



Impact Evaluation (IE) Concept Note Template
Microenterprise Supply Chain Intermediation Pilot

Colombia
IE-P158250-IMPE-TF0A1582
March 10, 2016

JEL codes: O12, O17, D22, L23,

Keywords:¹ **Microenterprises, Supply Chain, Pricing, Randomized Control Trial**

Choose one of more keywords/categories that describe your IE.² (R)

¹ Please refer to JEL classification codes <http://papers.ssrn.com/sol3/displayjel.cfm> .

² The concept note is aligned to Ethical clearance (E) and Registry (R) indicative requirements. These indicative requirements are referenced throughout the document.

Table of Contents

IE PROFILE INDICATORS	3
1. EXECUTIVE SUMMARY	4
2. BACKGROUND AND KEY INSTITUTIONAL FEATURES.....	5
3. LITERATURE REVIEW (E).....	5
4. POLICY RELEVANCE	6
5. THEORY OF CHANGE (E).....	7
6. HYPOTHESES/EVALUATION QUESTIONS (E,R)	8
7. MAIN OUTCOMES OF INTEREST (E,R).....	9
8. EVALUATION DESIGN AND SAMPLING STRATEGY (E,R).....	10
8.1 TREATMENT AND CONTROL GROUPS.....	12
8.2 SAMPLE SIZE CALCULATIONS	12
9. DATA COLLECTION (E,R).....	15
9.1 QUANTITATIVE INSTRUMENTS.....	16
9.2 MANAGEMENT OF DATA QUALITY	16
9.3 ETHICAL ISSUES.....	17
9.4 QUALITATIVE INSTRUMENTS.....	17
9.5 IE IMPLEMENTATION MONITORING SYSTEM (R)	17
10. DATA PROCESSING AND ANALYSIS	17
10.1 DATA CODING, ENTRY, AND EDITING (E).....	17
10.2 MODEL SPECIFICATION FOR QUANTITATIVE DATA ANALYSIS.....	18
11. STUDY LIMITATIONS AND RISKS (E).....	19
12. IE MANAGEMENT (E,R).....	20
12.1 EVALUATION TEAM AND MAIN COUNTERPARTS.....	20
12.2 WORK PLAN AND DELIVERABLES	20
12.3 BUDGET	21
13. PLAN FOR USING DATA AND EVIDENCE FROM THE STUDY	21
REFERENCES.....	22
APPENDIX	ERROR! BOOKMARK NOT DEFINED.
I2I INDICATORS OF IE INFLUENCE ON PROGRAM/POLICY	ERROR! BOOKMARK NOT DEFINED.
THEORY OF CHANGE EXAMPLE.....	ERROR! BOOKMARK NOT DEFINED.

IE PROFILE INDICATORS

No.	Indicator	Description
1	IE code	IE-P158250-IMPE-TFOA1582
2	IE Title	Microenterprise Supply Chain Intermediation Pilot
3	IE TTL	David McKenzie
4	IE Contact Person	David McKenzie, DECFP
5	Region	LCR
6	Sector Board/Global Practice	T&C, F&M
7	WBG PID (if IE is evaluating a WBG operation)	P144510
8	WBG Project Name (if IE is evaluating a WBG operation)	Colombia Innovation and Competitiveness Programmatic Approach
9	Project TTL (if IE is evaluating a WBG operation)	Leonardo Iacovone
10	Intervention	SMS/phone ordering and order aggregation for small vendors to reduce the supply chain
11	Main Outcomes	Firm unit costs, Firm prices, Firm profits and sales
12	IE Unit of Intervention/Randomization	Geographic block within Bogota
13	Number of IE Units of Intervention	60 blocks
14	IE Unit of Analysis	Individual Firm
15	Number of IE Units of Analysis	1800 firms
16	Number of Treatment Arms	2 (Treatment and Control)
17	IE Question 1 (Treatment Arm 1)	What is the impact of supply chain intermediation on the costs, prices, and profits of small vendors?
18	Method IE Question 1	Random Assignment at the Block level
19	Mechanism tested in IE Question 1	Package
20	IE Question 2 (Treatment Arm 2)	N/A
21	Method IE Question 2	N/A
22	Mechanism tested in IE Question 2	N/A
23	IE Question 3 (Treatment Arm 3)	N/A
24	Method IE Question 3	N/A
25	Mechanism tested in IE Question 3	N/A
25	Gender-specific treatment (Yes, No)	No
27	Gender analysis (Yes, No)	TBD (we don't know what percent of vendors will be female)
28	IE Team & Affiliations	Name 1 (Organization/Unit Affiliation, Role); Name 2 (Organization/Unit Affiliation, Role)...David McKenzie (World Bank, DECFP, PI and IE TTL); Leonardo Iacovone (World Bank, GTC04, co-PI and operational TTL); IPA Colombia (survey team and field assistant).
29	Estimated Budget (including research time)	487,000
30	CN Review Date	4-2016
31	Estimated Timeframe for IE	3-2016 to 6-2018
32	Main Local Counterpart Institution(s)	Agruppa; MinCit.

1. EXECUTIVE SUMMARY

(1 page)

- Describe the proposed IE **in non-technical language in one paragraph or less**. This could be an abstract of your IE. Include broad motivation/background and policy/research contribution. (E,R)
- Present IE questions and main outcome(s) the intervention aims to affect.
- Briefly explain how you are proposing to test your main evaluation question(s).

Small-scale vendors in urban slums around the world face considerable inefficiencies in their value chains for inputs that result in high travel costs and higher input prices, and limited product diversity for consumers. For example, vendors in the slums of Bogota spend an average of 15 hours per week and 20 percent of their weekly revenue travelling to make inventory purchases, and because they purchase small quantities at any point in time, do not benefit from bulk discounts. There is a coordination failure amongst such firms, since competition and lack of trust between each other prevents them from banding together to share travel costs and combine discounts. Agruppa, a start-up based in Colombia, aims to overcome this problem by leveraging mobile phone technology to aggregate small vendors' demand for produce, creating collective orders which add up to wholesale quantities, and delivering the goods directly to firms from farms, cutting out middlemen in the supply chain. This impact evaluation will test the effectiveness of this new technology.

This approach will be tested in a pilot program being implemented by Agruppa in Bogota. A census of 2,500 vegetable vendors in poor neighborhoods in the south of Bogota was conducted, and these firms were subdivided into 60 geographic blocks. Within each block, the baseline survey assesses whether or not firms would be interested in using the services of Agruppa. We then randomly assign 30 of these blocks to be treatment blocks, and 30 to be control. Within each treated block, those firms which indicated their interest in the Agruppa program will be marketed the service. High-frequency surveys will then measure the prices, unit costs, travel time, and profits and sales of both interested and uninterested firms in treatment and control blocks over time.

