis. Indeed, these two types of data are complementary: whereas satellite and phone data make it possible to observe unprecedented breadth (in time, space, and number), they lack the depth and nuance captured in household surveys. And as we have seen in earlier sections of this chapter, many of the seminal papers in this growing field rely critically on household survey data, such as DHS and LSMS surveys, to train the machine learning algorithms that are then used to produce more granular and high-frequency welfare indicators. There is also room for innovation in how traditional data are collected – for instance via more lightweight surveys (e.g., Yoshida and Aron 2024) or adaptive and strategic approaches to data collection (Soman et al. 2022). As we move from cross-sectional poverty estimates to real-time poverty dashboards, it will be essential to find ways to support the collection and integration of both types of data.

Second, and related: Reliable mechanisms and protocols must be implemented to ensure the ethical and appropriate use of non-traditional data sources, while also enabling responsible access and sharing (Blumenstock 2018). Mobile phone, satellite, and other data may create privacy risks (de Montjoye et al. 2013, McKenna et al. 2019); clear frameworks will be required to address privacy concerns (Oliver et al. 2020), as well as issues related to data ownership and data stewardship (Ademuyiwa and Adeniran 2020, Abebe et al. 2021, Blumenstock & Kohli 2023). And as we have noted throughout this chapter, each dataset has inherent biases in who and what is represented (and under-represented) in the data (Blumenstock et al. 2010, ITU 2024, GSMA 2023); care is needed to ensure that inferences and decisions based on the data are fair and representative (Wesolowski et al. 2013, Aiken et al. 2023). Facilitating ethical and responsible access to both digital and traditional survey data are fundamental to building trust among stakeholders and researchers and promoting a sustainable model of collaboration and knowledge dissemination.

Third, methodological and technological innovations should be driven by explicit policy objectives. This, in turn, requires that policymakers and practitioners develop a more nuanced understanding of how non-traditional data sources can be meaningfully integrated into project design and evaluation. Each approach has strengths and weaknesses that may be more or less appropriate for certain applications. For instance, some settings (such as crisis response) may require rapid estimates, with a higher tolerance for error, whereas others (such

as impact evaluation) may prioritize quantitative accuracy over cost and speed. Unfortunately, to date, the vast majority of published papers demonstrate proof-of-concept ideas of how a particular measure of well-being can be estimated using a particular new dataset, without a clearly articulated policy rationale. Such studies often speculate how the new measurements might influence policy, but rarely document the policy or downstream impacts that occur as a result of the new measurements. Looking forward, we expect that the most influential research in real-time poverty measurement will rigorously document not only methodological advances but also explicitly assess the operational advantages and limitations of these approaches, including trade-offs in accuracy, cost, timeliness, transparency, and robustness to manipulation. Aiken et al. (2025) takes a first step towards rigorously evaluating some of these trade-offs, developing a framework for cost-benefit analysis in comparing targeting with traditional and nontraditional data sources

Fourth, sustained investments and partnerships are necessary to unlock the potential of real-time monitoring. As outlined above, the most impactful innovations will require multi-year collaborations between policymakers and implementing organizations, who dictate the objectives and constraints, with interdisciplinary research teams, who can bring to bear state of the art methods in both computation and economic analysis. Practical applications will often require partnerships with data providers, who might be private companies or public trusts, and coordination with regulatory agencies, to ensure the data are used responsibly and ethically. To build a robust evidence base from such projects requires long-term, strategic investments in research and knowledge production.

In summary, realizing the potential of real-time poverty measurement through digital data sources demands a deliberate, structured, and collaborative approach. Over the past decade, key innovations now make it possible to measure the static distribution of poverty with impressive scale and granularity, at a fraction of the cost of traditional methods. With sustained and strategic investments in partnerships, we expect that the next decade can produce similar progress toward measuring poverty and well-being in real-time.

