INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

PROJECT NOTE | MAY 2019

OPPORTUNITIES AND CHALLENGES IN FIELD DATA VALIDATION AND CORROBORATION

Matching household survey data with project monitoring data in Ethiopia

Simrin Makhija, David J. Spielman, Gashaw T. Abate, and Tanguy Bernard

The application of digital tools to agricultural extension and advisory services has attracted considerable attention in recent years. Among the many advantages it offers is the ability to capture and analyze large amounts of data on farmers, their farms, crops, and fields; and the choices they make about input use, technologies, management practices, and marketing.

As Digital Green and the Government of Ethiopia continue to refine and expand their video-mediated approach to agricultural extension provision[i] and larger digital platforms for agricultural advisory services, an important question being asked is whether the data being captured and analyzed through this approach can help inform decision-making in the extension system in a timely manner. This note examines the utility of monitoring and evaluation data by examining the relationship between Digital Green's monitoring data—generated to support and manage the video-mediated extension approach—and household survey data, which was collected in early 2018 as part of an impact evaluation being conducted by the International Food Policy Research Institute (IFPRI)[ii]. Findings highlight opportunities to strengthen interoperability between monitoring and evaluation data and offer insights for similar initiatives that integrate agricultural extension with information and communication technologies (ICTs).

THE VALUE OF MONITORING DATA

An integral part of Digital Green's model is its "Connect Online Connect Offline" (COCO) platform[iii], an open-source customer relationship management system. COCO is at the heart of Digital Green's back-end data and analytics infrastructure, integrating field-based data collection interfaces designed to work in situations where internet connectivity is low or variable, with databases and dashboards for monitoring program performance.

COCO generates a variety of metrics and indicators for program monitoring. It tracks the videos produced and video screenings held by extension agents, while tracking key attributes of the farmers who participate in screenings, including indicators such as age and sex of program participants, number of video screenings they attend, and number of recommended agricultural practices they adopt.

These data are also potentially valuable for the design and implementation of program evaluations, including evaluations designed to assess program impacts on key outcomes such as changes in farmers' awareness and knowledge about the agricultural practices being promoted, their adoption of those practices, and the productivity or welfare effects resulting from adoption of the practices. At very least, COCO data and evaluation data can be used together to validate whether video screenings were, in fact, conducted, and whether participants derived any value from the information contained in the screenings. More ambitiously, COCO data and evaluation data used together can help identify operational weaknesses in the program and inform the design of possible remedies. This is particularly important as the Government of Ethiopia explores ways to enroll farmers in a digital platform that improves service delivery through input markets, extension services, and commodity-specific value chains.

In the current context, we focus on matching COCO data with household survey data being collected for the IFPRI impact evaluation (discussed in more detail below). There are at least two specific reasons why matching monitoring data with evaluation data can be useful in the current context. One reason is that matching exercises allow project implementers and evaluators to cross-check the validity of field data, identify where differences in data might originate from, and make operational changes to improve data capture protocols in the field. Another reason is that the use of COCO data in conjunction with multiple household survey rounds allows implementers and evaluators to monitor what happens to farmers between survey rounds, for example, in terms of participation in video screenings, exposure to the knowledge transmitted during these sessions, and their decisions to adopt (or not adopt) new practices.

DATA AND DATA SOURCES

Data for the matching exercise described here was collected as follows. In early 2018, IFPRI conducted a survey of farm households as part of an ongoing cluster randomized controlled trial designed to evaluate the impact of the video-mediated extension approach being piloted by Digital Green and the Government of Ethiopia. The trial itself was implemented in the four most agriculturally important regions of Ethiopia during the 2017 meher (rainy) season, with the aim of exploring impacts on three priority crops (teff, wheat, maize) and three key agricultural practices (row planting, more precise seeding rates, and urea fertilizer dressing).

The associated household survey conducted as part of this trial obtained detailed information from 2,422 house-holds, of which 1,610 resided in kebeles in (and were members of development groups with) which development agents (DAs) screened videos on the focal crops and technologies. The remaining households were allocated to the control group and received information from DAs via the conventional extension approach.

During the household survey in 2018, each field enumerator was given a list containing the name and COCO identification number of farmers in kebeles that were assigned to the trial. During the enumerator's interview with each farmer, the enumerator checked the list for that farmer's name, and entered the farmer's COCO ID into the survey questionnaire response form[iv]. Where a COCO ID could be identified for the farmer, IFPRI then matched the house-hold in the survey against the COCO data to check for correspondence between responses. Emphasis was placed on checking data on video screening attendance and adoption of practices. Keep in mind, however, that while Digital Green's goal is to reach all farmers in its program area, not all farmers are likely to be documented in the COCO database at this stage of operations.

