

IPA and MCC response to comments from DSWD, MCA-P and MCC

Comments from DSWD

1. From Table 1 about the SCA.... Community form match hardware store form (ie, what the community says they intend to purchase is what is purchased from the hardware store)
2. Project type selected in SCA matches baseline ranking of projects in HH survey (Project type options taken from list of options from project implementers.)

Comment [kda1]: Please clarify.
IPA: hope this is clear now

Comment [kda2]: Grant may not be enough to finance projects identified in HH survey
IPA: yes, this is a good point. See revisions.

We will provide thorough explanation of this analysis by 1) discussing first whether communities preferences in baseline were met by KC (through the quantitative surveys), and 2) attempting to see if there is a loose correlation between SCA projects and baseline preferences. We would not use these data to conclude that the needs of the community were not met.

3. May we be clarified what does "intent-to-treat" estimates mean and how does it relate to program attrition.

IPA: this paper has a nice overview.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3159210/>

Basically, this means that one wouldn't want to exclude municipalities that don't take up KC since there could be something special about municipalities that do not take up KC, eg, they are not high capacity, violent, corrupt, etc. If one excludes these municipalities, one will not get a comprehensive picture of the impacts of the project since the sample was chosen to be representative of municipalities that qualified for KC.

4. The same with Julien's comment, in case of contamination, how does the impact evaluation shall be interpreted?
IPA: hope this is clear now.

Comments from Jansen from 12 August 2014
Main Hypothesis

<p>H1: Relevant SPs reduce travel time, distance and cost to key services</p>	<p>There were reports immediately after Yolanda struck that majority of the SPs constructed through KC in the typhoon path withstood minimal or no damages compared to other basic services infrastructure. How would this info affect the analysis? (we actually have a report of state of SP in the typhoon path but none of the non-KC areas).</p> <p>As part of the interim survey, the barangay qx included questions about the state (and funding source) of an infrastructure constructed since baseline, so we would know the extent of infrastructure destruction in both treatment and control areas.</p>
<p>H3: KC increases quantity and quality of participation in local</p>	<p>Not in the baseline but somehow in SCA</p> <p>I found it strange that testing this hypothesis only will require</p>

<p>governance around implementation</p>	<p>participation in local public projects. If memory serves me right, this section asks participation in project implementation as well as their contribution. Both of which are measures of quantity. How do we then measure quality? (at least H2 also include confidence and self-efficacy)</p> <p>Data on quality might come from SCA but I am not clear which.</p> <p>This is a good question about the distinction between quality and quantity. Quantity would include things like</p> <ul style="list-style-type: none"> • Participation in formal government structures • Participation in KC activities (compared pre-post as a way to document treatment) • Barangay Council (BC), Barangay Development Council (BDC), and Barangay Assembly (BA) meeting frequency and attendance • SCA meeting held or not • Number of SCA meeting attendees; gender of attendees; 4P/not attendees; age of attendees; IP self-identification; attendees as function of barangay size <p>Whereas quality would include</p> <ul style="list-style-type: none"> • Participation in and knowledge of formal structures • Barangay information sharing and inclusiveness • Confidence and self-efficacy • Inclusiveness of the BC, BDC and BA • For the SCA <ul style="list-style-type: none"> ○ 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 ○ Chose project ○ Decision made before meeting ○ Clear next steps ○ Meeting duration
<p>H4: KC increases knowledge & awareness of local governance</p>	<p>Both in the baseline and the SCA</p> <p>The recent barangay election might have influenced the fairly big awareness of elected officials captured by the Interim. This and the fact that barangays are fairly small in terms of population (which was validated by baseline figures which showed that roughly 70% knew brgy household and almost a quarter had at least one relative) could already affect this hypothesis even without the</p>

	<p>project.</p> <p>An advantage of the RCT is that if you think that the barangay election would have influenced awareness of local officials everywhere (ie, in treatment and control areas), then the change you would observe in KC areas is beyond the general awareness that increased everywhere.</p>
<p>H5: KC increases interactions among peers (novel opportunity to work together; equitable participation)</p>	<p>Not in the Baseline but in the SCA</p> <p>I cannot remember if “...frequency of interaction, whether discussed problems in barangay...” were asked in the Baseline Qx but I know this was asked in the interim. Do we then use the interim data as “baseline” for this hypothesis?</p> <p>If you are talking about the social networks module, yes, this was asked in the baseline qx. But in general, if a question was not asked in baseline but in the interim questionnaire (there are very few questions like this), then one would compare differences between the treatment and control groups.</p>
<p>H6: KC does not reduce participation in current community organisations or support</p>	<p>I find it strange that what we are measuring here is a negative, when all other impact we are measuring is stated as a positive? Why not just test if KC increase participation in current community organizations or support?</p> <p>This is actually a positive statement. We are not testing whether KC increases participation but rather that KC does not cause participation to go down. There was an indication from the first IE and anecdotal evidence that KC was displacing other kinds of community engagement. This is testing whether KC simply does not displace engagement (and keeping all hypotheses positive).</p>

For H3 and H5, I wonder how the analysis will be done if we cannot find matching items in the baseline. For H4, the effect of the recently held barangay election (and the heavy turnout if I may add) needs to be further explored (for proper attribution).

