A DISCRETE-EVENT SIMULATION MODEL TO ESTIMATE THE NUMBER OF PARTICIPANTS IN THE CICLOVIA PROGRAMS

Melisa Murcia
Raha Akhavan-Tabatabaei
Olga L. Sarmiento
Carlos F. Valencia
Universidad de los Andes
Bogotá D.C, Colombia

ABSTRACT

Open-Street Programs, also known as Ciclovia, are free multi-sectorial programs for people from different socio-economic backgrounds where public spaces and streets are closed to motorized traffic and open for leisure activities. Performance indicators of Ciclovia programs allow for the measurement and subsequent analysis of the programs' impact on public health. The number of participants is one of the key performance indicators, which makes its reliable estimation crucial to measuring the cost-effectiveness of the programs. Furthermore, a unified and flexible estimation methodology is required to make it applicable to every Ciclovia program, in order to be able to do comparisons between programs, taking into account their accelerated growth in the last three years.

As previous work, we had proposed a discrete-event simulation model to estimate the number of participants in such programs. We apply our approach to a case study in the city of Bogota (Colombia), with the largest program in the world. This paper constitutes an upgrade of this previous proposed model. In particular we focus on three aspects: upgrade the quality of the input data, propose a methodology for its collection and relax of some the model's main assumptions. This allowed the upgrade of the quality of the estimation of the daily number of participants and the METS, a relevant public health measure of the program.

1 INTRODUCTION

Ciclovia is defined as the temporary closure of streets to motorized traffic, creating a safe and pleasant space for walking, jogging, skating, running and cycling, among others (Sarmiento et. al 2011). This program was first created in Bogota on December 15th, 1974 and it has become popular in different countries, especially in the Americas, since then. For instance, in a two-year period (2010-212), almost 300 new programs were created (Diaz et. al 2013). This is due to two main reasons: the fact that the citizens recognize it as a space designed for their benefit and the flexibility of the program which is resource-efficient and easily adapted to every city (Montes et. al 2011). This way, citizens with no distinction of strata are given the opportunity of doing physical activity and the city infrastructure and resources are used efficiently, generating ample and secure spaces for the participants (Montes et. al 2011).

The Ciclovia of Bogota is open on Sundays and holidays from 7:00 a.m. to 2:00 p.m. It is currently the largest Ciclovia program in the world with an extension of 113.6 km and a geographical distribution across the city which enables the participation of people from different socio-economic backgrounds. Currently it is under the administration of the public entity *Instituto Distrital de Recreación y Deporte (IDRD).*

Given the potential impact of this program, it is required to have a good estimation of the number of participants in order to measure the health benefits on the population. Furthermore, due to the increasing number of Ciclovia programs, the estimation must be flexible so that it can be used by any Ciclovia, allowing comparisons between them. In a previous work (Murcia, Rivera et. al 2013), we developed a flexible methodology that allows us to accurately estimate the number of participants, applicable to different Ciclovia programs. In order to reach this goal, we built a counting model based on a discrete-event simulation. The purpose of this paper is to build upon our previous work, by addressing how to collect better quality input data, relaxing the main assumptions and thus upgrading the quality of the estimation of the model. Furthermore, we also adapt the previous methodology for the city of Bogota and other large Ciclovias in which more precision might be needed to better imitate the behavior of the system.

The rest of this document is organized as follows. In Section 2 we explain our previous work. In Section 3 we explain the motivation to continue working on this specific topic. Section 4 tackles the most challenging problem which consists of accurately estimating the arrival rate to the system. In Section 5 we tackle the relaxation of some assumptions of our previous work. In Section 6 we integrate the improved input data gathered by the proposed methodology for data collection and the relaxed assumptions to Bogota's Ciclovia. We obtain a better quality estimation of the number of participants as well as of the energetic consumption of the program and we point out the future work.

2 PREVIOUS WORK

As mentioned in the introductory section "A discrete-event simulation model to estimate the number of participants in the Ciclovia programs" (Murcia, Rivera et. al 2013) presents the general proposed methodology for the estimation of the number of participants. In this section, we are going to describe the findings of this paper, which constitute a starting point for the work presented here.

The paper first presents an extensive literature review on methodologies to tackle counting problems in several fields. This literature review section concluded that given the actual conditions of the Ciclovia, the existent technologies for counting and the non-existent historical information available for new programs, the proposed method to use is discrete-events simulation. Simulation allows imitating the behavior of the real system, understanding how the different parts interact with each other. The modeled system can be visualized through animation enabling the understating of the different actors who take an active role in this type of programs (e.g. medical community, local authorities). Additionally, it is possible to gather with good quality the input data required for the discrete-event simulation model at any Ciclovia program and built upon this tool for further key performance indicators of the system.

The proposed method was divided into two phases: data collection for the inputs and the construction of the discrete-event simulation model. The simulation model requires three types of input information: the percentage of the participants in each category (i.e., cyclist, walker, etc.) and their velocities, the distribution of the sojourn time in Ciclovia and the number of participants that enter the Ciclovia from each point of entry at every hour. For the case study of Bogota, we had three reliable sources of information available: previous work carried out by Universidad Nacional (Universidad Nacional 2005), a health and physical activity survey conducted in 2011 (Sarmiento et al. 2011), and our own fieldwork in 2013.

From the first study (Universidad Nacional 2005) we use their proposed velocities of participants as input data. Additionally, their estimate of 1'400,274 participants per month with an average participation of twice a month per participant, or an average of 700,000 participants on a given day of Ciclovia, is used as validation. In 2011 an extensive survey on health and healthy habits was conducted in Bogota's Ciclovia (Sarmiento et al. 2011). The key questions of this survey used in our model as input data and findings are: the sojourn time in Ciclovia, modelled like an empirical distribution, the arrival time to Ciclovia and the departure point in Ciclovia. At last, we estimated the arrival rate of participants to the Ciclovia per meter at a specific hour at a specific track through fieldwork. The arrival rate we obtained was 0.113 arrivals per meter in 15 minutes; we will be referring to this as the *reference arrival rate*. This name is due to the fact that fieldwork was carried out in key a time and place. For instance, fieldwork was done at 11 a.m. which IDRD identified as the rush-hour of the system and we worked on one average Ciclovia's street (Carrera 15), that could be taken as representative of the Ciclovia according to Ciclovia's director for IDRD.

For the structure of the simulation model we identified three required components: entities, tracks and intersections. Each entity when created is assigned an activity category and a sojourn time in Ciclovia. The tracks are the set of streets that make up the Ciclovia. They are connected by a set of intersections, and for modeling purposes we assume that the arrival of participants occurs only at intersections and that this has no impact on the number of entities who join the system. Every time entities arrive at an intersection their time left at Ciclovia is evaluated. According to that process, they leave the system or are routed to another track. In the case of Bogota, given that we had no data to estimate the proportion of people that choose one route or the other at an intersection, we assume that all the connecting tracks had the same probability of being chosen. With these elements in mind, Ciclovia of Bogota was divided into nineteen bidirectional tracks and sixteen intersections, as shown in Figure 1.

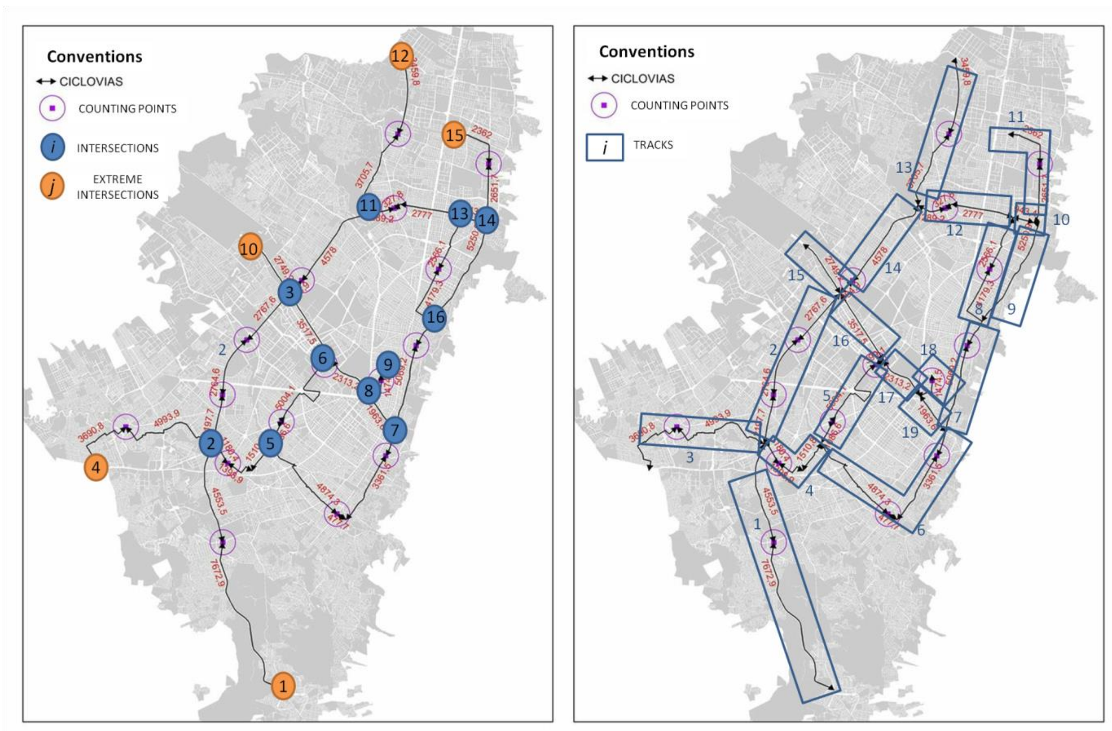

Figure 1: Defined intersections and tracks in the simulation model.