The Primary Research Questions and Main Outcomes are:

- 1) Can the technological solution offered by Agruppa overcome coordination issues among firms to reduce the amount of time they spend travelling to purchase inventories? [travel time measured in surveys]
- 2) Does the technology lower the cost firms pay for their inventories? [measured by surveys and market price collection]
- 3) Are lower costs passed onto consumers in terms of lower prices? [measured by surveys of good prices]
- 4) Do the lower costs and incomplete pass-through or elastic demand result in higher sales and higher profits for firms? [measured by surveys]
- 5) Do non-participating firms lose sales as a result of their competitors having lower prices or longer opening hours? [measured by surveys]

Questions 1)-4) will be tested by comparing outcomes for interested firms in the treatment blocks to outcomes for interested firms in the control blocks. Question 5) will be tested by comparing outcomes for uninterested firms in the treatment blocks to uninterested firms in the control blocks.

2. BACKGROUND AND KEY INSTITUTIONAL FEATURES

(1 page)

- Present an overview of the local context.
- Identify and define the problem: what is the policy/research problem this IE is proposing to study? Which groups are affected by the problem?
- Describe the intervention whether existing or new, implementing organization, institutional setting and any important consideration.
- Describe the intervention geographic/demographic scale and scope: Does it represent the “mode” of delivery in the country? (R, E)

Small, food retail, businesses are the last link in a long and fragmented supply chain in many urban cities. The vendors are the main source of food consumed in low-income neighborhoods, and inefficiencies in this supply chain can result in limited dietary diversity and high prices for poor consumers (Ruel et al, 1998). In Colombia alone, there are over 340,000 small stores of this type, of which 85,000-110,000 are located in the capital of Bogotá (FAO 2010).

These small vendors face several key challenges in running their businesses. Because they sell small volumes daily of a perishable product, they must make frequent, small, purchases of vegetables to sell. The small size of these purchases prevents them from being able to take advantage of bulk discounts, resulting in high unit costs. Moreover, the current supply chain centers around one central marketplace (*Corabastos*), and vendors must travel there each day to purchase goods. This travel can be time-consuming and expensive, with estimates from Agruppa’s field research suggesting that vendors spend an average of 15 hours per week, and 20-30 percent of their weekly income, travelling to the marketplace. These added costs lower firm profits and increase prices for their consumers. In principle vendors could overcome both of these challenges by pooling their orders together (allowing them bulk discounts) and sharing travel costs. However, in practice competition amongst firms coupled with communication and coordination costs prevents them from working together in this way.

The intervention is to then to introduce a new technology developed by Agruppa, which leverages mobile phone technology to create virtual buyer groups in poor urban areas of the world. The system aggregates demand from small vendors, creating daily collective orders which together add up to wholesale quantities. Thanks to economies of scale, these orders can be purchased at prices significantly below (an estimated 30%) what vendors currently pay when sourcing individually. By bulk buying from farmer co-operatives and delivering produce directly to the vendors’ doorstep, Agruppa aims to reduce supply chain inefficiencies, lowering the cost of fresh produce and saving vendors the time and money currently spent travelling to central markets. The goal is also that people living in these low-income communities then benefit from sustainable nutritious food at accessible prices.

The implementing organization, Agruppa, is an NGO that is currently in the piloting stage of deploying this technology. In its initial pilot, 80% of those offered the product signed up for it, and 50-60% became weekly users. The intervention will be rolled out in poor socio-economic neighborhoods of Bogota, following Agruppa’s standard mode of delivery. As a result, it will be scalable should the results show that the program succeeds. As discussed below in the policy section, there is broader interest from the Ministry of Trade, Industry and Commerce (MINCIT

3. LITERATURE REVIEW (E)

(1 page or less)

- Describe most relevant literature/scientific background specifically linked to your problem/evaluation question(s).

There are two strands of literature that are related to the problem being investigated here: a literature on inserting firms into value chains, and a literature on supply chain intermediation and disintermediation.

On the policy side, while there has been increasing attention given to the role of value chains or supply chains, most of this emphasis has been on trying to link small producers (such as farmers or artisans) to global value chains and larger buyers. The underlying constraints seen in this work are the fixed costs of investment and information needed for small firms to connect to these value chains, and whether small firms can produce at the quality levels needed for global markets (OECD and World Bank, 2015). Evidence on the effectiveness of policy efforts to connect firms to these value chains is still scarce, with a recent systematic review of SME programs by Piza et al. (2016) only finding two evaluations of these programs (using non-experimental methods). Arráiz et al. (2013) find that Chile's supplier development program, by linking suppliers to large firm customers, was able to increase the sales and employment of these small firms. In contrast, in our context the small firms are the retailers, rather than producers, and the challenge is not inserting them into value chains, but shortening the supply chains that they face.

While we have not found any impact evaluation literature which measures the effectiveness of attempting to shorten supply chains for firms, there is an economic literature on supply chain intermediation which is summarized in a handbook chapter by Wu (2004). There are two insights from this literature that are useful for our work. The first is work which analyzes why intermediaries exist in the first place, and what changes in market conditions can make these intermediaries no longer needed. In a supply chain with retailers, wholesalers, and farmers, the wholesaler is seen to create value by aggregating orders from multiple retailers, reducing the retailer's shortage risk by offering products from multiple farmers, and lowering transaction costs as a result. However, in this environment, a new technology which aggregates orders and also sources from multiple farmers to manage shortage risk may lower transaction costs enough that the wholesaler will be disintermediated. The literature has then discussed how this process has shortened supply chains in business to business e-commerce by cutting out middlemen. An important insight for our work from this process is that intermediaries are still important in electronic commerce, just the form of intermediation changes. In our context, it may therefore be that the number of parties in the supply chain may not necessarily shrink with the introduction of Agruppa, but rather the transaction costs along this supply chain.

4. POLICY RELEVANCE

(1/2 page or less)

- Assess the extent to which the study may influence policy and institutional capacity at the national, regional, and international level. Explain how you plan to track the policy influence of your study (see Appendix on i2i sample indicators of IE influence on program/policy. These indicators, which are currently under revision, will be collected through Grant Monitoring and Reporting on annual basis from all i2i supported IEs).

This project responds directly to three key objectives of the Colombian National Development Plan 2014-2018 recently approved: (1) improve productivity of Colombian firms, (2) contribute to productive development through [science,] technology and innovation, (3) promote ICT as a platform for [equity, education and] competitiveness. Accordingly, the Director for MSMEs of the Ministry of Industry, Commerce, Services and Tourism (MINCIT) has been involved in the discussion about the design of the pilot since the beginning and the MINCIT is very interested in the results of this pilot and would consider expanding it at national level depending on the results of the impact evaluation.

