

# **The Colombia 2010 Enterprise Surveys Data Set**

## **I. Introduction**

1. This document provides additional information on the data collected in Colombia between June 2010 and October 2010 as part of the Latin America and Caribbean (LAC) Enterprise Survey 2010, an initiative of the World Bank.

The Enterprise Surveys, through interviews with firms in the manufacturing and services sectors, capture business perceptions on the biggest obstacles to enterprise growth, the relative importance of various constraints to increasing employment and productivity, and the effects of a country's business environment on its international competitiveness. They are used to create statistically significant business environment indicators that are comparable across countries. The Enterprise Surveys are also used to build a panel of enterprise data that will make it possible to track changes in the business environment over time and allow, for example, impact assessments of reforms.

The report outlines and describes the sampling methodology, the sample structure as well as additional information that may be useful when using the data, such as information on non-response cases and the appropriate use of the weights.

## **II. Sampling Structure**

2. The sample for Colombia was selected using stratified random sampling, following the methodology explained in the *Sampling Note*<sup>1</sup>. Stratified random sampling<sup>2</sup> was preferred over simple random sampling for several reasons<sup>3</sup>:

a. To obtain unbiased estimates for different subdivisions of the population with some known level of precision.

b. To obtain unbiased estimates for the whole population. The whole population, or universe of the study, is the non-agricultural economy. It comprises: all manufacturing sectors according to the group classification of ISIC Revision 3.1: (group D), construction sector (group F), services sector (groups G and H), and transport, storage, and communications sector (group I). Note that this definition excludes the following sectors: financial intermediation (group J), real estate and renting activities (group K, except sub-sector 72, IT, which was added to the population under study), and all public or utilities-sectors.

c. To make sure that the final total sample includes establishments from all different sectors and that it is not concentrated in one or two of industries/sizes/locations.

d. To exploit the benefits of stratified sampling where population estimates, in most cases, will be more precise than using a simple random sampling method (i.e., lower standard errors, other things being equal.)

---

<sup>1</sup> The complete text can be found at [http://www.enterprisesurveys.org/documents/Implementation\\_note.pdf](http://www.enterprisesurveys.org/documents/Implementation_note.pdf)

<sup>2</sup> A stratified random sample is one obtained by separating the population elements into non-overlapping groups, called strata, and then selecting a simple random sample from each stratum. (Richard L. Scheaffer; Mendenhall, W.; Lyman, R., "Elementary Survey Sampling", Fifth Edition).

<sup>3</sup> Cochran, W., 1977, pp. 89; Lohr, Sharon, 1999, pp. 95

e. Stratification may produce a smaller bound on the error of estimation than would be produced by a simple random sample of the same size. This result is particularly true if measurements within strata are homogeneous.

f. The cost per observation in the survey may be reduced by stratification of the population elements into convenient groupings.

3. Three levels of stratification were used in this country: industry, establishment size, and location. The original sample design with specific information of the industries and locations chosen is described in Appendix E.

4. Industry stratification was designed in the way that follows: the universe was stratified into 5 manufacturing industries, 1 service industry -retail -, and 1 residual sector as defined in the sampling manual. The four identified manufacturing sectors each had targets of 160 interviews, with other manufacturing having a target of 120 interviews. Both retail and other services had targets of 120 interviews each.

5. Size stratification was defined following the standardized definition for the Enterprise Surveys: small (5 to 19 employees), medium (20 to 99 employees), and large (more than 99 employees). For stratification purposes, the number of employees was defined on the basis of reported permanent full-time workers. This seems to be an appropriate definition of the labor force since seasonal/casual/part-time employment is not a common practice, except in the sectors of construction and agriculture.

6. Regional stratification was defined in four locations (city and the surrounding business area): Bogota, Cali, Medellin, and Barranquilla.

### **III. Sampling implementation**

7. Given the stratified design, sample frames containing a complete and updated list of establishments as well as information on all stratification variables (number of employees, industry, and region) are required to draw the sample. Great efforts were made to obtain the best source for these listings. However, the quality of the sample frames was not optimal and, therefore, some adjustments were needed to correct for the presence of ineligible units. These adjustments are reflected in the weights computation (*see below*).

8. TNS Opinion was hired to implement the LAC 2010 enterprise surveys roll out. In Colombia the local subcontractor was Centro Nacional de Consultoría S. A.

9. For Colombia, two sample frames were used. The first was supplied by the World Bank and consists of enterprises interviewed in Colombia in 2006. The World Bank required that attempts should be made to re-interview establishments responding to the Colombia 2006 survey where they were within the selected geographical regions and met eligibility criteria. That sample is referred to as the Panel. The second sample frame was produced by Centro Nacional de Consultoría S. A. using the Confecámaras and Chamber of Commerce Bogotá, 2008. Each database contained the following information:

- Coverage;
- Up to datedness;
- Availability of detailed stratification variables ;
- Location identifiers- address, phone number, email;
- Electronic format availability;
- Contact name(s).

Counts from sample frames are shown below.

## Panel sample counts

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services	Grand Total
Bogota	Small	22	56	72	0	4	29	65	248
	Medium	24	49	22	0	4	16	31	146
	Large	1	4	4	0	2	7	9	27
Total		47	109	98	0	10	52	105	421
Cali	Small	22	27	8	0	4	9	25	95
	Medium	10	18	3	2	2	11	11	57
	Large	0	1	1	0	0	0	5	7
Total		32	46	12	2	6	20	41	159
Medellin	Small	22	30	24	0	14	10	19	119
	Medium	16	41	13	0	32	11	16	129
	Large	14	28	3	0	0	7	11	63
Total		52	99	40	0	46	28	46	311
Barranquilla	Small	18	12	6	0	0	11	16	63
	Medium	5	9	3	0	0	10	9	36
	Large	0	0	0	0	0	1	8	9
Total		23	21	9	0	0	22	33	108
Grand Total		154	275	159	2	62	122	225	999

