DISCLOSURE ANALYSIS PLAN
AND RECORD OF IMPLEMENTATION
FOR THE TAJIKISTAN 2012-2013 POPULATION-BASED SURVEY
FTF FEEDBACK PUBLIC USE DATASET

Under the terms of the Executive Order *Making Open and Machine Readable the New Default for Government Information* and the Administration's *Open Data Policy*, the USAID Bureau of Food Security (USAID/BFS) is making data from Feed the Future FEEDBACK (FTF FEEDBACK) performance monitoring surveys available as open and machine-readable public use datasets after publication of the country reports.

An essential step in the preparation of FTF FEEDBACK datasets for public use is ensuring respondent privacy and confidentiality. Respondent privacy and confidentiality is ensured and protected by removing identifying information from the datasets, including *direct identifiers* (information such as names, addresses, Global Positioning System (GPS) coordinates, or any other personally identifying number or characteristic) and *indirect identifiers* (data that do not specifically identify a person or location, but that can be used to do so, one variable at a time or in combination, by uniquely describing a person or household).

The FTF FEEDBACK Data Manager or designee is responsible for developing and implementing the Disclosure Analysis Plan (DAP) for each FTF FEEDBACK dataset intended for public use.

Activity Name: **TAJIKISTAN Baseline PBS Public Use Datasets Preparation**

Prepared By: **Amy Biggi**

Date: **June 20, 2014**

SURVEY DESCRIPTION

FTF FEEDBACK seeks to capture data in the geographic areas targeted by Feed the Future interventions, known as Feed the Future Zones of Influence (ZOI). This population-based survey (PBS) is a product of FTF FEEDBACK, which is responsible for supporting performance monitoring of the Feed the Future initiative.

In Tajikistan, the PBS fieldwork was conducted by the Centre for Sociological Research (Zerkalo). Westat and TANGO International provided technical and training support. The fieldwork took place from December 23, 2012 to January 23, 2013. The ZOI in Tajikistan was comprised of 12 of 24 districts in Khatlon province. A total of 2,000 households in the ZOI were interviewed for the PBS data collection activity. These households were selected randomly from within 100 sampled clusters which were selected based on their probability in proportion to size (PPS).

STEPS TAKEN TO ENSURE COMPLIANCE WITH DATA USE RESTRICTIONS

FTF FEEDBACK reviewed all agreements with USAID/BFS, Government of Tajikistan (GOTJ) Agency on Statistics under the President of the Republic of Tajikistan and Zerkalo as well as the questionnaire informed consent statement to ensure that the public release datasets are in compliance with those agreements and statements. In Tajikistan, Zerkalo does not require survey implementers to undergo Institutional Review Board (IRB) clearance. In countries without IRB requirements, the Westat IRB requires clarification of the rights of children and a clear understanding of how their rights will be protected under local law. Consequently, Zerkalo requested parental consent for the participation of their children under age 18.

Neither the Zerkalo nor local laws place any further restriction on the collection and dissemination of PBS data.

STEPS TAKEN TO PROTECT IDENTITIES OF STUDY PARTICIPANTS

ASSESSMENT OF POTENTIAL LINKAGE TO EXTERNAL DATASETS

The survey was conducted borrowing from the sample design of the 2009 Micronutrient Status Survey (MSS). The sampling frame for that survey was based on Standard Enumeration Areas (SEAs) from the 2010 Tajikistan Census. For the public-release PBS datasets, the household and cluster SEAs are survey-specific, such that it will not be possible to link or merge the PBS dataset with the 2009 Micronutrient Status Survey (MSS), with census data, or with any other existing dataset.

ELIMINATION OF DIRECT IDENTIFIERS
All names of individuals and their addresses have been removed from the data. Only region and district geographic identifiers will be included by name. No telephone numbers, identification numbers, or GPS coordinate data were collected during this survey.

MANAGEMENT OF INDIRECT IDENTIFIERS
Step 1. Production of crosstabulations for selected data items.
Crosstabulations were generated to identify low-frequency, high-disclosure risk variables by the most disaggregated geographic identifier for which names are provided to data users. For FTF FEEDBACK TAJIKISTAN 2012-2013 PBS, this is the district name.

The production of crosstabulations focused on several types of variables:
- Household characteristics readily discernible by casual observation; these would include household construction materials, size of the household, and ownership of assets such as cars or livestock that set the household apart from others.
- Other respondent characteristics that might be documented in other data sources, such as education, occupation and other economic activities.

