FSDZ Zambian Financial Diaries Project 2015 User Guide

Version 1.2

Alex Montgomery, DataFirst

December 2016

Contents

List of Figures

List of Tables

1 Introduction

In 2014, Financial Sector Deepening Zambia (FSDZ) commissioned Microfinance Opportunities (MFO) to implement the Zambia Financial Diaries Project (FDP). The intention of this was to understand how low-income people in Zambia managed their cash flows and how they utilised transfers, savings, loans and insurance to do so. The approach adopted by MFO was broadly based on the financial diaries methodology introduced in Rutherford (2000).

The financial diaries approach involves sending fieldworkers to a cohort of respondents every week for a year and documenting the various transactions performed during that week. These transactions can include, "purchases and sales of goods, sources of income, uses of financial goods, and in-kind transactions (Stuart *et al.*, 2016a, p. 3)." It also requires the recording of any significant events in the lives of respondents during the week. Examples of such events include things like funerals, weddings, and medical issues. Between November 2014 and mid-December 2015 eleven fieldworkers visited 355 respondents in four Zambian provinces (see Figure 1) conducting 16,510 interviews over the course of 58 weeks using the financial diaries approach. Complementary cross-sectional and in-depth interviews were also conducted to better, "understand respondents' attitudes towards asset building, risk management, and financial service providers (Stuart *et al.*, 2016b, p. 3)." This guide aims to achieve the following:

- Document the relationship between all of the aspects of the data generating process and the data files.
- Concisely convey the kinds of information available in each data file
- Empower users of the data to quickly and easily connect the various data files with one another.
- Provide a single resource for users interested in understanding and using this study for their own research.

2 Financial Diaries Methodology

2.1 Sampling Methodology

The sampling frame for the FDP was developed under certain logistical constraints. Stuart *et al.* (2016b) note that while there was no firm rubric for their selection, "the priority was to develop a sample that, while not statistically representative, was still reflective of the varying levels of financial service access and livelihoods of low-income Zambians. We selected four provinces Copperbelt, Eastern, Lusaka, and Western Provinces that contained a diverse mix of urban and rural respondents, various levels of financial access, and a preponderance of individuals involved in informal businesses (Lusaka Province), the mining sector (Copperbelt Province), or farming (Eastern and Western Provinces)." The choice of provinces were made with input from MFO, Ipsos, and FSDZ.

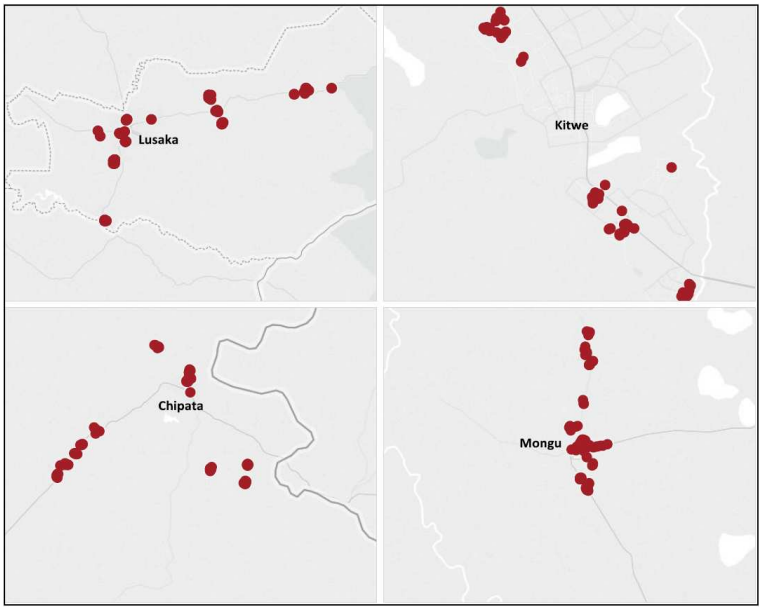

Figure 1: Location of respondents (Stuart *et al.*, 2016b)

From those provinces districts were selected based on further conditions. First, any selected district needed to have a town with sufficient services to support a fieldwork team for a year that was within an hour and a half of all field sites. Once those constraints were satisfied standard enumerator areas (as defined in the master sample developed in the Zambian 2010 Census) were randomly selected from the pruned set of districts in the provinces mentioned above. Households were then selected within those enumerator areas using a random walk. Respondents within households were chosen using a Kish grid as per Kish (1949) and screened for eligibility with an enrolment questionnaire. This questionnaire had certain requirements that needed to be fulfilled for the respondent to be included in the sample. For example, if the respondent was going to be away for the majority of the year the interview was terminated and the respondent was excluded.