As previously mentioned, the rate at which entities are created in the source of the intersection is the key input to this model. However, based on results of the survey and fieldwork observations, we deduced this arrival rate is not constant in the different Ciclovia tracks throughout the day. Because of this, we needed to adjust our *reference arrival rate* for every hour of the Ciclovia as well as in every track. This was done with the information of the survey (Sarmiento et al. 2011) by obtaining the relative weights of arrivals at every hour compared to the number of arrivals at which fieldwork took place, expressed as a percentage. The rate at the different hours of the day was calculated by adjusting the *reference arrival rate* and its relative weight. Finally, this was multiplied by four, assuming that arrivals are homogenous during the hour (i.e., between 8:00 and 9:00 am) to estimate the hourly rate.

To weight the arrival rates per track, we used the survey (Sarmiento et al. 2011) data on the point of entry in order to calculate the relative weights of the other tracks compared to our reference track. This way, we obtained a relative percentage of participants that join the Ciclovia at a given track. By multiplying the weight of each track, by the *reference arrival rate* per meter and by the total length of the given track, we had the arrival rate per track, assuming that the track has a homogeneous arrival rate along its entire length. To be able to find an estimation of the arrival rate in each intersection, 50% of the arrival of the track was assigned to each of the two adjacent intersections.

The model was verified following the recommendations for verification techniques in (Law and Kelton 2000). The simulation model was built using Simio Simulation Software, Academic Edition Version 5.83 and it estimates a total of 675,500 participants joining the Ciclovia on one given day. As observed, there was no confidence interval in the estimation due to the deterministic nature of the input data. For the purpose of validation, we used the study from Universidad Nacional carried out in 2005 (Universidad Nacional 2005) due to the fact that it was the only available data. Although this value is very close to our estimation, we are conscious of the limitations of this comparison due to the time lag between the two studies. Since the run conditions involved the use of a non-homogeneous Poisson Process with a deterministic changing rate per hour, we performed only five replications to estimate the number of participants. The final output was found to be sensitive to the arrival rate to the system, taking into account that its summation over the operating hours of Ciclovia gives the total number of participants.

3 MOTIVATION

From the previous work presented, we proposed a methodology from which we could estimate the number of participants on a day of Ciclovia given the required input data. However, we found ourselves limited in overall data availability due to the lack of up to date information for Bogota's Ciclovia. We believe this estimation can be upgraded with better input data. In this paper we propose a methodology for input data collection, in particular for the arrival rate, which we had defined as the main input to the model. We worked on a new approach to collect this input by estimating the arrival rate from a methodology for direct observation tested in field and an auxiliary model that estimates the arrival rate based on matching the flow of people with the rush-hour flow. For the city of Bogota, we believe the auxiliary model using IDRD's database gives better quality data; however, a new program or a program without historical data could benefit from the direct-observation methodology.

Additionally, because this model was built with the only goal of calculating the number of participants, we had strong assumptions that did not allow the calculation of other important performance measures. This is why, in this paper we explore how to relax the assumptions of the model to better imitate the actual system and tackle a broader scope. On the other hand, these assumptions could indeed be a closer representation of the behaviors of small Ciclovia users rather than in the case of large Ciclovias. As an example we can evaluate the routing; in around 70% of the Ciclovias, their extension of smaller than 10 km and routes tend to have a linear circuit (EpiAndes and U. Rosario, 2013). Routing in small programs is not a problem; however we still have 30% of the Ciclovias to account for. We seek to explore these assumptions to evaluate which ones can be kept and which ones must be relaxed.

The main objective of this research is to upgrade the quality of the inputs and relax main assumptions in the simulation model to improve the outputs. These outputs are an updated estimation of the number of participants who use the system daily and a measure of energetic expenditure (METS). As secondary objectives we want to:

1) Improve the collection of the input-data arrival rate. By proposing a direct observation methodology for the arrival rate, we adjust it through tests in fieldwork and suggest an auxiliary model that based on less volatile observations, such as the flow of people, yields the arrival rate.
2) Identify main variables that influence users' behavior:
    a. Routing (e.g. parks, extremities, etc.),
    b. Sojourn time (e.g. activity, main reason of usage, etc.)
    c. Arrival rate (e.g. holidays, weather, etc.)
3) Upgrade the model by integrating the relaxation of assumptions, the improvement of the input data and the new desired outputs.

4 ARRIVAL RATE

The main input of this model is the arrival rate of the people to the system, taking into account that its summation over the operating hours of Ciclovia yields the total number of participants. Therefore, the objective of this section is to estimate the arrival rate of participants to Ciclovia programs as a key input to estimate the number of participants using the discrete-event simulation model presented on Section 3. We start by introducing the proposed methodologies, followed by input data collection required to test the proposed methodologies. Then, we show the results and conclusions. Finally, we evaluate the impact of factors such as holidays, special events and weather on the arrival rate, which, as we have explained previously, will impact the number of people attending the Ciclovia program.

4.1 Proposed methodologies for estimating the arrival rate

In this section we present three suggestions for estimating the arrival rate. One relies on direct observation and the other two make an estimation of the arrival rate through an auxiliary method based on discrete-event simulation and the flow of people. First, we propose a methodology for direct observation, which can be particularly relevant for new programs or programs that do not have a database of historical data. However, this direct observation can be very volatile and reducing the variance would require a lot of fieldwork which would be costly. To overcome the volatility implied by direct observation, we seek to find the arrival rate needed so that the flow of people observed is matched, through an auxiliary discrete-event simulation model. For this, two approaches are suggested: one based on the time at which people join Ciclovia and the 15-minute flow observed at 11:00 a.m. which is historical data available from IDRD's database and another based on the 15-minute flow collected at the beginning of each hour, which constitutes the new data from fieldwork (Section 4.2.1). However, we discourage using the second model, unless the trajectory is mostly linear or different routes taken by the participants are known.

4.1.1 Proposed methodology for direct observations

This proposed methodology consists of the direct observation of the people joining a specific Ciclovia track. In general, it consists of having one observer at each side of the street counting the people who join the system over a given distance in a given period of time. Having two observers is important so that both sides can be counted by a different observer and thus the quality of the observation is not compromised. Next, we describe each of the components:

1) Distance:

   The aim is to cover the largest distance possible as more information could be obtained from large rather than short distances. Additionally, due to the fact that the number of participants changes throughout the day, the selected distance either could be changed or kept constant, taking into account for the second scenario that it must support capability of the observer even at the rush hour. We recommend having only one distance and keeping it constant, so that this selected distance is not an additional source of noise for the observation.

2) Time:

   Since technologies for automated counts do not apply (Murcia, Rivera et. al 2013), the only way of collecting information on people joining the system are manual counts as which makes this method highly dependent on the human observer. Ideally, we would like to have the longest possible time, to have more information on the arrivals of people. However, since we rely on persons, the longer time they do this task, the more they will get tired, which would make us lose accuracy on the data. To reduce this error, counting intervals must be kept short. A fairly good time to use is 15 minutes. This time choice is in our experience from Fieldwork (Section 4.2.1), the longest range that allows observers to remain sharp and record good observations.

3) Location:
   A key aspect in the selection of this point is the choice of the spot where the counting will take place. In each track the following aspects should be considered:
   - Choose an average point: A point that can be representative of the selected track as a whole. This means that spots such as malls or churches should be avoided because they might influence people in terms of leaving or joining the system.
   - Include an intersection: Since an important proportion of the participants are cyclists and they can only join the system in connecting streets, we must make sure the distance considered includes a perpendicular street.
   - In Figure 2 we show an example of a good counting point. In certain cases, even a public spot like a university that would normally influence the influx of people becomes an regular counting point because it gets closed on a Sunday or a holiday. Additionally, the map shows where the observation will take place. We can see how a street joins Ciclovia (Calle 26 in yellow), allowing the passing of bikers.

Figure 2: Example of counting spot

This methodology is made up of two phases. Phase 1 to set the distance, taking into account that smaller programs could benefit from covering a larger distance and Phase 2 which is the actual arrival rate determination.

Phase 1:
The aim of this phase is to determine the distance to cover. Phase 1 should be done on the rush hour, where more people participate in the program. Furthermore, the selected distance is encouraged to be kept constant throughout the day.

1. Place the two observers together on the same sides of the street. Each one is in charge of counting the total number of people who join the system within a 15 minute time frame.
2. Record the distance to be covered (i.e. 20 meters) and define how the ending point is going to be recognized. For example, a cone can be used for marking (e.g. count until the cone) or a place can be defined (e.g. count until a given tree/house/etc.).

3. After the time has elapsed, have the counters verify their observations. If they match, distance can be increased.

Phase 2:
This phase covers the actual observation of the arrival rate. Besides the steps here, it can be "corrected" by subtracting the number of participants who did not join the system but instead merely crossed the street. Furthermore, if one participant enters more than once in this distance, it can also be removed to avoid double counting.

1. Place the two observers on different sides of the street.
2. Each is in charge of counting one flow (north-south/south-north or east-west/west-east depending on the orientation of the street).
3. Divide the number of arrivals on the chosen distance to have the 15-minute arrival rate per meter.
4. Multiply by four to have the hourly arrival rate per meter.

In the case of Bogotá, this is to be done on the *reference track* (Carrera 15) and by having an average of these observed rates at 11:00 a.m. from fieldwork, we would have estimated the *reference arrival rate* to use as input in the simulation model. However, if Phase 2 is carried on throughout the day, instead of doing the weigh by time and location, we could keep each hour's arrival rate and only multiply it by the location weight to assign it to each track.

