TOOLS TO ASSIST IN DETERMINING BUSINESS VALUES OF INDIVIDUAL MINIBUS-TAXI OPERATIONS IN RUSTENBURG, NORTH-WEST, SOUTH AFRICA

J COETZEE*, A MULLA* and N OOSTHUIZEN*

*GoMetro, South Africa
*MESH Integrated Solutions, *GladAfrica, South Africa*
justin@gometroapp.com, ahmed@mesh-is.co.za, nicola@solvem.co.za

ABSTRACT

A multi-phase municipal informal public transport survey was undertaken in the city of Rustenburg to collect public transport operational and business data of minibus taxi associations running their services within the municipal boundary. The methodology followed consisted of cordon counts coupled to onboard surveys. The onboard surveys were executed by means of the GoMetro Pro mapping application, a mobile application that enables field collectors to digitally map respective transport networks and operations. Operational data was captured over a full operating day by a representative sample of vehicles of the larger population on pre-determined days and periods of the month that were selected in collaboration with the taxi associations, to accurately reflect the monthly operations. The data collected formed part of an operations analysis of the existing minibus-taxi transport system for the spatial and business planning of the Rustenburg Rapid Transport (RRT) Yarona Bus Rapid Transit (BRT) system. Revenue data and operational data per vehicle for a full day of operations was collected by the mobile application, repeated on selected days in a month. The data was inferred by post-processing of the data and applied to individual business models for each sampled minibus taxi, as representative drivers of a value in a system. The individual business models were extrapolated to the population as a representative sample of the population to determine the systems revenue and systems operations plan of an association.

1 INTRODUCTION

In South Africa the majority of people are transported by bus services and paratransit, a large-scale informal public transport service and for the vast majority of cities in the developing world, information on the public transport industry is not easily accessible (Johnson *et al.*, 2016). This is evident from the objectives of the Digital Matatus project, to create General Transit Feed Specification (GTFS) data about minibus taxi operations for commuters of Nairobi, Kenya (Williams *et al.*, 2015), and the surveys that are needed to be conducted to understand the reliability of informal public transport in Accra, Ghana (Johnson *et al.*, 2016).

During the implementation of a new bus rapid transit system, the project team was required to conduct transport surveys that inform the municipality how minibus taxis, serving public transport corridors in Rustenburg, operate and generate revenue. The novel work that this paper would like to present is a mobile application collection tool coupled to a specific technique to use mapping results of a statistically-relevant sample of vehicles, with the aim of providing a tool to assist with determining operational characteristics of formal or informal transport operations. The collection tool and the fundamental technology behind it, has already been presented in the paper by Ndibatya, et al., (2016). From this information a business model of the taxi operation can be inferred for business valuation of minibus taxi operations that would be replaced by the IRPTN operations or provide a basis for the development of a hybrid transport operations systems plan. A hybrid paratransit system sacrifices some of the flexibility of demand-responsive transportation to attain improved productivity and cost savings but retains some of that flexibility to achieve the levels of service necessary for adequate market penetration.

1.1 Conducting Informal Public Transport Surveys

The process of conducting informal public transport surveys consists of the following steps illustrated in figure 1 below:

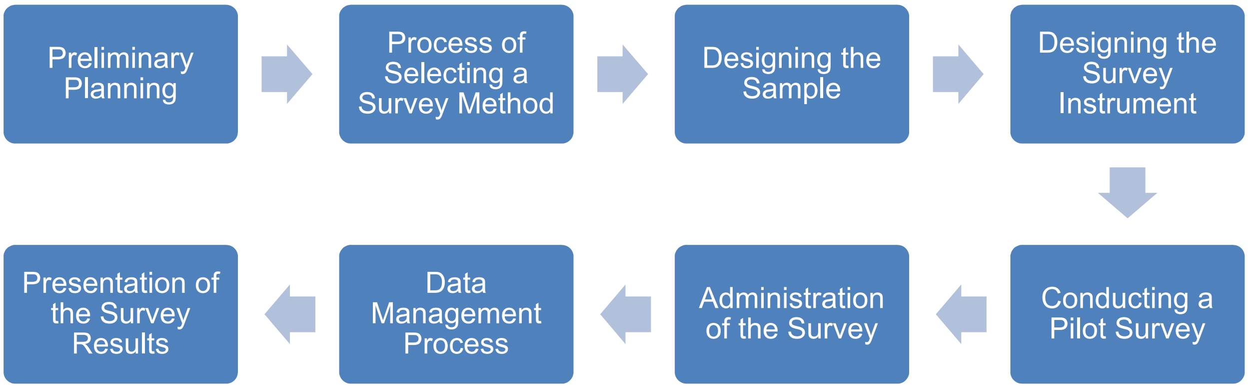

Figure 1: Methodology of Conducting Informal Public Transport Surveys

The methodology presented in this paper focuses on the administration, data management and presentation of survey results.

1.2 Measuring Public Transport Business Operations

The cost of the provision of public transport service consists of capital costs and operational costs (Del Mistro and Aucamp, 2000). The scope of this paper will focus only on the operational costs incurred by public transport services, as this is the dataset that is the most unclear or hard to measure when delivered by informally optimized private operations. Vehicle operating costs consist of fixed and variable costs. The variable costs are a function of the distance travelled by the vehicle or veh-km (Del Mistro and Aucamp, 2000). For a minibus taxi operator, the biggest variable cost in relation to the distances a vehicle travels is the fuel cost in Rand per kilometer (R/km). Without referring to any data, theoretically a minibus-taxi with the optimal ratio between Rand per kilometer is reached when the number of trips completed, and distance travelled per day returns the highest amount of revenue. In section 3.2, graphs are presented that illustrate actual data collected from the Rustenburg project which provide a visual understanding of the minimum, average, and maximum of distances, trips and revenue. The objective of the surveys conducted during this project through this survey methodology was to collect data from the minibus taxi industry and bus operations that would then inform planning authorities, supply data that could allow commuters to conduct trip planning and how the operators can operate more efficiently. A previous paper published by Del Mistro and Berehns (2012) investigates the impact of service type and route length minibus taxi cost model using data inferred from a Household Travel Survey. Although there is reference in utilising Household Travel Survey data to plan for minibus-taxi and bus operations, the data is limited to aggregated household travel behaviours. This paper reports on the survey of a statistically representative sample of a minibus fleet currently running in Rustenburg with data that represents the actual operations of the industry which was confirmed by the minibus-taxi fleet operators and owners.

To accomplish this a survey methodology is presented that allow for large scale data collection necessary to produce results that will inform business models for minibus taxi operations.

2 SURVEY ADMINISTRATION

The success of the RRT IRPTN onboard survey was reliant on successfully implementing certain practices at different phases of the project. The following areas, illustrated in figure 2, of the survey methodology are reported on:

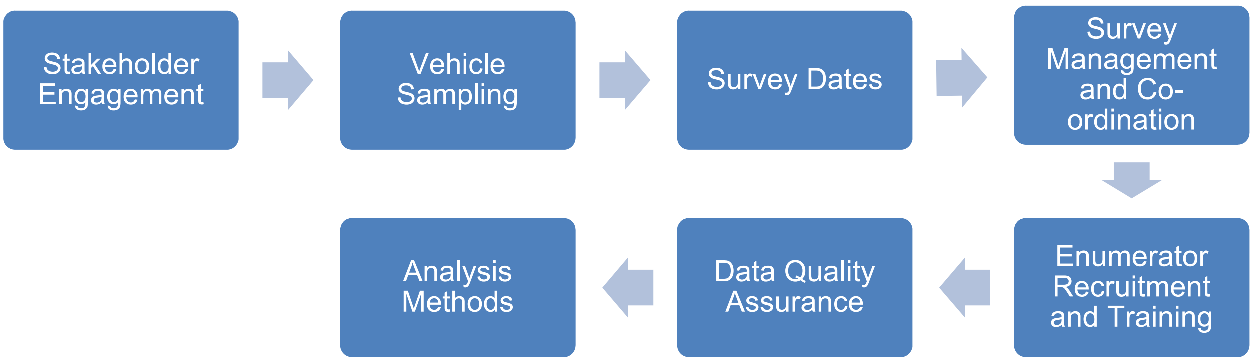

Figure 2: Areas Reported on for the Survey Methodology

2.1 Minibus Taxi Stakeholder Engagement

The minibus-taxi industry has a deep understanding of their operations and therefore are key stakeholders in successfully planning out the surveys to capture an accurate picture of the industry operations. Once agreements were reached, trust was earned, and the taxi associations cooperated with the survey team. They assisted the survey team in gathering insights about what to expect during the survey and were available to assist the survey team overcome challenges in conducting the surveys. Engaging with the taxi industry was one of the most critical parts of planning the survey. This engagement involves comprehensively informing the executive committee of the minibus taxi association about the survey, i.e. why the survey is being conducted and what data will be collected. The trust built during the initial engagements with the associations is important for ensuring that the surveys proceed with full cooperation from daily operational staff of the association and the facilitation by the local authority.

One of the lessons learned was the importance of communication between the survey team and taxi association through the appointed industry representative to solicit information of the taxi associations' day-to-day operations before beginning with the surveys and to work with the drivers of the vehicles and make sure they are aware and supportive of the endeavour. These are insights that can assist with various aspects of data processing of the survey as well as the business modelling. The last lesson learned from the stakeholder engagements was that the results from the data collection should be illustrated and swiftly made available to minibus-taxi owners, which enables for errors in the results to be quickly addressed and trust to be retained. The key advantage here for the GoMetro mapping tool is the fast turnaround and live feedback during and after the survey process.

2.2 Vehicle Sampling

A representative sample size of the total vehicle population of the minibus-taxi industry in Rustenburg is required to adequately capture the variability of the operations. To select a representative sample size from the vehicle population for surveying, a simple random sampling technique was employed. First, the total vehicle population dedicated to serving the corridor in question was determined by a list of registered vehicles listed under both taxi associations operating in the survey area. Sample sizes were then determined for associations that were identified who currently serve within the study area. Additional vehicles were sampled as a safety measure to ensure that the accuracy of the results were within acceptable limits. A total of 363 individual vehicles comprised the population from which a random sample was to be drawn. A sample size calculation was used to calculate an approximate sample size required to yield the specified levels of confidence and accuracy. A sample size of 23 was required to satisfy the confidence level (90%) and margin of error (10%) specified. A margin of error of 10% translates to a 20% confidence interval about the population mean.

Once the design sample size for the association was determined, the appointed supervisor, discussed in the next section, was then required to select willing and available drivers to participate in the survey. The vehicle sampling procedure ensures that enough vehicles are surveyed to capture the variable behavior of the association's daily operational fleet. One of the key outcomes of this exercise was the use of a design sample size, which is typically larger than the minimum sample size. This allows the survey to achieve a sampling rate that results in an increased probability of the required accuracy in the data. Oversampling also proved useful in minimizing driver influence or manipulation. The accuracy of the data is determined by the amount of resources available to the survey. High accuracies require larger samples which results in more mappers, more coordinating staff and more funds to pay them.

2.3 Determining the Survey Calendar

The survey calendar was prepared before the stakeholder engagements and finalised once all stakeholders agreed to the final calendar. The minibus taxi associations were key during this stage of the project, as their insights and knowledge of the industry meant adapting the calendar to correctly represent the industry. The dates that needed to be determined were: dates for screening and the preliminary selection of mappers. The next set of dates that were determined were for the training of mappers, dates for dry run(s) and dates for survey execution. The survey team and taxi associations wanted to capture the variability of the demand and supply of minibus taxi operations. Most of the survey dates are selected to cover all variables in the operating system to accurately portray the daily operations of the minibus taxi industry. The days for surveying selected fell within a midweek time-period (e.g. Monday, Tuesday, Wednesday, Thursday) and weekend days such as a Friday and Saturday. Both sets of time periods were surveyed mid-month and month-end. A noteworthy lesson from having implemented the survey calendar was that, once accepted by stakeholders, the survey calendar acts as an effective survey administration tool and aids in minimising errors (Franklin, S., & Walker, 2003). The survey calendar allowed for sufficient time to recruit mappers from within the communities that fall within the survey study area and provided sufficient time for mappers to be trained. Varying the days surveyed for a taxi association with weekdays and different time periods of a month, ensures that operational variation is adequately captured. This was important for the survey results from this methodology as they were being presented to all stakeholders for the first time and had to be defendable.