## Sample Frames

Source: Confecámaras and Chamber of Commerce Bogotá, 2008

Location	Size	15	17-18	24-25	27-29	Manuf.	Retail	Services	Grand Total
Barranquilla	Small	21	27	24	35	85	247	782	1221
	Medium	13	18	25	11	24	45	230	366
	Large	18	4	15	4	18	16	57	132
<b>Barranquilla Total</b>		<b>52</b>	<b>49</b>	<b>64</b>	<b>50</b>	<b>127</b>	<b>308</b>	<b>1069</b>	<b>1719</b>
Bogota	Small	787	1151	1255	1347	3703	6238	21596	36077
	Medium	253	352	475	410	1119	1567	5856	10032
	Large	104	116	150	107	319	430	1599	2825
<b>Bogota Total</b>		<b>1144</b>	<b>1619</b>	<b>1880</b>	<b>1864</b>	<b>5141</b>	<b>8235</b>	<b>29051</b>	<b>48934</b>
Cali	Small	67	75	69	88	164	572	1448	2483
	Medium	21	26	25	21	53	118	289	553
	Large	16	14	12	1	22	14	56	135
<b>Cali Total</b>		<b>104</b>	<b>115</b>	<b>106</b>	<b>110</b>	<b>239</b>	<b>704</b>	<b>1793</b>	<b>3171</b>
Medellin	Small	76	271	79	113	292	786	1748	3365
	Medium	27	89	38	40	103	202	492	991
	Large	9	35	17	13	22	37	117	250
<b>Medellin Total</b>		<b>112</b>	<b>395</b>	<b>134</b>	<b>166</b>	<b>417</b>	<b>1025</b>	<b>2357</b>	<b>4606</b>
<b>Grand Total</b>		<b>1412</b>	<b>2178</b>	<b>2184</b>	<b>2190</b>	<b>5924</b>	<b>10272</b>	<b>34270</b>	<b>58430</b>

10. The two sample frames were then used for the selection of a sample with the aim of obtaining interviews at 1000 establishments with five or more employees.

11. The quality of the frame was assessed at the onset of the project through visits to a random subset of firms and local contractor knowledge. The sample frame was not immune from the typical problems found in establishment surveys: positive rates of non-eligibility, repetition, non-existent units, etc. In addition, the sample frame contains no telephone/fax numbers so the local contractor had to screen the contacts by visiting them. Due to response rate and ineligibility issues, additional sample had to be extracted by the World Bank in order to obtain enough eligible contacts and meet the sample targets.

12. Given the impact that non-eligible units included in the sample universe may have on the results, adjustments may be needed when computing the appropriate weights for individual observations. The percentage of confirmed non-eligible units as a proportion of the total number of sampled establishments contacted for the survey was 13.20% (448 out of 3395 establishments)<sup>4</sup>.

<sup>4</sup> Based on out of target contacts and impossible to contact establishments

#### **IV. Data Base Structure:**

13. The structure of the data base reflects the fact that 3 different versions of the questionnaire were used. The basic questionnaire, the Core Module, includes all common questions asked to all establishments from all sectors (manufacturing, retail, and other services). The second expanded variation, the Manufacturing Questionnaire, is built upon the Core Module and adds some specific questions relevant to the sector. The third expanded variation, the Services Questionnaire, is also built upon the Core Module and adds to the core specific questions relevant to either retail or IT. Each variation of the questionnaire is identified by the index variable, *a0*.

14. All variables are named using, first, the letter of each section and, second, the number of the variable within the section, i.e. *a1* denotes section A, question 1. Variable names preceded by a prefix “LAC” indicate questions specific to LAC, therefore, they may not be found in the implementation of the rollout in other countries. All other suffixed variables are global and are present in all country surveys over the world. All variables are numeric with the exception of those variables with an “x” at the end of their names. The suffix “x” denotes that the variable is alpha-numeric.

15. There are 2 establishment identifiers, *idstd* and *id*. The first is a global unique identifier. The second is a country unique identifier. The variables *a2* (sampling region), *a6a* (sampling establishment’s size), and *a4a* (sampling sector) contain the establishment’s classification into the strata chosen for each country using information from the sample frame. The strata were defined according to the guidelines described above.

16. There are three levels of stratification: industry, size and location. Different combinations of these variables generate the strata cells for each industry/location/size combination. A distinction should be made between the variable *a4a* and *d1a2* (industry expressed as ISIC rev. 3.1 code). The former gives the establishment’s classification into one of the chosen industry-strata, whereas the latter gives the actual establishment’s industry classification (four digit code) in the sample frame.

17. All of the following variables contain information from the sampling frame. They may not coincide with the reality of individual establishments as sample frames may contain inaccurate information. The variables containing the sample frame information are included in the data set for researchers who may want to further investigate statistical features of the survey and the effect of the survey design on their results.

- a2* is the variable describing sampling regions

- a6a*: coded using the same standard for small, medium, and large establishments as defined above. The code -9 was used to indicate units for which size was undetermined in the sample frame.

- a4a*: coded using ISIC codes for the chosen industries for stratification. These codes include most manufacturing industries (15 to 37), other manufacturing (2), retail (52), and (45, 50, 51, 55, 60, 63, 72) for other Services.

18. The surveys were implemented following a 2 stage procedure. Typically first a screener questionnaire is applied over the phone to determine eligibility and to make appointments. Then a face-to-face interview takes place with the Manager/Owner/Director of each establishment. However, the phone numbers were unavailable in the sample frame, and thus the enumerators applied the screeners in person. The variables *a4b* and *a6b* contain the industry and size of the establishment from the screener questionnaire. Variables *a8* to *a11* contain additional information and were also collected in the screening phase.

19. Note that there are additional variables for location size by population (*a3*) and firm size by no. of workers (*l1*, *l6* and *l8*) that reflect more accurately the reality of each establishment. Advanced users are advised to use these variables for analytical purposes.

20. Variables *l1*, *l6* and *l8* were designed to obtain a more accurate measure of employment accounting for permanent and temporary employment. Special efforts were made to make sure that this information was not missing for most establishments.

21. Variables *a17x* gives interviewer comments, including problems that occurred during an interview and extraordinary circumstances which could influence results. Please note that sometimes this variable is removed due to privacy issues.