References

- Abebe, Rediet, Kehinde Aruleba, Abeba Birhane, et al. 2021. “Narratives and Counternarratives on Data Sharing in Africa.” *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, March 3, 329–41. <https://doi.org/10.1145/3442188.3445897>.
- ACM Conferences. n.d. “Poverty on the Cheap | Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.” world. <https://doi.org/10.1145/2556288.2557358>.
- ACM Other Conferences. n.d. “On the Relationship between Socio-Economic Factors and Cell Phone Usage | Proceedings of the Fifth International Conference on Information and Communication Technologies and Development.” world. <https://doi.org/10.1145/2160673.2160684>.
- Ademuyiwa, Idris, and Adedeji Adeniran. 2020. *Assessing Digitalization and Data Governance Issues in Africa*. no. 244. https://www.cigionline.org/static/documents/documents/no244_0.pdf.
- Aiken, Emily, Anik Ashraf, Joshua Blumenstock, Raymond Guiteras, and Ahmed Mushfiq Mobarak. 2025. “Scalable Targeting of Social Protection: When Do Algorithms Out-Perform Surveys and Community Knowledge?” Working Paper 33919. Working Paper Series. National Bureau of Economic Research, June. <https://doi.org/10.3386/w33919>.
- Aiken, Emily, Suzanne Bellue, Joshua E. Blumenstock, Dean Karlan, and Christopher Udry. 2025. “Estimating Impact with Surveys versus Digital Traces: Evidence from Randomized Cash Transfers in Togo.” *Journal of Development Economics* 175 (June): 103477. <https://doi.org/10.1016/j.jdeveco.2025.103477>.
- Aiken, Emily, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E. Blumenstock. 2022. “Machine Learning and Phone Data Can Improve Targeting of Humanitarian Aid.” *Nature* 603 (7903): 864–70. <https://doi.org/10.1038/s41586-022-04484-9>.
- Aiken, Emily L., Sarah F. McGough, Maimuna S. Majumder, et al. 2020. “Real-Time Estimation of Disease Activity in Emerging Outbreaks Using Internet Search Information.” *PLOS Computational Biology* 16 (8): e1008117. <https://doi.org/10.1371/journal.pcbi.1008117>.

- Aiken, Emily, Tim Ohlenburg, and Joshua Blumenstock. 2024. "Moving Targets: The Role of Model and Data Recency in Proxy Means Test Accuracy." Paper presented at NEUDC.
- Aiken, Emily, Esther Rolf, and Joshua Blumenstock. 2023. "Fairness and Representation in Satellite-Based Poverty Maps: Evidence of Urban-Rural Disparities and Their Impacts on Downstream Policy." arXiv:2305.01783. Preprint, arXiv, May 2. <https://doi.org/10.48550/arXiv.2305.01783>.
- Aji, Alham Fikri, Genta Indra Winata, Fajri Koto, et al. 2022. "One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia." arXiv:2203.13357. Preprint, arXiv, March 24. <https://doi.org/10.48550/arXiv.2203.13357>.
- Alberro Encinas, Luis Inaki, Sebastian Geschwind, and Sarah Nirvana Patella. n.d. *Dynamic Social Registries for Adaptive Social Protection*. World Bank, Washington, DC. <http://documents.worldbank.org/curated/en/099810103242514178>.
- Antenucci, Dolan, Michael Cafarella, Margaret Levenstein, Christopher Ré, and Matthew D. Shapiro. 2014. "Using Social Media to Measure Labor Market Flows." Working Paper 20010. Working Paper Series. National Bureau of Economic Research, March. <https://doi.org/10.3386/w20010>.
- Ayush, Kumar, Burak Uzgent, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. 2021. "Efficient Poverty Mapping from High Resolution Remote Sensing Images." *Proceedings of the AAAI Conference on Artificial Intelligence* 35 (1): 12–20. <https://doi.org/10.1609/aaai.v35i1.16072>.
- Balashankar, Ananth, Lakshminarayanan Subramanian, and Samuel P. Fraiberger. 2023. "Predicting Food Crises Using News Streams." *Science Advances* 9 (9): eabm3449. <https://doi.org/10.1126/sciadv.abm3449>.
- Beers, Andrew, Joseph S. Schafer, Ian Kennedy, Morgan Wack, Emma S. Spiro, and Kate Starbird. 2023. "Followback Clusters, Satellite Audiences, and Bridge Nodes: Coengagement Networks for the 2020 US Election." arXiv:2303.04620. Preprint, arXiv, May 30. <https://doi.org/10.48550/arXiv.2303.04620>.
- Blanchard, Paul, and Stefania Rubrichi. 2025. "A Highly Granular Temporary Migration Dataset Derived from Mobile Phone Data in Senegal." *Scientific Data* 12 (1): 1051. <https://doi.org/10.1038/s41597-025-04599-4>.