In addition, this project is very well aligned to the currently ongoing Programmatic Approach for competitiveness and innovation which objective is to stimulate increased innovation and technology development. A key indicator of this programmatic approach is to strengthen the capacities of the Government of Colombia to design and rigorously evaluate interventions to improve the productivity and profitability of MSMEs.

5. THEORY OF CHANGE (E)

(1 Figure and 2-3 paragraphs)

- Describe the main elements of the intervention, and the hypothesized causal chain from inputs, through activities and outputs, to outcomes.
- Describe the main assumptions and other factors underlying the causal chain (internal and external).

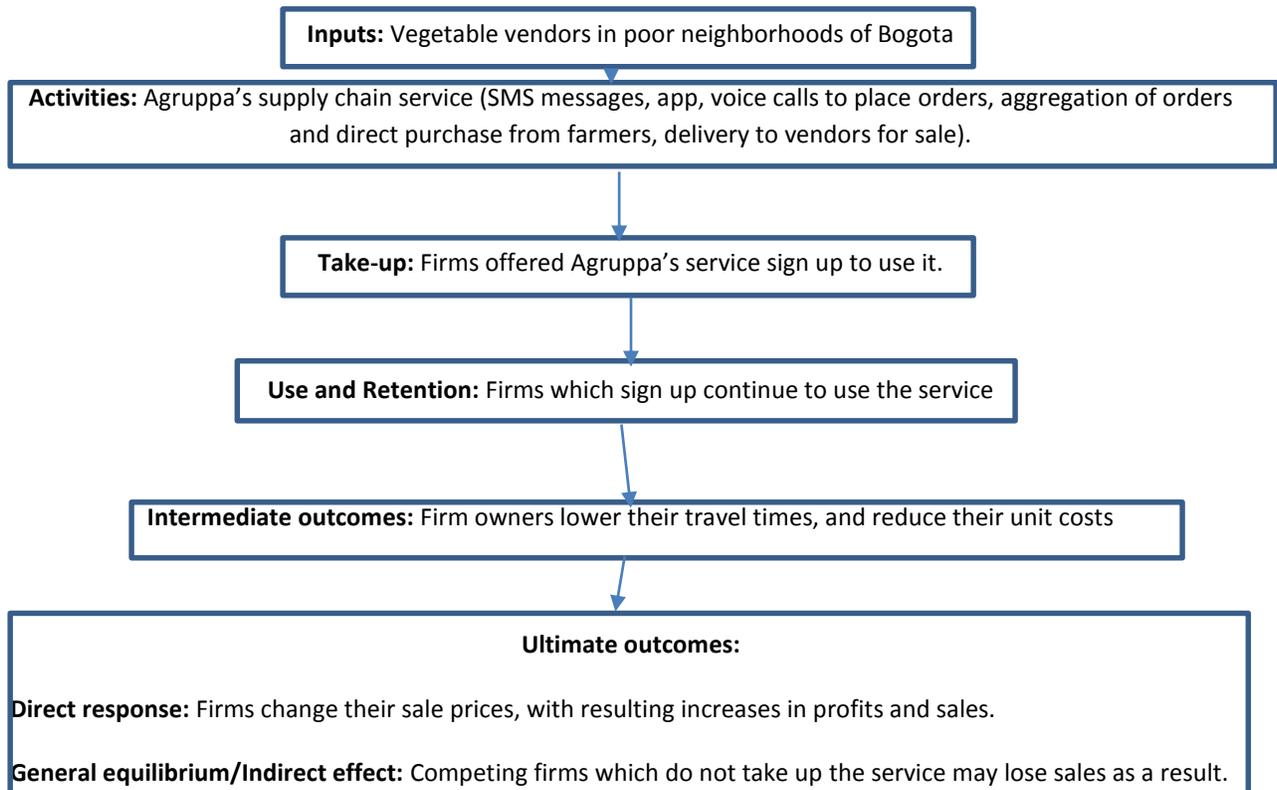
A baseline survey will screen firms for eligibility (to be discussed below), and for interest in Agruppa's product. Firms will then be classified as interested or uninterested, and blocks of firms randomly assigned to treatment or control.

The intervention will consist of Agruppa offering its service to interested firms in the treated areas to allow them to purchase the four main fresh vegetable products sold in these areas: tomatoes, onions, plantains, and potatoes. The service will then operate as follows:

- 1) Vendors place their daily order of products through SMS text messaging, an app, or a phone call (all three options will be available to them).
- 2) Agruppa's software receives the incoming orders and creates daily collective order per product, with the individual orders of vendors in a certain area.
- 3) Agruppa then bulk purchases the order from farmers, access the wholesale price and eliminating mark-ups typically charged by middlemen in the value chain. This can result in 5 to 30 percent savings on products, and also potentially lead to higher quality, fresher, produce.
- 4) Food is then delivered by truck to Agruppa's Distribution outlet, which is located close to the targeted neighborhoods. It is then delivered directly from there to the vendors' stores within a specified time window.

Figure 1 then sets out a simplified theory of change.

Figure 1: Theory of Change



In terms of initial inputs, the underlying assumption is that there are many vegetable vendors who would be interested in a service that would help lower their unit costs and reduce their travel times. The theory is then that the Agruppa service will attract them and get sufficient take-up and retention to be able to measure the impact of this program on these firms. The service will need to lower travel time and reduce unit costs for it to achieve these goals, and for it to ultimately affect the profits and sales of the firms. The assumption is also that firms who do not initially express interest in the program may experience a drop in sales if their competitors are able to achieve lower unit costs, which in the longer-term may also induce them to take up the service. Ultimately the assumption is that this competition will result in lower prices for consumers in these neighborhoods, while increasing the incomes of the small vendors.

6. HYPOTHESES/EVALUATION QUESTIONS (E,R)

(1/2 page)

- List the hypotheses derived from your theory of change.
- List the main evaluation question(s) to be addressed by the proposed study. Evaluation questions connect the specific intervention/treatment variation to the outcomes of interest, and end with a question mark. They should be in the following format: What is the impact of <intervention/intervention variation> on <outcomes>? E.g., What is the impact of a parental workshop on financial literacy on student knowledge, attitudes and behavior?

- **You may have a broad evaluation question** based on the knowledge gap and the strategy proposed. **However, the number of specific questions in this section should be perfectly aligned to the number of your treatment arms** (i.e., if you have 3 treatment arms you should have three specific evaluation questions). Each question can be evaluated on a vector of outcomes (i.e., you may organize them as sub-questions). Methods to answer sub-questions on heterogeneous treatment effects and spillovers should be described in the methods section.
- Describe how the evaluation questions were derived.

