KALAHI-CIDSS Impact Evaluation
Third-Round Data Analysis
Pre-Analysis Plan

Innovations for Poverty Action
Amanda Beatty Ariel BenYishay, Elisabeth King, Aniceto Orbeta, Menno Pradhan

Contents

I. Overview of intervention
II. Overview of the evaluation design
    a. Experimental design
    b. Sample
    c. Assignment of treatment
    d. Replacements
    e. Compliance
    f. Attrition
III. Overview of data sources
    a. Baseline and Second-Round Data Collection
    b. Third-Round Data Collection
IV. Hypotheses
V. Longer Term Outcomes
VI. Regression specifications
    a. Unit of analysis
    b. Average treatment effects
    c. Treatment effects for individual sub-projects
    d. Subgroup analyses
    e. Multiple hypothesis testing
    f. Robustness checks
    g. Treatment duration
Appendix 1: Matching of Survey Modules with Subproject Types

Introduction

This pre-analysis plan provides the methodologies and specifications to be employed in testing hypotheses related to the short-term outputs and longer-term impacts of the the Kapit-bisig Laban sa Kahirapan-Comprehensive and Integrated Delivery of Social Services (Kalahi-CIDSS or KC) community-driven development (CDD) project in the Philippines. The baseline data were collected from April to July 2012, and the second-round data were collected from February to June 2014. The third-round data are to be collected in July to Oct 2015.

The results from the planned analysis will be described in evaluation report as part of the impact evaluation of the KC project that aims to provide an independent assessment of the impact of KC generally, and specifically of the returns to the Millennium Challenge Corporation's (MCC) US$120 million investment. Furthermore, the impact evaluation aims to contribute to broader research about the impacts of CDD programs.

I. Overview of intervention

The Kalahi-CIDSS or KC is a community-driven development (CDD) project implemented by the Government of the Republic of the Philippines' (GRP) Department of Social Welfare and Development (DSWD) as a major poverty reduction initiative. At the provincial level, KC targets the poorest 48 provinces, out of a total of 81 provinces. KC aims to improve welfare in rural areas by targeting communities with a poverty incidence greater than the national average through small-scale, community-driven development subprojects (SP) aimed at addressing their most pressing needs.[1]

The first phase of KC, known as KC1, took place from 2003-2009 with the support of the International Bank for Reconstruction and Development – World Bank, providing roughly US$100 million in lending, the GRP financing US$31 million, and communities and local governments contributing US$51 million in training and grants to 4,583 barangays (villages) in 183 municipalities across the country's 42 poorest provinces.

In 2011, KC was expanded, providing grants and technical support to 362 municipalities within the original 42 provinces served by KC and six new provinces. This phase of KC was financed with a renewed US$59 million in loan funding from the World Bank and a US$120 million grant from MCC. In addition, local governments (region, municipality and/or barangay) are required to contribute 30% of subproject costs. Because there was not enough money to fund all municipalities, the project was randomized and municipalities were selected via a lottery so that each municipality had an equal chance of being selected. This phase of KC is still ongoing and is the phase being evaluated. Each participating municipality is allocated approximately PHP 450,000 (about US$11,250) times the number of barangays in the municipality, although not every barangay in participating municipalities ultimately get subprojects.

Teams composed of barangay resident volunteers develop proposals for infrastructure and services to meet poverty reduction goals. Proposals are then evaluated by individual municipalities. Representative teams from each barangay in the municipality vote for which subprojects are most deserving of funding; the funds are then designated to the barangays according to their ranking in the voting until the municipal allocation is used up.

In parallel, the KC program trains the communities and their local governments, both at the barangay and municipal level, in choosing, designing and implementing the subprojects. This is done through a five-stage program known as the KC Community Empowerment Activity Cycle (CEAC). The main stages of the process are: (1) Social preparation; (2) Subproject identification and conceptualization; (3) Subproject prioritization; (4) Subproject implementation; and (5) Transition. Further details on each stage can be found in the baseline report (Beatty et al. 2013).

Roughly one third of barangays receive subprojects each year, although some barangays may receive multiple subprojects and others none. Most subprojects are programmed to be implemented within six months, thus the stages of preparation, funding and implementation generally take nine to twelve months and are called a cycle. The same process is repeated over three one-year cycles, with cycles two

[1] Projects at the barangay level are called subprojects because they are part of the overarching KC "project".

and three having a condensed social preparation phase since communities have already become familiar with the project and process.

Over the past several years, the Government of the Philippines (GoP) has also advanced two additional community driven development initiatives with similar aims and features to KC: the KALAHI-CIDSS National Community Driven Development Program (KC-NCDDP) and the Grassroots Participatory Budgeting (GPB) Program. Because this roll-out has overlapped with the KC evaluation sample, there is a risk that these initiatives could confound the treatment effects identified in the evaluation, thereby altering the results. We have therefore significantly revised our original evaluation design to account for these confounds and generate appropriate conclusions from the evaluation results. We discuss these features in the Compliance subsection of the Design section below (section III.E).

II. Overview of the evaluation design

A. Experimental design

In order to isolate KC's effects, a matched-pair randomized design was implemented, with municipalities as the unit of assignment. This design offers both causal identification and the needed statistical power at an acceptable cost.

B. Sample

The evaluation sample frame of municipalities was determined by both program eligibility and the subset of municipalities that were not automatically participants or non-participants. At the municipal level, municipalities with 70% poverty incidence or above automatically received the project, and municipalities with less than 33% poverty incidence automatically did not receive the project. The impact evaluation thus focuses on municipalities with poverty incidence between 33-69%.

In addition, DSWD allocated funding to each province based on the number of municipalities in that province (a province received funding for half of the municipalities in the province minus one). In some provinces, there were at least as many municipalities guaranteed funding (those with poverty rates above 69%) as available funding based on the aforementioned rule. These provinces were excluded from the evaluation sample. A total of 26 provinces (out of 48 total) were thus included in the evaluation sample frame.