4.1.2 Proposed auxiliary model for determining the arrival rate through discrete-event simulation and based on surveys weighing and rush-hour flow

The logic behind this option is to build a discrete-event simulation model based upon the *reference track* (Carrera 15 for the city of Bogota) and considering arrivals at the hour. Each hour has a weight assigned to it from the surveys. Every person joining the system will have their corresponding duration. Then we create an output-statistic that considers the net flow from 11:00 a.m. to 11:15 a.m. and experimentally vary the *reference arrival rate* to get the desired output. Then this is run for each observation (date) and the average is estimated to be used as the input in the estimation model.

4.1.3 Proposed auxiliary model for determining the arrival rate through a discrete-event simulation and based on hourly flows

The basic idea of this model is that the flow of people counted in a given time is composed of the difference between arrivals and exits over the previous hours of operations of the Ciclovia and the new people who joined at this time. Thus, we seek both to count the flow of every hour, since it is a more reliable piece of data and to combine this data with the fitted sojourn time, so that we can infer the arrival rate. As we can see, the previous statement implies that there exists a discretization of the time in periods. These periods correspond to each hour of the program, for which we had proved that people arrive homogeneously during an hour (Section 5.1).

Let Y_t be the number arrivals, Z_t be the number of exits and N_t be the net flow, at time *t*. For every hour, this must be estimated, taking into account the observed flows and the sojourn times. The exits are the disposals of participants by the simulation model once their time is elapsed. The net flow is counted at fieldwork during 15-minute counts at the beginning of each hour along the duration of the program. Additionally, we take into account that the people counted at the first hour (N_0) are not "*Net flow*" but can be considered as only "*Arrivals*" ($N_0 = Y_0$). We therefore have the following equations expressing the relationship between arrivals, exits and net flows:

$$\sum_{0 \le i < t} (Y_{t-i} - Z_{t-i}) + Y_t = N_t \quad \rightarrow \quad Y_t = N_t - \sum_{0 \le i < t} (Y_{t-i} - Z_{t-i})$$

The left-hand-side of the equation shows that the difference of arrivals and exits, or the people remaining from previous hours plus the arrivals of a given hour, are equal to the flow observed. From this equation, we isolate the variable of the arrivals Y_t, which is the observed flow minus the people who remain from previous hours of Ciclovia. With this in mind, we construct the model adding an hourly flow statistic that must be matched to the one observed during the fieldwork by varying the arrival rate at each hour.

4.2 Data Collection

In order to test the methodologies above we perform a data collection phase. We test and refine the direct-observation methodology in fieldwork. Additionally, we improve the quality of the input data for Ciclovia of Bogota. For instance, we propose a survey specially designed for Ciclovias for a better estimation of key information identified in the Previous Work, Section 2, such as the weights of arrivals per hour. We also perform hourly flow counts, for the auxiliary model based on hourly flows. In the Appendix section, we include the survey that can be directly applied to any Ciclovia program and the formats used for the observation of the arrival rate and the flow of people. Finally, these activities must be applied across the entire Ciclovia network and during every hour of operation of the system. In the case of Bogota, this is performed in IDRD's counting spots and the different locations of the city from 7:00 a.m. to 2:00 p.m. so that they reflect the behavior of the entire Ciclovia system.

4.2.1 Fieldwork

We perform fieldwork during the whole Ciclovia day (seven hours from 7:00 a.m. to 2 p.m.) five times. During these visits, randomly sampled in the location, we perform at the beginning of every hour, a count of participants' flow divided according to the activities which they do in Ciclovia. Additionally, we test and improve the methodology to estimate the arrival rate based on direct observation, which is explained in detail in Section 4.1.1, as well as some practical guidelines for the selection of the spot to perform the fieldwork.

Besides these visits to a particular spot described above, we did a large-scale fieldwork as the closure activity of a national workshop of Ciclovia programs *"Taller nacional para programas de vías activas y saludables"*. The Sunday we spread ourselves all over IDRD's counting spots from before 7 a.m. to 2 p.m. (the whole duration of the program) with the company of the directors, collaborators and advisers of Ciclovias from 58 cities of Colombia. Here we collected the same information of the arrival rate and flow of people, and we applied the survey described in the next subsection. The formats used can be consulted in the Appendix section. From this session we were able to refine our formats that can be consulted in the Appendix section of this document.

The main findings of this fieldwork were that the flow of people follows the same behavior at the different tracks of the Ciclovia, as can be seen in Figure 3. This suggests the independence from location and people's behavior (arrival rate and flow). On the other hand, the weighing from 2011 survey (Sarmiento et. al, 2011) does not match fieldwork observations and it seems off. This may be because their original design was not made for this purpose and the conditions on which surveys were performed were not the best (e.g. times of the day, coverage of Ciclovia network, rain, etc.). This leads us to the next subsection in which we explain the new surveys we performed.

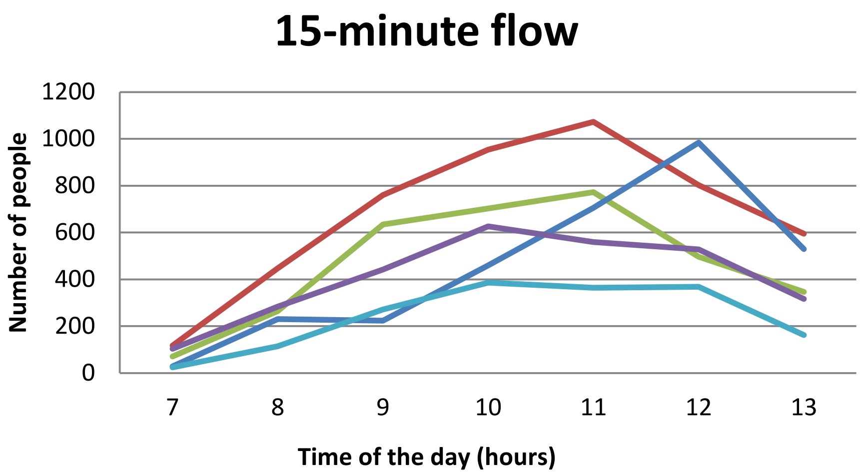

Figure 3: 15-minute flow throughout the day. Each colored line represents the observation of a day of fieldwork.

4.2.2 Survey design

As a part of the proposed methodology, we also suggest a systematic way for data collection. In this section we will explain how the surveys should be designed and performed, what the content (questions to be asked) should be and how to apply it in the case of a large Ciclovia where flow is different depending on the street. This is important for the application in any Ciclovia and specifically improves the data we had used in the Ciclovia of Bogota. This is because throughout the fieldwork explained in Section 4.2.1 we have found some discrepancies in regard to the weighing per hour from the survey (Sarmiento et. al 2011). This makes sense because this was not a survey originally designed for Ciclovia programs, but a general survey on *physical activity and healthy habits* which included a subset of questions on Ciclovia. Additionally, the conditions in which the surveys were performed were not the best (e.g. times, rain, locations, etc.) for obtaining a representative sample of Ciclovia users and thus a new properly done survey can help upgrade the input data to be used in the model for the Ciclovia of Bogota.

The survey can be consulted in the Appendix section. The survey starts with general questions to identify general information specific to the survey (id of the surveyor and surveyed, climate and time), then it asks about the demography (sex and age), then we move on to key information of the model: arrival location and time. We then ask for the sojourn time and at last the activity performed at Ciclovia. Additionally, this survey is encouraged to be personalized for each Ciclovia program in order to get more information as needed. For example, in the extension for Ciclovia of Bogota we included questions on company, the main objective of usage and the frequency (factors analyses for the sojourn time in Section 5.6) when using Ciclovia. Other Ciclovias, for example, might be interested in ask about complementary activities or schedules, for example.

In order to decide on the type of sampling, we used as reference Sampling techniques (Cochran, 2010) and we considered the questions on the survey and in particular information on arrivals. At different hours and locations of the program we will encounter a different population. This means we need to segment by clusters, since the answers will be heterogeneous on time and place. The idea is to distribute the surveys in time and location, so that we get homogeneous responses. The first limitation is how to allocate this survey as we do not know the variances of the different spots or times, and the density of the neighborhood is not a good measure as the target population is Ciclovia users rather than city habitants.

Therefore, the distribution must be based on the only measure available; the flow of people. Based on IDRD's database, we were able to explore 1 year worth of data which contains 15 minute counts at 11:00 a.m. at 17 different spots of the city and calculate their relative weight. Furthermore, relying on our fieldwork (Section 4.2.1) we could do this same analysis for the time of the day. Since for this type of system it is not possible to calculate an optimum number of surveys, we did the analysis based on the available resources. This gave us the optimum way to allocate these surveys so that they would be representative of the Ciclovia population and no area or time of the day was favored. In Table 1, we can see the allocation of 900 surveys in the city of Bogota.

Table 1: Sojourn time (hours) in Ciclovia

Place/time weights		7	8	9	10	11	12	13
		0.04	0.09	0.15	0.20	0.20	0.20	0.12
Carrera 24 X Calle 41	0.02	1	1	2	3	3	3	2
Av. Boyaca X Cll 72 Sur	0.02	1	2	3	4	4	4	2
Calle 42 Sur X Av. Carrera 68	0.03	2	2	4	5	5	5	3
Carrera 10 X Calle 17 Sur	0.03	2	3	5	6	6	6	4
Carrera 50 X Calle 3	0.06	2	4	8	10	10	11	6
Calle 116 X Av. Suba	0.07	3	6	10	13	13	13	8
Av. Boyaca X Calle 3	0.09	3	7	12	17	17	17	10
Calle 26 X Carrera 40	0.10	3	7	13	17	17	18	11
Av. Boyaca X Calle 53	0.10	4	8	13	18	18	18	11
Carrera 7 X Calle 60	0.10	4	8	13	18	18	18	11
Av. Carrera 9 X Calle 140	0.13	5	10	17	22	23	23	14
Av. Boyaca X Calle 138	0.13	5	10	17	22	23	23	14
Carrera 7 X Av. Jiménez	0.13	5	10	17	23	23	23	14

The findings of the surveys were:

- The most important finding of these surveys was the correction of the behavior of the arrivals as shown in Figure 4. This is an important contribution for a better estimation of participants and it will be useful for the analysis of assumptions in the next section.