Evaluation Question 1: What is the impact of using Agruppa’s supply chain service on the costs, travel time, prices, profits, and sales of firms which are offered the service?

Hypothesis 1a: Use of Agruppa’s service will result in a result in travel time and travel costs for participating firms.

Hypothesis 1b: Use of Agruppa’s service will result in a reduction in unit costs for vegetables for participating firms

Hypothesis 1c: Participating firms will pass through some of the reduction in unit costs to prices, so they will charge lower prices for their products.

Hypothesis 1d: Participating firms will sell more products and make higher profits. This will come through longer opening times (since less time is spent travelling), through attracting more customers (due to lower prices), or to make higher profits on the same amount of sales (due to reduced costs)

Evaluation Question 2: What is the impact of competitors using Agruppa’s service on firms not offered the service to begin with?

Hypothesis 2a: In the short-term non-participating firms will suffer a drop in sales or in profits as a result of their competitors reducing prices.

Hypothesis 2b: Over the longer-term, some of these initially uninterested firms will respond to the loss of sales by signing up for Agruppa’s service.

These evaluation questions follow naturally from the theory of change outlined above.

7. MAIN OUTCOMES OF INTEREST (E,R)

(1 table)

- Briefly list and define main outcomes of interest (primary and secondary/intermediate) as in Table 1.
- Further details on how the outcomes will be measured/collected will go in the data collection section.

Table 1. Main Outcomes of Interest

Outcome Type	Outcome Name	Definition	Measurement Level
Primary/Secondary			
Primary	Travel time	Time spent by firms travelling to purchase inputs	Firm level from surveys

Primary	Unit cost	Unit cost for purchasing each of the four main vegetables offered by Agruppa	Administrative data from Agruppa; Firm level data from surveys
Primary	Price	Price per unit charged by firm for the four main vegetables offered by Agruppa	Firm level data from surveys
Primary	Sales	Vegetable sales made by the firm in a week	Firm level data from surveys
Primary	Profits	Profits made by the firm from vegetable sales	Firm level data from surveys
Secondary	Take-up rate	Percent of those offered the service who take it up	Administrative data from Agruppa
Secondary	Profits and Sales of Uninterested firms	Profits and Sales for Vegetables sold by uninterested firms	Firm level data from surveys

8. EVALUATION DESIGN AND SAMPLING STRATEGY (E,R)

(2 pages or less)

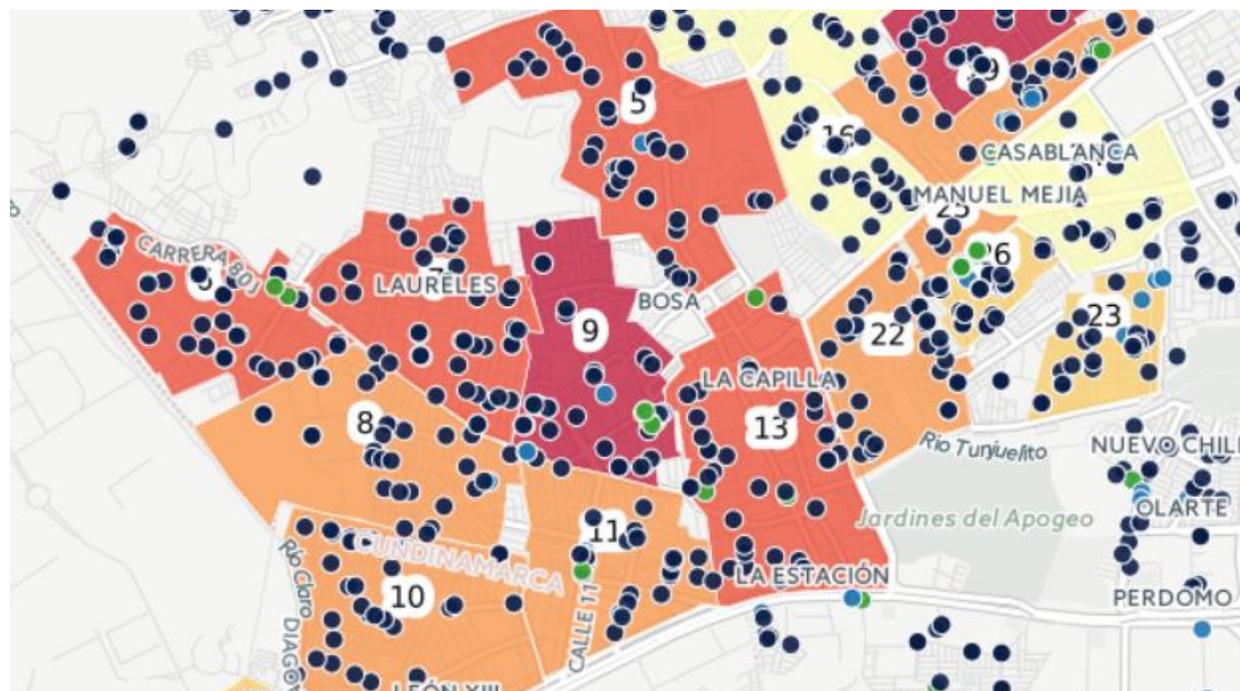
- Present the main features of the proposed evaluation design to address the evaluation question(s).
- Describe precisely the identification strategy (e.g., trial design including clustering, factorial, stratification details) for each evaluation question.
- Report all inclusion/exclusion criteria to define the target population/population studied, providers, settings, and clusters (as relevant).
- Report any **ethical issues** that may arise concerning the evaluation design and the sampling strategy (not related to data collection).

This evaluation is designed as a randomized controlled trial.

Step One: Take a Census of the Potential Firms of Interest

Agruppa selected the south of Bogota as the region where it will operate, in neighborhoods of socioeconomic status 1, 2 and 3 (where socioeconomic status is defined by the government on a 1 to 6 scale, where 1 is the poorest). Two neighborhoods known to be notoriously unsafe when then excluded, as were purely industrial zones, and neighborhoods too close to their current piloting area. Then a door-to-door listing census was taken of shops, carts, and stalls selling at least one of the the products (potato, plantain, tomato and onion). Altogether more than 2,500 firms were listed.

Step Two: Divide the 2500 firms into approximately 60 equal blocks, using major roads/natural geographic breakpoints to divide geographically. The Figure below shows an example of some of the first such blocks formed, with an average of 35 firms per block.



- BODEGA AGRUPPA
- CARRETA
- CORABASTOS
- PUESTO
- TIENDA



Colors indicate number of firms per block.