It is important to note that for the purposes of the IFPRI evaluation described above—to allow for credible analysis of outcomes attributable to the video-mediated approach—the sample of households selected for the survey was randomly drawn from among members of development groups in treatment kebeles and not purposefully drawn from household lists found in the COCO database.

The COCO data collection process worked as follows. The IFPRI team first sent a data request to Digital Green. Digital Green's technology team then queried the MySQL database that stores the incoming COCO data based on IFPRI's data request. The result was person-level data that included information on region, woreda, kebele, village, group name, person id, person name, age, gender, number of adoptions, and number of screenings attended in the 2017 meher growing season.

FINDINGS

We report findings for those farmers who (i) resided in kebeles and were members of development groups in which video screenings were conducted, and (ii) were interviewed as part of the household survey (n=1,610). These are farmers who were sampled from the treatment groups of the IFPRI impact evaluation.

Table 1 shows that COCO IDs were matched between the COCO data and the household survey data for 49 and 46 percent of the trial's two treatment groups, respectively. We posit that this is a decent match rate when we account for the fact that Digital Green and the Government of Ethiopia are still in the process of scaling the COCO platform and the video-mediated extension approach.

Table 1 Digital Green COCO data matched to household surveydata

	No., % of observations	Not matched	Matched	Total
Treatment group	**n**	**848**	**762**	**1,610**
	%	**52.67**	**47.33**	**100**

Next, we examine the match rate among farmers who reported watching a video in the household survey (**Table 2**). As expected, we find a higher match rate for farmers who watched a video when compared to those who did not. 56 percent of farmers in the treatment group who watched a video had COCO ids that matched, while 44 percent of farmers who watched a video do not have a corresponding COCO id. Of those that did not watch a video, about 41 percent of farmers have COCO ids.

Table 2 Relationship between matched households and those that watched a video for treatment group

Watched a video	No., % of observations	Not matched	Matched	Total
No	n	552	388	940
	%	58.72	41.28	100
Yes	n	296	374	670
	%	44.18	55.82	100
Total	N	848	762	1610
	%	52.67	47.33	100

A possible explanation for the imperfect match observed in the data is the frequent turnover and changes in mem-bership of development groups, which are small groups of farmers that are organized within kebeles and which are the forums in which video screenings are conducted.

To determine what predicts a match on COCO ID we run an ordinary least squares estimation with "COCO ID matched" as the dependent variable and a vector of household-level characteristics as the explanatory variables (

Table 3). COCO IDs are more likely to have matched for male-headed households and households that cultivate maize, and less likely to have matched for households further away from a dry season road, farmer training center, and a DA house/office. This suggests that these variables predict the concordance between COCO data and house-hold survey data. One interpretation of this is that there are systematic biases in how data are collected, suggesting greater accuracy for certain types of households.

Table 1 Table 3 Determinants of a COCO ID match

Explanatory variables	COCO ID matched	
	Coefficient	p-value
Household size	0.00931	(0.00584)
Male household head	0.0962***	(0.0339)
Household head age	-0.00112	(0.000917)
Household head received formal education	-0.0316	(0.0250)
Main roof material is corrugated metal	0.00567	(0.0314)
Number of distinct rooms	-0.00528	(0.00470)
Distance to nearest asphalt/tar road (one-way minutes)	-0.000158	(0.000200)
Distance to nearest dry season road (one-way minutes)	-0.000686***	(0.000222)
Distance to nearest all-weather road (one-way minutes)	0.000545	(0.000387)
Distance to nearest market place (one-way minutes)	0.000243	(0.000221)
Distance to woreda administrative center (one-way minutes)	-0.000133	(0.000291)
Distance to nearest agricultural cooperative (one-way minutes)	-2.62e-05	(8.68e-05)
Distance to nearest agro-input dealer (one-way minutes)	7.63e-05	(0.000192)
Distance to nearest farmer training center (one-way minutes)	-0.000561**	(0.000278)
Distance to nearest DA house/office (one-way minutes)	-0.000845*	(0.000434)
Distance to nearest savings and credit coop (one-way minutes)	0.000108	(0.000277)
Distance to microfinance institution (one-way minutes)	-0.000234	(0.000146)
Distance to a bank (one-way minutes)	8.47e-05	(0.000139)
Number of parcels	0.00678	(0.00664)
Did your household cultivate teff in meher 2017-18?	0.00288	(0.0340)
Did your household cultivate wheat in meher 2017-18?	-0.0173	(0.0345)
Did your household cultivate maize in meher 2017-18?	0.0584**	(0.0292)
Constant	0.425***	(0.0840)
Observations	1,610	
R-squared	0.298	

Robust standard errors in parentheses clustered at the *kebele* level. Woreda fixed effects.
***** p<0.01, ** p<0.05, * p<0.1**

Next, we compare data from these two data sources for video attendance and adoption. For this analysis, we focus on observations for which we have COCO data.