## **V. Universe Estimates**

22. Universe estimates for the number of establishments in each cell in Colombia were produced for the strict, weak and median eligibility definitions. The estimates were the multiple of the relative eligible proportions.

23. Appendix B shows the overall estimates of the numbers of establishments in Colombia based on the sample frame.

24. For some establishments where contact was not successfully completed during the screening process (because the firm has moved and it is not possible to locate the new location, for example), it is not possible to directly determine eligibility. Thus, different assumptions about the eligibility of establishments result in different adjustments to the universe cells and thus different sampling weights.

25. Three sets of assumptions on establishment eligibility are used to construct sample adjustments using the status code information.

26. Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. The resulting weights, which include adjustments applied to panel firms (see below), are included in the variable *w\_strict\_panadj*.

Strict eligibility = (Sum of the firms with codes 1,2,3,4,&16) / Total

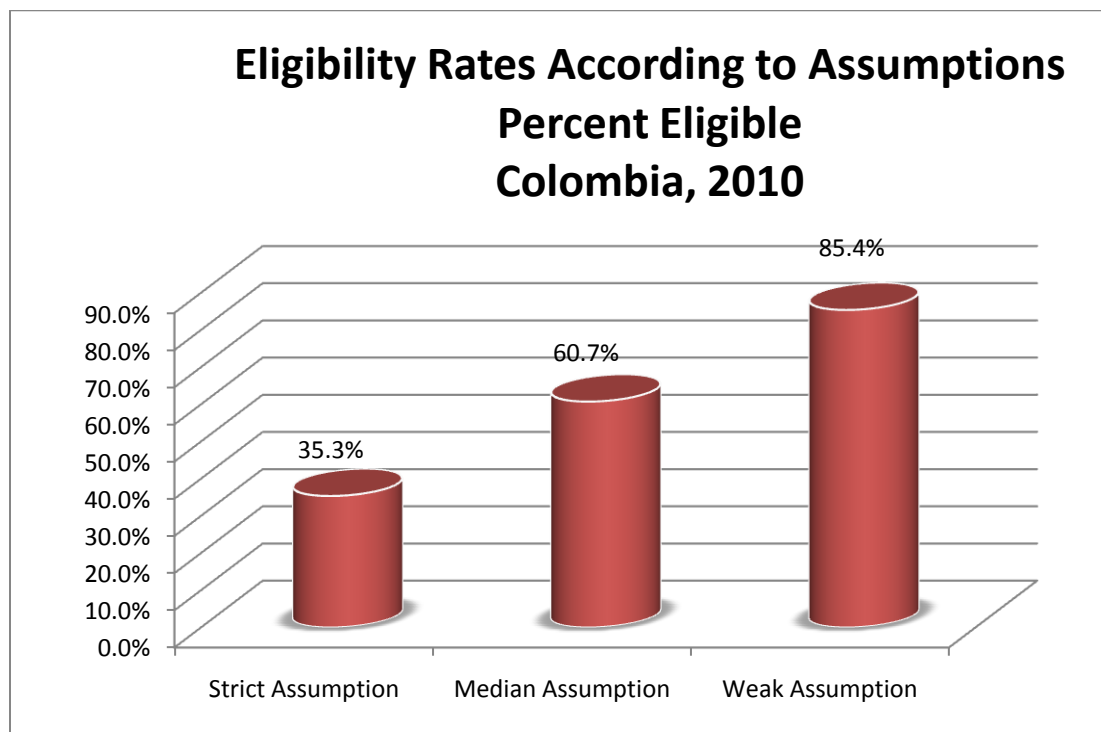
27. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *w\_median\_panadj*.

Median eligibility = (Sum of the firms with codes 1,2,3,4,16,10,11, & 13) / Total

28. Weak assumption: in addition to the establishments included in points a and b, all establishments for which it was not possible to contact or that refused the screening questionnaire are assumed eligible. This definition includes as eligible establishments with dead or out of service phone lines, establishments that never answered the phone, and establishments with incorrect addresses for which it was impossible to find a new address. Under the weak assumption only observed non-eligible units are excluded from universe projections. The resulting weights are included in the variable *w\_weak\_panadj*.

Weak eligibility= (Sum of the firms with codes 1,2,3,4,16,91,92,93,10,11,12,&13) / Total

29. The indicators computed for the Enterprise Survey website use the median weights. The following graph shows the different eligibility rates calculated for firms in the sample frame under each set of assumptions.



30. Universe estimates for the number of establishments in each industry-region-size cell in Colombia were produced for the strict, weak and median eligibility definitions.

Appendix D shows the universe estimates of the numbers of registered establishments that fit the criteria of the Enterprise Surveys.

31. Once an accurate estimate of the universe cell projection was made, weights for the probability of selection were computed using the number of completed interviews for each cell.

## **VI. Weights**

32. Since the sampling design was stratified and employed differential sampling, individual observations should be properly weighted when making inferences about the population. Under stratified random sampling, unweighted estimates are biased unless sample sizes are proportional to the size of each stratum. With stratification the probability of selection of each unit is, in general, not the same. Consequently, individual observations must be weighted by the inverse of their probability of selection (probability weights or  $pw$  in Stata.)<sup>5</sup>

33. Special care was given to the correct computation of the weights. It was imperative to accurately adjust the totals within each region/industry/size stratum to account for the presence of ineligible units (the firm discontinued businesses or was unattainable, education or government establishments, establishments with less than 5 employees, no reply after having called in different days of the week and in different business hours, no tone in the phone line, answering machine, fax line<sup>6</sup>, wrong address or moved away and could not get the new references) The information required for the adjustment was collected in the first stage of the implementation: the screening process. Using this information, each stratum cell of the universe was scaled down by the observed proportion of ineligible units within the cell. Once an accurate estimate of the universe cell (projections) was available, weights were computed using the number of completed interviews.

The selection of panel firms required additional adjustments to account for varying probabilities of selection between fresh and panel sample universes. For additional information on this methodology, please refer to Enterprise Survey documentation of weighting methodology.

34. Appendix C shows the cell weights for registered establishments in Colombia.

## **VII. Appropriate use of the weights**

35. Under stratified random sampling weights should be used when making inferences about the population. Any estimate or indicator that aims at describing some feature of the population should take into account that individual observations may not represent equal shares of the population.