2.2 The Field Entry Data Sheet

Most of the data generated in this dataset is the result of respondents filling out notebooks on a weekly basis. Those notebooks were then used to create official records under the supervision of fieldworkers. The intervention of fieldworkers every week ensured that notebooks were filled out diligently, with respondents probed regarding the accuracy of their entries.

The purpose of the financial diaries methodology is to get a picture of all transactions and events that occurred within the household in each week of observation. Multiple adult household members were interviewed with the intention of capturing the gamut of transactions that occurred within the household during any given week.

2.2.1 Definition of a Household

Households were defined as people who usually live and eat together out of the same home from the same kitchen. (Microfinance Opportunities, 2014) go on to further describe what constitutes a household as follows:

> Two households can share the use of a building but be separate because they cook their food in separate kitchens. Include only the people who have lived and eaten here for *all of the past 6 months.* Also include all spouses, children who are away at school (who are still supported by the household head) and the household head even if he or she has not lived in the household for the past 6 months. Do not include married sons or daughters living separately. Do not include household servants or domestic laborers
>
> Exceptions:
>
> - Always include new spouses, even if they have lived in the household less than 6 months
> - Always include infants, even if they were born less than 6 months ago
> - Always include household members living in an institution elsewhere, but who are still dependent on the household for support (e.g. boarding school students)
> - Always include adopted children even if they have lived in the household for less than 6 months
>
> Note: a single household may live in more than one dwelling (e.g. compound).

2.2.2 Household Members Interviewed

Originally the plan was to interview all adults and youths (people between the ages of 13 and 18) that regularly conducted transactions. In practise, however, only those ages 18 and above were interviewed. Usually, this was the head of household and his or her spouse.

Household heads that did not regularly reside within the domicile were excluded, but adult members of the household who were "economically active and intermittently at home [were] interviewed. (Microfinance Opportunities, 2014, p. 1)" If fieldworkers missed a week's data they were instructed to gather that information the following week and include them on separate survey instrument data sheets.

2.2.3 Information Captured by the Paper Instrument

The paper instrument captures even the smallest transactions which are aggregated with similar goods in the week and recorded as one transaction in the data. Enumerators were instructed to progress through the instrument instructions methodically. To aid with this the paper instrument is broken down into several sections. Section A captures cash outflows. All transactions in which there was an outflow of money - goods or services

purchased, loans repaid, loans made, savings deposited, or cash gifts given are recorded here.

Section B covers all inflows - goods or services sold, loans repaid to the subject by a borrower, loans received, savings withdrawn, and cash gifts received. There is also a section on barters, in-kind gifts, and transfers from storage. Summarily, the financial diaries paper instrument captures the following exchanges:

- External to household
  - purchases of goods and services
  - exchanges of goods and services for other goods and services (bartering)
  - gifts
  - loans
  - savings deposits
- Within household
  - transfers of money between household residents (intra-household transfers)
  - withdrawals or inputs into food storage

The paper instrument also aims to capture any unusual events that may have occurred during the week. These events could be things like funerals, weddings, social events, or family illnesses.

3 Data Files in the Dataset

There are five data files in the dataset generated by three separate survey instruments. These are summarised in Figure 2 below. Note here that the panel datafile is derived from the events and transactions data files. These were themselves generated from information recorded in the financial diaries paper instrument Microfinance Opportunities (2014). Each datafile is described in turn below.