Finally, municipalities that had previously received funding from the KC project were also excluded from the sample, as our analysis initially focused on comparisons of KC treatment effects relative to an untreated reference group.

To summarize, a municipality was included in the evaluation sample frame if it:

1. Had between 33-69% poverty incidence
2. Was located in a province in which guaranteed municipalities have not been allocated all of KC funding based on the 50% minus one rule
3. Did not receive KC1 funding

From this sample frame of municipalities, we randomly sampled 198 municipalities, spread over 26 provinces. The sample size was determined via power calculations and was expected to be large enough to be able to detect MCC's projected eight percentage point change in household income, as well as effects on other variables[2].

Municipalities were informed that they would be excluded from the program if the mayor or his/her representative was not present at the public lottery event at which treatment was assigned (described in further detail below). In practice, 23 municipalities out of the 313 invited to participate in the randomization were disqualified for not attending the public lotteries or because they declined to participate.

Within each municipality, one barangay was randomly sampled in which data collection activities (described below) have been carried out.

C. Assignment of treatment

In order to maximize the precision of the evaluation results, municipalities were paired based on readily available characteristics, and treatment was randomly assigned within each pair. The 198 municipalities were paired using nearest neighbor matching within each province based on a composite measure of their poverty incidence, population size, land area and number of barangays.[3] Ninety nine pairs were formed. One member of each pair was then randomly assigned into the treatment group and one member into the control group through public lotteries. A total of 12 public municipal selection events took place in separate provinces (with municipalities from multiple provinces attending each event).

D. Replacements

Of the original 99 pairs selected for the impact evaluation sample, one pair was dropped, but we replaced it with another pair in a manner consistent with the randomized design and in advance of the baseline data collection. Thus, we were able to collect baseline data on 99 pairs. The baseline report provides additional details on the dropout pair and its replacement (Beatty et al. 2013, table 4.1).

E. Compliance

[2] The outcome variables considered in the sample size computation include: total expenditure per capita, family income per capita, distance from main water source, proportion with safe water source, proportion of children 6-17 years attending school and proportion of mothers in the labor force. Other household level and individual level outcome indicators also considered include: the proportion of households trusting others in the village, trusting local officials, attending village assembly, joining barangay development planning, having difficulty fetching water, located within 30 minutes of the post office, visiting a health professional, and located within 30 minutes of a school.

[3] Municipal poverty incidence was naturally included as this is a key variable in project eligibility. The number of barangays was used to help balance the pairings since this is the unit of intervention (i.e. grants are made at the barangay level). Population and municipality land area were included because they are factors in determining the Internal Revenue Allotment (IRA) of a municipality, which largely determines the financial resources available to the local government unit (LGU), and affects counterpart contributions.

Compliance with assignment to treatment group

To date, there have been five municipalities assigned to the treatment group that have not complied with the randomization and have not participated in the KC program. The reasons for non-compliance were inability to raise counterpart funding (3 municipalities) and governance issues such as conflict (2 municipalities). In these cases, we continue to conduct intention-to-treat analysis and include these municipalities (and their paired control municipalities) in our evaluation sample.[4]

Compliance with assignment to control group

In addition, three municipalities originally assigned to the control group were nonetheless allocated funding by DSWD in the first year of program implementation. As in the case of non-compliance with treatment group assignment, we continue to conduct intent-to-treat analysis using the original group assignment as an indicator for treatment.

Several exogenous factors also serve as threats to compliance with assignment. Of particular threat are two large-scale Government of the Philippines (GoP) community driven development initiatives with similar aims and features to KC: the KALAHI-CIDSS National Community Driven Development Program (KC-NCDDP) and the Grassroots Participatory Budgeting (GPB) Program. GPB represents an effort to encourage municipalities to produce annual development budgets through greater barangay-level engagement. GPB has been implemented in a portion of the treatment and control communities since 2012, although monitoring done to date indicates GPB is not sufficiently similar to KC as to constitute a threat to the evaluation.[5] In contrast, KC-NCDDP rolled out in the majority of treatment and control communities in August 2014 and was anticipated to be very similar to KC. Because this roll-out has overlapped with the KC evaluation sample, there is a risk that these initiatives could confound the treatment effects identified in the evaluation, thereby altering the results. We continue to monitor the roll-out based on administrative sources and will further study its extent using specific instruments in the third round data collection, discussed in the data section below.

To date, roll-out of KC-NCDDP across our study's treatment and control groups has been balanced. At the time of the second follow-up data collection, treatment municipalities also served by KC-NCDDP will have initiated four cycles of CDD (three funded by KC, one by KC-NCDDP). By contrast, control group municipalities served by KC-NCDDP will have completed one cycle. Those municipalities affected by Typhoon Yolanda will have received an accelerated roll-out of KC-NCDDP, with two (sped up) cycles of CDD potentially provided. Thus, our control group of municipalities will now be partially treated. Nonetheless, given the differences in treatment intensity, our control group still provides a useful comparison group.

We thus plan to compare the effects of a large dose of a community-driven development program to a small dose using our original treatment assignment as a valid source of exogenous variation. The exact

[4] We considered whether our pairwise randomization could be used to drop comparable non-compliers from both the treatment and control group, then proceeding with estimating treatment effects only on treated communities. However, our pairings were based on municipality-level poverty and population conditions and likely do not reflect major factors in selective compliance. That is, non-compliance is likely selective on a variety of unobserved characteristics on which our paired municipalities are not necessarily balanced.

[5] Data from the interim data analysis indicated that GPB was implemented at similar funding levels in the evaluation sample's treatment and comparison groups.

difference in dosages will be determined via both DSWD administrative data and independently collected municipal, barangay, household, and area facilitator interviews. In the event we find significant differences in dosages, we can compare these impacts across the treatment and comparison groups and attribute them to the additional cycles that the treatment municipalities will have experienced. Such a dose-response approach offers a causally valid evaluation strategy. It also addresses an important policy question (how many cycles of KC (or similar CDD programs) are in fact needed for impacts to occur?) and tests a widespread belief in the field that more rounds of CDD programming are likelier to bring about the intended effects. Figure 1 below graphs an illustrative dose-response comparison, in which treatment and control municipalities are compared and observable differences in outcomes are detectable.