---

<sup>5</sup> This is equivalent to the weighted average of the estimates for each stratum, with weights equal to the population shares of each stratum.

<sup>6</sup> For the surveys that implemented a screener over the phone.



36. However, there is some discussion as to the use of weights in regressions (see Deaton, 1997, pp.67; Lohr, 1999, chapter 11, Cochran, 1953, pp.150). There is not strong large sample econometric argument in favor of using weighted estimation for a common population coefficient if the underlying model varies per stratum (stratum-specific coefficient): both simple OLS and weighted OLS are inconsistent under regular conditions. However, weighted OLS has the advantage of providing an estimate that is independent of the sample design. This latter point may be quite relevant for the Enterprise Surveys as in most cases the objective is not only to obtain model-unbiased estimates but also design-unbiased estimates (see also Cochran, 1977, pp 200 who favors the used of weighted OLS for a common population coefficient.)<sup>7</sup>

37. From a more general approach, if the regressions are descriptive of the population then weights should be used. The estimated model can be thought of as the relationship that would be expected if the whole population were observed.<sup>8</sup> If the models are developed as structural relationships or behavioral models that may vary for different parts of the population, then, there is no reason to use weights.

#### **VIII. Non-response**

38. Survey non-response must be differentiated from item non-response. The former refers to refusals to participate in the survey altogether whereas the latter refers to the refusals to answer some specific questions. Enterprise Surveys suffer from both problems and different strategies were used to address these issues.

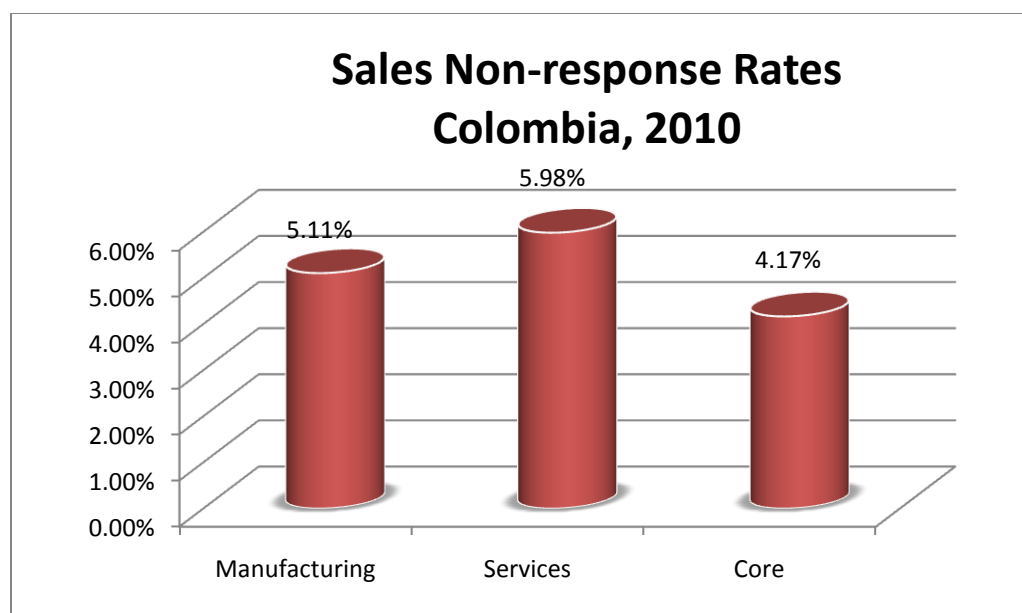
39. Item non-response was addressed by two strategies:

- a- For sensitive questions that may generate negative reactions from the respondent, such as corruption or tax evasion, enumerators were instructed to collect the refusal to respond as a different option from don't know (-7).
- b- Establishments with incomplete information were re-contacted in order to complete this information, whenever necessary. However, there were clear cases of low response. The following graph shows non-response rates for the sales variable, *d2*, by sector. Please, note that the coding utilized in this dataset does not allow us to differentiate between "Don't know" and "refuse to answer", thus the non-response in the chart below reflects both categories (DKs and NAs).

---

<sup>7</sup> Note that weighted OLS in Stata using the command regress with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands svy will provide appropriate standard errors.

<sup>8</sup> The use weights in most model-assisted estimations using survey data is strongly recommended by the statisticians specialized on survey methodology of the JPSM of the University of Michigan and the University of Maryland.