2.4 Survey Management and Coordination

During this stage of the survey process it was found that queue marshals are the best candidates for mapper supervisors and that the role of a supervisor requires an individual with excellent reliability with good communication skillsets as continuous liaison between the supervisor, drivers and mappers. A mapper will be prohibited to board a vehicle without communication between the supervisor and driver, which could potentially lead to a delay in the data collection process. Queue marshals coordinate vehicle queues at ranks as part of their day-to-day job, from the first taxi that leaves the rank to the last. In instances where a vehicle being surveyed needs to be replaced, the queue marshal is in the best position to identify a replacement vehicle quickly. A queue marshal also has on-the-ground knowledge that can be incorporated in the planning processes.

The IRPTN surveys were coordinated by two onsite project managers who were based in the study area for the duration of the surveys, and a back-office team based in Cape Town. Each taxi association provided a supervisor on the ground with the primary responsibility to manage the mappers, drivers and assist mapper supervisors where needed. The appointed supervisors from the taxi association managed and monitored the boarding and departure times of the mappers to ensure that the mapping of routes were completed successfully. The industry appointed supervisors provided insights that informed discussions between the city and the minibus taxi association thereby also playing mediation role. The project managers stayed in close communication with the supervisors to address issues timeously and to make sure the survey ran as planned.

2.5 Mapper Recruitment, Training and Management

As part of the agreements that were reached to build trust, fieldworkers for the survey were sourced from residents that were local to the area. Each association was requested to present a list of candidates who reside in areas they respectively serve. As far as was practical each field worker surveyed the minibus taxis serving the areas they reside in. A competency test was applied to each mapper before field work commenced, ensuring proficiency in the application. Training was conducted for all mappers who passed the competency test. The training of the roles and responsibilities for the mappers and further provides a theoretical understanding of the GoMetro Pro App. It is key that the training ensures that the mappers understand the entire survey process to provide a level of background and mappers can feel confident in their skills, which all aids in the accurate collection of robust data (Richardson, Ampt and Meyburg, 1995). The mappers underwent a practical, in-vehicle mapping exercise which was followed by an assessment of their practical capabilities. Depending on the criteria of the assessment process, the best candidates according were selected for the available enumerator positions on the survey team which includes a final selection process that enhances the skills quality of the reserve mappers available.

2.6 Data Collection

Mappers boarded the allocated minibus taxis with a mobile device equipped with the GoMetro Pro App at the start of the operations and continuously throughout the daily operations. Operations of the sampled minibus taxis were surveyed for the entire day. Separate trips were captured and logged for the both the forward and return trips of each individual route, i.e. one trip from the CBD taxi rank to Tlhabane and one trip back from Tlhabane to the CBD taxi rank. Vehicles being surveyed operated from the early morning to the evening. Two mappers captured a vehicle's operations in one day by working two shifts, one in the morning and the other in the evening. The mappers changing shifts would communicate to determine the handover location. This was then communicated to the project managers who would confirm and keep record of the devices being handed over. Due to cooperation between the managers, supervisors and mappers, no devices were lost or damaged during the execution of this survey, which is a remarkable achievement.

The back-office team found that a total of 7950 trips were uploaded during the survey for the IRPTN. At any one point 20 - 30 mappers equipped with devices were in the field collecting data. All the trips collected were successfully uploaded to the back-office with route audits being done on devices, field sheets and the data manager back end tool. The provision of power banks meant that devices could be recharged during the day, which prolonged the battery life of mobile devices in the field, allowing the same device to be used to survey a vehicle for the whole day. It was also found that it is good practice for trips to be uploaded as frequently as possible during the survey day to prevent data from being lost due to device issues. Moreover, trips captured by mappers should remain saved on the phone until uploaded, which allows for trips to be recovered in case a device is damaged or was unable to upload trips.