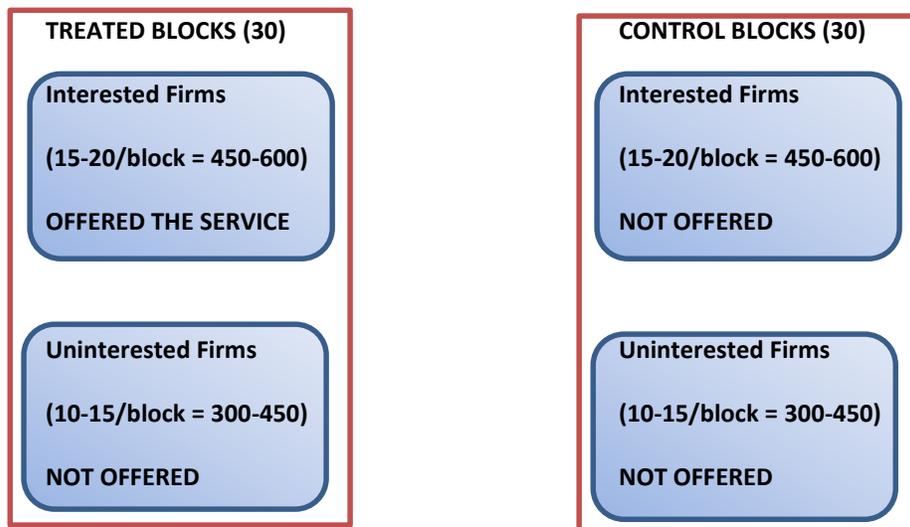
Step Three: Randomly assign blocks to treatment and control, stratifying on cluster size (number of firms in the block) and geography. This will be done in batches, giving us the ability to redefine the boundaries of the remaining blocks if the first few blocks indicate take-up rates which differ from our assumptions.

Step Four: Conduct a baseline survey of the firms in the selected treatment and control blocks. This baseline survey will collect key information on where the firm sources products, travel time, unit costs and prices for the main products, other products sold, total sales and profits, etc. At the end of the survey, we will explain the service that Agruppa provides and explain that it is currently being slowly rolled out into different neighborhoods in Bogotá. We then ask whether they would be interested in using the service. This is used to classify firms into “interested” and “uninterested” firms within each block.

Step Five: Screen firms to reduce heterogeneity. Using the baseline information, we will screen out firms with sales or profits which differ dramatically from other firms in the survey, in order to reduce the baseline heterogeneity

amongst firms and increase power. (Exact cutoffs to be determined upon seeing the baseline data and how heterogenous firms are). We aim to reduce the 2500 firms to approximately 1800 in this way.

Step Six: Firms in treated blocks that indicated that they were interested in Agruppa’s service and meet our screening criteria will then be visited by Agruppa who will explain the service. They will use marketing techniques such as a heavy discount on the first order and frequent customer discounts to encourage take-up and retention. This will give us the following four groups:



We then measure the direct impact of the program by comparing outcomes for the interested firms in the treated blocks to the interested firms in the control blocks, and the indirect impacts by comparing uninterested firms in the treated blocks to those in the control blocks.

8.1 TREATMENT AND CONTROL GROUPS

- Provide specific description of features of each control and treatment arm (one paragraph per arm).

As discussed above the **treatment** group will consist of the 450-600 firms in the treated blocks who indicate interest in the Agruppa service in the baseline survey. The **indirectly treated** group will consist of the remaining (300-450) uninterested firms in the treatment blocks. The 450-600 interested firms in the control blocks will serve as a **control** group for the treatment group, and the 300-450 uninterested firms in the control blocks will serve as a control group for the indirectly treated firms.

8.2 SAMPLE SIZE CALCULATIONS

- Present the sample size estimates. Describe how the sample size was determined, including the sampling frame, and main assumptions including Minimum Detectable Effect (MDE), variance estimates, intra-cluster correlation, and units per cluster (if applicable).

Since these firms are all operating in the same sector, selling similar products at the same scale, they will be more homogeneous than firms in many standard PSD interventions. This should improve the power of our

experiment to detect impacts. Moreover, we believe that unit costs and travel time should be highly correlated over time, and will use multiple weeks of data to further improve power. The baseline data is required to fine-tune these power calculations, as we stand ready to adjust the design as needed if we find that based on the baseline data and initial take-up results that power is lower than anticipated. That said, here are some initial power calculations based on two outcomes that we have used to help guide our preliminary choices on sample size.

We assume take-up of the intervention will be 65% among those who express interest—this is conservative, given Agruppa have found approximately 80% take it up in their initial pilots. Our design at present is then random assignment at the market block level of 30 blocks to treatment and 30 to control, with each block containing 20 firms interested in the intervention, for a total of 600 treated and 600 control firms among the interested (we also have uninterested firms in each block).

Our starting point assumptions are that i) key outcomes are likely to be highly autocorrelated (firms that have long travel times today will have long travel times in a month due to geography, mode of transport, etc.); ii) key outcomes will have strong intra-cluster correlations in the cross-section within market blocks (e.g. firms that are all within the same block will charge similar prices and have similar travel times to market); and iii) the intra-cluster correlations will be much weaker in terms of changes (firms within blocks experience different shocks and react differently to the intervention).

To see this, consider the time in hours per week spent to travel to market. In the cross-section at time t , we model this for firm i in block b as:

$$hours_{i,b,t} = \alpha_i + \beta block_b + u_{b,t} + \varepsilon_{i,b,t}$$

Then if we consider the correlation between two firms in the same block in the cross-section, we have:

$$Cov(hours_{i,b,t}, hours_{j,b,t}) = \beta^2 Var(block_b) + Var(u_{b,t})$$

This can be high if block-level time-invariant factors are important determinants of travel time. But then with panel data we can write:

$$hours_{i,b,t+1} = hours_{i,b,t} + (u_{b,t+1} - u_{b,t}) + (\varepsilon_{i,b,t+1} - \varepsilon_{i,b,t})$$

Or more generally in our ANCOVA framework with a treatment effect:

$$hours_{i,b,t+1} = a + bTreat_{i,b} + c hours_{i,b,t} + v_{b,t+1} + u_{i,b,t+1}$$

We see that if the autocorrelation is high (so c is close to one), then the intra-cluster correlation in the error term

depends only on $u_{b,t+1} - u_{b,t}$ and not on $block_b$. We now apply this to two key outcomes.

Outcome 1: Reduction in travel time spent by firms to buy goods from the central market

Assumptions: Mean weekly travel time of 15 hours (estimated by Agruppa), standard deviation 5 hours, intra-cluster correlation of 0.6 (since firms within markets will have similar travel times). The intervention aims to reduce hours by an average of 5. The ITT is thus 3.25 hours (5×0.65).

