

Addendum to Pre Analysis Plan for the ‘Randomized Impact Evaluation of the Community Auxiliary Police, Bougainville, Papua New Guinea’

Jasper Cooper - Columbia University, New York

August 26, 2016

The purpose of this addendum is to pre-register changes to the research design of the ‘Randomized Impact Evaluation of the Community Auxiliary Police, Bougainville, Papua New Guinea’. The first registration took place on August 21, 2016. This addendum outlines changes to the measurement of behavioral outcomes in the candidate survey, and to the sampling strategy in the household survey. This addendum was pre-registered prior to data collection.

Changes to Behavioral Measurements in Candidate Survey

The original Pre-Analysis Plan (PAP) pre-specified three behavioral games to be played with respondents. The addendum includes the three main changes to the design of these measures.

First, only two games will be played now, with 16 rounds each: the game of the respondent against the community with no monitoring, and the game of the respondent against the community with monitoring (i.e. random assignment on each roll to have a roll revealed or not).

Second, the payoffs are structured in a slightly different manner so that dishonest respondents are incentivized to maximize the number of evens they report rolling, rather than to roll just above half even as in the previous setup. This is achieved in the following manner: in each roll of the dice, the respondent reports the number they rolled. If they roll an even, they put one of their tokens into a bowl. If the respondent reports an odd roll, the other player from the community puts a token (differently colored) into the bowl. This is repeated 16 times. At the end, the enumerator jumbles the bowl, and draws one of the tokens at random. The person whose token is drawn receives the prize. In this manner, the respondent has a strictly higher expected payoff the more they cheat.

Third, the order of play of the two games is randomized, blocking on the COE and attempting to achieve balance between both possible orders (in COEs with an odd number of candidates a binomial draw breaks the tie). In this way, the causal effect of the order of play can be identified independently from the features

of the game played (monitoring or not).

The following code was used in R to randomize the order of rolls:

```
# Order alphabetically by COE
cands <- cands[order(cands$COE),]

# Complete assignment function
assign_order <- function(size){
  if(size %% 2 > 0){
    start <- rep(0:1,floor(size/2))
    end <- rbinom(n = 1,size = 1,prob = .5)
    assignment_vector <- c(start,end)
  } else {
    assignment_vector <- rep(0:1,size/2)
  }
  assignment_vector <- sample(assignment_vector)

  return(assignment_vector)
}

# Candidates in each COE
COE_n <- table(cands$COE)

# Set random number generator seed
set.seed(1988)

# Random assignment:
cands$monitoring_first <- unlist(lapply(COE_n,assign_order))
```

The survey now includes a question that takes the result of this random assignment and orders the games accordingly.

Implications for estimands, hypotheses and outcomes: The estimand reported in tables 55 and 60 can no longer be estimated due to the removal of one of the games (that played against the enumerator with no

monitoring). The game previously referred to as ‘Game 2’ will now be referred to as the ‘No Monitoring’ game, and the game previously referred to as ‘Game 3’ will now be referred to as the ‘Community Monitoring’ game. The estimands in tables 58 and 59 will simply be estimated using the two games now.

Changes to sampling strategy for household members

Section 4.1 of the PAP specified ‘The male or female head of household is interviewed in each household, depending on which gender was randomly assigned to be interviewed there’. The method for selecting respondents at the household level has been altered to allow for random selection of adult members within households.

Each enumerator carries a grid with numbers 1 - 10 as column headers. Denoting columns j and rows i , the cell (i,j) contains an integer from 1 to the j 'th column header, selected at random using the R code given below.

Each enumerator carries a unique grid generated at random. When a male enumerator arrives at a household randomly sampled as a male household, he asks the head of household how many men between the ages of 18-65 reside there. He looks up the corresponding integer in the j^* 'th column header of the grid, and goes to the cell x in row i^* , where row i^* is the next row that has not been crossed off the grid. He then asks to interview the $x_{i^*j^*}$ 'th oldest male member of the household aged between 18-65. If the selected respondent is not available. He decreases j^* by one to get j^{\dagger} and asks for the $x_{i^*j^{\dagger}}$ 'th oldest member, repeating this process until he finds an available respondent. The same procedure is followed by female enumerators, with female respondents. Each time a respondent is interviewed in a given household, that row is crossed off the grid.

The procedure thus creates a representative sample of the population of respondents available on the day of surveying. All other sampling procedures remain as specified in the original PAP.

The code used to generate grids is as below:

```
library(xtable)
library(dplyr)

village_numbers <- as.numeric(village_names$village)

enumerator_numbers <- 1:9
```

```

random_seeds <-
  unlist(lapply(enumerator_numbers,
    function(x)as.numeric(paste0(x,"0",village_numbers))))

sample_number <- function(x) sample(1:x,1)

sample_vector <- function(x) sapply(x,sample_number)

HH_sizes <- 1:10

make_grid <- function(seed) {
  set.seed(seed)
  t(replicate(n = 60,sample_vector(HH_sizes)))
}

grid_list <- lapply(random_seeds,make_grid)
names(grid_list) <- 1:length(grid_list)

make_table <- function(
  grid,grid_number,
  folder_path = "11_CAP_endline/11_CAP_field_documents/hh_member_grids/tex_files/"
){
  grid_table <- data.frame(grid)
  colnames(grid_table) <- 1:10

  tab_out <-
    xtable(
      grid_table,
      align = c(
        "l",
        "p{.5cm}|",
        "p{.5cm}|",

```

```

    "p{.5cm}|",
    "p{.5cm}|",
    "p{.5cm}|",
    "p{.5cm}|",
    "p{.5cm}|",
    "p{.5cm}|",
    "p{.5cm}|",
    "p{.5cm}|"
  )
)

file_path <- paste0(folder_path,"grid_",grid_number,".tex")

bold <- function(x) {paste('\textbf{\large ',x,'}', sep =')}

sink(file_path)
cat("\documentclass[8pt,english]{article}
    \n\usepackage[margin = 1cm,a4paper]{geometry}
    \n\usepackage{float}
    \n\usepackage{graphics}
    \n\begin{document}\n\n")
cat("\vspace*{\fill}\n")
cat(paste0("\centering Household Member Grid ",grid_number,"."))
cat("\scriptsize\n")
print(
  tab_out,
  include.rownames = F,
  include.colnames = T,
  caption.placement = "top",
  hline.after = -1:60,
  table.placement = "H",
  sanitize.text.function = function(x){x},

```

```
sanitize.colnames.function = bold,  
scalebox = 1.2,  
size = "\\footnotesize"  
)  
cat("\\vspace*{\\fill}\\n")  
cat("\\end{document}")  
sink()  
}  
  
mapply(make_table,grid = grid_list,grid_number = names(grid_list))
```