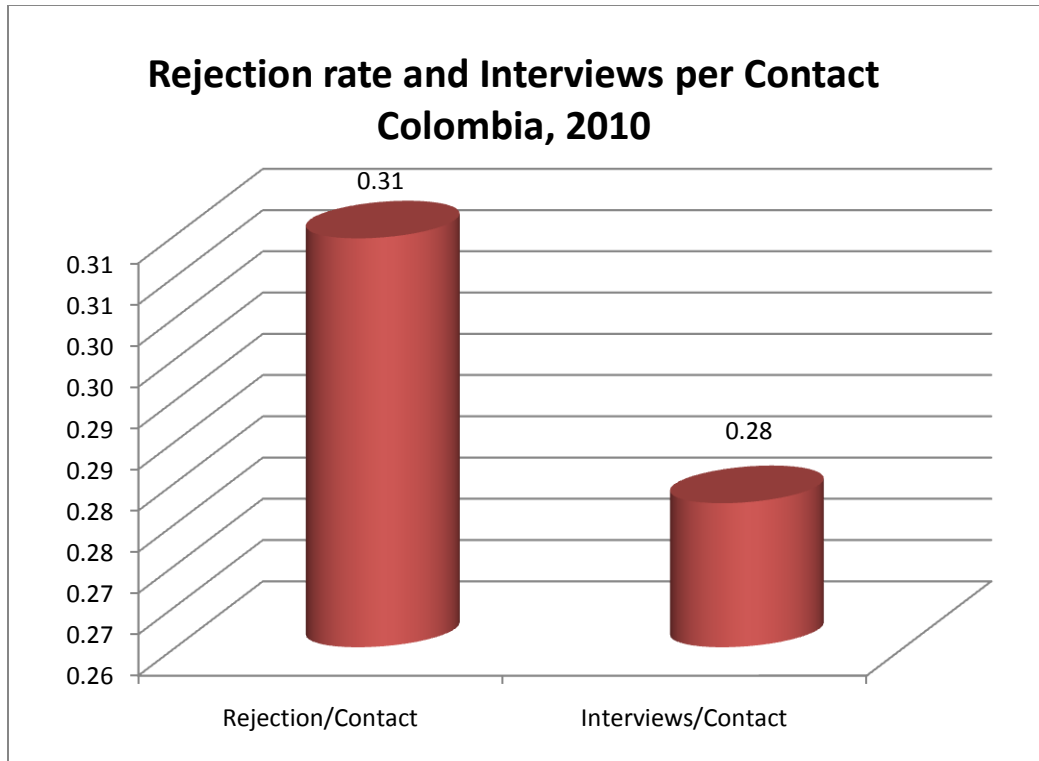


40. Survey non-response was addressed by maximizing efforts to contact establishments that were initially selected for interview. Attempts were made to contact the establishment for interview at different times/days of the week before a replacement establishment (with similar strata characteristics) was suggested for interview. Survey non-response did occur but substitutions were made in order to potentially achieve strata-specific goals. Further research is needed on survey non-response in the Enterprise Surveys regarding potential introduction of bias.

41. As the following graph shows, the number of realized interviews per contacted establishment was 0.28<sup>9</sup>. This number is the result of two factors: explicit refusals to participate in the survey, as reflected by the rate of rejection (which includes rejections of the screener and the main survey) and the quality of the sample frame, as represented by the presence of ineligible units. The number of rejections per contact was 0.31.

---

<sup>9</sup> The estimate is based on the total no. of firms contacted including ineligible establishments.



42. Details on the rejection rate, eligibility rate, and item non-response are available at the level strata. This report summarizes these numbers to alert researchers of these issues when using the data and when making inferences. Item non-response, selection bias, and faulty sampling frames are not unique to Colombia. All enterprise surveys suffer from these shortcomings, but in very few cases they have been made explicit.

**References:**

Cochran, William G., Sampling Techniques, 1977.

Deaton, Angus, The Analysis of Household Surveys, 1998.

Levy, Paul S. and Stanley Lemeshow, Sampling of Populations: Methods and Applications, 1999.

Lohr, Sharon L. Sampling: Design and Techniques, 1999.

Scheaffer, Richard L.; Mendenhall, W.; Lyman, R., Elementary Survey Sampling, Fifth Edition, 1996.

## Appendix A

### Status Codes Fresh:

	ELIGIBLES	
Eligible	1. Eligible establishment (Correct name and address)	786
Eligible	2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	2
Eligible	3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	3
Eligible	4. Eligible establishment (Moved and traced)	47
		0
Ineligible	5. The establishment has less than 5 permanent full time employees	88
Ineligible	6 The firm discontinued businesses	159
Ineligible	7. Not a business: Private household	51
Ineligible	8. Ineligible activity: Education, Agriculture, Finances, Government, etc.	3
Ineligible	151 Out of target - outside the covered regions	49
Ineligible	152. Out of target - moved abroad	3
Unobtainable	91. No reply after having called in different days of the week and in different business hours	178
Unobtainable	92. Line out of order	342
Unobtainable	93. No tone	15
Unobtainable	10. Answering machine	32
Unobtainable	11. Fax line- data line	20
Unobtainable	12. Wrong address/ moved away and could not get the new references	142
	13. Refuses to answer the screener	639
	14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	48
	Total	2607

### Response Outcomes Fresh:

Target	
Complete interviews	636
Incomplete interviews	0
Elegible in process	0
Refusals	202
Out of target	353
Impossible to contact	729
Refusal to the Screener	639
Total	2559

### Status Codes Panel:

	ELIGIBLES	
Eligible	1. Eligible establishment (Correct name and address)	304
Eligible	2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	5
Eligible	3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	6
Eligible	4. Eligible establishment (Moved and traced)	47
Eligible	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	0
Ineligible	616 The firm discontinued businesses - (Establishment went bankrupt)	30
Ineligible	618 The firm discontinued businesses - (Original establishment disappeared and is now a different firm)	23
Ineligible	619 The firm discontinued businesses - (Establishment was bought out by another firm)	5
Ineligible	620 The firm discontinued businesses - (It was impossible to determine for what reason)	11
Ineligible	621 The firm discontinued businesses - (Other: SPECIFY in COMMENTS)	1
Ineligible	7. Not a business: Private household	16
Ineligible	8. Ineligible activity: Education, Agriculture, Finances, Government, etc.	0
Ineligible	151 Out of target - outside the covered regions	9
Ineligible	152. Out of target - moved abroad	0
Unobtainable	91. No reply after having called in different days of the week and in different business hours	38
Unobtainable	92. Line out of order	77
Unobtainable	93. No tone	6
Unobtainable	10. Answering machine	6
Unobtainable	11. Fax line- data line	2
Unobtainable	12. Wrong address/ moved away and could not get the new references	39
	13. Refuses to answer the screener	163
	14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	0
	Total	788

### Response Outcomes Panel:

Panel	
Complete interviews	306
Incomplete interviews	0
Elegible in process	9
Refusals	47
Out of target	95
Impossible to contact	168
Refusal to the Screener	163
Total	788