See comments above about consistency between the baseline and interim questionnaires, and the recent barangay elections. There are very few items that weren't asked in both questionnaires.

In “4. Barangays receiving subproject funding and those not yet receiving subproject funding”, using RD is really very tricky because of the disparity in the number of brgys per muni. The number of SPs that could be funded in a municipality in a cycle is also largely dependent on the type of subprojects that are being proposed by the top-ranked brgys (which makes any ranking over 50% increases the likelihood of not getting SP).

We appreciate that using RD here could be challenging but preliminary indications are that by the interim survey, approximately 2/3 of barangay had received SPs. Analysis of the interim data will help us understand the possibility of using this strategy.

In Contamination

BUB is now called another name.

We have noted in the report that this is now called Grassroots Participatory Budgeting Process (GPB).

“BUB as implemented until the interim data collection has not been run by DSWD or in a KC-like manner. ” needs further investigation

We have changed this to say that this question will be explored through the interim analysis.

“collected data in the interim household survey as to whether individuals have heard of BUB and or attended BUB meeting” will probably result in a low score but it does not necessarily mean that BUB being implemented by KC or have adopted the priority list from KC is not there.

All BUB items, can we discuss again?

As noted in the pre-analysis plan revisions, in the barangay survey, we collect information on the funding source of every project that has taken place in the barangay since the baseline, including BUB, so we will know if the barangay received any BUB funding. The barangay captain and officers are also asked whether they have heard of BUB and whether BUB has been implemented in the barangay.

Start of Julien's comments

1. To Publish or Not To Publish - My main concern at this stage regards the decision to publish (or not) the finding from the interim data analysis. While there's a clear case for the findings of the final survey to be published, there's a number of particular challenges surrounding the interim survey that make the case far less clear cut. I take those challenges (in addition to the contamination threats already discussed in the pre-analysis plan) in turn below:

1.1 Lack of Power - One of the main issues when the evaluation was being designed was that, even with 198 pairs, the evaluation might lack sufficient power to detect some of the expected effects. There's a risk that with only 40 percent of the initial sample being re-surveyed during the most recent data collection exercise, the evaluation will only have enough power to detect very large effects. That's failure to reject the null might be due to the fact that the team will not have enough data. This is the kind of no results that are always difficult to communicate to the public and could negatively impact support for the project.

The evaluation was powered to detect changes in income, a variable which required a significant sample to detect a reasonable MDI (see baseline or design report for details). However, we choose the interim sample size to be able to detect changes in variables involved in hypotheses about participation, community engagement and empowerment, and project implementation, which require a much smaller sample. In addition, ex-post power calculations using baseline data confirm the earlier results using national surveys and KC1 data that most of the outcomes of interest 40 pairs will be able to identify 20 percent standardized effect size at 80 percent power.

1.2 Treatment Intensity - As discussed earlier, we have concerns that - stemming from similar CDD evaluations worldwide - one cycle will be insufficient to have sufficiently large effects. Usually, between 2 and 3 cycles are necessary before impacts start materializing.

Overall, the PAP doesn't touch on this issue which might reflect the fact that the plan was drafted by IPA but the decision rests with MCA-P/MCC. But, the plan should clearly indicate how the decision regarding which results to publish will be made by DSWD, MCC and the World Bank.

It would be helpful to know on what variables you expect to see impacts on at what stages. We think there is an understanding by all involved parties that certain variables, for example participation in barangay assemblies, would change even from the first cycle. Please see Table 1 in the pre-analysis plan for hypotheses that will be tested at this stage.

MCC is committed to transparency and the independence of its evaluations. MCC's approach to publishing reports and datasets from MCC-funded evaluations is outlined in the attached document. This document is consistent with the information that MCC has shared with the World Bank in previous conversations about this evaluation. MCC does not restrict where the independent evaluators or other individuals publish papers that use MCC-funded datasets, which are made publicly available on MCC's website.

2. Contamination Threats - The pre-analysis plan notes a number of potential threats to the proposed analysis. I just want to flag two things at this stage:

2.1 BUB - The PAP notes that "*in the event that we observe more than 30% contamination....*" - I'm unclear as to what the "30% contamination" means in practice and the pre-analysis plan ought to be much more specific. In addition, it would not only affect "our interpretation of the impact evaluation" but also (i) how it is carried out and (ii) how it is communicated to the public.

This has been clarified in the report.

2.2 Typhoon Haiyan - The issue isn't so much whether they were equally affected or not, which given the random allocation is unlikely not to be case, but whether, the typhoon prevented KC impacts from materializing. That is the issue is whether some of sub-projects were destroyed. Let's imagine that a given barangay, the community prioritized a road. The barangay was prioritized and the road was built just before the typhoon and was washed away by the typhoon. If that's the case then we shouldn't expect to observe any impact.