3.1 Transactions

This (and the events datafile) are the most raw version of the data available to users representing straight captures of the financial diaries paper instrument. Each row entry in this datafile represents some transaction that happened either once or on multiple occasions during the preceding week. If the transaction happened only once during the week, there is a variable to identify the day on which it happened. If the transaction occurred multiple times during the week the number of transactions that occurred during that week is represented instead.

An abbreviated example of the sort of information one would expect to see in the transactions datafile is given in Table 1. We observe a single purchase of 25kg of mealie meal on the Monday of that week and multiple purchases of relish and groceries throughout

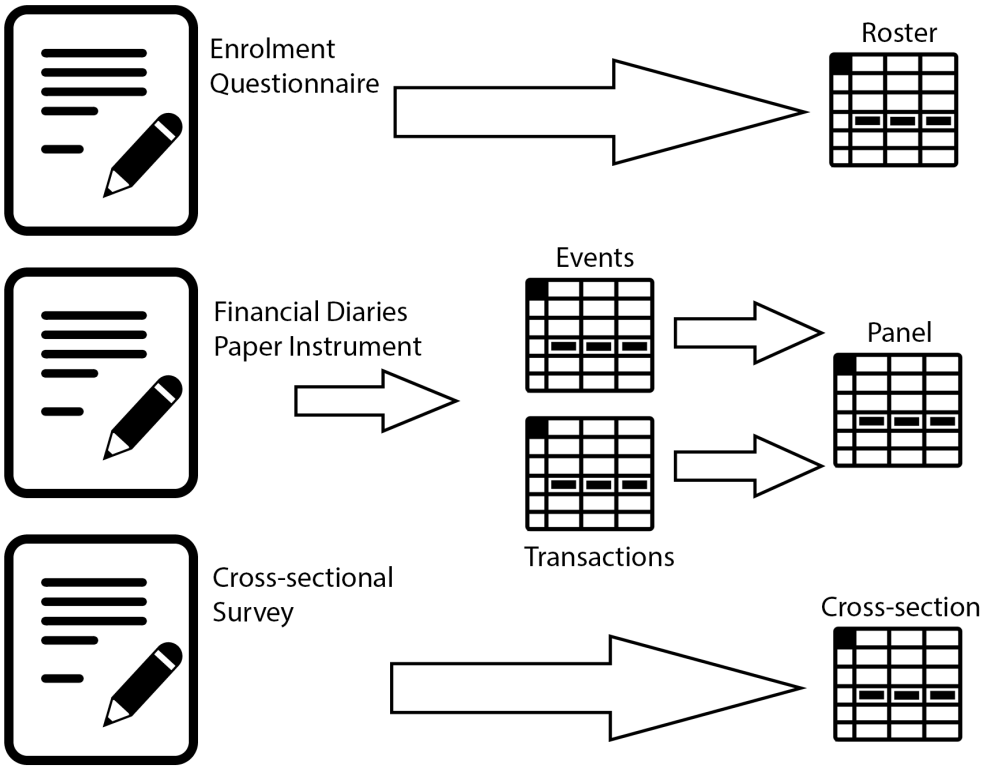

Figure 2: The relationship between the survey instruments and data

the week. Note here that the amount, in ZMW, has already been aggregated for those multiple transactions. We also see an inflow of money from what the sale of firewood that happened on multiple occasions over the week.

Table 1: Stylised example of a week of transactions

RespID	**week**	**day**	**daycount**	**type**	**item**	**qty**	**unit**	**amt**
1001.1	30	Monday	N/A	Outflow	Mealie Meal	25	Kg	80
1001.1	30	Multiple	7	Outflow	Relish	15	N/A	60
1001.1	30	Multiple	7	Outflow	Groceries	12	N/A	30
1001.1	30	Multiple	7	Inflow	Firewood	33	Bundle	175

Roughly 97% of the 124,017 transactions captured are simple inflows and outflows of the type represented in Table 1, but there is also information on the other transaction types described in Section 2.2.