3 DATA MANAGEMENT

The data captured is uploaded to a server that converts and stores the collected trips to a General Transit Feed Specification (GTFS) data is summarised in the GoMetro Pro cloud software, which is a tool that allows the captured data to be viewed, edited and downloaded in .csv and .kml format. Data is stored on a trip by trip basis, with the following variables:

- Vehicle Registration Number
- TripID
- Route Description
- Company
- Start Time
- Travel Time
- Distance
- Revenue
- Date Mapped
- Discrepancies
- Fares
- Number of Stops
- Passengers

Once downloaded, the data can be summarised further in terms of operational parameters for each vehicle surveyed and analysed to estimate the system's operations for each day surveyed. In table 1 below the totals for each variable acquired through the surveys conducted during this project, is illustrated.

Table 1: Totals of each data variable collected during the survey process of the IRPTN

Variable	Total
Number of trips	7 955
Number of routes	3 865
Number of unique vehicles	488
Number of km	70 464.055 km
Number of passengers	879 643
Number of stops	5 8350

3.1 Data Quality Assurance

Data QA is an iterative process that involves summarising and analysing data on a daily basis. Quality assurance for the data (data QA) was done continuously to ensure that all trips are uploaded and correctly captured. The data QA process consisted of ensuring that the association names and vehicle registrations are correctly captured, ensuring the trip distances and trip durations are within reasonable limits given the trip route (viewable on the cleaner), ensuring that stop data captured for each trip (including passenger numbers and fares for each stop) is correctly captured following a logical sequence checking for discrepancies between the boarding and alighting totals. This has subsequently been corrected in the software through a self-check.

A summary of the number of times a tripID appears in the data assisted the survey team in identifying duplicate trips. Duplicate trips occurred when there were network issues and the upload was attempted again by the mobile application. This has subsequently been improved in the software to manage network issues on the phone side to ensure that duplicate trips do not occur. A summary of trips with extremely high or low trip distances and trip durations indicated a need to adjust the trip's starting and ending points or combining or breaking trips up into more typical origins and destinations. A summary of trips comparing the maximum, minimum and average fare was indicative to the correction and cleaning process of actual fare costs. Using GoMetro Pro, a tool that was optimised to reduce the time is takes to modify trip data on the database directly, was more efficient than using third party spreadsheet software to edit, manipulate and then re-upload the data. Data QA was done daily from the first day of data collection on the field, which meant that the data came in manageable quantities, reducing resources required to ensure high quality data.

3.2 Results of Fleet Analysis - Analysis Methodology

The output of the survey includes the interpreted data in the form of a data warehouse and online mapping report with exportable spreadsheets and shape files. To estimate the association's daily operations, each vehicle's operational data was summarised for all trips captured for every survey day. The operational data that was summarised for each vehicle was: the number of trips per day and the total distance travelled per day, illustrated below in figure 3, and the gross revenue generated per day illustrated in figure 4. Figures 3 and 4 demonstrate the range of theoretically expected means for each day surveyed. These curves were developed using the raw data collected from the on-board surveys and illustrated as a frequency measure.

3.2.1 Trips

The crest of each curve indicates the mean of the data. In figure 3 the mean for the normal distribution of the trip data amounts to 11 trips per day. The range of values recorded were confirmed by each association as a representation of their operations and the number of trips were also confirmed by the rank counts.