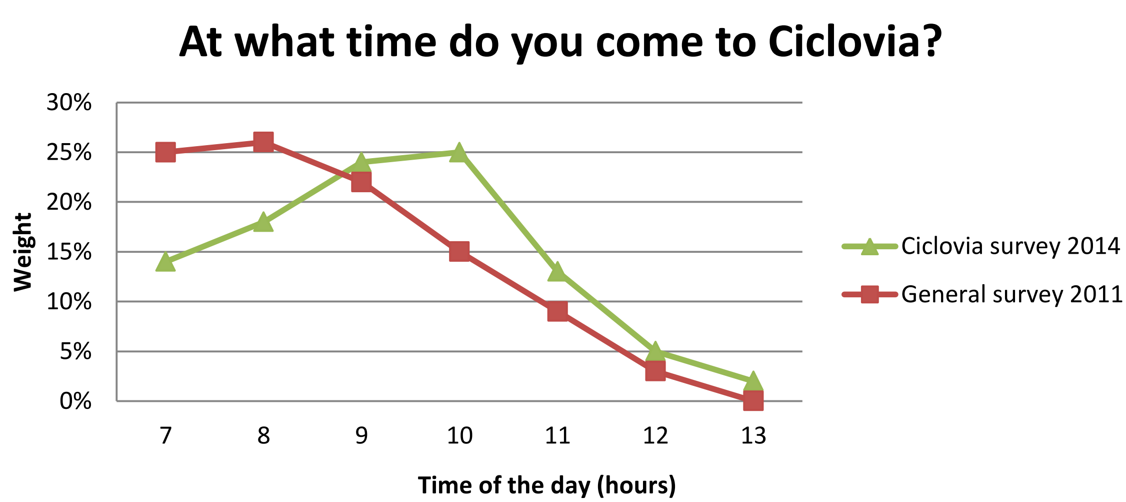

Figure 4: Weight of the hour of arrival

- Regarding the point at which people join Ciclovia, we believe this input data was also improved since we ask for it directly in the survey. This is an improvement compared to the previous survey (Sarmiento et. al 2011) in which respondents were asked to give their addresses and were assumed to join Ciclovia from their homes. In Figure 5 we can see that most of the survey points fit directly in the Ciclovia network and there are less points that need to be assigned to the closest Ciclovia street as in the first estimation for Bogota (Murcia, Rivera et. al 2013).

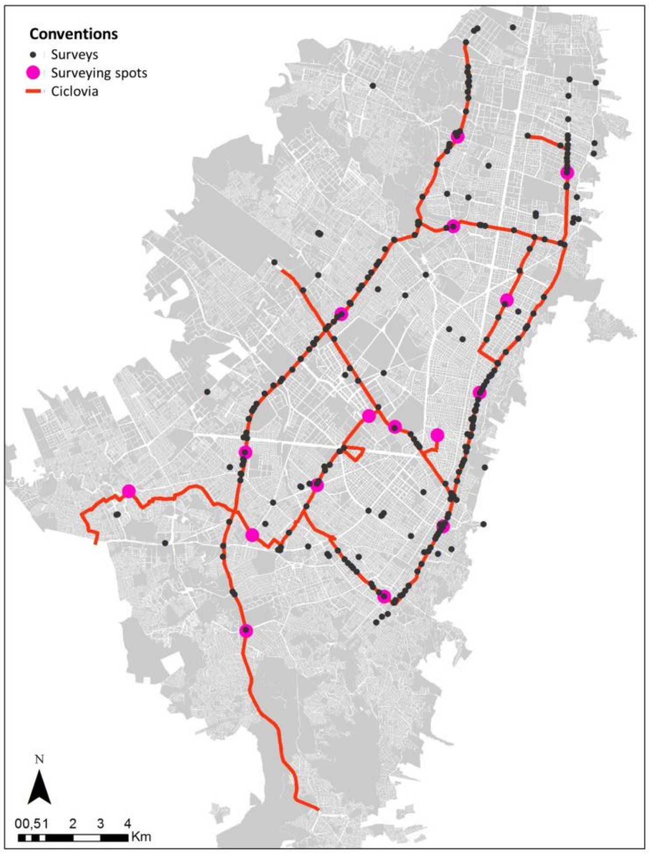

Figure 5: Geo-referentiation of the starting point of participants

- Concerning the demographics of the participants at Ciclovia: there is a 65% male participation against a 34% female participation.
- Regarding the company, 31% of the surveyed people tend to come alone. On the other hand, people come with one person and two persons 26% and 21% of the time, respectively. From this point on, it starts to decrease having 11% of the participants coming in groups of 4 people; 22% of the participants come in groups of 4 or more, recording up to 20 people, who belong to the same biking club.
- We divided the possible reasons for which people join Ciclovia into two main categories physical activity (e.g., sports, a healthy life style, training, etc.) and amusement (e.g., spend time with friends and family, enjoy time outdoors, etc.), with a percentage of participants of 60% and 40% respectively.

- Finally, we examined the frequency of participation in the program. We obtained that 64% of the participants surveyed come every Sunday into the Ciclovia, 23% once a month, 10% once and month and the remaining 3% come with less regularity.

4.3 Results

In this section we present the results for the three proposed methodologies. Finally after showing the limitations of every suggested approach, we recommend that an auxiliary model be used whenever possible. This is based on survey weighing and rush hour flow, coming from the flow and the weights of the surveys. Applying this methodology to the city of Bogota, we obtained an estimation of the *reference arrival rate* with a mean of 0.166 persons/15-minutes, which will be taken as the parameter. Finally, we encourage Ciclovia programs to perform a survey which covers several spots of the Ciclovia network and is distributed throughout time, as based on our fieldwork observations. We believe they are indeed a good representation of the participants' behavior, which means that there is no need to relax the assumption of the weighted arrival times and auxiliary model based on hourly flows is not necessary.

4.3.1 Results from direct observations

This direct observation methodology seems to work on the field. Based on the fieldwork, we suggest an observation distance of 50 meters. This recommendation, based on the Ciclovia of Bogota, the longest program in the world and therefore the most visited, should hold for other smaller cities and municipalities. Thus, if a given Ciclovia program wants to follow the recommended 50 meters, phase one of the method can be skipped. However, cities with special geography such as curved streets or larger Ciclovia programs are encouraged to find their own distance. Likewise, smaller Ciclovia programs could benefit from determining a larger distance in order to capture more information.

This methodology for fieldwork would only take place in one track that is representative of the Ciclovia to determine the *reference arrival rate.* This way, it neither underestimates nor overestimates the number of participants. However, these observations are very volatile, sensitive to the observer and sensitive to the selection of the observation distance and spot and even with practical recommendations and guidelines on how to choose the spot, this is still an approximation. This means that a high number of observations would be required and according to the case study this could be very costly. This volatility would make this input vary between different Ciclovia programs, making the output of the model non comparable. Instead we advise the use of the auxiliary models that is based on more reliable observations: the flow of people, obtains the arrival rate. This leads us to the next section.

4.3.2 Results from the auxiliary model for determining the arrival rate through a discrete-event simulation based on surveys weighing and rush-hour flow

This model was verified following the recommendations for verification techniques in (Law and Kelton 2000) of incremental built of the model by adding divisions of the track, checking for mean, min, max of sojourn time and having colored entities per hour of arrival allowing us to check the composition at each hour and check whether it made sense. It was validated by taking into account that the flow statistics matched IDRD's database and comparing arrival rate gotten by the model and the one observed at fieldwork. Information on 2014 IDRD's flow target was always met using a 90% C.I. and 10 replications, while from the six performed fieldworks the mean difference was of 10% between the observed rate at 11 a.m. in the reference track and the estimation using this auxiliary model, taking into account that both are estimations and thus have some error.

In the case of Bogota, this auxiliary model is run for our reference track Carrera 15 using IDRD's database for 2013, obtaining a mean of 0.166 persons per 15-minutes per meter. Additionally, a **s**ensitivity analysis is done to the number of partitions of the track under consideration knowing we want to imitate infinity and the randomly selected points where counting took place. The model proved not to be sensitive to the number of partitions as confidence intervals with a different number of partitions overlapped and was highly sensitive to randomly selected counting spots. This is why one of the model limitations is to use only IDRD's counting spots

As a conclusion, this model relies on the weights suggested by the surveys. Thus it becomes interesting to relax this assumption and take into account the flow at the different moments of the day. This helps us introduce the second model based on hourly flow.

4.3.3 Results from auxiliary model for determining the arrival rate through a discrete-event simulation model based on hourly flow

This model was verified using the same techniques as the previous model. However the model was not validated. This is because in order to match the flow, the resulting hourly weights did not make sense. First because relative weights were very high (e.g. 200% of arrival rate at 7 a.m.) and also because arrival rate chart does not look as expected (peak at an early hour to match the shape of the flow in Figure 3). Figure 6 illustrates this situation.