- Blondel, Vincent D, Adeline Decuyper, and Gautier Krings. 2015. "A Survey of Results on Mobile Phone Datasets Analysis." *EPJ Data Science* 4 (1): 10. <https://doi.org/10.1140/epjds/s13688-015-0046-0>.
- Blumenstock, Joshua. 2018. "Don't Forget People in the Use of Big Data for Development." *Nature* 561 (7722): 170–72. <https://doi.org/10.1038/d41586-018-06215-5>.
- Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science* 350 (6264): 1073–76. <https://doi.org/10.1126/science.aac4420>.
- Blumenstock, Joshua E., Nathan Eagle, and Marcel Fafchamps. 2016. "Airtime Transfers and Mobile Communications: Evidence in the Aftermath of Natural Disasters." *Journal of Development Economics* 120 (May): 157–81. <https://doi.org/10.1016/j.jdeveco.2016.01.003>.
- Blumenstock, Joshua E., and Nitin Kohli. 2023. *Big Data Privacy in Emerging Market Fintech and Financial Services: A Research Agenda*. <https://doi.org/10.26085/C3WK53>.
- Blumenstock, Joshua, and Nathan Eagle. 2010. "Mobile Divides: Gender, Socioeconomic Status, and Mobile Phone Use in Rwanda." *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development, December 13, 1–10*. <https://doi.org/10.1145/2369220.2369225>.
- Blumenstock, Joshua Evan, and Nathan Eagle. 2012. "Divided We Call: Disparities in Access and Use of Mobile Phones in Rwanda." *Information Technologies & International Development* 8 (2): 1–16.
- Bolch, Kimberly Blair, Maria Eugenia Genoni, and Henry Walter Scott Stemmler. 2024. *Real-Time Welfare Monitoring: A Typology of Approaches*. September 5. <https://policycommons.net/artifacts/16410333/measuring-welfare-when-it-matters-most/17295101/>.
- Borge-Holthoefer, Javier, Walid Magdy, Kareem Darwish, and Ingmar Weber. 2015. "Content and Network Dynamics Behind Egyptian Political Polarization on Twitter." *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (New York, NY, USA), CSCW '15, February 28, 700–711*. <https://doi.org/10.1145/2675133.2675163>.

- Burke, Marshall, Anne Driscoll, David B. Lobell, and Stefano Ermon. 2021. "Using Satellite Imagery to Understand and Promote Sustainable Development." *Science* 371 (6535): eabe8628. <https://doi.org/10.1126/science.abe8628>.
- Chakraborty, Amartya, and Nandini Mukherjee. 2023. "Analysis and Mining of an Election-Based Network Using Large-Scale Twitter Data: A Retrospective Study." *Social Network Analysis and Mining* 13 (1): 74. <https://doi.org/10.1007/s13278-023-01081-0>.
- Chi, Guanghua, Han Fang, Sourav Chatterjee, and Joshua E. Blumenstock. 2022. "Microestimates of Wealth for All Low- and Middle-Income Countries." *Proceedings of the National Academy of Sciences* 119 (3): e2113658119. <https://doi.org/10.1073/pnas.2113658119>.
- Dang, Hai-Anh H., Talip Kilic, Kseniya Abanokova, and Calogero Carletto. 2025. "Poverty Imputation in Contexts Without Consumption Data: A Revisit With Further Refinements." *Review of Income and Wealth* 71 (1): e12714. <https://doi.org/10.1111/roiw.12714>.
- Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30 (4): 171–98. <https://doi.org/10.1257/jep.30.4.171>.
- Dwi Prasetyo, Nugroho, and Claudia Hauff. 2015. "Twitter-Based Election Prediction in the Developing World." *Proceedings of the 26th ACM Conference on Hypertext & Social Media - HT '15*, 149–58. <https://doi.org/10.1145/2700171.2791033>.
- Eagle, Nathan, Michael Macy, and Rob Claxton. 2010. "Network Diversity and Economic Development." *Science* 328 (5981): 1029–31. <https://doi.org/10.1126/science.1186605>.
- Ebener, Steeve, Christopher Murray, Ajay Tandon, and Christopher C Elvidge. 2005. "[No Title Found]." *International Journal of Health Geographics* 4 (1): 5. <https://doi.org/10.1186/1476-072X-4-5>.
- Elvidge, C. D., K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, and C. W. Davis. 1997. "Relation between Satellite Observed Visible-near Infrared Emissions, Population, Economic Activity and Electric Power Consumption." *International Journal of Remote Sensing* 18 (6): 1373–79. <https://doi.org/10.1080/014311697218485>.