Video attendance

The household survey collected data on the number of videos watched from a subset of videos directly relevant to the crops and technologies that were the focus of the impact evaluation. The COCO data contains information on the number of screenings attended by a farmer. While these data are not precisely comparable, they still allow us to explore correlations. We find that these two variables are positively correlated with a correlation coefficient of 0.1135 (significant at the 1 percent level). While this coefficient is not very high, it can be attributed to the fact that these variables are measuring relatively different phenomena.

Adoption

The household survey collected data on the adoption of the three focal technologies described earlier. The COCO platform provides data on the number of adoptions by a farmer, which is calculated as the total number of times a farmer reported having adopted any technology promoted in the videos. Thus, once again, these two variables are not perfectly comparable, but give us the opportunity to examine correlations. We estimate two correlation coefficients based on two different assumptions. The first excludes missing adoption data in COCO: the correlation coefficient equals 0.1440 and is significant at the 5 percent level. The second includes missing adoption data in COCO and assumes that they are equivalent to non-adoption: the coefficient equals 0.1682 and is significant at the 1 percent level. This shows fairly close concordance between the household data and COCO data.

CONCLUSION

The main takeaway from this analysis is that in order to effectively capitalize on the complementarities between monitoring data and household survey data, both the survey questionnaire and the monitoring framework need to be in sync such that they collect information on the same indicators. In the absence of this, the best we can do is run simple correlations to estimate if similar variables are related. Given that IFPRI's study is focused on specific technologies and crops, collecting data on the whole gamut of technologies and screenings to mirror the data col-lected by COCO would have been a challenge to implement.

To improve the match between COCO monitoring data and household survey data and to create better opportuni-ties for comparing similar key indicators across the two data sources, IFPRI modified its approach in the 2019 household survey, by introducing the following steps in its survey procedures.

1. Verified the accuracy of the list of development group members in the kebele by consulting with Digital Green staff, DAs, and development group heads prior to conducting the household surveys, with specific reference to a new sub-sample of households that were not in treated development groups but may have benefited from within-kebele spillover effects.

2. Emphasized the importance of recording COCO IDs to enumerators during the survey training to try and achieve the maximum number of matches possible while enumerators are in the field.

By employing these improved data collection practices, it is expected that the match rate will improve considerably, thereby allowing for comparisons to be made between a larger sample of COCO data and household survey data. This analysis will, in turn, strengthen the continuous monitoring and analysis of the video-mediated extension ap-proach and the evaluation of its impact on its target population.

Simrin Makhija *is a senior research analyst at IFPRI in Washington, DC, USA.* ***David J. Spielman*** *is a senior research fellow at IFPRI in Washington, DC, USA.* ***Gashaw T. Abate*** *is a research coordinator at IFPRI in Addis Ababa, Ethiopia.* ***Tanguy Bernard*** *is a senior research fellow at the International Food Policy Research Institute (IFPRI), and professor in the Groupe de Recherche en Économie Théorique et Appliquée (GREThA), University of Bordeaux, France.*

[i] The Digital Green approach is being undertaken in partnership with the Ethiopian Ministry of Agriculture (MoA), the Agricultural Transformation Agency (ATA), the Ethiopian Institute of Agricultural Research (EIAR), and regional bureaus of agriculture, with funding from the Bill & Melinda Gates Foundation, and the U.S. Government's Feed the Future program via USAID's Developing Local Extension Capacity (DLEC) project.

[ii] Abate, G., T. Bernard, S. Makhija, and D.J. Spielman (forthcoming in 2019), Accelerating Technical Change through Video-mediated Agricultural Extension: Evidence from Ethiopia. IFPRI discussion paper. Washington, DC: International Food Policy Research Institute.

[iii] "COCO: a web-based data tracking architecture for challenged network environments" discusses more details about the platform and can be accessed here—https://dl.acm.org/citation.cfm?id=1926189

[iv] The survey questionnaire response form itself developed using SurveyCTO, so all data captured by the enumerators was immediately entered in a digital format.

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

A world free of hunger and malnutrition

1201 Eye Street, NW, Washington, DC 20005 USA | T. +1-202-862-5600 | F. +1-202-862-5606 | Email: ifpri@cgiar.org | http://www.ifpri.org/