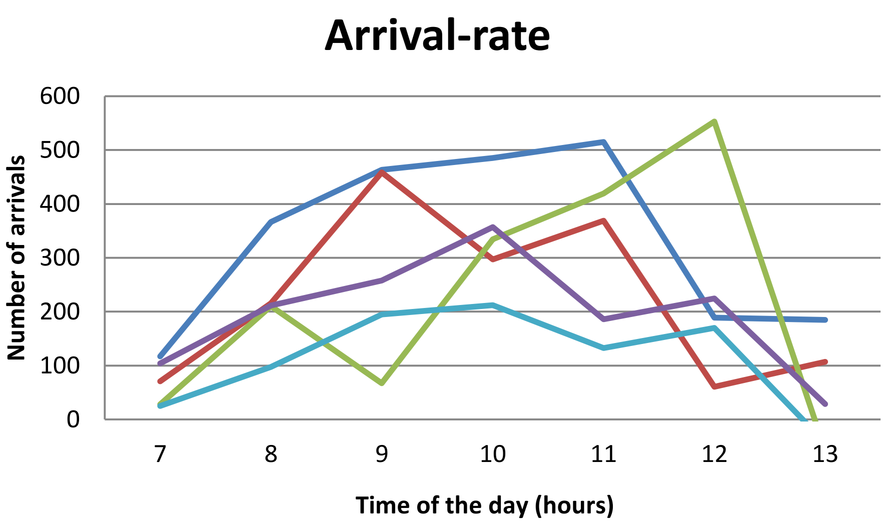

Figure 6: Resulting arrival rate from auxiliary model based on hourly flow. Each colored line represents the observation of a day of fieldwork.

As a conclusion, we believe the main reasons the model fails to represent reality are that the routing from other tracks also influences the model and that there might be problems associated with using a discrete sojourn time distribution. In order to solve this issue we modify the original counting model for adjusting routing probabilities (Section 5.4) and smooth the sojourn time distribution (Section 5.6). Furthermore, given a good survey plan, we believe the weighing per hour is reliable and advisable, making it a suitable assumption to keep.

4.4 Analysis on the arrival rate according to external factors

Finally, having established the auxiliary model based on survey weighing and rush hour flow to estimate arrival rate, we proceed to examine how other factors affect (increase/decrease) this rate and therefore the number of participants. The main factors we identified which could potentially impact the arrival rate and for which there is data available for analysis are the following:

1) Holidays: these are especially important in the context of Ciclovia, because it operates on Sundays and holidays, which are mostly on Mondays. Furthermore, Colombia is the country with the most holidays in the world, with around 17 holidays every year.
2) Weather: IDRD records the climate along with the flow of people into four categories. These are sunny, cold, rainy and cloudy.
3) Races and other special events: Ciclovia is a dynamic program in which both the city authorities and the private entities organize events for the integration of the people and the support of certain causes. Some examples of these organized races are Fundación Matamoros for amputees of the civil conflict, Nike 10km race to support breast cancer or the food festival Alimentarte.

We analyze these factors using a design of experiments, taking into account that the variable are categorical. As we can conclude by examining the significance of each factor, the arrival rate on holidays is statistically different, while the factors races and climate do not affect the arrival rate. Our initial hypothesis was that taking into account the participants of the race they were going to be more people coming to Ciclovia. However, people who do not enroll in the race do not modify their behavior. Thus, for taking this change into account, it would be enough to use the estimated number of participants from the simulation model and add the number of people who enrolled in the sporting event. Regarding the weather, it seems that the impact on the change in the rate is not as significant as expected. Finally, the difference between a regular day and a holiday accounts for a mean reduction of 0.062 people per hour and is statistically significant with a p-value smaller than 0.01. The estimation of the number of participants for regular days and holidays is included in Section 6.

5 ASSUMPTIONS RELAXATION

In this section we examine the assumptions of the previous model, whether they need to be relaxed and how this is done. As an introduction we present a summary of the assumptions used in the implementation of the model. It is important to highlight that, except for the first three assumptions, none of these assumptions have an impact on the estimated total number of participants in the Ciclovia of a given day, but they facilitate the design and implementation of the simulation model.

1) The arrivals are homogeneous during an hour. Thus, the 15-minute *reference arrival rate* is transformed into an hourly rate by being multiplied by four.
2) The arrivals are homogeneous within a track. This means that the behavior of arrivals is the same in the whole length of the track.
3) The change in the arrival rate by hour is independent of the track.
4) As the routes followed by people are unknown, participants randomly choose the next track to follow.
5) All tracks have the same proportion of participants in each category.
6) Sojourn time modeled as an empirical distribution equal for every person regardless of their characteristics.

5.1 Arrivals are homogeneous during an hour

We test the homogeneity of arrivals in two ways: the homogeneity on the flow and on the arrival rate according to the proposed methodology of direct observation. We are able to test this assumption based on the homogeneity of the flow because of the existent relationship between arrivals and flow explained in Section 4.1.3. For this matter, we use the fieldwork data: flow and arrival rates during different hours of the program and different tracks. All of these were randomly sampled. During an hour we group three sets of 5-minutes observations. Then we compare the means using ANOVA. The null hypothesis of equality is not rejected. This means this assumption is statistically reasonable.

Both approaches yield the same conclusion and it is that this assumption can be held. Furthermore, as the main goal of this Ciclovia counting is the ease of implementation and replicability in different Ciclovia programs directly by Ciclovia directors, who might not have an engineering background, it is better to discretize the time than to do an approach as a continuous variable. Additionally, since arrival rate is weighted according to survey results, a discrete approach is more suitable. Finally, this assumption of discretizing the time offsets a continuous approach and does not change the desired outputs.

5.2 Arrivals are homogeneous within a track

From the original methodology, tracks were defined as segments between two intersections. Our partition of the Ciclovia of Bogota, as shown in Figure 1 from Section 2 is compared with the way IDRD divides it. The partitions they have implemented are called *corridors*. A corridor is defined as streets with similar behaviors on the flow on people, but not exactly homogeneous. This becomes evident with their selection of counting spots, as each has a different level of attendees but some are located on the same corridor. Both corridors and counting spots can be seen in Figure 7.

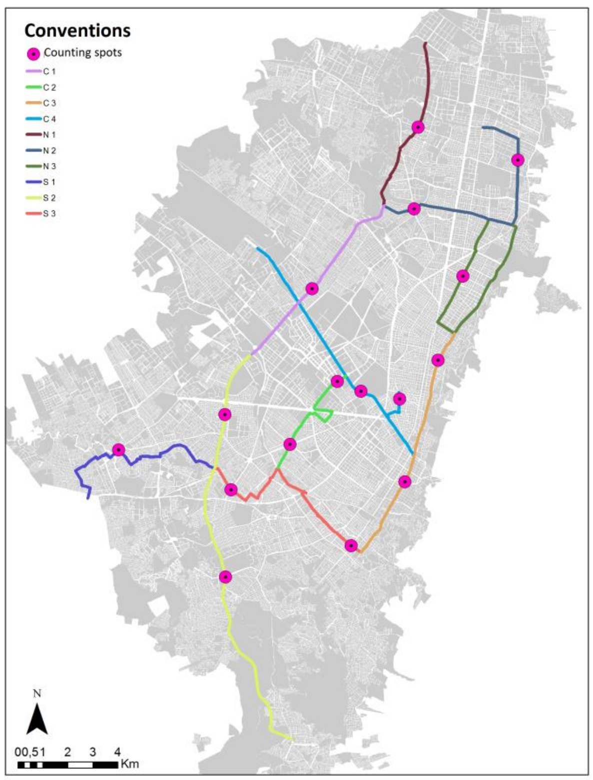

Figure 7: Partition of Ciclovia of Bogota according to IDRD's corridors

As a conclusion we can see that IDRD also divides Ciclovia in smaller street segments with similar behavior, we just go further and ensure homogeneity of arrivals within the track. Since we both use the same assumption, taking into account the relationship between the flow and the arrivals, it is reasonable to say it is correct. Furthermore, if the model were modified to use the IDRD partition, only the arrival location of the people would change. The number of respondents, who said they arrived to specific tracks that make up a corridor, would be added together and there would be no implications on the total number of participants.

On the other hand, this assumption is actually relaxed because now tracks are defined as a street segment with similar behavior and not necessarily based on forks, as in the previous version of this work. This is due to the fact that most Ciclovia programs in Colombia follow a linear trajectory (EpiAndes and U. Rosario, 2013) and the feedback by working with directors of Ciclovia program at a national level at a Workshop organized by Coldeportes on how to divide a Ciclova network. This assumption, that similar behavior relies on geographical conditions, is confirmed by expert opinion. In the case of Bogota the trigger of the partitions is forks since participants have a routing choice. While in other cities this trigger corresponded to geographical conditions of the city in general. Some examples are: where the city starts, a double-lane street, a neighborhood and a street going up a hill. As a final remark, this assumption does not change the desired outputs but would enable us to relax the assumption number 6 on the routing.

5.3 Independence of hour and location of the arrivals

Let I be the set of hours in which a particular Ciclovia program operates and J the set of tracks in which we had partitioned this Ciclovia. We know the arrivals depend on both the time and the location. However, we have assumed so far that the change in the arrival rate by hour is independent of the track. This assumption is what had allowed us to calculate the arrival rate of any track at any hour using the formula below:

$$Arrival\ rate_{ij} = w_i \cdot w_j \cdot "reference\ arrival\ rate"$$

This assumption was tested by the data collected on the fieldwork and by the analysis of the surveys performed, where we analyzed the behavior of arrivals. We obtained the percentage of these arrivals occurring at each of the locations visited on different days for fieldwork. Same came from the surveys, for each of the surveying spots. We built a confidence interval of the percentage of the arrivals at each hour compared to all the arrivals throughout the day. The variance was estimated from the difference of the fieldwork days for the first case and the different locations for the surveys. We found that these intervals overlapped for the different locations. This indicates that statically speaking, we could keep this assumption.