- Fatehkia, Masoomali, Isabelle Tingzon, Ardie Orden, et al. 2020. “Mapping Socioeconomic Indicators Using Social Media Advertising Data.” *EPJ Data Science* 9 (1): 22. <https://doi.org/10.1140/epjds/s13688-020-00235-w>.
- Gaumont, Noé, Mazyar Panahi, and David Chavalarias. 2018. “Reconstruction of the Socio-Semantic Dynamics of Political Activist Twitter Networks—Method and Application to the 2017 French Presidential Election.” *PLOS ONE* 13 (9): e0201879. <https://doi.org/10.1371/journal.pone.0201879>.
- GIZ. n.d. “On-Demand and Up-to-Date? Dynamic Inclusion and Data Updating for Social Assistance | Socialprotection.Org.” Accessed August 21, 2025. <https://socialprotection.org/fr/discover/publications/demand-and-date-dynamic-inclusion-and-data-updating-social-assistance>.
- Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. “Fake News on Twitter during the 2016 U.S. Presidential Election.” *Science* 363 (6425): 374–78. <https://doi.org/10.1126/science.aau2706>.
- GSMA. n.d. *The Mobile Economy* 2023.
- Gualavisi, Melany, and David Newhouse. 2025. “Integrating Survey and Geospatial Data for Geographical Targeting of the Poor and Vulnerable: Evidence from Malawi.” *The World Bank Economic Review* 39 (2): 377–409. <https://doi.org/10.1093/wber/lhae025>.
- Gupta, Ritwik, Bryce Goodman, Nirav Patel, et al. 2019. “Creating xBD: A Dataset for Assessing Building Damage from Satellite Imagery.” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 10–17.
- Henderson, J. Vernon, Adam Storeygard, and David N Weil. 2012. “Measuring Economic Growth from Outer Space.” *American Economic Review* 102 (2): 994–1028. <https://doi.org/10.1257/aer.102.2.994>.
- Huang, Luna Yue, Solomon M. Hsiang, and Marco Gonzalez-Navarro. 2021. “Using Satellite Imagery and Deep Learning to Evaluate the Impact of Anti-Poverty Programs.” Working Paper 29105. Working Paper Series. National Bureau of Economic Research, July. <https://doi.org/10.3386/w29105>.
- Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. “Combining Satellite Imagery and Machine Learning to Predict Poverty.” *Science* 353 (6301): 790–94. <https://doi.org/10.1126/science.aaf7894>.