The transactions datafile also contains information on the party conducting the transaction. While the RespID (respondent ID) identifies the person completing the interview that generated the diary entry for that week, the *transact_person_transacting* variable identifies the member of the household who actually conducted the transaction. One can also find information on the characteristics of the respondent (their livelihood, age, gender, etc.) as well as the channel through which the transactions were conducted. Other aspects of the transaction captured by the datafile are the characteristics of the person or institution with which the exchange occurred and the place of that exchange. For more detailed information on exactly how all the different transactions were captured please consult the financial instrument instruction sheet (Microfinance Opportunities, 2014).

3.2 Events

The events datafile aims to capture any unusual events that occurred during the week. Examples of such events are things like medical issues, children getting sick, funerals, parties, and weddings. Each row entry represents one of these events and one can identify the week in which that event occurred. They were sorted into various categories that are captured by a number of different variables. These variables convey information on whether or not the event,

- was a shock,
- happened to the respondent directly,
- led to a loss of income,
- resulted in expenditure,
- required travelling,
- brought about income gain,

- or happened to a peripheral party (i.e. did not happen to the respondent or their family but someone in the community).

There are 2,571 such events in the datafile.

3.3 Panel

The panel datafile is a refined version of the panel and events datafiles. It provides a summary of each week for each household combining information about the various events and transactions that occurred. It also relates some information on the household itself.

The variables break down the various sources of inflow and outflow in detail and provide a count of each type of transaction and event that occurred. This separation of transactions is quite granular, distilling the more complex information in the transactions and events data files into usable numeric quantities while preserving relevant detail by collapsing the information onto a single row.

3.4 Roster

The roster datafile contains the data generated from the initial enrolment survey conducted by Ipsos Zambia. This questionnaire was constructed as a way to screen respondents according to several criteria, most of which were informed by the reasons mentioned in Section 2. The variables captured in this datafile naturally reflect its intended function, with information on the material conditions, educational attainment, financial service usage, occupation, and income of respondents.

3.5 The Cross Sectional Survey

At the end of the financial diaries project (December 2015) enumerators conducted a cross-sectional survey which was used to generate the cross-section data file. This was a far more in-depth survey instrument than the enrolment questionnaire. In this dataset users can find information on,

- the make-up of the household,
- demographic characteristics of respondent,
- asset ownership,
- use of assets as investments,
- asset financing decisions,
- insurance purchasing behaviour and knowledge,
- retirement savings decisions,
- perspectives on savings groups,

- perceptions of shocks,
- depositing patterns,
- earnings and occupation (for all members of household),
- general savings behaviour,
- farming in the household,
- micro-retail behaviour,
- propensity to permit credit,
- and energy and water consumption.

Note that there are only 321 observations in this datafile as some households were missed in the cross-sectional interview stage. These missing respondents are present in the panel datafile, however.

4 Linking the DataFiles

Users will most likely want to combine information from the different data files in their analysis. This is quite easy to do using whichever merging functionality available in their preferred statistical analysis software. To do this successfully users need to know which variable(s) uniquely identify observations in each dataset. This is summarised in Table 2 below.

Table 2: Unique identifiers in each data file

Datafile	**Unique identifying variable(s)**
Roster	RespID
Cross-section	RespID
Panel	RespID, Week
Transactions	transact_id
Events	events_id

Table 3 summarises how all of the data files fit together. The bottom-left half of the matrix relates the ratio of the datafile in the row heading to the datafile in the column heading. For example, the cross-section datafile can be merged with the roster data file on a 1:1 basis. This is because every row entry in both the roster datafile and the cross-section datafile is uniquely identified by the variable *RespID*. To provide a further example of how to read Table 3: the panel datafile, which is uniquely identified by *RespID* and *Week*, can be merged with the cross-section datafile on a m:1 (many-to-one) basis. There are multiple panel data rows which must be **each** be assigned single row entries from the cross-sectional data file. The upper-right half of the table tells us that the variable to use to merge on this m:1 basis is *RespID*.

Finally, it is worth noting that the transactions and events datafiles should be appended to one another if users would like to combine them. This is because they're effectively different kinds of the same type of data entry - uniquely identified only by their specific identifiers and potentially occurring multiple times within the same week for the same respondent. For any merging related issues users may contact DataFirst support at support@data1st.org for further instruction.