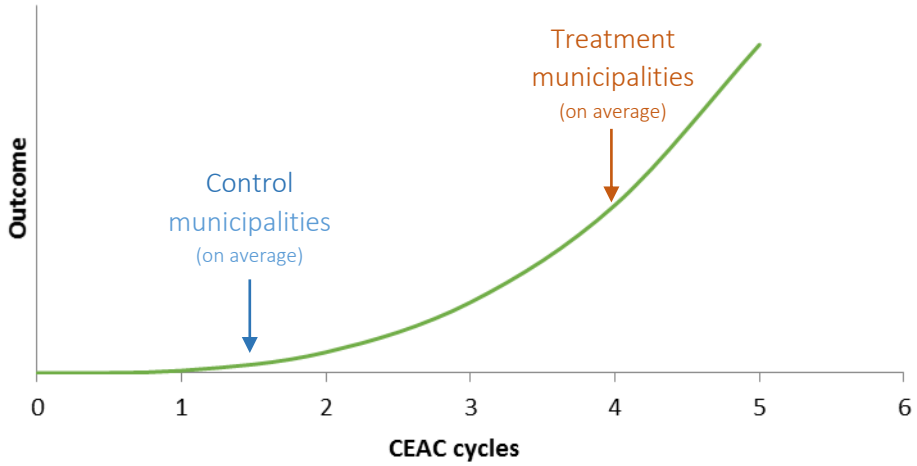

Figure 1: Illustrative dose-response comparisons

However, this dosage-response approach is subject to several potential complications: First, it is possible that treatment effects materialize quickly from low dosages and then saturate. If that is the case, we would not observe impacts in changing dosages from 1-2 cycles to larger dosages. The interim data collection indicated very few significant gains due to treatments lasting approximately 1-2 years (or 1-2 CEAC cycles), consistent with findings in other CDD studies. Nonetheless, we plan to verify this by comparing changes in outcomes in our control group between baseline and interim (when they were almost entirely untreated), and interim and follow-up (when treatment is scaled up in our controls). Finally, even if effects materialize quickly, there would continue to be differences in cumulative outcomes unless the treatment effects saturate at 1-2 cycles. In that case, our evaluation design would remain a valid way of attributing non-differences in outcomes to these (ineffective) differences in dosages.

Second, smaller differences in dosages would likely require large sample sizes to detect impacts, and our actual sample may not be large enough to detect such impacts for dosage differences below a certain minimum threshold, even though such impacts may exist. We plan to use the estimated standard errors to describe the minimum detectable effect (MDE) size in our sample from the actual treatment intensity differences. Whether this ex post MDE is policy-relevant is a substantive issue and partly dependent on benefit-cost comparisons. If the MDE is very large because treatment differences were too small and

swamped out by other noise even in our large sample, we will clearly state this. Such a result would make the evaluation much less informative.

However, we prefer to make such conclusions using the actual standard errors rather than pre-specifying a cut-off value for funding or CEAC cycle differences, as such a cut-off would be arbitrary and not well-informed. This is largely because there are no prior estimates of impacts due to dosage differences on which to rely in making power calculations. Thus, we plan to estimate standard errors and the associated MDE and be transparent about these values in our results and interpretation.

F. Attrition

Although significant efforts are being made to minimize attrition, we expect attrition of between 5 and 8%, largely due to migration (attrition in the first interim follow-up was 6.5%). We will use the analytic sample to assess whether attrition affected the comparability of treatment and control groups. These models will have the same structural form as the models that will be used to estimate impacts. We will test whether attrition rates and attriter characteristics are comparable across treatment and control status. To assess this comparability more formally, we will re-estimate the baseline levels of outcomes across treatment and control groups across both attrited and non-attrited households via both seemingly unrelated regression (SUR) and ordinary least squares (OLS) specifications.

III. Overview of data sources

Data for the estimation of outcomes and impacts of KC comes from three data collection exercises, each involving several instruments:

A. Baseline data collection: Administered April to July 2012

Household survey: Within each barangay, 30 households were randomly selected from among all households to comprise the household survey sample. For part of the survey questionnaire (questions relating to perceptions and empowerment), half of the 30 selected households were randomly assigned a male target respondent and half a female target respondent. All other parts of the survey could be answered by the principal respondent with the support of any other household member 15 years of age or above. A total of 5,940 households were surveyed across the 198 barangays.

Barangay survey: A barangay-level survey was conducted for each of the 198 barangays, with barangay captains as principal respondents.

Qualitative focus groups: Qualitative focus groups were conducted in a sub-sample of 24 municipalities (12 municipality pairs) from the study's 198 municipalities. The qualitative sample covered 12 provinces (one pair per province) spread over the three island groups of Luzon, the Visayas and Mindanao. For each barangay, there were three FGDs: one with male participants, one with female participants, and one with both male and female participants, for a total of 72 FGDs. Each group had an average of 15 participants.

Qualitative key informant interviews (KIIs): Qualitative KIIs were conducted in the same sub-sample as the qualitative focus groups. A total of eight key informants (KIs), (six municipal and two barangay officials) per municipality were targeted for interviews. At the municipality level, the KIs were the municipal mayor, vice-mayor, representatives from the Municipal Development Council (MDC), municipal engineer (ME), municipal planning and development officer (MPDO), and municipal social welfare and development officer (MSWDO). At the barangay level, the KIs interviewed were the Barangay Captain (BC), and representatives of Barangay Development Council (BDC).

B. Second-round data collection (February – June 2014)

Note that social preparation for cycle 1 began in approximately April 2012. At the time of the second-round survey, treatment municipalities had been through approximately two rounds of social preparation, and were in the process of implementing or beginning to implement a second round of sub-projects.