## Appendix B

### Universe Estimate, Colombia:

Region name	Size	15	17-18	24-25	27-29	Manuf.	Retail	Services	Grand Total
Barranquilla	Small	32	34	27	35	88	253	798	1267
	Medium	13	19	26	11	24	46	235	374
	Large	18	4	15	4	18	16	57	132
<b>Barranquilla Total</b>		<b>63</b>	<b>57</b>	<b>68</b>	<b>50</b>	<b>130</b>	<b>315</b>	<b>1090</b>	<b>1773</b>
Bogota	Small	810	1195	1308	1348	3711	6254	21652	36278
	Medium	256	368	490	411	1124	1575	5880	10104
	Large	104	122	153	107	319	433	1607	2845
<b>Bogota Total</b>		<b>1170</b>	<b>1685</b>	<b>1951</b>	<b>1866</b>	<b>5154</b>	<b>8262</b>	<b>29139</b>	<b>49227</b>
Cali	Small	71	86	75	90	168	580	1464	2534
	Medium	23	27	26	21	54	121	294	566
	Large	17	14	12	1	22	14	56	136
<b>Cali Total</b>		<b>111</b>	<b>127</b>	<b>113</b>	<b>112</b>	<b>244</b>	<b>715</b>	<b>1814</b>	<b>3236</b>
Medellin	Small	87	301	94	113	296	794	1765	3450
	Medium	37	133	45	40	107	210	505	1077
	Large	19	56	20	13	24	42	121	295
<b>Medellin Total</b>		<b>143</b>	<b>490</b>	<b>159</b>	<b>166</b>	<b>427</b>	<b>1046</b>	<b>2391</b>	<b>4822</b>
<b>Grand Total</b>		<b>1487</b>	<b>2359</b>	<b>2291</b>	<b>2194</b>	<b>5955</b>	<b>10338</b>	<b>34434</b>	<b>59058</b>

## Appendix C

### Strict Cell Weights Colombia:

#### Panel

Average Strict Cell Weights, Panel firms

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services
Bogota	Small	1.00	1.00	1.00		1.00	1.12	1.43
	Medium	1.00	1.36	1.00		1.59	1.57	1.40
	Large	1.00	1.17	1.00		1.05	1.00	1.81
Cali	Small	1.00	1.28	1.00			1.00	1.00
	Medium	1.00	1.00		1.42	1.00	1.00	
	Large							1.35
Medellin	Small	1.00	1.00	1.00			1.00	1.34
	Medium	1.00	1.01	1.00		5.81	1.00	1.00
	Large	1.00	1.01	1.00			1.00	1.00
Barranquilla	Small	1.00	1.00	1.00				2.08
	Medium	1.10	1.20	1.00			1.05	1.00
	Large							1.90

#### Fresh Firms

Strict Cell Weights\*

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services
Bogota	Small	16.14	26.26	35.37	12.77	68.46	119.75	697.82
	Medium	2.62	5.68	4.96	3.35	30.60	18.79	217.70
	Large	1.31	1.09	1.23	1.45	6.09	3.52	44.04
Total								
Cali	Small	8.25	22.41	31.84	3.15		119.75	697.82
	Medium	4.46	2.23	2.65	2.81		39.39	61.57
	Large	1.32	1.40	1.23	1.00	3.44	1.09	13.89
Total								
Medellin	Small	25.51		11.04	3.68	68.46	119.75	697.82
	Medium	5.21	9.26	2.29	1.77	18.86	15.32	190.50
	Large	1.00	4.24	1.06	1.00	3.13	2.18	17.58
Total								
Barranquilla	Small	4.67	1.62	13.33	1.31	68.46	119.75	697.82
	Medium	1.33	1.24	3.99	2.36	11.64	9.04	58.97
	Large	1.34	1.00	3.71	3.00	3.39	3.74	17.01

\*Collapsed cells used in certain cases

## Weak Cell Weights Colombia:

### Panel Firms

Average Weak Cell Weights, Panel firms

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services
Bogota	Small	2.03	1.00	2.96		1.00	2.97	4.93
	Medium	2.18	2.70	1.96		2.44	2.57	3.24
	Large	1.00	1.51	1.22		1.01	1.11	2.26
Total								
Cali	Small	1.00	2.12	1.56			1.00	1.18
	Medium	2.05	1.00		8.24	2.43	1.00	
	Large							1.41
Total								
Medellin	Small	2.60	2.16	1.83			1.29	2.31
	Medium	1.43	1.41	2.33		10.93	1.54	1.12
	Large	2.10	2.24	1.00			1.00	3.60
Total								
Barranquilla	Small	1.67	1.00	1.50				4.95
	Medium	1.10	1.19	1.00			1.51	1.58
	Large							3.71

### Fresh Firms

Weak Cell Weights\*

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services
Bogota	Small	61.85	120.94	107.59	39.95	223.86	450.87	2061.85
	Medium	8.87	23.07	13.31	9.24	107.74	76.17	692.54
	Large	3.95	3.95	2.95	3.58	19.17	12.75	125.18
Total								
Cali	Small	18.22	59.49	55.82	5.68		450.87	2061.85
	Medium	8.68	5.23	4.10	4.47		92.02	112.88
	Large	2.29	2.93	1.71	1.00	6.24	2.27	22.75
Total								
Medellin	Small	68.98		23.70	8.12	223.86	450.87	2061.85
	Medium	12.42	26.53	4.33	3.44	46.85	43.81	427.58
	Large	1.72	10.85	1.79	1.60	6.94	5.57	35.26
Total								
Barranquilla	Small	7.52	3.14	17.05	1.72	223.86	450.87	2061.85
	Medium	1.89	2.12	4.50	2.74	17.24	15.41	78.90
	Large	1.70	1.47	3.75	3.11	4.48	5.70	20.34

\*Collapsed cells used in certain cases



## Median Cell Weights Colombia:

### Panel Firms

Average Median Cell Weights, Panel firms

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services
Bogota	Small	1.51	1.00	1.49		1.00	1.72	2.36
	Medium	1.19	1.89	1.04		2.43	2.10	2.05
	Large	1.00	1.17	1.00		1.00	1.00	2.01
Total								
Cali	Small	1.00	1.65	1.56			1.00	1.00
	Medium	1.74	1.00		3.35	2.36	1.00	
	Large							1.42
Total								
Medellin	Small	1.23	1.26	1.00			1.00	1.69
	Medium	1.00	1.05	1.58		8.62	1.27	1.00
	Large	1.68	1.53	1.00			1.00	3.02
Total								
Barranquilla	Small	1.17	1.00	1.50				3.50
	Medium	1.20	1.16	1.00			1.41	1.25
	Large							2.71