5.4 Random routing

The objective of this section is to fix the routing on the model, in order to accomplish a better imitation of the real system and the way participants move across the different Ciclovia tracks. This will be done by comparing the 15-minute flow of people at 11:00 am of IDRD database with the one simulated by the model. This way, we change routing probabilities at each intersection until the flow matches.

It is important to consider that since this is not a one-to-one problem, several sets of probabilities could lead to the same answer. In order to avoid obtaining an unreasonable probability set, a previous analysis on conditions was done (e.g. more chance to take a track including parks, less chance to take tracks at the extremities of the network). With these ideas in mind, a first step is to identify the proportion of the total counted participants that each track accounts for and divide the flow into three categories: low, medium and high. The basic idea is to reduce the probability of choosing a track in which there is a low-flow counting spot and likewise, increase the probability of choosing a track of high-flow. Medium counting spots will not be modified at this stage. This classification was assigned as follows:

Table 2: Classification of flow levels

ID	Mean flux of people	Weight	Cumulative weight	Classification
13	60.2	0.0142	0.0142	Low
11	66	0.0155	0.0297	Low
14	73.2	0.0172	0.047	Low
7	91.3	0.0215	0.0685	Low

17	101.7	0.024	0.0924	Low
16	120.2	0.0283	0.1208	Low
15	141.4	0.0333	0.154	Low
8	206.2	0.0486	0.2026	Low
3	259.5	0.0611	0.2637	Medium
5	333.1	0.0784	0.3422	Medium
12	343.5	0.0809	0.4231	Medium
6	352.8	0.0831	0.5062	Medium
10	356.3	0.0839	0.5901	Medium
4	394.9	0.093	0.6831	Medium
2	445.9	0.105	0.7881	High
1	446.5	0.1052	0.8932	High
9	453.3	0.1068	1	High

Out of the low categories we have chosen counting spots 7 and 8, counting spots with medium flow 6 and 10 and from the high category, spots 2 and 9. This selection is due to their location in the Ciclovia network, in other words, spots at which changes seem to be intuitive and reasonable. In Table 3, we can see the mean flux of people without adjusting the flux with routing and the associated error, with regard to the IDRD database. From this we can identify that counting spots 2 and 6 with -82% and 210% error are the most distant from the database value. Furthermore, we have an overall error of -123% meaning that random routing simulated the flow of less people than it should have in the counting spots, and on average each point has a -21% error.

Table 3: Mean flow of people with random routing

ID	Mean flow of people	Error
2	80.0	-82%
6	1093.7	210%
7	132.6	45%
8	235.9	14%
9	302.0	-33%
10	245.8	31%

In order to fix this problem, we used the add-in complement OptQuest for Simio. The control variables we are allowed to change are the probabilities of choosing a given path, under the constraint that the different weights for a given intersection add up to one. The response variables were the flow at each of the counting spots specified above in Table 3, and the objective was to minimize the error between the database and the simulated flow.

Finally, based on conditions analysis, relevant counting spots and the OptQuest, the best configuration gotten has an overall error of -38% and an average at each counting spot of -6%. The routing changed in seven intersections. The remaining eleven intersections include 5 extreme intersections were no routing is possible as entities must take the only available track and six other intersections which continue to have random routing with the same probability of choosing any of the connection paths. The changed intersections were: intersections 2, 3, 5, 6, 7, 11 and 14 and their change was as shown on Figure 8.

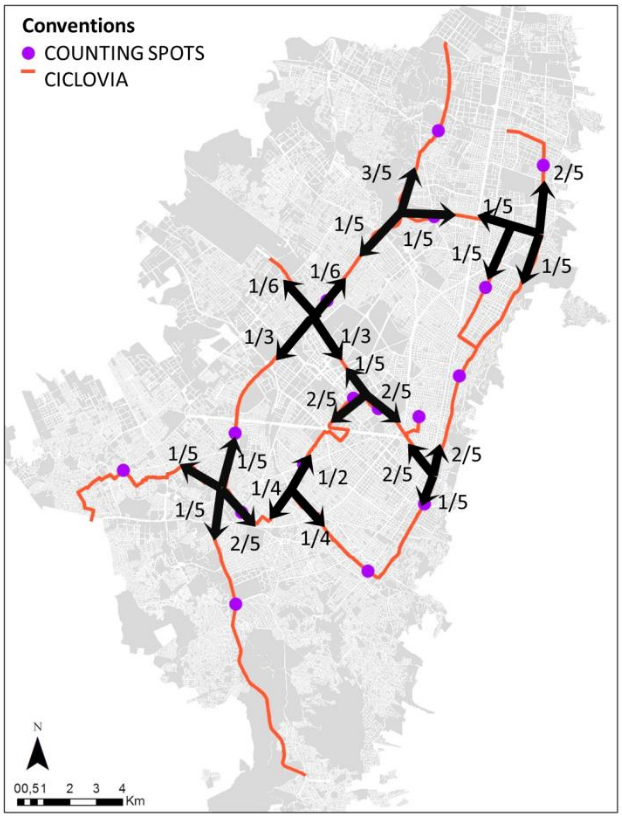

Figure 8: New routing probabilities.

We can see that now more people tend to go to the extremities in the north of the city, where two of the counting spots which report the most flux are located. Furthermore, the rest of the extremity tracks in the model have reduced their selection probabilities. We can see that the most important changes occurred at the center of the network, where the participants have more routing choices. Finally, this improvement in the routing helps the model to better imitate the actual behavior of participants in the Ciclovia of Bogota.

5.5 Same percentage per category on each track

At the arrival of participants, each is assigned with a category according to the activity they perform in Ciclovia. The same proportion for each activity was assigned in every source of the model. This assumption was tested using the IDRD database, from which the percentage of the four observed categories at each counting spot was calculated. This allows us to test the assumption by comparing the percentages' mean for each activity using an analysis of variance (ANOVA). These four models which included only the proportion per activity and the location were non-significant. This means that this assumption can be kept.

5.6 Sojourn time

In the previous approach of the methodology we had used an empirical distribution to model the sojourn time of the participants since this did not affect the overall number of participants. However, in order to add more detail on the behavior of participants and better mimic the system, we smooth this distribution as seen in Figure 9.

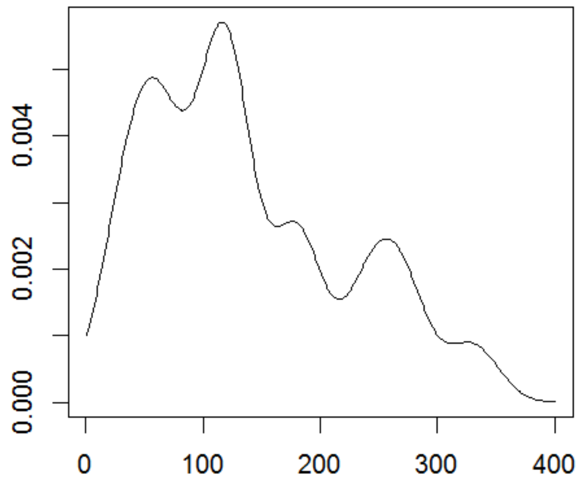

Figure 9: Sojourn time at Ciclovia as an empirical distribution and smoothed as a linear combination of normal distributions (Sarmiento et al. 2011).

On a first step, we carefully observed and analyzed this data. For instance, during the first three hours we can see a peak every half an hour according to the answers of the participants who tend to discretize the time as an answer to a survey, but also to behave accordingly. After this time, it all seems to collapse as respondents varied significantly from their answers. Then, as a means of smoothing this distribution, we modeled this data, using a weight w_i, as a linear combination of normal distributions with mean μ_i and constant variance σ^2, based on a set of time ranges I.

$$\sum_{i \epsilon Time\ Range} w_i N(\mu_i, \sigma^2)$$

This was done by keeping a half an hour mean during the first three hours, and for the last part, by grouping the rest of the time together and calculating its mean. As for the variance, it was first computed by taking into account the S_p^2. Nevertheless, due to an over fitting of the distribution, variance was increased arbitrarily such that the distribution was smoothed altogether. The same variance was kept for all the normal distributions. Finally, for each individual normal distribution a weight was computed according to the frequency and the density is truncated at zero. Table 4 shows all the parameters mentioned above.

Table 4: Sojourn time (hours) in Ciclovia

i	From	To	μ	S^2	w_i
1	0.00	0.75	0.50	0.004	13.9%
2	0.75	1.25	1.00	0.002	17.2%
3	1.25	1.75	1.50	0.002	5.8%
4	1.75	2.25	2.00	0.000	28.1%
5	2.25	2.75	2.50	0.003	2.1%
6	2.75	3.25	3.00	0.000	19.6%
7	3.25	7.25	4.48	0.367	13.4%

Additionally, in the previous version of this methodology we had that every person, regardless of their type of activity, has the same sojourn time distribution in the system. To upgrade this information, we performed a design of experiments (DOE) to analyze the impact of some factors that could affect the sojourn time of participants. The factors considered are believed to contribute to the main variation of the time people spend in Ciclovia. These factors were chosen based on expert opinion (Dr. Sarmiento and Francisco Cañon, Director of Ciclovia's Program in Bogota). These are:

1) Factor A: Objective (Perform physical activity, Amusement)
   National survey (Sarmiento et al. 2011) includes many more categories such as "share time with friends and family" and so on. However these can be collapsed into two groups. We might think the low level on this factor is *physical activity* as these people are going to exercise, while *amusement* is considered in the high level of the factor, since people participating in Ciclovia for fun can potentially make pauses to eat or relax, staying longer in the system.
2) Factor B: Type of activity (Walking, Cycling, Jogging)
   Beforehand it is not possible to determine if a certain type of user lasts longer in the Ciclovia. The activities considered here are based on the ones used in the model (i.e., cyclists, joggers, walkers, others). However, we are not considering the fourth category because the percentage is very small and the composition not homogenous (i.e. skaters, wheelchairs, etc.) thus it will be difficult to add in the experiment and could result in adding more variability.
3) Factor C: Company (Alone, With company)
   People who come with company are expected to last longer in Ciclovia, as they might want to stay longer and enjoy the time out. This way, the low level of the factor is *alone*, while the high level is *company*.
4) Factor D: Frequency of participation in Ciclovia (Regulars users, Random users)
   Regular users are defined as people who participate in Ciclovia at least twice a month. We might think the *regular users* may have a known expected time in Ciclovia because they have generated a schedule and a routine every Sunday or holiday. This way, the low level for this factor is *regular users* and the high level *random users*.

