Article

It's Only a 5-minute Walk from Here: Comparing Respondent Travel Distance Perceptions in Social Surveys to Alternative Measures

Asia Pacific Journal of Evaluation,
Vol. 01, Issue 01,
January : 2023

Morchan Karthick[1] **Shruti Kakade**[1]
Nelson Mathews[1] **Kartikeya Bhatotia**[1]

Abstract

Self-reported information on individuals' access to infrastructure and services relevant to distance and travel time is commonly recorded in social surveys. Accuracy issues and rounding-off errors afflict the data obtained through this technique. This study explores the use of Google Places API as an alternate approach for estimating the distance and travel time between respondents' locations and the closest infrastructural facilities. Along with socio-demographic information, we gathered self-reported data on the time taken to access the closest amenities from 1,238 entrepreneurs in four districts of India. Using the Google Places API, we determined the distance and time it took to go from the business site GPS coordinates collected during the survey to the closest infrastructural facilities and compared both sets of data. A majority of respondents overestimated the time to reach the facilities, whereas one in four provided accurate figures. The results indicate that self-reported data on time and distance are frequently inaccurate and subject to rounding-off biases. We recommend that time and distance information be estimated via alternate data sources to offer precise real-time data for improved decision making.

Key Words: GPS; Self-reporting; Travel surveys; Google Places API

Introduction

Modern, efficient and reliable infrastructure is the backbone of a healthy economy. Infrastructure links urban and rural residents to higher-quality employment, education, and healthcare options. However, most studies concentrate on investigating the impact of supply side determinants on access to employment, with little importance attached to concerns including access to infrastructure (Lei, Desai, & Vanneman, 2019). Some studies suggest that infrastructure investments benefit households, reduce poverty, and boost economic growth (Briceno-Garmendia, Estache, & Shafik, 2004) (Kumari & Sharma, 2017). Over the years, India has gathered infrastructure-related spatial data in a methodical manner, yet data and applications are seldom shared with consumers and the private sector. Therefore, researchers rely on the respondents' subjective and perceived answers to determine the distance and travel time between their location and infrastructural facilities.

[1] LEAD at Krea University, IIT-Madras Research Park, Kanagam, Chennai, India

Perceptions play a major role in decision-making process. For instance, distance perceptions are strongly linked to social and personal significance. A person prone to ailment is more likely than a healthy individual to appropriately estimate hospital distance (MacIntyre et al., 2008). Environment-related perceptions are also influenced by lifestyle behaviours, beliefs, and cultural values. For instance, the accuracy of a "usual" travel route is projected to be greater than that of a non-usual route, which is prone to high inaccuracies owing to a lack of familiarity.

Historically, social science surveys have heavily relied on individual responses to determine the spatial disposition of individuals, households, communities and businesses. A popular line of questioning is "how far" or "how near" a facility, place or resource is from an individual's reference point, usually their household or workplace. These self-reported distance and access records are employed as a valid proxy in a variety of analyses. However, issues like poor recall, adherence, and personal judgment are likely to render self-reported data inaccurate (Stopher P. F., 2011), (Bricka et al., 2012). Advances in Geographical Information Systems (GIS) and geoinformatics have resulted in the emergence of approaches for investigating or replacing these proxy indicators. The Google Maps Application Programming Interface (API) is one of the applications used to determine origin-destination travel time based on a real-time updated transportation network (Wang & Xu, 2011). Numerous studies (Ribeiro et al., 2014), (Kelly et al., 2013), (MacIntyre et al., 2008), (Janz, 2006) , (Stopher P S. L., 2007) have demonstrated significant variability in self-reported distances and actual, objective distances over the years. According to the studies (Stopher & Bullock, 2002), (Stopher P F. C., 2009), and (Hallo J. C., 2005), GPS devices may offer accurate location and temporal data and can be used to adjust the time and distance stated by survey respondents. Self-reported proxy distances in social science surveys are consistently under or over-reported when compared to GPS-recorded distances, depending on where one is located. Emerging research has examined the basis of people's judgements about distances and where they locate resources.

