

# Testing the Effectiveness of Mobile Phone Data Collection for Microenterprises in Africa: **ADDENDUM TO PRE-ANALYSIS PLAN**

Robert Garlick\*, Kate Orkin<sup>†</sup>, Simon Quinn<sup>‡</sup>

June 28, 2016

We registered a pre-analysis plan on 7 April 2014 at the AEA's RCT Registry. The plan is available at <https://www.socialscisearch.org/trials/346>. The pre-analysis plan was registered after the repeated interviews started but before the endline interviews started or we received any data from the repeated interviews. This addendum lists and explains differences between the analysis reported in the paper and the analysis proposed in the pre-analysis plan.

1. We listed seventeen outcome of interest in the pre-analysis plan. We omit results for two measures of reporting accuracy in item 3.1(x) of the pre-analysis plan. We could not verify that respondents received airtime payments on the same date that we made them. Hence we are not confident interpreting the difference between when respondents report receiving airtime and when we made airtime payments as a measure of respondent accuracy. We constructed a measure of consistency in sales reporting (absolute value of the difference between week  $t$  sales reported in week  $t$  and week  $t$  sales reported in week  $t + 1$ ). However, this measure was only available for enterprises in the two weekly interview groups and only for sequential successful interviews. The resultant sample size was low and the results are noisy.
2. We proposed standardizing all outcome measures to have mean 0 and standard deviation 1 in the monthly interview group. For ease of interpretation, we do not standardize binary outcome measures: respondent is still operating enterprise, enumerator's assessment of respondent's honesty, enumerator's assessment of respondent's carefulness, respondent's use of notes or records in the survey.
3. We proposed estimating quantile regressions of repeated and endline outcomes on treatment indicators, stratum fixed effects, and (for the repeated interviews) survey week fixed effects.

---

\*Department of Economics, Duke University; [robert.garlick@duke.edu](mailto:robert.garlick@duke.edu).

<sup>†</sup>Department of Economics, Centre for the Study of African Economies and Merton College, University of Oxford; [kate.orkin@merton.ox.ac.uk](mailto:kate.orkin@merton.ox.ac.uk).

<sup>‡</sup>Department of Economics, Centre for the Study of African Economies and St Antony's College, University of Oxford; [simon.quinn@economics.ox.ac.uk](mailto:simon.quinn@economics.ox.ac.uk).

We omit the stratum and survey week fixed effects because estimating quantile regressions with approximately 110 right-hand side variables proved computationally burdensome.

4. We proposed estimating systems of simultaneous quantile regressions at percentiles 5, 10, 25, 50, 75, 90, and 95. We then proposed to conduct joint hypothesis tests on all 7 quantile treatment effect estimates for each outcome. We now believe that there is no available standard error estimator for simultaneous estimation of quantile regressions that allows clustering. We therefore estimate the regressions independently across the 7 quantiles. We use sharpened  $q$ -values (Benjamini, Krieger, and Yekutieli, 2006) instead of  $p$ -values to conduct separate hypothesis tests on the 7 quantile treatment effect estimates for each outcome.
5. We proposed measuring differential attrition in the repeated interviews by regressing an attrition indicator on treatment indicators, stratum fixed effects and survey week fixed effects using enterprise-by-week-level data and clustering standard errors by firm. We instead collapse the data to firm-level data and regress the fraction of missed interviews (out of 4 or 12) on treatment indicators and stratum fixed effects, clustering standard errors by firm. The original approach we proposed required an unbalanced panel regression because the enterprises assigned to monthly interviews would have fewer observations. The results from our original and revised methods are almost identical.
6. We proposed estimating Lee bounds for the mean regressions using repeated interview data, standard deviation regressions using repeated interview data, and mean regressions using endline interview data. We do estimate Lee bounds for the mean regressions but not for the standard deviation regressions. The theory underlying Lee bounds is not applicable to the standard deviation regressions that we run.
7. We proposed estimating heterogeneous treatment effects on eight dimensions: gender, education, digit recall, whether the firm reported keeping financial records, the number of employees, financial literacy, numeracy, and whether the firm commonly uses a phone for business. We report heterogeneous treatment effects on the first five dimensions. We aggregate financial literacy and numeracy into a single measure and report heterogeneous treatment effects on this dimension.
8. We proposed testing if the variance of each endline outcome measure is equal across treatment groups. We instead report quantile regressions of endline outcomes on treatment fixed effects, which we believe provide more information.
9. We proposed reporting  $p$ -values, sharpened  $q$ -values to control false discovery rates (Benjamini, Krieger, and Yekutieli, 2006), and family-wise error rates based on a stepdown bootstrap (Westfall and Young, 1993) for all hypothesis tests. We report only  $p$ -values and sharpened  $q$ -values. Tests that control family-wise error rates are generally more conservative but this is not obviously desirable given our research design and research questions. For the false discovery rates, we use the outcome families defined in the pre-analysis plan.
10. We proposed investigating predictors of enterprise closure. We omit this analysis because enterprise closure is relatively rare: 5.4% from baseline to endline.

## References

BENJAMINI, Y., A. M. KRIEGER, AND D. YEKUTIELI (2006): “Adaptive Linear Step-Up Procedures that Control the False Discovery Rate,” *Biometrika*, 93(3), 491–507.

WESTFALL, P. H., AND S. S. YOUNG (1993): *Resampling-Based Multiple Testing: Examples and Methods for  $p$ -value Adjustment*.