Power using just a single round of follow-up data:

If randomization was at the individual level, power is 1:

`sampsi 15 11.75, sd(5) n1(600) n2(600)` gives power of 1

and we would need only 38 firms in each group to get 80% power

`sampsi 15 11.75, sd(5) power(0.8)`

But with an intra-cluster correlation of 0.6, and 20 firms per market, we need a minimum of 48 blocks (and hence 472 in each treatment group) to achieve 80 percent power.

Power using the baseline to improve power:

We assume now that using the baseline hours we have an autocorrelation of 0.7 with follow-up hours, so the residual variance becomes $\sqrt{1-0.72} \times 5 = 3.57$ hours. We assume that this then reduces the intra-cluster correlation in hours to 0.3 (once the block component has been removed). Then we have:

`sampsi 15 11.75, sd(3.75) power(0.8)`

`sampclus, obsclus(20) rho(.3)`

Gives a minimum of 15 blocks, and 141 treated and 141 control in total needed.

Outcome 2: Reduction in Price Charged by Firms for The Products Sold to them by Agruppa

Assumptions: we currently have no price data, so we are using data from tomato vendors in a Kenya microenterprise study one of the PIs has underway to provide some basic numbers here. Mean price is 15, s.d. is 6, intra-cluster correlation is 0.1, autocorrelation is 0.8. The intervention aims to lower direct costs by at least 5%, plus to lower travel costs. Consider then a 5% reduction in prices, which accounting for partial take-up, is a 3.25% reduction.

Power using just a single round of data:

`sampsi 15 14.475, sd(6) power(0.8)`

`sampclus, obsclus(20) rho(0.1)`

This gives a minimum of 595 blocks needed, and 5,948 firms in each treatment group.

Power using the baseline to improve power:

The residual variance is now $\sqrt{1-0.82} \times 6 = 3.6$. We assume the intra-cluster correlation in prices is reduced to 0.02. Then we have:

sampsi 15 14.475, sd(3.6) power(0.8)

sampclus, obsclus(20) rho(0.02)

This gives us a minimum of 102 blocks needed, with 1020 firms in each treatment group. So we will still not quite have enough power with 60 blocks and 600 firms in each treatment group. There are two approaches we will use to further boost power. The first is to screen firms at baseline to reduce the residual standard deviation. For example, if we consider firms only with prices between 10 and 20 pesos per unit, we could reduce the standard deviation from 6 to 4. This would then reduce the minimum number of blocks to 46, with a minimum sample size of 455 treatment and 455 control.

sampsi 15 14.475, sd(2.4) power(0.8)

sampclus, obsclus(20) rho(0.02)

Moreover, given that total costs are likely to reduce by more than 5%, the effect size we are hoping for should be larger than what is done for power here. To the extent that prices are not as highly autocorrelated as we expect here, we can then further boost power by collecting multiple rounds of post-treatment data and pooling the effects, as in McKenzie (2012).

Bottom line

The bottom line from these calculations is that it appears that with an appropriate combination of screening to reduce heterogeneity, use of the baseline data to improve power, and potentially multiple rounds of follow-ups, we will have sufficient power with the anticipated sample sizes and randomization at the block level. We will of course update these calculations once baseline data become available, and as noted here, have several margins to potentially adjust the design in order to ensure we have enough power once the study gets underway.

9. DATA COLLECTION (E,R)

(1 page if basic, 1-2 pages if include all sections for registration and ethical clearance)

- Describe main instruments for data collection

The main instruments for data collection will be surveys taken of the firms, coupled with administrative data from Agruppa on sales made. In terms of surveys, we plan the following:

- A detailed baseline survey which covers identifying information of the owner, business activities including the importance of the products Agruppa sells, business characteristics (opening hours, travel time, employees, accounting, degree of competition faced), purchases and sales of the Agruppa products (including place of purchase, costs, quantities, etc.), income and expenses of the firm, personal information on the owner (education, age, household composition, household income, and work/life balance), and finally interest in Agruppa

- High-frequency phone/SMS surveys to collect information on prices and costs of the Agruppa products, and means of travel. These will be 5 minute surveys conducted bi-weekly for the first three months after the product is launched to them.
- Detailed follow-up surveys: at 6 months and 12 months after the baseline – these will revisit the baseline questions and ask more detailed questions to provide information on the mechanisms behind any effects.

9.1 QUANTITATIVE INSTRUMENTS

- Describe how primary and secondary outcomes (from section 7) will be measured, their timing and frequency.

The take-up rate, and the cost and quantity of purchase from Agruppa can be measured using the administrative data being collected by Agruppa. The remaining primary and secondary outcomes will be collected through surveys, as discussed above. For example, the section below of our draft baseline questionnaire asks for each product, in terms of standardized units, the quantity purchased and price paid, and then the price and quantity they were selling the items for.

H	Consider...	A) Did you buy ... when you went to Corabastos on (date of question 3 / 3. 1)?	B) How much quantity of ... did you buy when you went to Corabastos on (date of question 3 / 3. 1)?	C) What price did you pay for...?	OPERATOR: DEPENDING ON HOW THEY ANSWER FILL IN E-F FOR PRICE PER WEIGHT AND G-I FOR PRICE PER BAG OR MOUNTAIN!	D) At what prices were you selling... yesterday?	E) How much quantity of ... did you sell yesterday?	F) At what prices did you sell each bag or mountain yesterday?	G) How many bags or mountains of ... did you sell yesterday?	H) OPERATOR PLEASE WEIGH BAG OR MOUNTAIN.
H	onion - "cabezona"	yes no => go to next product	arroba half arroba quarter arroba bulto in total doesn't know	per arroba per half arroba per quarter arroba per bulto in total doesn't know		per kg per pound doesn't know	kg pounds doesn't know			grams per bag/mountain
H	potato "pastusa-gruesa"	yes no => go to next product	per bulto per half bulto in total doesn't know	per bulto per half bulto in total doesn't know		per kg per pound doesn't know	kg pounds doesn't know			grams per bag/mountain
H	potato - "pastusa- pareja"	yes no => go to next product	per bulto per half bulto in total doesn't know	per bulto per half bulto in total doesn't know		per kg per pound doesn't know	kg pounds doesn't know			grams per bag/mountain
H	tomate "chonto"	yes no => go to next product	per canastilla doesn't know	per canastilla doesn't know		per kg per pound doesn't know	kg pounds doesn't know			grams per bag/mountain

This will give us unit costs, prices, and quantities for each product. We can then look at impacts on a per-product basis, and then aggregate them together to get a weighted price index and total quantity, where baseline levels of the products sold will be used as the weights.

As noted before, we will measure prices and unit costs at a bi-weekly frequency over the first three months post-treatment, as well as 6 months and 12 months post-treatment.