Household survey: 80 municipalities (40 pairs) of the original 198 municipalities that are part of the full baseline sample were resurveyed during the second-round.[6] Within the 80 municipalities, we re-surveyed the same 80 barangays and the same 30 households per barangay. The 80 municipalities were chosen with a simple random sampling of municipality pairs roughly proportional to the baseline sample. The household survey instrument was modified to focus primarily on empowerment, participation, and governance outcomes, as described below (treatment effects on other outcomes were not expected to have materialized).

Barangay survey: Within the same sample of 80 barangays, we administered a barangay survey with the current barangay officials as the key respondents.

Structured Community Activity (SCA): The SCA was implemented in the same 80 barangays as the household and barangay surveys. Each community was given a small sum of money and the discretion to use it to construct or repair a public building, space, or infrastructure. The SCA is a data collection tool that entails collecting quantitative and qualitative data on community participation and decision-making by observing how a community carries out this simple community-driven task.

Barangay Assembly Observations: Qualitative barangay assembly (BA) observations took place in 10 barangay (5 pairs) taken from the original 24 barangays (12 pairs) that were part of the baseline qualitative focus group sample. The sample of 10 barangays included at least one barangay from each of the three main islands (Visayas, Mindanao and Luzon). Observations took place during the mandatory Barangay Assembly on or around March 29, 2014.

C. Third-round data collection (July – Oct 2015)

[6] Power calculations using community empowerment and social capital outcomes, the focus of the second round data collection, indicated that only 80 communities were required given the smaller in-sample variance in these outcomes.

Household survey: All 198 municipalities (99 pairs) in the evaluation sample are to be resurveyed. Within these municipalities, we re-survey the same individual barangays and the same 30 households per barangay.

The household survey in the third-round included the all the modules in the second-round plus modules related to consumption, agriculture, livestock, fisheries, health, education, water and sanitation (as treatment effects on outcomes in these domains are more likely to have materialized). In each barangay, three of these modules are administered (partially randomly and partially based on KC sub-projects chosen) in an attempt to both limit interview length and capture outcomes in contexts where these were specifically likely to have been impacted. To do so, we mapped specific types of subprojects to modules capturing relevant outcomes. We then identified the specific subprojects funded in treatment group barangay and mapped these to the relevant modules. We randomly sampled one of these modules (when there are more than one), and then randomly sampled two additional modules. Each control group barangay is administered the same set of modules as its paired treatment group barangay. The share of barangay in which each module is administered is thus correlated with the prevalence of the related SP types in the population. Appendix Table 1 shows the shares of barangay in which each module is administered, broken down by subproject type.

Barangay survey: As in the second round data collection, a barangay survey is administered with the current barangay officials as the key respondents.

Structured Community Activity (SCA): The SCA is implemented identically to the process carried out in the second-round data collection.

KC Area Coordination Team (ACT) survey: The data collection includes a new set of interviews conducted with the area coordination team (ACT) responsible for each municipality in which KC is active (nearly all treatment group municipalities and the subset of control group municipalities receiving KC-NCDDP). These interviews aim to document the KC implementation process, including the dates of and extent of participation in each social mobilization phase, as well as any modifications made to the process in the municipality. The survey also details the rankings of subproject proposals assigned during the Municipal Inter-Barangay Forums (MIBFs) and validates the list of subprojects selected and implemented.

Municipal official survey: The data collection also includes a new set of interviews of municipal officials aimed at describing the full range of community-driven development programs active in each municipality. The survey covers municipal-level involvement with KC, KC-NCDDP, and GPB (these programs are discussed in section II.E). It also captures the current state of municipal development planning, including the status of the local development investment program (LDIP) and annual investment plans (AIPs). Officials are asked to complete (or provide) tables detailing local subprojects approved under the Local Poverty Reduction Action Plan (LPRAP). These modules are simultaneously administered to the Municipal Planning and Development Coordinator (MPDO) and the Municipal Local Government Operations Officer (MLGOO). Finally, a module on gender and development planning is administered to the Municipal Gender and Development Focal Person.

Qualitative Focus Group Discussions (FGDs): In 12 purposively selected barangay, an independent researcher will conduct focus group discussions with KC volunteers and those not directly involved in KC, as well as FGDs with IPs specifically (as applicable) to address six primary questions: (a) Can we reconcile discordance between generally favorable baseline perceptions of and participation in barangay

governance with important gaps in public good provision? (b) Are there specific mechanisms or measurement issues behind the first follow-up data collection findings that perceptions of commissions are higher in treatment barangays; (c) How can we explain the interim report finding that contributions to community decrease with treatment? Is this a substitution effect, as previous research finds? (d) Do residents report governance and participation gains from KC treatment that we may not have captured with the quantitative instruments? (e) What mechanisms underlie differential gains in barangays with indigenous persons (IP) present (found in the first follow-up data collection)? (f) What more can we learn about calamities (i.e. Yolanda) and our interim findings that Yolanda-affected treatment communities were better able to respond than Yolanda-affected control communities? As is common practice in qualitative research, we may also pursue other questions that may arise over the course of the qualitative study.

IV. Hypotheses

We categorize our hypotheses to be tested into three primary domains: socio-economic, institutional, and community empowerment. Below, we lay out each hypothesis and the related data sources that will be used to test it.