In the TAJIKISTAN PBS, the indirect disclosure risks that were discovered are summarized in Table 1, which indicates the changes made to reduce disclosure risks.

Step 2. Confirmation of outlier status in larger population-based surveys or censuses.
External data sources were examined for evidence that potentially identifying outliers in the PBS data are actually common among the general population. When this is found to be the case, an outlier in the PBS cannot serve as an identifying characteristic and thus will not be subjected to further disclosure prevention procedures. This approach preserves respondent confidentiality while maintaining data accuracy and usefulness. In the absence of such confirmatory data, all PBS outliers identified in Step 1 would be considered potential disclosure risks.

There are 2 external sources of information that could provide information on whether data suppression or recoding needs to be done for the variables describe in the Step 1: the 2012 Tajikistan Demographic and Health Survey (TDHS) and the 2010 census.

The TDHS did ask questions to elicit information about many of the items listed in Table 1. However, the TDHS could not be used for the present purpose because the districts were not labelled in the

TDHS data. This precluded a district-to-district comparison of TDHS variable values to the PBS variable values.

Step 3. Suppressing or recoding high-risk data items.
When cases were found that required either data suppression or recoding, they were inserted into a table to document these findings and provide the statistical programmer with sufficient information to implement suppression or recoding (see Table 1). Each row of this table represents a change made to the data.

Individual cases were identified as having outlier values for variables that could be readily identified by observation of the physical household or by a casual acquaintance of the household. The values of these variables were then recoded to mask the outlier value while retaining the case in the data.

Households with large numbers of members can potentially be identified because they are unique or rare in the survey district. To effectively avoid revealing the identity of any large household that may be the only one or among a very few of its size in its district, a number of household members have been removed from several households. The numbers of households involved and household members removed will not be disclosed. The members who were removed were not eligible for interview elsewhere in the questionnaire.

The FTF Feedback Poverty and Expenditure indicators are person-level measures that are based on household consumption patterns. We have included intermediate variables of total household consumption and per capita consumption in the public release data. Because large households may have had one or more household members removed (to prevent disclosure based on household size), the intermediate variables related to expenditures and poverty have been altered to prevent users from determining the original household sizes. These variables include pcc_weight, hhexp, hhexp_usd, pcc and pcc_usd.

REVISION OF CODEBOOK AND README FILES FOR PUBLIC USE
The codebook distributed for public use will reflect the values of variables that were revised to prevent disclosure. In addition, any revision processes will be documented in the README file that is also distributed with the data to include the revisions, but without sufficient detail to permit reconstruction of the original data files.

Table 1. Management of Indirect Identifiers in FTF FEEDBACK Public Use Population-Based Survey Datasets

	Variable	Variable Description	Modifications
HOUSEHOLD DATA FILE		
1	d01	Roof top material	For the all districts where the frequency was less than 5, recoded to 7 ("Other")
2	d02	Floor Material	For the all districts where the frequency was less than 5, recoded to 7 ("Other")
3	d03	Exterior Wall	For the all districts where the frequency was less than 5, recoded to 9 ("Other")
4	d04	Number of rooms	Created a new category called "5 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of rooms was 5 and above, recoded to "5 or more."
5	d05	Main type of toilet	For the all districts where the frequency was less than 5, recoded to 9 ("Other")
6	d06	Source of drinking water	For the all districts where the frequency was less than 5, recoded to 12 ("Other")
7	d07	Electricity	For the all districts where the frequency was less than 5, recoded to 5 ("Other")
8	d08	Source of cooking fuel	For the all districts where the frequency was less than 5, recoded to 9 ("Other")
9	ZIA_ModC_Wrk_abroad	Working abroad	Created a new category called "2 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where working abroad was2 and above, recoded to "2 or more".
10	d10	Secondary source of electricity	For the all districts where the frequency was less than 5, recoded to 5 ("Other")
11	d11	Secondary source of cooking fuel	For the all districts where the frequency was less than 5, recoded to 9 ("Other")
12	d09	Source of water for drinking, bathing, washing, and cooking	For the all districts where the frequency was less than 5, recoded to 12 ("Other")

