Burkina Faso ADP Baseline DRB Data Package
Transparency Statement

Datasets made available for use outside the evaluation require the data to be de-identified to decrease the likelihood that individuals or households can be identified by external users. However, the type of access external users have to the data determines the level of de-identification required. In particular, data submitted for public use requires the data to be fully de-identified, which typically includes randomizing all identifying ID variables, removing all PII and direct identifiers, and removing or masking indirect identifiers (e.g., trimming outlying values, and masking rare responses as well as any combinations of responses that together could identify respondents). Because access to data submitted for restricted use is controlled, the de-identification of restricted-use data is typically limited to randomizing all identifying ID variables, and removing all PII and direct identifiers.

Fully de-identifying data for public use often alters variable values used to generate analysis indicators. Thus, the degree to which previous results can be replicated is often reduced with public-use data. In contrast, data prepared for restricted use typically allows the user to fully replicate previous results because the information in the removed PII and direct identifiers are infrequently used to generate analysis indicators. As such, we anticipate that our baseline analysis results generated with the data we are submitting for restricted use are fully replicable with the following caveats:

- The restricted-use files only contain baseline survey data. The consent language and IRB guidance do not give us permission to share administrative data with the de-identified baseline survey data. The analyses of the Di PAP and Di Lottery both relied, in part, on administrative data.[1] As such, only baseline analysis results generated using survey data alone are replicable with the submitted baseline restricted-use data files. The interim data delivery will include Di PAP and Di Lottery administrative data because we have IRB permission to share administrative data for interim respondents, as consent language for the interim survey asks for permission to link survey data with administrative data. Thus, we anticipate that at the time of interim data delivery the user will be able to fully replicate Di Lottery baseline analysis results for those of the baseline sample represented in the interim sample. We anticipate that differences in results will be small due to low levels of attrition in the Di Lottery sample from baseline to interim. The Di PAP baseline and interim samples are not the same meaning the Di PAP baseline analysis results generated using administrative data cannot be replicated.

- We are delivering our baseline analysis programs so that the user may replicate our baseline analysis results. However, it would require a significant amount of work to alter the analysis programs to run on the restricted-use data due to the lack of administrative data in the restricted-use files, and alterations to the names of and the data in the data files we are submitting for restricted use as a result of the data de-identification process.[2] As such, we anticipate that there is a significant burden for the user to adapt our baseline analysis programs for use with the baseline restricted-use files.

[1] Di PAP administrative data includes the amount of land PAPs lost and amount of money received in compensation. Di Lottery administrative data includes lottery eligibility, candidate, and land allocation data.

[2] For example, the data file names in the baseline analysis programs will need to be updated (e.g. all restricted-use versions of the survey data files carry the suffix "_ruf" in their files names), and any references to variables dropped or altered for data de-identification reasons will need to be removed or updated in the analysis programs (e.g. indicators or covariates created from non-region geographic variables will need to revised using the randomized codes in place of location names). Note that any PII and directly identifying information referenced in the analysis programs have been scrubbed from the programs delivered in this data package.