Socio-economic domain

H1a: SPs improve access to related key services

Household Surveys

- Reductions in travel time and cost to nearest schools, health center, and water sources due to education, health, and water SPs
- Reductions in travel time to all facilities outside of barangay due to roads SPs

H1b: Roads SPs reduce agriculture, fisheries and livestock transport costs

Household Surveys

- Crop-specific transport costs (AG26-32)
- Share of crops sold at market (relative to farmgate)

H1c: Roads SPs improve productivity in agriculture, fisheries, and livestock sectors

Household Surveys

- Crop-specific yields (AG22-25 divided by AG17)
- Fish and livestock production

H1d: Daycare SPs increase daycare enrollment and female labor market participation

Household Surveys

- Daycare enrollment, female labor market participation for all adult females in the household (formal and informal)

Barangay Survey

- Number of daycare facilities

H1e: School SPs increase school enrollments and improves student/teacher ratios

Barangay Survey

- Number of facilities, enrollment, student/teacher ratio (section EDF)

H1f: Health SPs increase visits to health facilities and adult labor force participation

Household Surveys

- HH visits to health facility (HS1, 2), adult labor force participation (total)

H1g: Health SPs improve prenatal and birth services

Household Surveys

- Access to improved birth facility (PG3), number of visits to prenatal care (PG7), improvement in delivery assistance (PG4)

H1h: Water SPs reduce time and costs spent obtaining water

Household Surveys

- Improved water source (WS2), location of water source (WS3), time spent collecting water (WS8,9), time savings for women and children (WS10), household water costs (WS11,12, 13)

H1i: Disaster Preparedness SPs reduce extent of damage in case of natural disasters

Barangay Survey

- Share of HH with losses due to disasters (floods, earthquakes, typhoons) decreases (environmental projects – seawalls, better drainage, etc)

Institutional domain

H2: KC increases quantity and quality of participation in local governance around decision-making and implementation *related to KC activities*

H2a: Participation in and knowledge of formal structures related to KC

Household Survey

- Proportion of households (HHs) participating in a governmental group or institution or purok/barangay organization in the last 12 mos.
- Total instances of participation in sweeping, cleaning, construction, repair, maintenance of different activities by HHs in the last 12 mos. (of instances outside of the house?)
- Total days participated in sweeping, cleaning, etc. by HHs in the last 12 mos. (not in individual's house?)
- Total contributions of cash and/or goods to sweeping, cleaning, etc. by HHs in the last 12 mos. (PHP) (not in individual's house?)
- Number and proportion of HHs that attended a barangay council (BC) meeting.
- … attended a barangay assembly (BA) meeting.
- … attended a barangay development council (BDC) meeting.
- Proportion of HHs who were informed that there was a BA in the last 12 months.
- Proportion of HHs who know when the next BA will be.
- Proportion of HHs who think BAs should be open to the public.
- Proportion of HHs who attended a BA in the last 12 months.
- Proportion of HHs who spoke publicly at the last BA attended.

- Number of times HHs spoke publicly at the last BA attended.
- Proportion of HHs who said that barangay resident attendees of the BA were consulted when a decision needed to be made on any problems or issues.

H3: KC increases quantity and quality of participation in local governance around decision-making and implementation *beyond KC activities*

H3a: KC increases participation in and knowledge of formal structures beyond KC

Household Surveys

- Number and proportion of HHs who
- … attended a municipal council meeting.
- … met with, called, sent a letter or text message to a politician.
- … participated in a protest or demonstration.
- … participated in an information campaign.
- … informed a newspaper, radio or TV station to a local problem.
- … notified police about a local crime.
- filed a complaint to the *"lupong tagapamayapa."*[7]

SCA

- SCA meeting held or not
- Number of attendees; number and proportion of female attendees, 4P attendees, IP attendees; average attendee age
- Number of interventions (times people spoke)
- Number and proportions of interventions by gender, BC member (each by type of intervention and target audience)
- BC member dominance (perceptions of BC influence)
- Voting occurred
- Project chosen
- Decision made during meeting
- Clear next steps
- Meeting duration
- Follow-up forms available and complete
- Implemented planned project
- Community contributions to project
- Community form match hardware store form

H3b: KC improves barangay information sharing and inclusiveness beyond KC

Household Surveys

- Proportion of HHs who said there were NO issues where the community should have been consulted but was not, at the last BA.

[7] This is a barangay-level council that administers resolutions of disputes among barangay members

- Proportion of HHs who said NO particular group usually dominates decision-making in BAs.
- Barangay captain's responsiveness to recommendations by HHs of different projects that would benefit the barangay. (scale 0 to 1, where 0 is the barangay captain is not very responsive and 1 is very responsive)

Barangay Survey

- Barangay Council (BC), Barangay Development Council (BDC), and Barangay Assembly (BA) meeting frequency and attendance
- Inclusiveness of the BC, BDC and BA, including the shares of members and meeting attendees who are female,
- The number of different ways people were informed about the last BA.
- Whether any decisions were made on problems or issues in the barangay at the last BA. (1=Yes)
- The extent to which each decision related to problems or issues at the last BA was determined by all attendees (as opposed to Barangay Captain).
- Barangay officials believe that no particular group usually dominates decision-making in BAs. (1=no groups dominate)
- Total number of organizations/people who were involved in creating the barangay development plan
- Total number of unique types of organizations registered with the barangay.
- Number of times the BDC met with the different organizations in the last year.
- Number of organizations that participate in the decision making process of the BDC.

H3c: KC increases Confidence and Self-efficacy beyond KC

Household Survey

- HH confident to participate collectively in community development activities. (scale 0 to 3, where 0 is not confident at all and 3 is very confident)
- HH agreement or disagreement with statement: "I feel comfortable speaking during BAs." (scale 0 to 3, where 0 is strongly disagree and 3 is strongly agree)
- If there were any issues in the community that the HH felt very strongly about, the likelihood the HH would talk to the barangay captain about it. (scale 0 to 3, where 0 is very unlikely and 3 is very likely)
- HHs opinion on how much impact they think they have in making the barangay a better place. (scale 0 to 3, where 0 is no impact and 3 is a big impact)

H4: KC increases knowledge and awareness of local governance

Household Surveys

- Self-reported familiarity with municipal and barangay government officials
- Correctly named members of BC
- Knowledge of whether there is a BDC in the barangay

Barangay Survey

- Number of BDC and BC meetings with KC volunteer present

H5: KC improves the degree to which local projects correspond to ex-ante preferences