	Variable	Variable Description	Modifications
		HOUSEHOLD MEMBERS DATA FILE	
13	ZIC_Edu_lvl	Highest grade of education	Created a new category called "1 = None or Koranic/religious only (no formal education)" combining existing codes 1 and 10 in all districts. Then created a new category called "11 = Other," and recoded to 11 ("Other") for all districts where the frequency was less than 5.
		WOMEN'S EMPOWERMENT IN AGRICULTURE DATA FILE	
14	c01b_a	Agriculture land	Created a new category called "12 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the agricultural land was 12 and above, recoded to "12 or more."
15	c01b_b	Large livestock	Created a new category called "3 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of large livestock was3 and above, recoded to "3 or more."
16	c01b_c	Small livestock	Created a new category called "1 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of small livestock was 1 and above, recoded to "1 or more."
17	c01b_d	Chickens	Created a new category called "1 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of chickens was 1 and above, recoded to "1 or more."
18	c01b_e	Fish pond or fishing equipment	Created a new category called "1 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of fish pond or fishing equipment was 1 and above, recoded to "1 or more."
19	c01b_f	Farm equipment (non-mechanized)	Created a new category called "6 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of farm equipment (non-mechanized) was 6 and above, recoded to "6 or more."

	Variable	Variable Description	Modifications
20	c01b_g	Farm equipment (mechanized)	Created a new category called "1 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of farm equipment (mechanized) was 1 and above, recoded to "1 or more."
21	c01b_h	Nonfarm business equipment	Created a new category called "1 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of Nonfarm business equipment was 1 and above, recoded to "1 or more."
22	c01b_i	House	Created a new category called "2 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of houses was 2 and above, recoded to "2 or more."
23	c01b_j	Large Consumable	Created a new category called "5 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of large consumables was 5 and above, recoded to "5 or more."
24	c01b_k	Small Consumable	Created a new category called "6 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of small consumables was 6 and above, recoded to "6 or more."
25	c01b_l	Cell phones	Created a new category called "4 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of cell phones was 4 and above, recoded to "4 or more."
26	c01b_m	Other land not used for agriculture purpose	Created a new category called "1 or more" when there were fewer than 5 occurrences of each value in any district. For all districts, where the number of other land not used for agriculture purpose was 1 and above, recoded to "1 or more."
WOMEN'S DATA FILE			
27	e701/e701a	Tandoor/earthen stove	Create a label for value = 1: "Owns 1 or more Tandoor/earth stoves." In households where any Tandoor/earth stoves are owned, set this variable to 1.

	Variable	Variable Description	Modifications
28	e701/e701a	Sandalee	Create a label for value = 1: "Owns 1 or more Sandalee." In households where any Sandalee are owned, set this variable to 1. In districts where the number of households with 1 or more Sandalees is less than 5, recode "Sandalee" to ("999") = ("Other").
29	e701/e701a	Generator	Create a label for value = 1: "Owns 1 or more Generator." In households where any generators are owned, set this variable to 1. In districts where the number of households with 1 or more Generators is less than 5, recode "Generators" to ("999") = ("Other").
30	e701/e701a	Air Conditioner	Create a label for value = 1: "Owns 1 or more Air Conditioner." In households where any Air Conditioners are owned, set this variable to 1. In districts where the number of households with 1 or more Air Conditioners is less than 5, recode "Air Conditioner" to ("999") = ("Other").
31	e701/e701a	Computers	Create a label for value = 1: "Owns 1 or more Computers." In households where any Computers are owned, set this variable to 1. In districts where the number of households with 1 or more Computers is less than 5, recode "Computers" to ("999") = ("Other").
32	e701/e701a	Satellite Dish	Create a label for value = 1: "Owns 1 or more Satellite Dish." In households where any Satellite Dishes are owned, set this variable to 1.
33	e701/e701a	Motorcycle/ scooter	Create a label for value = 1: "Owns 1 or more Motorcycle/scooters." In households where any Motorcycle/scooters are owned, set this variable to 1. In districts where the number of households with 1 or more Motorcycle/scooters is less than 5, recode "Motorcycle/scooter" to ("999") = ("Other").
34	e701/e701a	Car	Create a label for value = 1: "Owns 1 or more Car." In households where any Cars are owned, set this variable to 1.

	Variable	Variable Description	Modifications
35	e701/e701a	Trucks	Create a label for value = 1: "Owns 1 or more Trucks." In households where any Trucks are owned, set this variable to 1. In districts where the number of households with 1 or more Trucks is less than 5, recode "Trucks" to ("999") = ("Other").
36	e701/e701a	Bicycles	Create a label for value = 1: "Owns 1 or more Bicycles." In households where any Bicycles are owned, set this variable to 1.