



WORKING PAPER

CGAP Smallholder Financial Diaries

Crop Tracker Datasets User Guide

Laura Cojocaru

July 2017

CGAP launched the year-long Financial Diaries with Smallholder Families to elucidate the financial lives of smallholder households and build the evidence base on this important client group. The study, conducted between June 2014 and June 2015, captured the financial and in-kind transactions of about 270 households in impoverished northern Mozambique, the fertile farmlands of western Tanzania, and Punjab province, the breadbasket of Pakistan. CGAP retained the services of Bankable Frontier Associates (BFA) to manage the project. For in-country data collection, BFA worked with International Capital Corporation in Mozambique, Digital Divide Data in Tanzania, and RCons in Pakistan.

The Diaries methodology combines in-depth quantitative and qualitative research. Research teams met participating families about every two weeks to collect granular data on cash flows in and out of the household, their financial tools, assets, major life events, and attitudes toward agriculture and financial services.

The [Smallholder Financial Diaries: Datasets User Guide](#) (2016) details the research methodology and the interpretation of the datasets. Building on that foundational document, this User Guide to the Crop Tracker Datasets details the use and interpretation of these unique datasets.

THE CROP TRACKER

Sales of crops by households were captured in the cash flow datasets. The Smallholder Diaries crop tracker, however, allows for a better understanding of timing related to crop transactions and offers quantitative details on harvests, consumption, sales, losses, and amounts given away for all the crops grown by a household. It also paints a picture of crop fluctuations over time and illuminates the extent of dependence on self-consumption and the magnitude of crop losses.

Working with the datasets

The crop tracker was an addition to the Financial Diaries database developed specifically for the Smallholders Diaries project. The cash flows datasets record information on crops that were sold over the duration of the study or had related input investments. The crop tracker datasets, however, go further by providing information on all crops, including those that were not sold by the house-

hold. Furthermore, the crop tracker records a larger set of crop transactions, including those that were not monetized, such as harvests that were not yet sold, amounts consumed, lost, etc. All data was self-reported. For transactions other than sales, households were asked to value the transaction based on their knowledge of current market prices.

See Section 9 in the [Smallholder Financial Diaries: Datasets User Guide](#) (2016) for further information on working with the datasets.

Note: Due to a technical issue, in the Smallholder Diaries Pakistan crop tracker, dataset transactions related to crops given away lack both the interview date and the value. The values could theoretically be imputed by using prices for the same crop as recorded by sales in the cash flows datasets around the same time.

See Section 9.1 in the [Smallholder Financial Diaries: Datasets User Guide](#) (2016) for further information on the datasets.

Datasets overview

Dataset name	Tanzania crop tracker.dta Mozambique crop tracker.dta Pakistan crop tracker.dta
Description	Tracked over time each crop grown by the household and recorded what quantities (in kgs) were harvested, consumed, sold, lost, and given away, as well as the corresponding self-reported value in the local currency. The data did not record each transaction that occurred; instead, it recorded the total quantities and amounts, by transaction type, over periods of roughly two weeks—the time between researchers' visits.
Number of variables in the dataset	7
Number of observations	Mozambique: 8,127 Tanzania: 3,571 Pakistan: 3,222
Period	Mozambique: May 2014-June 2015 Tanzania: May 2014-July 2015 Pakistan: July 2014-June 2015

General structure of the datasets

The crop tracker datasets are unbalanced panels that record changes affecting crop amounts and corresponding values.¹ Observations are by household, crop, and transaction type. The data are attributed to a household, not to specific household members. Each observation represents a change that affected a crop over a period of approximately two weeks—the time between two interviews. The date connected to each data point is the date when the interview took place, not the actual date when the transaction occurred.