H5a: KC improves the degree to which barangay projects correspond to ex-ante preferences

Barangay Survey

- KC-funded projects match baseline ranking of projects in HH survey
- Impacts on barangay governance: Non-KC-funded projects match baseline ranking of projects in HH survey

Community empowerment domain

H6: KC increases interactions among peers

Household Survey

- Proportion of other randomly sampled households a respondent knows
- Frequency of discussions of barangay problems with other households

H7: KC increases participation in community organisations

Household Survey

- HH member participation in civic, political, volunteer groups (frequency and total time)
- Value of labor and goods contributions to these groups

Barangay Survey

- Number of civic, political, volunteer groups in the barangay

H8: KC improves how well communities deal with natural disasters and other hardships

Household survey

- For people experienced financial hardship as a result of a natural disaster…
- … the types of people that helped them (scale: 0 to 21)
- … whether they were helped by someone inside the bgy (scale: 0 to 2)
- … the number of different types of support received (scale: 0 to 6)

V. Longer Term Outcomes

We define longer-term outcomes as those we expect to change after the project's completion. We do not propose these as explicit hypotheses testing the KC program logic as we did the hypotheses presented earlier because we do not hypothesize that changes to these variables would have taken place by this phase of data collection. Nonetheless, we collect data on these outcomes and thus assess whether treatment effects have already materialized. Thus, should our estimates indicate no statistically significant differences between treatment and comparison groups, we will not interpret these as failure of the program to achieve expected improvements.

As before, we divide these outcomes into socio-economic and institutional domains:

Socioeconomic Domain:

Longer Term Outcome 1: KC raises household consumption and asset holdings

Household survey:

- Aggregate household consumption per adult equivalent
- Aggregate value of household assets per adult equivalent

Longer Term Outcome 2: KC raises household labor force participation and earnings

Household survey:

- Aggregate household labor force participation rate per adult (total labor hours by all adults, employment rate among all adults)

Institutional Domain:

Longer Term Outcome 3: KC improves perceptions of local governance

Household Survey

- Quality of officials (President/mayor/bgy captain/BC) carrying out their duties (scale: 01to 4)
- Perceived corruption of above officials (scale: 1 to 4)
- Perceived honesty of officials and police (scale: 1 to 4)
- Proportion of respondents believing it is typical for officials to get a commission and typical commission*
- Perception of acceptable commission*
- The proportion of HH that would turn to people outside the HH if there's a dispute about money, including barangay officials, courts, or other households
- Number of support systems in case of tragedy (scale: 0 to 4)
- Number of ways HHs would solve problems (outside the HH) (scale: 0 to 3)

* Qualitative work as part of this data collection will help determine whether this question reliably reflects objective corruption. Based on the findings of this qualitative work, the question may be excluded from the index.

Longer Term Outcome 4: KC raises capacity of barangay government

Barangay Survey:

- Whether the barangay had a record of projects or a barangay development plan in preceding two years
- Value of …
- … Internally generated funding in preceding two years
- … Donations in preceding two years

VI. Regression Specifications

A. Unit of Analysis

Although we collect both barangay and individual-level outcome data, our primary interest is in the outcomes aggregated to the relevant unit of assignment. However, to maximize precision, we may use covariates at the individual or household level. To do so, our primary unit of analysis would be the individual household or respondent, with appropriate clustering of standard errors ensuring our results are precise yet not over-stated. When using barangay-level data, our primary unit of analysis would be the barangay.

B. Average treatment effects

Most of our hypotheses relate to the average treatment effects associated with a municipality's participation in KC (socioeconomic domain hypotheses are exceptions, discussed below). For these hypotheses, we will estimate the average treatment effects across our full baseline sample. These should be interpreted as intent-to-treat estimates due to program attrition discussed previously. Moreover, among those municipalities receiving treatment, all of our sample barangays will have completed the initial portion of the KC CEAC (known as "social preparation"), but only a portion will have received funding for their requested subprojects. The average treatment effects across our full sample should thus be interpreted as averaging the effects across these sets of barangays, irrespective of whether a SP was funded or which type (below, we discuss sub-group analysis that attempts to separately identify the effects of each treatment).

Furthermore, as noted earlier, the number of CEAC cycles that each municipality will have received will vary across the full sample, with some receiving no treatment at all and some receiving as many as four cycles of KC. Similarly, funding amounts from KC and KC-NCDDP will vary across the full sample as well, both across treatment and controls groups and within each group. We therefore characterize the average treatment effects as being due to a barangay having on average Δ_c more CEAC cycles and Δ_{SP} more SP funding. (We plan to estimate Δ_c and Δ_{SP} as described below).

The following OLS regression will be used to estimate the average treatment effects:

$$y_{ist} = \alpha + \beta_1 T_{is} + \beta_2 y_{ist-2} + D_s + \epsilon_{ist}$$

where *i* indexes either *barangay* or individual (as discussed above), *s* indexes strata, and *t* indexes data round.

Following Bruhn and McKenzie (2009), we include strata (pair/triplet) dummies (based on the matched pairing completed prior to randomization), and baseline values (y_{ist-1}) as regressors (where the latter are available). We cluster standard errors at the municipality level (the unit of treatment assignment) to account for correlation of residuals within treatment units.

Our main coefficient of interest is β_1, the average treatment effect.

To estimate the differences in CEAC and SP funding treatments, we adopt a regression approach that accounts for sampling errors due to our evaluation sample. We estimate the follow specifications at the barangay level

$$CEAC_{is} = \alpha + \Delta_c T_{is} + D_s + \epsilon_{is}$$

$$SPFunding_{is} = \alpha + \Delta_{SP} T_{is} + D_s + \epsilon_{is}$$

Where $CEAC_{is}$ denotes the cumulative cycles of CEAC completed by barangay *i* and $SPFunding_{is}$ denotes the cumulative SP funding received by the barangay from KC. Our parameters of interest in this specification are Δ_c and Δ_{SP} .