### Fresh Firms

Median Cell Weights\*

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services
Bogota	Small	34.55	63.77	65.13	24.85	133.56	250.92	1206.05
	Medium	5.59	13.72	9.09	6.49	67.42	44.46	424.90
	Large	2.75	2.60	2.22	2.78	13.24	8.21	84.74
Total								
Cali	Small	14.22	43.82	47.20	4.94		250.92	1206.05
	Medium	7.64	4.34	3.91	4.38		75.03	96.74
	Large	2.23	2.68	1.71	1.00	6.02	2.04	21.51
Total								
Medellin	Small	43.72		16.28	5.73	133.56	250.92	1206.05
	Medium	8.88	17.91	3.36	2.74	33.27	29.02	297.68
	Large	1.36	8.08	1.53	1.41	5.44	4.07	27.08
Total								
Barranquilla	Small	5.73	2.25	14.08	1.46	133.56	250.92	1206.05
	Medium	1.62	1.72	4.19	2.62	14.71	12.27	66.01
	Large	1.61	1.31	3.75	3.11	4.22	5.01	18.78

\*Collapsed cells used in certain cases

## Appendix D

### Strict Universe Estimates

#### Colombia

Strict Universe Estimates (Fresh + Panel)

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services	Grand Total
Bogota	Small	191.77	235.84	1066.55	386.48	617.10	505.61	3507.76	6511.11
	Medium	71.33	100.72	160.32	128.75	278.60	314.75	1546.94	2601.40
	Large	32.38	30.75	58.55	37.80	93.51	102.53	495.27	850.79
Total		295.48	367.31	1285.42	553.03	989.21	922.89	5549.96	9963.30
Cali	Small	19.50	31.40	33.84	27.69	0.00	242.49	1401.80	1756.71
	Medium	12.57	12.93	18.22	21.27	2.00	42.39	127.38	236.75
	Large	15.50	5.60	7.41	4.15	10.32	4.35	29.12	76.46
Total		47.56	49.92	59.47	53.11	12.32	289.23	1558.30	2069.92
Medellin	Small	29.51	8.00	38.13	44.14	342.28	725.47	705.86	1893.39
	Medium	16.41	33.62	23.37	23.13	111.99	67.27	193.50	469.29
	Large	11.00	34.13	11.48	8.00	9.38	17.07	55.74	146.79
Total		56.92	75.75	72.98	75.27	463.64	809.81	955.10	2509.48
Barranquilla	Small	15.66	12.71	15.33	13.08	68.46	249.53	701.21	1075.97
	Medium	10.43	9.76	13.96	9.95	11.64	4.19	119.93	179.87
	Large	6.69	4.48	15.13	5.36	10.16	16.52	37.83	96.17
Total		32.78	26.95	44.42	28.39	90.26	270.24	858.97	1352.01
Grand Total		432.75	519.93	1462.29	709.80	1555.43	2292.17	8922.33	15894.70

## Weak Universe Estimates

Weak Universe Estimates (Fresh + Panel)

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services	Grand Total
Bogota	Small	716.38	1013.55	3182.06	1207.69	2015.76	1900.48	10373.30	20409.22
	Medium	227.89	341.69	422.58	354.80	974.57	1241.87	4905.90	8469.29
	Large	95.79	103.39	138.02	93.20	289.59	361.34	1390.57	2471.90
Total		1040.06	1458.63	3742.66	1655.69	3279.92	3503.68	16669.77	31350.41
Cali	Small	39.44	74.32	58.94	48.72	0.00	904.75	4132.94	5259.11
	Medium	26.66	24.90	33.81	51.70	4.85	95.02	233.72	470.67
	Large	28.83	11.71	10.23	6.68	18.72	9.09	46.92	132.18
Total		94.93	110.93	102.99	107.10	23.57	1008.85	4413.58	5861.95
Medellin	Small	79.39	17.28	80.26	97.43	1119.31	2714.29	2075.69	6183.65
	Medium	33.40	74.24	45.76	45.98	328.50	184.47	430.94	1143.29
	Large	21.23	85.95	17.34	12.80	20.82	37.39	116.58	312.12
Total		134.02	177.47	143.36	156.21	1468.63	2936.15	2623.21	7639.06
Barranquilla	Small	25.29	21.81	20.05	17.21	223.86	918.16	2068.52	3294.89
	Medium	13.75	13.24	15.50	12.36	17.24	6.05	160.96	239.11
	Large	8.50	7.17	15.74	5.85	13.45	26.81	48.10	125.62
Total		47.54	42.22	51.29	35.42	254.55	951.02	2277.58	3659.62
Grand Total		1316.55	1789.26	4040.30	1954.42	5026.67	8399.70	25984.14	48511.04

## Median Universe Estimates

Median Universe Estimates (Fresh + Panel)

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services	Grand Total
Bogota	Small	404.85	545.52	1879.73	752.02	1203.00	1060.16	6060.91	11906.19
	Medium	141.64	210.08	285.98	249.29	611.68	730.36	3011.17	5240.20
	Large	66.92	68.39	104.16	72.17	200.55	233.85	944.13	1690.17
Total		613.41	823.98	2269.87	1073.49	2015.23	2024.37	10016.21	18836.56
Cali	Small	31.43	55.36	50.32	43.33	0.00	504.83	2420.03	3105.31
	Medium	21.66	21.37	28.65	36.76	4.72	78.03	200.46	391.64
	Large	26.43	10.74	10.23	5.94	18.05	8.17	44.45	124.01
Total		79.52	87.47	89.20	86.02	22.76	591.03	2664.94	3620.95
Medellin	Small	48.64	10.04	53.84	68.77	667.78	1512.49	1216.17	3577.72
	Medium	23.76	51.56	34.71	35.89	208.72	123.71	300.68	779.03
	Large	16.89	60.51	15.26	11.25	16.31	28.40	90.31	238.93
Total		89.29	122.11	103.80	115.91	892.81	1664.60	1607.16	4595.68
Barranquilla	Small	18.95	16.53	17.08	14.60	133.56	515.10	1211.01	1926.81
	Medium	12.53	11.51	14.57	11.08	14.71	5.63	134.52	204.55
	Large	8.06	6.07	15.43	5.73	12.66	22.28	42.98	113.20
Total		39.54	34.11	47.08	31.41	160.93	543.00	1388.51	2244.56
Grand Total		821.76	1067.67	2509.96	1306.83	3091.73	4823.00	15676.82	29297.76