9.2 MANAGEMENT OF DATA QUALITY

- Describe methods used to enhance the quality of measurements (e.g., multiple observations, training of surveyors), electronic data collection, protocols for quality assurance.

Data collection will be carried out by Innovations for Poverty Action, Colombia, using a team of trained enumerators. Data will be collected using handheld tablets, which will have consistency checks pre-programmed in them to enhance data quality. Prices and Costs which differ substantially from those charged by other firms will be flagged

and re-checked. Standard survey team management methods in terms of the use of random spotchecks and supervisor visits will be employed.

9.3 ETHICAL ISSUES

- Describe if this IE will require ethical approval, informed consent procedures, and important ethical considerations related to data collection.

In terms of ethical clearance, the PI has done the NIH and CITI human subjects training programs. IPA is in the process of determining whether the study will need to be cleared through its IRB. The surveys will use standard informed consent procedures for collecting data on small firms.

The one ethical consideration to note is that Agruppa feels that ethically, once it offers its service to some firms in a small block, it should be available to all firms in this block to prevent concerns about unfair competition. For this reason we are not randomizing within blocks. However, they are ok with firms who say they are not interested not being offered the service at first, which will provide us with a means of measuring spillovers by comparing uninterested firms in treated blocks to uninterested firms in control blocks.

9.4 QUALITATIVE INSTRUMENTS

- Provide a description of all qualitative instruments (if applicable).

The key qualitative information to be collected will be i) information from pilot interviews and discussions with Agruppa clients to understand the demand for the product and refine the baseline survey; ii) a planned qualitative scoping study of the full supply chain to understand who the other parties in the supply chain are, and what their characteristics and reactions to the Agruppa service are.

9.5 IE IMPLEMENTATION MONITORING SYSTEM ^(R)

- Describe the IE implementation monitoring system, particularly, what specific indicators and system will be used to follow up the studied population, their treatment participation, treatment *actually* delivered and received based on activities, and outputs (see the theory of change section).

The key information here will come from Agruppa's monitoring records (cross-checked by IPA). This will provide information on which firms actually sign up for the service, how much they pay for it, whether they continue to use the service, etc. We will also work with Agruppa to obtain their costs of providing the service, in order to be able to provide cost-benefit calculations.

10. DATA PROCESSING AND ANALYSIS

(~ 1-2 pages)

10.1 DATA CODING, ENTRY, AND EDITING³ ^(E)

³ This subsection is optional

- Describe planned methods for data entry, and for handling missing data, imputations.

Electronic data collection with pre-specified consistency checks will be used. No imputation for missing data will be done, except as a robustness check. Instead we will check for selective attrition and item non-response, and if this differs with treatment, employ bounding approaches to examine the robustness of our findings to these issues.

10.2 MODEL SPECIFICATION FOR QUANTITATIVE DATA ANALYSIS

- Describe the statistical method(s) that will be used to compare groups for primary and secondary outcomes (the specific equation should be included), any transformations to quantitative data. Specify whether the standard errors will be clustered or corrected.
- Specify what IE parameter of interest will be estimated (e.g., ITT, TT, MTE, LATE).
- Describe how you plan to address multiple hypothesis testing.
- Describe methods for additional analyses, including spillovers and subgroup analyses.
- Provide a list of any variables to be collected to check balance and correct for potential selection due to attrition, non-response, take-up rate issues (all theoretically important variables to be measured at baseline, including, those thought to be related to participation/dropout/non-response and the outcomes of interest).
- Lay out a strategy to follow up, test and correct for (if required) sources of bias (e.g., non-random attrition, non-response, endogenous take-up).
- State if you plan to register this IE (see selected links below)
 - AEA RCT Registry (<https://www.socialscienceregistry.org/>)
 - 3ie Registry (<http://www.3ieimpact.org/evaluation/ridie/>)

We plan on running the following ANCOVA specification as our main specification for firm i in block b in randomization strata s :

$$Outcome_{i,b,t} = \alpha + \beta Treat_b + \gamma Interested_{i,b} + \delta Treat_b Interested_{i,b} + \mu Outcome_{i,b,0} + \sum_{s'=1}^S \theta_s 1(s' = s) + \varepsilon_{i,b,t}$$

Where γ gives the difference in outcomes between interested and uninterested firms in the control blocks, β gives the indirect impact of treatment on the uninterested firms, and $\beta + \delta$ provides the overall impact of being treated on interested firms in the treatment block. We control for the baseline value of the outcome, and for randomization strata dummies. Since randomization is at the block level, standard errors will be clustered at the block level. However, recent work (e.g. Young, 2016; ongoing work by Abadie et al.) has suggested alternatives to clustering the standard errors such as randomization inference and bootstrapping. Since this literature is still emerging, we plan to wait and see what the recommendations of this new literature are, but at the least, to explore sensitivity to using randomization inference or bootstrapping for the standard errors.

This equation will estimate the intent-to-treat (ITT) parameters - the impact of being offered Agruppa's service for the interested firms. If take-up is less than 100 percent, this will differ from the treatment effect on the treated. We will then instrument actual use with assignment to treatment to get the LATE. Since none of the control group should be able to use the service, the LATE effect will equal the treatment effect on the treated (TOT).

As noted above, spillovers will be examined within blocks by comparing the uninterested firms in treated blocks to the uninterested firms in control blocks. For robustness we will also use the GPS coordinates of each firm together with the random assignment to examine the impact of the percent of firms in a 1 km radius around each firm that are treated.

Multiple hypothesis testing will be addressed in four ways. The first is to specify a set of primary outcomes. The second is to aggregate together the different prices and unit costs to form a single measure of prices and single measure of costs. Third, we can use the omnibus test of Young (2016) to test for overall statistical significance. Fourth, we can form averages of standardized z-scores to combine different indicators together.

The exact set of variables used to check for baseline balance will be refined once the baseline data starts getting collected and we refine the questions. However, it will include key owner characteristics (age, education), firm characteristics (number of employees, age of firm) and baseline values of the key outcomes (travel time, prices, unit costs, profits and sales). We will use a joint test of orthogonality to test balance.

As discussed previously, our approach to dealing with attrition and item non-response will be to test for imbalance in these by treatment group, and to use bounding approaches to examine robustness to this.

We plan on registering this study in the AEA randomized trial registry once the baseline data collection is underway and we can see whether we need to make any changes in what is planned here.

11. STUDY LIMITATIONS AND RISKS (E)

(1/2 page)

- Provide an assessment of risk and threat to internal validity (related to previous section)
- Discuss issues related to external validity, particularly (i) representativeness of the sample; (ii) representativeness of the institution(s) delivering the intervention, and (iii) feasibility that the intervention can be scaled up.