C. Treatment effects for individual sub-project types

The aforementioned average treatment effects relate to the institutional and community empowerment domains. These outcomes were potentially affected by the KC CEAC and funding for a SP (of any type). However, socioeconomic outcomes of interest are likely to be affected by only a subset of SP types (or to have been disproportionately affected by certain SPs). For example, prenatal care may have improved in barangays that received funding for health care improvements, but such an improvement is less likely in barangays where SP funding was for an agricultural project. In fact, by estimating the average treatment effects on such outcomes across the full sample, we are likely to dampen these effects. This approach has been prevalent in prior CDD evaluations (King 2013), and our method thus makes an important advance over this limitation.

We therefore seek to estimate the impacts of funding for certain SP types on specific sets of related outcomes, in addition to the average effect across all treated barangay. The challenge in doing so is that funding for SPs of specific types was endogenously determined in the KC program. While eligibility for KC participation was randomly determined via lottery, funding for specific SPs was potentially shaped by which SP the community requested and whether the MIBF prioritized this particular proposal. In fact, the needs assessment aspects of KC suggest that SP proposals *should* have differed across communities based on the conditions in these communities. This could bias our estimates, if we limit our analysis to SP-specific treatment effects only in the communities that received each SP type (and their previously matched control barangay).

To address this potential endogeneity problem, we plan to follow a multi-step process, as described below. The approach essentially attempts to test whether the original control barangay are the best match for each treatment barangay in terms of their likely SP type; if this is not the case, we can re-pair the treatment and control barangay to better control for differences in SP types (in any cycle). This re-pairing will use the baseline characteristics of barangay to construct a likelihood that the barangay received each type of SP, based on the observed relationship between these characteristics and SP type. The re-pairing will also require that matched pairs are administered the survey modules related to each SP type, such that outcomes are observed in both treatment and paired control. Because of this restriction, we therefore allow matching with replacement in the re-pairing. This approach follows Imbens' (2015) best practices in propensity score matching methods.

1. Estimate a model of SP type on baseline conditions in the barangay.

   Formally, we will estimate a multinomial logit model of the form:
   $\Pr(S_{ik}) = f(\alpha_{0k} + X_{ikt-1}\mathrm{B}_{0k} + \delta_k Y_{kt-1})$
   where S_{ik} is an indicator of whether barangay *i* received funding for SP type *k*, X_{ikt-1} is a vector of barangay-level baseline variables (including summary measures of household preferences for SP type *k* and baseline governance conditions), and Y_{kt-1} is a vector of baseline outcome variables related to this SP type.

   The sample for this estimation would be restricted to barangay that ever received treatment (including many control barangays that will have received funding for at least one? CEAC cycle by the date of analysis). We then predict among our full sample the likelihood that a given barangay received funding for an SP of each type (formally, $\widehat{\Pr(S_{ik})}$).

2. We check whether the treatment and control groups of barangay are similar in their predicted likelihood of receiving each SP type (because they are balanced on the observable characteristics used in our multinomial logit model to predict this likelihood, we expect them to be balanced on the predicted probabilities).

3. We check whether the original pairs of barangays are sufficiently similar in their predicted probabilities. Specifically, we test whether the average distance in predicted probabilities for each SP type is less than or equal to 0.2 standard deviations. Formally:

   $$\frac{1}{N}\sum_{i=1}^{N}(\widehat{S_{ik}} - \widehat{S_{jk}}) \leq 0.2\,\sigma_{\hat{s}}\ .k$$

   Where $\widehat{S_{jk}}$ is the predicted probability of the paired control.

   If the original pairing does not satisfy this condition, we re-match barangay (with replacement) separately for each SP type, again conditioning on the requirement that both members of the pair are administered the modules relevant for that SP type. Following Imbens (2015), we trim the sample for each SP type so that the predicted probabilities for all barangay are on a common support. We then form pair dummies P_{2ik}. We use these new pairs in step #5 below.

4. If the original T/C assignment and barangay pairing achieves both balance and satisfies the aforementioned pair distance condition, we proceed to estimate the following regression for each outcome:

   $$y_{ik} = \alpha_{1k} + \beta_1 T_i + \lambda_1 S_{ik} + \gamma_1 T_i * S_{ik} + P_{1i} + \delta_{1k} y_{ikt-1} + \epsilon_{ik}$$

   Where β_1 is the average effect on outcome *k* across all treatment barangay, λ_1 reflects the effect of SP type k among control bgy, γ_1 now reflects the specific effect of SP type k differentially for the treatment barangay, and P_{1i} is the original pair dummies.

5. If the original T/C assignment or barangay pairing do not achieve both balance and sufficiently small pair distance, we instead estimate the same regression as above but instead use the new pair IDs and trimmed samples. Formally:

$$y_{ik} = \alpha_{2k} + \beta_2 T_i + \lambda_2 S_{ik} + \gamma_2 T_i * S_{ik} + P_{2ik} + \delta_{2k} y_{ikt-1} + \epsilon_{ik}$$

Where again, the coefficient that reflects the specific effect of SP funding of type k for treatment communities is γ_2.

In this approach, we estimate not only average treatment effects of *any* KC treatment but also heterogeneous effects of specific types of KC SP funding. The average effects of *any* KC treatment may be induced through a variety of channels, including greater community mobilization around all public goods or indirect effects of gains in wealth, education, etc on other outcomes. The heterogeneity in effects based on funding for the relevant SP types captures the (more) direct effects of SPs on specific outcomes.

Moreover, for most of the SP-specific outcomes, we expect gains to accumulate over time, so that control barangay that may have only recently been treated with such SPs would not have yet experienced gains. Thus, given that we have conditioned on pair dummies that predict S_{ik} , we expect $\lambda_1 = 0$ and $\lambda_2 = 0$.