If we know that there weren't differential impacts between treatment and control communities from the typhoon or any other natural disaster, then we can be confident that any additional impact in treatment areas is from KC. Moreover, it is possible that both treatment and control communities fared worse in some variables after weather-related disasters, but KC could have made communities less worse off. This would be considered a positive impact. We know from the interim data that natural disasters outside of typhoon Yolanda (eg basic flooding) destroyed KC projects. Given that weather is a common phenomenon across Philippines, any project that seeks to improve social welfare would be expected to do so in the face of such challenges.

3. Analysis

3.1 The team intends to use the sample of the 80 barangays regardless of whether or not the assignment was complied with. Given that the randomization was done within pairs, why not drop the affected pairs?

Yes, you are correct that we intend to use intent to treat analysis. Dropping municipalities that didn't comply would bias the results since non-compliance is endogenous. The practicalities of national project implementation mean that some municipalities might not comply and the objective in policy research is to deliver an impact result relevant to a variety of municipalities, not just the higher performing ones.

3.2. Coming back the issue of power highlighted above, the team might be able to use data from the baseline to get specification that could improve power. The basic idea, inspired by Barrios (2013), is to optimally chose from a set of Xs the variables that best explain Y and to commit to controlling for them in the analysis to improve power. Indeed, Barrios (2013) proposes an interesting method to use a set of Xs before the randomization is carried out to optimally carry out the stratification. While it's too late for that (and in any case the method can't be used when

the evaluation focuses on more than one outcome) but one of the paper's key finding is that his approach is actually similar to controlling for Xs in the regression:

This paper considers the gain in efficiency from effective stratification. We show that stratifying, in the case of matched pairs, leads to significant efficiency gains, that gains will be large if baseline variables are good predictors of the outcome of interest, and that it is optimal to stratify on the conditional expectation of the outcome given baseline variables. Simulations show that the gain in efficiency is comparable to having controlled for covariates in the analysis after randomization. That is, given a set of covariates X, matching on predictions based on X and estimating the difference in means ex-post gives estimators with mean squared error of the same size as performing a complete randomization and controlling for X with regression ex-post (Barrios 2013).

Put differently, the team could use the baseline data to, for each outcome of interest, select the Xs that best explains Y and run those regressions. I understand that the team might be reluctant to open itself to criticism of fishing but as long as the estimated regressions are listed in the pre-analysis plan before getting access to the interim data I don't see why this would be a problem.

The Barrios paper proposed the use of the predicted outcomes of interest as a function of baseline variables to identify pairs before random assignment into treatment. It is claimed that this has the effect controlling for the baseline variables after randomization. This result is mentioned as the justification for proposing the use of the best predicting baseline covariate in the estimation. It should be noted that the current specifications already include the lagged values of the outcome variable of interest. If the simulations in Bruhn and McKenzie (2009) are to be believed, they had pointed out that the lagged values of the outcome variable is one of the most strongly correlated with the outcome. This is likely to be true for many of the outcome variables, i.e. governance and trust, in the proposed analysis. If there are other candidate baseline variables that are known in the literature to determine the outcomes of interest distinctly well, this can be used as well.

3.3. Hypothesis 7 (page 7): IPA intends to use the match between sub projects selected during the SCA and preferences expressed at baseline as measure of community capacity to act collectively. There's two potential concerns with this. First, as indicated on page 5, the SCA provides each community with "a small sum of money" which might not be sufficient to deal with the most commonly identified problem. Second, the team needs to account for the possibility that some of the issues identified at baseline have been deal with by KC.

We appreciate that the SCA options are limited and that the SCA funding might not allow the community to realize their preferences expressed in the baseline. We will be sure to provide thorough explanation of this analysis by 1) discussing first whether communities preferences in baseline were met by KC (through the quantitative surveys), and 2) attempting to see if there is a loose correlation between SCA projects and baseline preferences. We would not use these data to conclude that the needs of the community were not met.

3.4 "Between-barangay subgroups" (page 9) - which barangay governance measure will the team use?

Are you referring to "Barangays where levels of baseline governance are in the top 50%."? If yes, then we will use the indices associated with H1-3 described in Appendix 1.

3.5. Effects of being prioritized (page 9/10) - The RDD approach (ie comparing barangays that were just above the threshold to be prioritized and those that were just below) is appealing but I'm wondering whether it will be practical in the current situation given the sample size and the fact that only one barangay was sampled by municipality. Indeed, the estimation sample is likely to be quite small as the team can only use the 40 barangays in the treatment group and only a few of them will be around the threshold. Also, since the team won't be able to use barangays in the control group you won't be able to include the pair-wise fixed-effects (Ds)

If the estimation is impossible due to sample size issues and SP implementation, we will not undertake this analysis.

References:

Barrios (2013) "Optimal Stratification in Matched Pairs Experiments " Harvard University, mimeo
<http://scholar.harvard.edu/tbarrios/publications/optimal-stratification-matched-pairs-experiments>

Hope this is useful. Happy to discuss further.

All the best,
Julien

End Julien's comments