- Jean, Neal, Sherrie Wang, Anshul Samar, George Azzari, David Lobell, and Stefano Ermon. 2018. "Tile2Vec: Unsupervised Representation Learning for Spatially Distributed Data." arXiv:1805.02855. Preprint, arXiv, May 30. <https://doi.org/10.48550/arXiv.1805.02855>.
- Kilic, Talip, Umar Serajuddin, Hiroki Uematsu, and Nobuo Yoshida. 2017. *Costing Household Surveys for Monitoring Progress Toward Ending Extreme Poverty and Boosting Shared Prosperity*. World Bank, Washington, DC. <https://doi.org/10.1596/1813-9450-7951>.
- Mahler, Daniel Gerszon, R Andrés Castañeda Aguilar, and David Newhouse. 2022. "Nowcasting Global Poverty." *The World Bank Economic Review* 36 (4): 835–56. <https://doi.org/10.1093/wber/lhac017>.
- McBride, Linden, Christopher B. Barrett, Christopher Browne, et al. 2022. "Predicting Poverty and Malnutrition for Targeting, Mapping, Monitoring, and Early Warning." *Applied Economic Perspectives and Policy* 44 (2): 879–92. <https://doi.org/10.1002/aepp.13175>.
- McBride, Linden, and Austin Nichols. 2018. *Retooling Poverty Targeting Using Out-of-Sample Validation and Machine Learning*. October. <https://doi.org/10.1093/wber/lhw056>.
- McKenna, Anne Toomey, Amy C. Gaudion, and Jenni L. Evans. 2019. "The Role of Satellites and Smart Devices: Data Surprises and Security, Privacy, and Regulatory Challenges." SSRN Scholarly Paper 3418420. Social Science Research Network, July 11. <https://papers.ssrn.com/abstract=3418420>.
- Montjoye, Yves-Alexandre de, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. "Unique in the Crowd: The Privacy Bounds of Human Mobility." *Scientific Reports* 3 (1): 1376. <https://doi.org/10.1038/srep01376>.
- Morales, A. J., J. Borondo, J. C. Losada, and R. M. Benito. 2015. "Measuring Political Polarization: Twitter Shows the Two Sides of Venezuela." *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25 (3): 033114. <https://doi.org/10.1063/1.4913758>.
- Mueller, Hannes, Andre Groeger, Jonathan Hersh, Andrea Matranga, and Joan Serrat. 2021. "Monitoring War Destruction from Space Using Machine Learning." *Proceedings of the National Academy of Sciences* 118 (23): e2025400118. <https://doi.org/10.1073/pnas.2025400118>.

Newhouse, David. 2024. “Small Area Estimation of Poverty and Wealth Using Geospatial Data: What Have We Learned So Far?” *Calcutta Statistical Association Bulletin* 76 (1): 7–32. <https://doi.org/10.1177/00080683231198591>.

OECD. 2024. *FinTech Lending in Sub-Saharan Africa*.

Oliver, Nuria, Bruno Lepri, Harald Sterly, et al. 2020. “Mobile Phone Data for Informing Public Health Actions across the COVID-19 Pandemic Life Cycle.” *Science Advances* 6 (23): eabc0764. <https://doi.org/10.1126/sciadv.abc0764>.

Ratledge, Nathan, Gabe Cadamuro, Brandon de la Cuesta, Matthieu Stigler, and Marshall Burke. 2022. “Using Machine Learning to Assess the Livelihood Impact of Electricity Access.” *Nature* 611 (7936): 491–95. <https://doi.org/10.1038/s41586-022-05322-8>.

Rolf, Esther, Jonathan Proctor, Tamma Carleton, et al. 2021. “A Generalizable and Accessible Approach to Machine Learning with Global Satellite Imagery.” *Nature Communications* 12 (1): 4392. <https://doi.org/10.1038/s41467-021-24638-z>.

Rußwurm, Marc, Konstantin Klemmer, Esther Rolf, Robin Zbinden, and Devis Tuia. 2024. “Geographic Location Encoding with Spherical Harmonics and Sinusoidal Representation Networks.” arXiv:2310.06743. Preprint, arXiv, April 15. <https://doi.org/10.48550/arXiv.2310.06743>.

Schmid, Timo, Fabian Bruckschen, Nicola Salvati, and Till Zbiranski. 2017. “Constructing Sociodemographic Indicators for National Statistical Institutes by Using Mobile Phone Data: Estimating Literacy Rates in Senegal.” *Journal of the Royal Statistical Society Series A: Statistics in Society* 180 (4): 1163–90. <https://doi.org/10.1111/rssa.12305>.

Serajuddin, Umar, Hiroki Uematsu, Christina Wieser, Nobuo Yoshida, and Andrew Dabalen. 2015. *Data Deprivation: Another Deprivation to End*. World Bank, Washington, DC. <https://doi.org/10.1596/1813-9450-7252>.

Sheehan, Evan, Chenlin Meng, Matthew Tan, et al. 2019. “Predicting Economic Development Using Geolocated Wikipedia Articles.” *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, July 25, 2698–706. <https://doi.org/10.1145/3292500.3330784>.