To create the datasets' structure, researchers initially recorded the following for each household:

- Each crop grown by the household
- The general crop category (e.g., rice, wheat)
- The initial store of each crop

In subsequent interviews, researchers recorded changes to these crops, as well as any new crops that households started to grow in the interview period, or that they failed to report initially. Changes were organized into the following types of transactions:

- Harvest
- Sale
- Consumption
- Loss
- Given away

For each transaction, the researchers recorded:

- The self-reported quantity
- The self-reported value of the quantity, in the local currency. Researchers instructed the respondents to base reported values on market prices, but did not verify these amounts.

¹ Households were interviewed on different dates and with frequency of approximately two weeks, but not with exact periodicity.

Variables

Variable name	Variable label	Description	Type Codebook
household_code	Household code	Unique code for households. It can be linked to other Financial Diaries datasets.	String
interview_date	Interview date	Date when the interview was realized.	Date
id	Crop name	An identifier given to each crop. The first part is CROP followed by a number given to each crop within each household. The number signifies only the order in which the crops were recorded by the interviewer. The word(s) in parentheses are free entries used to help the interviewers and the household members easily know which crop they are discussing. This crop identifier is unique only at the household level. Cannot be linked to cash flows data.	String
crop	Crop type	Crop category (e.g., rice, wheat). Customized for each country to capture the most common crops grown by the households.	Numeric, with labels Unique values Mozambique-12 Tanzania-13 Pakistan-7
trans	Type of transaction	Description of the type of change affecting the crop	Numeric, with labels 1 consumption 2 given away 3 initial store 4 loss 5 new harvest 6 sale
quantity	Quantity	Self-reported quantitative change, in kilograms (with exceptions).	Numeric
value	Value	Self-reported value of the change, in local currency.	Numeric

Note: Quantities were generally recorded in kilograms. However, there are several exceptions. For some crops, it was impossible for respondents to estimate quantities in kilograms since kilograms are not the usual measurement unit for these crops. Examples include bananas, spinach, and fodder. For such crops, quantities may be missing, or expressed as bunches, bales or other popular measures.

See Section 9.2 in the [Smallholder Financial Diaries: Datasets User Guide](#) (2016) for further information on the structure of the datasets.

See Section 9.4 in the [Smallholder Financial Diaries: Datasets User Guide](#) (2016) for further information on the variables.

household_code	interview_date	id	crop	trans	quantity	Value
TZFIL10	26-Jun-14	CROP01 (Rice)	Rice	initial store	2,000	600,000
TZFIL10	15-Jul-14	CROP01 (Rice)	Rice	consumption	40	12,000
TZFIL10	15-Jul-14	CROP01 (Rice)	Rice	sale	20	6,000
TZFIL10	15-Jul-14	CROP01 (Rice)	Rice	given away	20	6,000

Examples of how transactions are recorded

The table below shows a data excerpt from the Tanzania Smallholder Diaries crop tracker dataset. It refers to a rice crop of household TZILF10. When the interviewers first discussed the rice crop with the household, it already had in store 2,000 kg of rice. By the next interview, the household had consumed 40 kg, sold 20 kg and given away 20 kg over the previous two-week period. On both occasions, the household estimated that a kilogram of rice was worth TZS 300.

See Section 9.9 in the [Smallholder Financial Diaries: Datasets User Guide](#) (2016) for further information on how transactions are recorded.

Data cleaning

The crop tracker datasets were cleaned post data collection to the extent that errors and omissions were clearly identifiable. For example, where possible, miss-

ing quantities and values were filled in, using information from previous interviews or information extrapolated from data from other households. Clearly erroneous data (such as data points containing extra zeros) were corrected.

As such, the following number of corrections were made: Mozambique 171 corrections, Tanzania 106 corrections, and Pakistan 79 corrections. Nonetheless, the data were self-reported and no other in-the-field controls were implemented. The data may still have instances of erroneous data entry in cases where post data collection corrections were not obvious.

Note for researchers: Respondents may not have had a good knowledge of prevailing market prices (e.g., when estimating the value of self-consumption), may not have been familiar with estimating quantities precisely, have trouble calculating total amounts, etc. Even with post-collection cleaning, data should still be used with caution when trying to make any precise estimations.