The same experts to whom we have consulted stated some other possible factors, which we have not included as the *Design factors* because they are out of our control to run the experiment. Some examples are: gender, Ciclovia track and hour of the day. These factors will be allowed to vary with the assumption that the variability caused by them would not be significant so we rely on randomization to balance out these effects.

Nuisance factors may have large effects that must be accounted for, but in the context of the experiment they may not be of interest. It is important to block these factors. For instance, the day in which the experiment was run is a controllable factor, however inside it, it contains several sources of uncontrollable variation such as the weather or activities such as gastronomical festivals (e.g. Alimentarte) or sports events (e.g. Nike 10k race).

This design is going to be a factorial design with four factors. One factor has three levels (type of activity that the person performs at the Ciclovia) and the other factors have two levels. As we explained before, there are going to be blocks in this experiment according to the day of the run.

The data needed a transformation because the assumption of constant variance did not hold. After applying the squared root transformation, we evaluate the necessary assumptions with the transform model. There are no problems with the normality assumption from the normal plot of residuals, additionally; residuals do not behave in a particular or clear pattern thus the independence assumption holds. Finally, the transformation of the model seems to be adequate, because it corrects the tendency that the residuals followed and were dispersed. We can conclude that all assumptions hold.

Form the half-normal plot (Figure 10), we can see that the factor B (activity) accounts for the main source of variation in the sojourn time at Ciclovia. Our model is significant at 5% level; as well the individual factors also show to be significant.

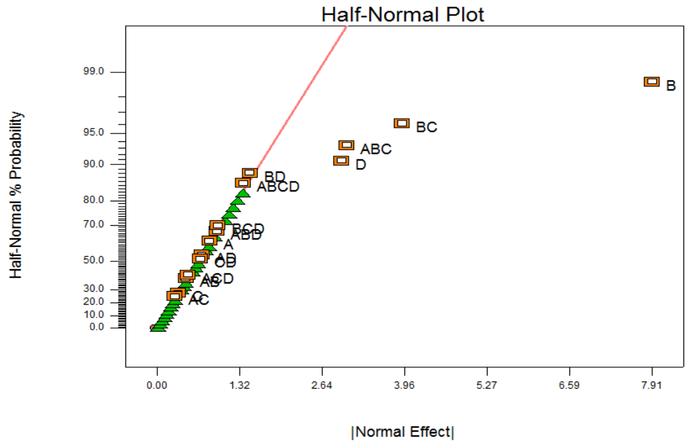

Figure 10: Half-normal plot for significant effects

In Figure 11 we present the results of the experiment and the factors that were significant in the ANOVA. As a conclusion, the sojourn time of bikers is longer than that of people who are walking or jogging. However, the participants doing these activities tend to spend a similar amount of time.

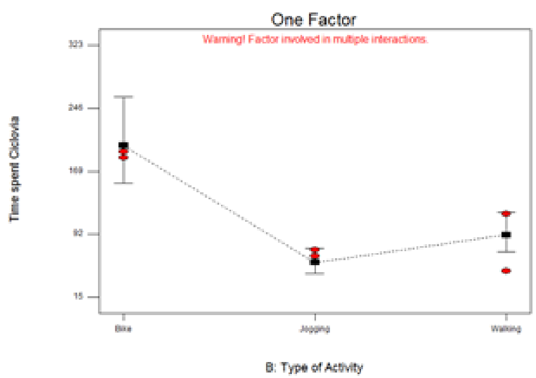

Figure 11: Factor B, type of activity

The contacted people were volunteers intercepted at Ciclovia. We are aware they might not be representative of the whole universe of participants in the Ciclovia, this is why, in order to fit the sojourn time distribution, we are not considering the actual times of the experiment, but the answers of the respondents from the survey. This way, two distributions are fitted; one for bikers as this level of the factor activity is statistically different from the others, and another one for walking and jogging as these two levels have overlapping intervals. These distributions are shown in Figures 12 and 13.

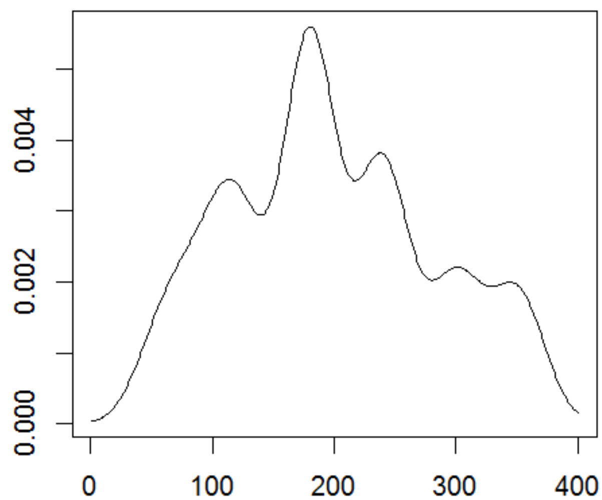

Figure 12: Sojourn time biking

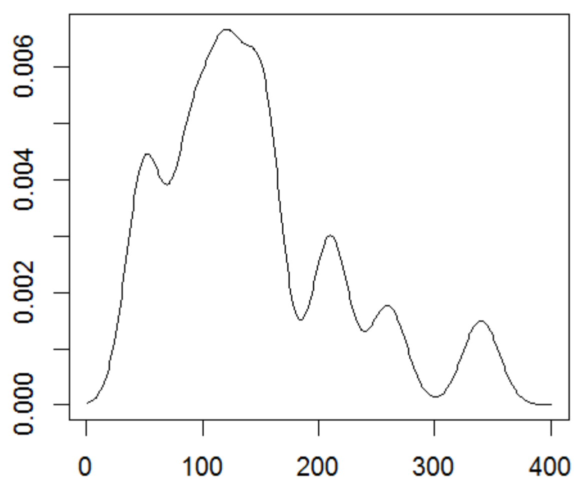

Figure 13: Sojourn time walking and jogging

6 CONCLUSIONS AND FUTURE WORK

This work contributes to the efforts started in 2012 inspired by the challenging problem of counting the number of participants in an open network like that of the Ciclovia programs and in which, due to the dynamism of the program with new programs opening sporadically, it is not possible to have a solid data base. In this work, we tackle the data collection of the arrival rate, the main input of the discrete-event simulation model. Additionally, the adaptation of the model through the relaxation of assumptions and construction of energetic expenditure statistics respond to the future work section of our previous work (2013), which contributes to the improvement of public health.

We gave insight on how to collect input data with proposed surveys and formats which can be consulted and used for other Ciclovias. We proposed a direct observation methodology tested in field as well as an auxiliary model for the collection of the arrival rate. We encourage programs to perform direct counting at an early stage of their counting initiatives both for validating their programs and for better understating this input. However, at a more advanced stage, programs should keep using the auxiliary model.

For the case of Bogota, we upgrade the available input data by performing fieldwork and surveys for more precise-estimation. This allowed us to improve the quality of the input data. Additionally, we review the assumptions made in the previous work. We kept the hour-homogeneity, the independence of time-location and the proportion of activities in the tracks, as they proved to be statistically non-significant. We redefined the definition of tracks based on a scanning of Colombia's Ciclovia programs and tackled the routing problem. In regard to the sojourn time, we smoothed this distribution and adjusted a probability distribution for the main activities performed in the Ciclovia, since this factor has the biggest influence on this input. Additionally, we explored important factors that influence the people's behavior: sojourn time, routing and arrival rate.

We finally integrated this and obtained the following results, in which the difference of a regular day and a holiday can be observed. This significant difference is around 60%.

$$Number\ of\ participants\ on\ a\ regular\ day = 1'019{,}024 \pm 49{,}110$$
$$Number\ of\ participants\ on\ a\ holiday = 650{,}703 \pm 46{,}284$$

It is important to mention that the confidence interval has a 95% confidence level and corresponds to a confidence interval from the model rather than from the actual system. It was constructed making 10 replications. The sources of variation are due to modeling the arrivals to the system as a non-homogeneous Poisson Process with the rate calculated in Section 4.

As means of comparison, using the data from fieldwork in which we have the 15-minute flow at the rush hour required to use IDRD's expansion formula to calculate the number of participants, we get 1,462,918 participants. These represents a difference of 30%. This results are as expected, because having encounter IDRD's staff, we were told they suspect their estimates to be too high. Furthermore, by analyzing their expansion formula, we saw that they count people during the rush hour and extend it without a correction factor through the whole duration of the program by multiplying the 7 hours of operation. Furthermore, IDRD told us the selection of their counting points ensured avoiding double counting by having them separated more or less 2.5 km, the maximum distance a biker could cover in 15 minutes. However, from analyzing their spots, this does not hold between every pair of spots.