This study uses a randomized experiment to evaluate the impact of the Agruppa service. Internal validity is a strength of such designs, although the following risks are possible: i) non-random attrition (discussed above); ii) low take-up limiting power (preliminary pilot evidence from Agruppa suggests demand for the product so we believe this risk to be reasonably low); and iii) spillovers between treatment and control. The concern with the latter is whether treated firms take sales away from control firms. The use of randomization at the block level rather than individual level is intended to mitigate this concern, and, as discussed, we can also use the GPS coordinates to assess sensitivity to how close the control firms are to treated firms at the border.

In terms of external validity, the initial sample will be representative of vendors in poor neighborhoods of Bogota, and the screened sample representative of the vast majority of such firms (excepting outliers). The technology being developed is one that is designed to scale, and Agruppa has also done pilots in Kenya, so believes it can use the results from Colombia to potentially scale to other countries as well as within Colombia.

12. IE MANAGEMENT (E,R)

(All tables)

12.1 EVALUATION TEAM AND MAIN COUNTERPARTS

- Provide list of all IE team members with their position, affiliation, and responsibilities (including lead researcher, other research team members, and all project staff involved in the IE work, and main implementing agency counterparts).

Table 2. IE Team and Main Counterparts

Name	Role	Organization/Unit
David McKenzie, Lead Economist	Principal investigator/Lead Researcher	World Bank DEC
Leonardo Iacovone, Senior Economist	Co-PI	World Bank, T&C
Sebastian Chaskel, Pablo Villar, and Maria Rodriguez Uribe	Field Supervision, Field Coordination, and Leading the Survey Team	Innovations for Poverty Action, Colombia
Verena Liedgens and Carolina Medina	Main implementing/policy counterparts	Agruppa

12.2 WORK PLAN AND DELIVERABLES

Table 3. Milestones, Deliverables, and Estimated Timeline

Milestones	Deliverables	Completion Date
Peer-reviewed Concept Note	Methodology note	April 4, 2016
Data collection plan and pilot	TORs Questionnaires	April 2016,
Data collection (Baseline)	Cleaned data Dictionaries	Baseline is rolling from April-August 2016, cleaned data October 2016
First data analysis	Presentation Data file Do files Baseline report	November 2016
Implementation of intervention aligned to evaluation	Rollout plan Monitoring reports verifying treatment and control status	April 2016 onwards
Follow-up data collection plan	TORs Questionnaire	May 2016, October 2016, April 2017

Data collection (Follow-up)	Cleaned data Dictionaries	September 2017
Final report and policy notes	Technical note Policy note Data file Do files	March 2018
Dissemination of findings	Presentations	March 2018 onwards

12.3 BUDGET

(1 paragraph)

- Present total budget and disaggregated by staff time, data collection, and travel. Include all sources of funding, both Bank-executed and client-executed (BB resources, trust fund and grants, FBS, EFO, project financing for the IE, such as data collection, and other client financing). Estimate and include all research/staff time (not only the time charged).

Table 4. Total Budget per Category

Category	USD	%
Staff	83,500	18.6
STC	84,500	18.9
Data Collection	246,000	54.9
Travel	33,800	7.5
Total	447,800	100

The main sources of funding are the DIME i2i preparation grant (\$25,000), three years of DIME i2i implementation grant funding (\$50,000 x 3), and the Jobs Multi-Donor trust funding (\$250,000). The remaining costs are covered by staff time allocations to the PIs. To the extent that more funding is obtained, it will be used to expand the sample size in either the cross-sectional or the time dimensions.

- Attach detailed budget (see excel file template). (Attached)

13. PLAN FOR USING DATA AND EVIDENCE FROM THE STUDY

(1 paragraph)

- Describe communication, participation, and dissemination strategy (potential users of findings, media channels) at all stages of the IE (design, baseline analysis, mid-corrections, follow-up analysis, and final results).

The baseline data will be used for internal purposes only within the team, to fine-tune our follow-up survey questionnaires and to provide Agruppa with more information on their possible clients. Take-up data will be used to assess whether our assumptions of take-up rates and retention are realistic, and as a basis for working quickly with

Agruppa to consider alternative marketing and retention strategies if necessary. The main form of dissemination will then be based around the full analysis, which will be written up as a research paper. This will be disseminated in the working paper series, presented in different conference and seminar settings, and submitted to an academic journal. In addition, dissemination will take place through a blog post on the Development Impact blog, and through a 2-page Finance & PSD Impact note. The results will be discussed with Mincit in Colombia through the World Bank country team. Anonymized data will be made available through the World Bank's Open data library and will be accessible also from the Lead Researcher's webpage.

REFERENCES

Arráiz, I., Henríquez, F., & Stucchi, R. (2013). Supplier development programs and firm performance: evidence from Chile. *Small Business Economics*, 41(1), 277-293.

Food and Agriculture Organization (FAO), (2010). "Análisis socioeconómico de tiendas de alimentos en áreas urbanas de bajos recursos en Latinoamérica: Informe del caso de estudio de Bogotá." Bogotá: FAO

McKenzie, David (2012) "Beyond Baseline and Follow-up: The Case for More T in Experiments", *Journal of Development Economics* 99(2): 210-21.

OECD and World Bank (2015) *Inclusive Global Value Chains: Policy options in trade and complementary areas for GVC Integration by small and medium enterprises and low-income developing countries*. Report prepared for G20 Trade Ministers Meeting. <http://www.oecd.org/trade/OECD-WBG-g20-gvc-report-2015.pdf>

Piza, Caio, Tulio Antonio Cravo, Linnet Taylor, Lauro Gonzalez, Isabel Musse, Isabela Furtado, Ana C Sierra, Samer Abdelnour (2016) "The Impacts of Business Support Services for Small and Medium Enterprises on Firm Performance in Low-and Middle-Income Countries: A Systematic Review", *Campbell Systematic Reviews* 12(1).

Ruel, Marie T. et al., (2005) "Why Is Child Malnutrition Lower in Urban Than in Rural Areas? Evidence from 36 Developing Countries", *World Development*, 33(8): 1285-1305.

Ruel, Marie, James L. Garrett, Saul S. Morris, Daniel Maxwell, Arne Oshaug, Patrice Engle, Purnima Menon, Alison Slack, and Lawrence Haddad (1998) "Urban Challenges to Food and Nutrition Security: A review of food security, health, and caregiving in the cities", IFPRI Food Consumption and Nutrition Division discussion paper no. 51

Wu, David (2004) "Supply Chain Intermediation", pp. 67-115 in David Simchi-Levi, S. David Wu, and Z. Max Shen (Eds.) *Handbook of Quantitative Supply Chain Analysis: Modeling in the e-Business Era* Kluwer Academic Publishers

Young, Alwyn (2016) "Channelling Fisher", Mimeo. LSE.