## Appendix E

### Original Sample Design, Colombia:

#### SAMPLE DESIGN

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services	Grand Total
Bogota	Small	22	28	44	24	24	24	24	190
	Medium	36	35	35	37	24	24	24	215
	Large	32	35	30	34	24	24	24	203
Total		90	98	109	95	72	72	72	608
Cali	Small	6	3	6	1	4	4	4	28
	Medium	8	6	5	6	4	4	4	37
	Large	7	3	3	10	4	4	4	35
Total		21	12	14	17	12	12	12	100
Medellin	Small	9	16	8	5	8	8	8	62
	Medium	10	10	11	12	8	8	8	67
	Large	10	8	8	13	8	8	8	63
Total		29	34	27	30	24	24	24	192
Barranquilla	Small	6	4	3	3	4	5	4	29
	Medium	7	7	5	5	4	4	4	36
	Large	7	5	2	10	4	3	4	35
Total		20	16	10	18	12	12	12	100
Grand Total		160	160	160	160	120	120	120	1000

### Completed Interviews, Colombia:

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services	Grand Total
Bogota	Small	23	34	38	45	21	37	24	222
	Medium	33	34	32	30	15	16	24	184
	Large	23	22	38	21	9	13	20	146
Total		79	90	108	96	45	66	68	552
Cali	Small	5	7	3	7	0	3	6	31
	Medium	5	9	4	9	1	6	4	38
	Large	8	4	7	2	5	3	3	32
Total		18	20	14	18	6	12	13	101
Medellin	Small	3	7	8	12	3	11	13	57
	Medium	10	16	11	9	4	7	6	63
	Large	9	17	13	10	5	8	9	71
Total		22	40	32	31	12	26	28	191
Barranquilla	Small	4	4	2	5	0	7	4	26
	Medium	10	10	5	10	3	2	6	46
	Large	6	6	4	3	3	3	1	26
Total		20	20	11	18	6	12	11	98
Grand Total		139	170	165	163	69	116	120	942

## Appendix F

### Local Agency team involved in the study:

Local Agency	Centro Nacional de Consultoría S. A.
--------------	--------------------------------------

### Sample Frame:

Characteristic of sample frame used:	Information obtained from Confecámaras ( Cali, Medellín y Barranquilla) and Chamber of Commerce Bogotá (for Bogotá). This database was acquired in 2006 and updated in 2007 and 2008.
Source:	Confecámaras and Chamber of Commerce Bogotá
Year:	2008
Comments on the quality of sample frame:	The database has excellent quality, but as far as number of employees only 60% of the companies report this data, therefore other parameters are used to estimate this, such as sales.
Year and organism who conducted the last economic census	The last economic census in Colombia was done by Departamento Administrativo Nacional de Estadística (DANE) in 1990.
Other sources for companies statistics	Associations such as ANDI; and Logistics Catalogues such Publicar.

### Sectors included in the Sample:

Original Sectors	<b>ECONOMIC SECTORS INCLUDED IN THE STUDY: LAC 2010 ENTERPRISE SURVEY</b> <ul style="list-style-type: none"> <li>• 15</li> <li>• 18</li> <li>• 24</li> <li>• 28</li> <li>• Other Manufacturing</li> <li>• 52</li> </ul> Other Services
Added Sectors	<ul style="list-style-type: none"> <li>• 15</li> <li>• 17-18</li> <li>• 24-25</li> <li>• 27-29</li> <li>• Other Manufacturing</li> <li>• 52</li> <li>• Other Services</li> </ul>

### Sample:

Comments/ problems on sectors and regions selected in the sample:	It is easier to get interviews with the manufacturing sector than in the Service sectors.
Comments on the response rate:	The rate of response was low, during the time of FW, there were ups and downs between 5 and 30 interviews for an average of 14 per week.
Comments on the sample design:	<p>The makeup of the sampling frame was achieved in the following way.</p> <ol style="list-style-type: none"> <li>1. Companies were selected from the CIU following lists provided by TNS Gallup</li> <li>2. When possible, companies were classified by size.</li> <li>3. When information about number of employees was not available, sales and actives were used to classify the companies,</li> <li>4. Companies with no information for size or CIU were not included fueron incluidas.</li> <li>5. The final database was the sampling for Bogotá, Cali, Medellín and Barranquilla. The clasification variables were: City, size and CIU.</li> </ol>

**Fieldwork:**

Date of Fieldwork	<ul style="list-style-type: none"> <li>- Beginning of FW May 18th 2010.</li> <li>- End of FW: March 31st 2011.</li> </ul>
Problems found during fieldwork:	<ul style="list-style-type: none"> <li>- Sometimes upon getting to the place where the interview was to take place, the interviewee refused to answer</li> <li>- Negative to answer Section N for confidentiality reasons.</li> <li>- Little compliance with sending the information via e-mail after the interview.</li> </ul>

**Questionnaires:**

Problems for the understanding of questions (write question number)	<ul style="list-style-type: none"> <li>-Q C11 No clarity about the term own or shared generators.</li> <li>- Q J30 same as above</li> </ul>
Comments on questionnaires length:	Some said that the length of the questionnaire is extreme for busy people such as them
Suggestions or other comments on the questionnaire:	<ul style="list-style-type: none"> <li>-Questions about sales and spending were not answered for reasons of confidentiality</li> <li>.</li> </ul>

**Country situation**

General aspects of economic, political or social situation of the country that could affect the results of the survey:	<p>Social aspects:</p> <ul style="list-style-type: none"> <li>- Mid year vacation (july –Agosto 2010)</li> <li>- End of year closing of companies 2010</li> <li>- different events in some regions, local fairs, carnivals, etc</li> </ul> <p>Economic aspects:</p> <ul style="list-style-type: none"> <li>- End of fiscal year fiscal 2010</li> <li>- Truckers strike in February 2011</li> </ul>
Relevant country events occurred during fieldwork:	Presidential elections 2010 -2014