On the other hand, this model also includes statistics related to public health. For example, taking into account the type of physical activities and the number of participants, we can calculate metabolic equivalents, or METs, to assess physical activity levels. A weighed MET score will be given to each activity category (sedentary = 1 MET, moderate = 3 METs, vigorous activity = 6 METs) according to Organización Panamericana de la Salud et al. (2009). These weighted MET scores will be multiplied by the observed number of participants in each type of category for physical activities at the moment of the observation. Then, this score will be divided by the total number of participants in the Ciclovia program. This calculation will produce a mean physical activity intensity score presented on Table 5.

Table 5: Average MET of the program

Activity	Average METS of Ciclovia day
Biking	943,038 ± 4,387
Jogging	75,768 ± 1,811
Other	22,771 ± 1,246
Walking	135,448 ± 526
Total	294,256 ± 1,992

We can see that these new results, account for a better estimation of participants in the city of Bogota, distinguished by regular day and holiday. This factor is indeed the most important one in terms of the variation of the arrival rate and the number of participants. Furthermore, a new output for the medical community is the MET consumption. This allows for a better measurement of the impact on public health.

As stated in the motivation and relevance of this work, we aimed to improve the imitation of behavior or participants at Ciclovia's network. This is why we initially thought that one of the most important contributions of fixing the routing would be that we could have an approximate average of the number of participants in each track. However, the obtained results were not exactly as expected since the average occupation of each track seemed somewhat unrealistic as they are some very empty tracks. This is due to the software's limitations; both for practical reasons and for internal constrains which made us have to keep model entities arrival only at intersections. These are: the statement of a very large number of arrival tables and violation of the maximum number of elements in Simio's version 5.83 University Edition, respectively. However, the development of open-source software specially built for counting in Ciclovia could allow for a better handling of this issue. For example, when creating entities in a track, they can be assigned a random number within the length of the track for the entity to start. This representation of the system would be closer to reality and potentially could allow a better calculation of the average number of participants per track output. The relevance of this output is its usefulness to the entities in charge of Ciclovia programs, who could use it as a factor to better allocate their resources (e.g. vigilance, collaborators, design of water spots, etc.).

Another one of the important limitations of this program is the need to consult experts on how to select the reference track. In the case of Bogota, we have been working with IDRD's coordinator who instructed us on the selection of a track representative of the Ciclovia as a whole. This selection is still arbitrary and thus, it would be very interesting to develop a correction factor between flow and/or arrival rate observer at fieldwork across the different spots of the Ciclovia network.

Additionally, there is still a lot of research to be done in the planning phase of Ciclovia; basically, to see which conditions a Ciclovia should have on the selected route, the opening-hours, amongst other factors to attract a given number of participants. Alternatively, to already establish Ciclovias in which changing the characteristics is not advisable, it would be interesting to explore how to optimize the resources to attract more people and have a more efficient program.

ACKNOWLEDGMENTS

We would like to thank all the entities who have enabled the completion of this project: Universidad de los Andes, Colciencias, Coldeportes and IDRD. Specially, we would like to thank the program directors, planners and advisers whose active engagement, generous contributions and feedback during Coldeportes workshop on April 25th-27th, made the project possible. We also thank Francisco Cañon director of the Ciclovia who authorized us to perform field work and the Ciclovia's team (guardians and monitors) for the interest shown in the project.

We would also like to thank Professor Rogeiro Ferminio from Pontifical University Catholic of Parana –PUCPR who helped us integrating the energetic expenditure of participants METS in our model.

REFERENCES

Ainsworht, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., Greer, J.L., Vezina. J., Whitt-Glover, M. C., Leon, A. S. 2011. "Compendium of Physical Activities: A Second Update of Codes and MET Values". American College of Sports Medicine.

Banks, J., J. Carson, B. Nelson, D. Nicol. 2010. *Discrete-Event System Simulation*. Prentice Hall; Forth Edition, 2010.

Besenyi, G. M., Kaczynski, A. T., Wilhelm, S. A., Vaughan, K. B. 2012. "Demographic variations in observed energy expenditure across park activity areas". Journal of Preventive Medicine, 2013, Volume 54, pg 79-81.

Cochran, W. G. 1971. *Sampling Techniques*. Compañia Editorial Continental.

Díaz, A., Pedraza, C., González, S., Díaz, J., Ibarra, L., Fernández, D., Pratt, M. Jacoby, E., Lozano, O., Almanza, A., Cañon, F., Brownson, R., Sarmiento, O. 2013. "Open Steets: a healthy epidemic. Fact sheet".

Grupo de Epidemiología de la Universidad de los Andes (EpiAndes) and Universidad del Rosario. 2013. *Manual para implementar promocionar y evaluar programas de Vías Activas y Saludables en Colombia.* Departamento Administrativo del Deporte, la Recreación, la Actividad Física y el Aprovechamiento del Tiempo Libre-COLDEPORTES.

Murcia, M., Rivera, M. J., Sarmiento, O. L., Akhavan-Tabatabaei, R. 2013. "A discrete-event simulation model to estimate the number of participants in the Ciclovia program of Bogota". Undergraduate thesis.

Sarmiento, O., Torres, A., Jacoby, E., Pratt. M., Schmid, T. L., Stiereling, G. 2010. "The Ciclovía-Recreativa: A Mass-Recreational Program With Public Health Potential". Journal of Physical Activity and Health, 2010, 7(Suppl 2), S163-S180.

APPENDIX

Appendix 1: Survey for input-data collection in Ciclovia programs

CICLOVÍAS	
ID Encuestador N1 N2 A1 A2	*ID Participante* P # - D D M M -

Estado del tiempo:	Soleado ☐ Despejado ☐ Nublado ☐ Lluvia ☐ Aguacero ☐ Otro, ¿Cuál?________________________
Hora en la que se realiza la entrevista	Horas: __ __ Minutos:__ __

Encuestador: preguntar si la persona ya respondió la encuesta. En caso de no haber respondido, proseguir con la misma.

Encuestador: Buenos días/tardes. Mi nombre es… estoy realizando una encuesta sobre la Ciclovía de Bogotá para un proyecto de investigación de la Universidad de Los Andes. Le agradezco si puede responder unas cortas preguntas. La información que suministre es totalmente confidencial y sólo será utilizada con fines académicos.

1. Sexo:

Masculino	01
Femenino	02

2. Edad: __________ años

Encuestador: pregunte por dónde ingresa el participante. En caso de que desconozca el punto, muéstrele el mapa y ubique con el participante el punto que debe ser registrado. Una vez identificado el punto, complete la dirección en la tabla escribiendo la intersección compuesta del número de calle, carrera, avenida, diagonal o transversal correspondiente.
Ejemplo:
Si un participante indica que ingresa por la Carrera 19 con calle 13, en el espacio correspondiente a carrera escriba el número 19 y en calle el número 13.
Si un participante indica que ingresa por el parque el Tunal, escríbalo en el espacio "Otro".

3. ¿Por qué lugar ingresa usualmente a la Ciclovía?

Calle
Carrera
Transversal
Avenida
Diagonal
Otro

4. El día de hoy, ¿A qué horas llegó a la Ciclovía?
Hora: __ __ Minutos: __ __

5. El días de hoy ¿A qué horas se va a ir de la Ciclovía? (hasta que horas va a estar en la Ciclovía)
Hora: __ __ Minutos: __ __

6. ¿Cuál de las siguientes actividades desarrolla principalmente en la Ciclovía? (marque sólo una opción)

Bicicleta	01
Patines	02
Camina	03
Trota	04
Scooter/monopatín	05
Patineta	06
Otro, ¿Cuál?______________________	

Gracias por su colaboración.

Appendix 2: Format for arrival rate data collection

CONTEO DE TASA DE ENTRADA

Hora Inicial _ _:_ _ Hora Final _ _:_ _

Nombre encuestador ______________ Tel ________

Clima

Soleado ☐ Despejado ☐ Nublado ☐

Llovizna ☐ Aguacero ☐

ID punto ______

Sentido

N-S ☐ E-O ☐

S-N ☐ O-E ☐

Tasa de entrada									
1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110
111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130
131	132	133	134	135	136	137	138	139	140
141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170
171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190
191	192	193	194	195	196	197	198	199	200
Total:									

Appendix 3: Format for flow per activity data collection

CONTEO DE FLUJO DE PARTICIPANTES DE LA CICLOVIA POR ACTIVIDAD

Hora Inicial _ _:_ _ Hora Final _ _:_ _

Clima

Soleado ☐ Despejado ☐

Llovizna ☐ Aguacero ☐

Nublado ☐

N-S ☐ E-O ☐

S-N ☐ O-E ☐

ID del grupo ________

Nombre encuestador ______________

Tel: __________

Montar bicicleta									
1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110
111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130
131	132	133	134	135	136	137	138	139	140
141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170
171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190
191	192	193	194	195	196	197	198	199	200
Total:									

Caminar				
1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30
31	32	33	34	35
36	37	38	39	40
41	42	43	44	45
46	47	48	49	50
51	52	53	54	55
56	57	58	59	60
61	62	63	64	65
66	67	68	69	70
71	72	73	74	75
76	77	78	79	80
81	82	83	84	85
86	87	88	89	90
91	92	93	94	95
96	97	98	99	100
Total:				

Trotar				
1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30
31	32	33	34	35
36	37	38	39	40
41	42	43	44	45
46	47	48	49	50
51	52	53	54	55
56	57	58	59	60
61	62	63	64	65
66	67	68	69	70
71	72	73	74	75
76	77	78	79	80
81	82	83	84	85
86	87	88	89	90
91	92	93	94	95
96	97	98	99	100
Total:				

Otros				
Patines				
Patineta				
Scooter/monopatín				
Silla de ruedas				
Coches				
Pasajero en una bicicleta				
...				
1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30
31	32	33	34	35
36	37	38	39	40
41	42	43	44	45
46	47	48	49	50
51	52	53	54	55
56	57	58	59	60
Total:				