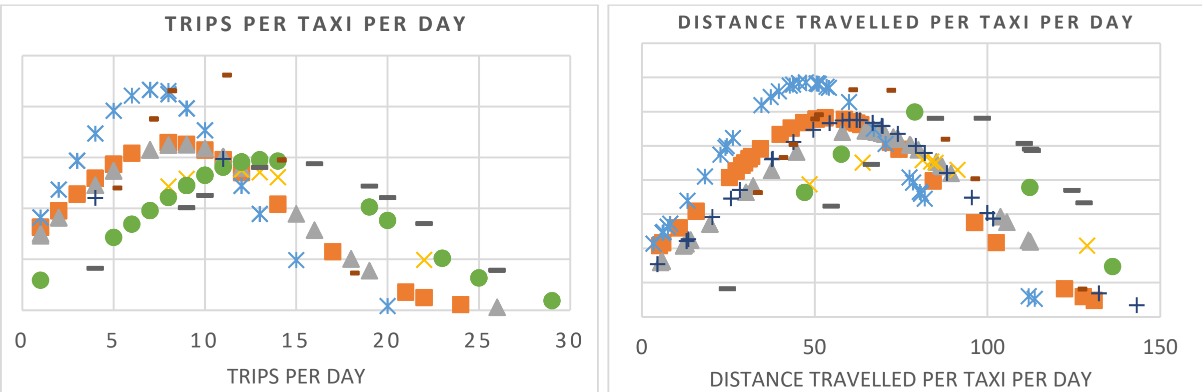

Figure 3: Theoretical normal distribution of number of trips made per vehicle and total distance for each day surveyed

3.2.2 Distance

A key factor that influences the profit margin of the total revenue per day is the total distances (veh-km) travelled per day. The number of trips is related to the distance driven, and so a relationship between revenue and distance can be determined. For the first time the revenue data is plotted on a graph that illustrates the revenue and number of trips made daily by a single minibus taxi operator.

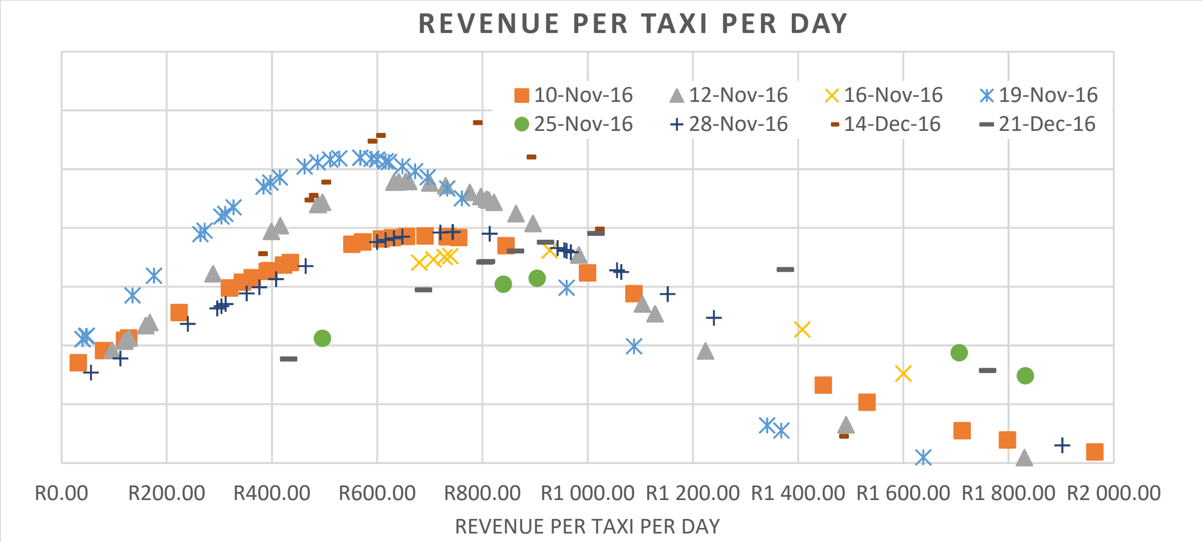

Figure 4: Theoretical normal distribution of revenue made per vehicle for each day surveyed