Because many outcomes we observe are based on modules that are sampled, we must correct for this sampling to achieve population-level treatment effects. We do so via inverse probability weighting. We will calculate the probability of module *k* being sampled as a function of the SPs implemented in barangay *i* (also considering that random modules are added). Then we use all observations for which module *k* was administered, but weight by 1/prob(module *k* sampled).

As an alternative strategy to the matching, we considered using the baseline preferences of barangays as instrumental variables (IVs) that predict SP type. In essence, such an approach would estimate the likelihood of each SP type based on the share of household respondents who ranked that SP type among their top preferences at baseline, and then use only the predicted SP type for each barangay in the main regression. However, since baseline preferences may be correlated with other barangay characteristics that affect outcomes (such as governance, education, information availability, etc.), they would be unlikely to serve as valid IVs (they would likely violate the exclusion restriction). Thus, we opt for a matching approach that attempts to use as much information as possible to predict SP type. If there remain major unobserved drivers of SP types that are not correlated with our predictors, our estimates may still be biased. Given the extensive baseline data we have at our disposal, we can minimize this threat to a reasonably small level.

D. Sub-group analysis

1. Within-barangay sub-groups

For all sub-groups of individuals within barangays, we will conduct our analysis at the household level. The sub-groups of interest will be:

1. Indigenous Persons (IP)[8]

[8] We will use the third round data collection to identify households as IPs, as IP status has been misinterpreted and miscoded in prior rounds.

2. Female respondents
3. Households officially classified as poor at baseline by falling below the official regional per capita income poverty threshold

Our OLS estimation for these sub-group effects will use the following specification:

$$y_{ist} = \alpha + \beta_1 T_{is} + \beta_2 z_{ist-1} + \beta_3 T_{is} * z_{ist-2} + D_s + \epsilon_{ist}$$

Where z_{ist-2} denotes a baseline indicator of this sub-group status, and $?_3$ denotes the differential treatment effect for this sub-group.

2. Between-barangay sub-groups

For all sub-groups of barangays, we will conduct our analysis at the barangay level. The sub-groups of interest will be the following:

1. Barangays with higher values of baseline outcome variables (which variables or types of variables?). These regressions will take the following form:

$$y_{ist} = \alpha + \beta_1 T_{is} + \beta_2 y_{ist-1} + \beta_3 T_{is} * y_{ist-2} + D_s + \epsilon_{ist}$$

2. Barangays with higher shares of poor households at baseline. These regressions will take the following form:

$$y_{ist} = \alpha + \beta_1 T_{is} + \beta_2 P_{ist-1} + \beta_3 T_{is} * P_{ist-2} + D_s + \epsilon_{ist}$$

where P_{ist-2} denotes the share of barangay households who are classified as poor at baseline.

3. Barangays where levels of baseline governance are in the top 50%. These regressions will take the following form:

$$y_{ist} = \alpha + \beta_1 T_{is} + \beta_2 G_{ist-1} + \beta_3 T_{is} * G_{ist-2} + D_s + \epsilon_{ist}$$

where G_{ist-2} denotes the share of barangay in the top 50% of the aforementioned baseline measure.

E. Multiple Hypothesis Testing

Because we test hypotheses based on multiple outcome variables, we estimate overall average treatment effects and overall sub-group-specific treatment effects pertaining to all variables related to each hypothesis following Kling & Liebman (2004). That is, we standardize all outcome variables, then estimate the aforementioned specification for each of these outcome variables. The overall treatment effect for each hypothesis is calculated as the mean of the variable-specific treatment effects. That is, the overall effect for the hypothesis averages the separate treatment effects estimated for each

individual outcome variable. The standard error of this effect is estimated using seemingly unrelated regression (SUR) estimation.

F. Robustness Checks

We will check for:

1. Estimate treatment effects with sampling weights incorporating the number of potential households to be sampled in each barangay, as well as the number of barangay within each municipality
2. Comparing effects in households where respondents changed between data collection rounds to those among households with constant respondents

G. Treatment Duration

Many of the hypotheses were also tested using the second round data, when average treatment duration in our treatment group was between one and two CEAC cycles (and one and a half calendar years). However, the hypotheses on impacts beyond KC were largely not statistically different from zero. When we test these hypotheses using the third round data, we will essentially be testing (equivalently) whether (a) these effects are different from zero, and whether (b) three to four CEAC cycles are required for these impacts to materialize.

We did observe impacts on outcomes related to Hypotheses 2 and 3b in the interim survey. Thus, we will also test whether there is a difference in impacts on these hypotheses using the third-round data. To do so, we will use the second-round outcome variables corresponding to these hypothesis and implement the same average treatment effect estimation as described in section VI.B. We will then estimate these impacts using the third-round data, limiting our sample to only those barangay included in the second round sample (thus ensuring differences in impacts are not due to sampling). We will then statistically test whether the treatment effects using the third-round differ from those using the second-round data.

Appendix 1: Matching of Household Survey Modules with Subproject Types

RELEVANT MODULES	KC SUBPROJECTS										
	School Building (SB)	Road (Concreting of Footpath - RFP Improvt. Of FMR - RR Rural Road Concreting - RRC Road Construction - RC Bridge Construction - BC)	Water System (WS-L1 WS-L2 WS)	Sanitation toilets (ST)	Drainage structures (culverts, overflow, spillway) (DRG)	Health Center/ Station (HS)	Day care/early childhood center (public or private) (DCC)	Post-harvest Facility (PHF)	Boat (Boat)	Warf (WRF)	Multi-Purpose Center (MPC)
04 EDUCATION	√	√					√				√
12 WATER		√	√								
12 SANITATION			√	√	√						
13 HEALTH		√				√	√				√
14 PREGNANCY RELATED QUESTIONS		√				√	√				√
15 AGRICULTURE		√	√					√	√	√	√
16 LIVESTOCK AND POULTRY CONSUMPTION		√							√	√	√
17 FISHERY AND AQUACULTURE		√							√	√	√