Table 3: How the Data Files Are Related

	Roster	**Cross-section**	**Panel**	**Transactions**	**Events**
Roster	**-**	RespID	RespID	RespID	RespID
Cross-section	1:1	-	RespID	RespID	RespID
Panel	m:1	m:1	-	RespID, Week	RespID, Week
Transactions	m:1	m:1	m:1	-	APPEND
Events	m:1	m:1	m:1	APPEND	-

5 Known Data Quality Issues

5.1 Transactions

Date start and date end variables are fuzzy (there are not always seven days in a week, the weeks don't always begin on the same day) which is most likely explained by data capturing errors on the part of the fieldworker. Most of the week lengths, when evaluated, come to seven days (as expected) but not all. For the user the more reliable measure of the week of observation is the transact_week variable.

Some transactions were incorrectly listed as occurring within weeks that started in 2105. This was assumed to be a number transposition error and corrected to 2015. Evidence for simple number transposition were bolstered by observing that these rows had their week endings listed as being 90 years prior but in seven days (e.g. transact_week_end was 22 July, 2015 but transact_week_start was listed as 15 July, 2105). Some transact_week_end cell values were also apparently incorrect in a similar fashion. In these cases the week end date was one year and seven days in the future (e.g. week starting 08 July 2015 and ending 14 July 2016). These were corrected by changing the year value from 2016 to 2015.

5.2 Roster

The wards variable seems to be imperfectly captured as many responses do not match lists of recorded Zambian electoral wards.Efforts have been made to make the entries more readable but are imperfect. The phone access variable is also bit misleading. There are 58 missing values for the variable *roster_phoneaccess_or_own* which seem to have a corresponding follow-up response with variable *roster_accessph_only*. It is unclear what

the *roster_accessph_only* variable is meant to represent because it is not on the enrolment questionnaire.

5.3 Cross-section

There were two sets of duplicates in terms of *RespID* in the original cross-sectional data file. Investigating further, it was discovered that one set of duplicates is the apparent result of the same fieldworker visiting the same household twice (the second visit occurring eight days after the first). This policy adopted here was to choose the latest observed row in the data. Notably, there were a few variables that had different values between the two observations. These are easily attributable to actual dynamics within the household. There were no substantive differences between static household characteristics.

The other set of duplicates was slightly more complicated. It involved the apparent incorrect assignment of the *RespID* code (that is, the *RespID* code was assigned correctly for one entry and incorrectly for the other). Fortunately, it was possible to check the correct *RespID* using the other data files. It turned out that the incorrectly assigned entry was meant to be represented by another code entirely. This is most likely a data capturing error. Correcting the incorrectly coded *RespID* yielded another set of two duplicates. The correctly coded entry of these two duplicates was the one used in the final file.

5.4 Reporting Further Data Quality Issues

The identification of data quality issues is an iterative process that relies on usage of the data to uncover anomalies. This is because, often, unusual patterns in the data that may not gel with prior expectations or qualitative insights generated elsewhere only become apparent once the data is scrutinised more closely during the process of analysis. Any users of the data that uncover what they, given their experience and specialised knowledge, deem to be unusual are encouraged to report them to the DataFirst support team which can then investigate the issue with those with more information regarding the data collection. This can be done by emailing DataFirst support (support@data1st.org). Efforts will be made to resolve any queries and add to the corpus of knowledge of data quality issues for this dataset. These insights will then be added to future versions of this document.

Bibliography

Kish, Leslie. 1949. A Procedure for Objective Respondent Selection within the Household. *Journal of the American Statistical Association*, **44**(247), 380–387.

Microfinance Opportunities. 2014. *Financial Diaries Instrument Instructions.*

Rutherford, Stuart. 2000. *The Poor and Their Money.* Oxford University Press New Delhi.

Stuart, Guy, Noggle, Eric, & Galagher, Conor. 2016a. *Zambia Financial Diaries: Interim Report.*

Stuart, Guy, Noggle, Eric, & Galagher, Conor. 2016b. *Zambia Financial Diaries: Managing Money in the Face of Risk and Uncertainty.*