3.2.3 Revenue

In figure 4 the mean amounts to R784.00 per day given the number of trips made. All the figures illustrated in figures 3 and 4 speak to each other in the sense that the number of trips and distance travelled daily has a direct impact on the revenue an operator makes on a given day. Revenue data is linked to fares collected per trip, one can reflect the daily gross revenue of taxi operators operating within the study area. The outputs created from the GoMetro Pro data cleaner allow for summaries of vehicles to easily be created using basic spreadsheets as the data is exported on a trip basis to .csv. The outputs of the cleaner can also be visualised using GIS software as trip routes and stops can also be exported in .kml or .shp file format. This made it relatively simple for additional operational parameters to be calculated in a post-processing step. One such example is dead distance, which is the distance a vehicle travelled without passengers onboard. This was not initially recorded in the field but was able to be calculated from the data as a post-processing calculation. The analysis used the number of trips the vehicle made per day, the start and end coordinates of every consecutive trip and the minimum distance between the start of $trip_{n+1}$ and end of $trip_{n}$.

The normal distribution of these graphs allows for planning authorities to identify operating norms in the system, which enables planning processes to be better informed and based on actual data of the study area, and not based only on industry knowledge and assumptions. There are many ways to infer the results of the operational data produced by the analysis. It is therefore important that the objective of the survey is used to guide the analysis process to ensure maximum value is derived.

4 CONCLUSION

The advances in mobile technology has made it cost-effective to design an accurate onboard transport survey instrument using a custom-built cellphone application. As mentioned earlier, the survey process has other demands such as engaging with stakeholders, designing the population sample, organising human resources and managing data that must be managed to deliver high quality results in a cost-effective way. Industry engagement and cooperation is important for fostering trust. The survey exercise provides for an opportunity to gain valuable insights about the minibus taxi industry and enables consultants to present results early to stakeholders. Vehicle sampling is important to ensure that the results provided can be used to accurately represent the behaviour of the minibus operations. A well-planned survey calendar creates urgency, a sense of purpose and accomplishment. The management and coordination of supervisors and mappers aided in ensuring that the sampled vehicles are sufficiently surveyed. Recruitment, assessment and training of mappers ensures that data is captured correctly, and less work is required to correct the data at a later stage. Therefore, the conclusion regarding the labour involved in the surveys is that the recruitment and training of mappers relates to the success and accuracy of the mapping execution and data collection phase of the project. The quality assurance process is a vital function in the data process, however it can be time consuming and repetitive. By using tools that are optimised to execute cleaning tasks efficiently, the quality of the data can be improved with minimal resources. The data results produced at the end of this survey process present statistically relevant data on operational information that is required to determine the business values of the respective minibus taxi associations. The next key area of improvement is the process of analysing the data collected to test the viability of a hybrid public transport system that combines both formal and paratransit operators, transforming public transport.

5 REFERENCES

Franklin, S., & Walker, C. (2003) *Survey Methods and Practices*. Ottawa: Minister of Industry, 2010. Available at: http://www.statcan.gc.ca/pub/12-539-x/12-539-x2009001-eng.pdf.

Johnson, A. *et al.* (2016) 'Fickle or Flexible? Assessing Paratransit Reliability with 2 Smartphones in Accra, Ghana', (November), pp. 1–17.

Del Mistro, R. and Aucamp, C. (2000) 'Development of a public transport cost model', (May), pp. 30–32.

Del Mistro, R. and Behrens, R. (2012) 'The Impact of Service Type and Route Length on the Operating Cost Per Passenger and Revenue of Paratransit Operations: Results of a Public Transport Cost Model', *CODATU XV -Le rôle de la mobilité urbaine pour (re)modeler les villes2Conference CODATU XV: The role of urban mobility in (re)shaping cities*, (October), p. 18.

Ndibatya, I., Coetzee, J. and Booysen, M. J. (2016) 'Mapping the Informal Public Transport Network in Kampala With Smartphones : Making Sense of an Organically Evolved Chaotic System in an Emerging City in Sub-Saharan Africa', 2040(Satc 2016), pp. 327–337.

Richardson, A. J., Ampt, E. S. and Meyburg, A. H. (1995) 'Survey Methods for Transport Planning', p. 459.

Williams, S. *et al.* (2015) 'The digital matatu project: Using cell phones to create an open source data for Nairobi's semi-formal bus system', *Journal of Transport Geography*. The Authors, 49, pp. 39–51. doi: 10.1016/j.jtrangeo.2015.10.005.