- Soman, Satej, Emily Aiken, Esther Rolf, and Joshua Blumenstock. 2022. “Can Strategic Data Collection Improve the Performance of Poverty Prediction Models?” arXiv:2211.08735. Preprint, arXiv, November 16. <https://doi.org/10.48550/arXiv.2211.08735>.
- Sutton, Paul C, Christopher D Elvidge, and Tilottama Ghosh. 2007. *Estimation of Gross Domestic Product at Sub-National Scales Using Nighttime Satellite Imagery*. 8.
- Tai, Xiao Hui, Shikhar Mehra, and Joshua E. Blumenstock. 2022. “Mobile Phone Data Reveal the Effects of Violence on Internal Displacement in Afghanistan.” *Nature Human Behaviour* 6 (5): 624–34. <https://doi.org/10.1038/s41562-022-01336-4>.
- Tatem, Andrew J., Scott J. Goetz, and Simon I. Hay. 2008. “Fifty Years of Earth Observation Satellites: Views from above Have Lead to Countless Advances on the Ground in Both Scientific Knowledge and Daily Life.” *American Scientist* 96 (5): 390. <https://doi.org/10.1511/2008.74.390>.
- Toole, Jameson L., Yu-Ru Lin, Erich Muehlegger, Daniel Shoag, Marta C. González, and David Lazer. 2015. “Tracking Employment Shocks Using Mobile Phone Data.” *Journal of The Royal Society Interface* 12 (107): 20150185. <https://doi.org/10.1098/rsif.2015.0185>.
- Tseng, Gabriel, Ruben Cartuyvels, Ivan Zvonkov, Mirali Purohit, David Rolnick, and Hannah Kerner. 2024. “Lightweight, Pre-Trained Transformers for Remote Sensing Timeseries.” arXiv:2304.14065. Preprint, arXiv, February 5. <https://doi.org/10.48550/arXiv.2304.14065>.
- UNSTATS. 2019. *Handbook on the Use of Mobile Phone Data for Official Statistics*. United Nations. <https://unstats.un.org/bigdata/task-teams/mobile-phone/MPD%20Handbook%2020191004.pdf>.
- Watmough, Gary R., Charlotte L. J. Marcinko, Clare Sullivan, et al. 2019. “Socioecologically Informed Use of Remote Sensing Data to Predict Rural Household Poverty.” *Proceedings of the National Academy of Sciences* 116 (4): 1213–18. <https://doi.org/10.1073/pnas.1812969116>.
- Wesolowski, Amy, Nathan Eagle, Abdisalan M. Noor, Robert W. Snow, and Caroline O. Buckee. 2012. “Heterogeneous Mobile Phone Ownership and Usage Patterns in Kenya.” *PLoS ONE* 7 (4): e35319. <https://doi.org/10.1371/journal.pone.0035319>.

Wesolowski, Amy, Nathan Eagle, Abdisalan M. Noor, Robert W. Snow, and Caroline O. Buckee. 2013. “The Impact of Biases in Mobile Phone Ownership on Estimates of Human Mobility.” *Journal of The Royal Society Interface* 10 (81): 20120986. <https://doi.org/10.1098/rsif.2012.0986>.

World Bank. 2025. *The Global Findex Database*. World Bank, Washington, DC.

Yang, Shihao, Samuel C. Kou, Fred Lu, John S. Brownstein, Nicholas Brooke, and Mauricio Santillana. 2017. “Advances in Using Internet Searches to Track Dengue.” *PLOS Computational Biology* 13 (7): e1005607. <https://doi.org/10.1371/journal.pcbi.1005607>.

Yeh, Christopher, Anthony Perez, Anne Driscoll, et al. 2020. “Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa.” *Nature Communications* 11 (1): 2583. <https://doi.org/10.1038/s41467-020-16185-w>.

Yoshida, Nobuo, and Danielle Victoria Aron. n.d. *Enabling High-Frequency and Real-Time Poverty Monitoring in the Developing World with SWIFT (Survey of Wellbeing via Instant and Frequent Tracking)*.

Zheng, Zhuo, Timothy Wu, Richard Lee, et al. n.d. *Dynamic, High-Resolution Wealth Measurement in Data-Scarce Environments*.



WORLD BANK GROUP