Another scope of variability in distance perception occurs due to specification errors and fundamentally unpredictable human behaviour while reporting distances. An example of this is the "rounding off" phenomenon - a systematic bias where respondents themselves introduce a high margin of error in distances. According to Rietveld (2001), rounding-off travel time is a prevalent issue in survey data, resulting in biases in the estimation of average transit times based on travel surveys. Outliers and rounding-off errors were found in self-reported distances and the accuracy of self-reporting is affected by socio-demographic factors, transportation mode used and the characteristics of the trip (Witlox, 2007), (Forrest TL, 2005), (Vanwolleghem, 2016), (Marcelle D, 2014).

Diverse human variables contribute to the variance between self-reported perception and objective distance metrics. Therefore, it is imperative to investigate why the variability exists and what strategies might be deployed to eliminate analytical inaccuracies. Investigating the variables that lead to the variability between objective and self-reported distance assessments might help to improve spatial approaches in the social sciences.

In this paper, we looked at a recent enterprise survey that we had conducted, in which we captured participants' self-reported distances and geo-locations. We compared self-reported distances to more objective distance values obtained by the Google API. We investigate the reliability of self-reported measures, and the factors that are found to be significant while studying the variability of self-reported and objective measures.

Methods

There were two primary components to the research. The first component was a quantitative survey conducted with enterprise owners to understand entrepreneur and enterprise characteristics and collect self-reported data on the time taken to reach major infrastructure facilities from the enterprise location. GPS coordinates were also acquired for the enterprise location data. In the second component, the distance and time taken to reach the major infrastructure points was calculated from Google Maps using the Location API services of Google.

Female Labour Force Participation-Enterprise Survey

The survey's main respondents were enterprise owners with at least one full-time salaried employee. The survey was conducted in-person in CAPI mode administered with the help of tablets. A state (Karnataka), with a high number of enterprises per 100,000 population and another (Jharkhand) with a low number of enterprises per 100,000 population were selected for the study based on the Economic Census 2013-14. On the basis of urban population ratio and total population, the districts of the two states were segregated into urban and rural districts, and one district from each stratum was picked at random. Between December 2021 and February 2022, we conducted surveys with 1,238 enterprise owners. The field data collection exercise was concluded in the districts of Bangalore and Mandya in the state of Karnataka, and Dhanbad and Garhwa in the state of Jharkhand. The sampling frame for the study comprised the District Industries Center's (DIC's) list of businesses with fewer than 100 workers. Data from the 2013-14 Economic Census was used to estimate the proportion of Production/Service/Trading enterprises in each district. The enterprises were chosen at random from the DIC list in the same proportion of Production/Service/Trading in the Economic Census 2013-14.

We surveyed enterprise owners about employment history, recruitment practices and hiring choices. We also collected data on infrastructural amenities accessible to respondents at work and the distance to the closest facilities by foot. Although owners may have used various means of transport (car/bike/walking/public transport), we requested everyone to input walking time to preserve consistency. The data was collected for infrastructural facilities like bus stops, railway stations, bank branches, and hospitals. Finally, we collected data on geo-locations of the enterprise through the CAPI survey instrument.

Google Places API Data Integration

We combined the survey data with the Google Places API data using the Distance Matrix API technique in Python**.** The Google API (Application Programming Interface) enables users to scrape data from the Google cloud platform services, including Google Maps. The Places API is a service that responds to HTTP queries with information about locations. This API defines places as establishments, geographic locations, and notable points of interest.

The Google API's nearby search function delivers a list of all instances of the specified keyword (bus stop, railway station, hospital, or bank) within a fixed radius. The API's free version limits each request to 60 results. Since the maximum number of results per request is limited, we chose to conduct the search across a narrower geographical region. For each district, we determined the GPS coordinates of all the block headquarters and incorporated them into the nearby search tool. Using the retrieved information, we were able to calculate the average walking time (in minutes) and the average distance (in kilometres) for our respondents to the closest facilities from their company sites. For example, out of 10 hospitals available for our respondent within a radius of 5 kms, the hospital closest to the enterprise location was selected.

The first step for using Google API services is to sign-up and get an API key. An API key is a unique code associated with the project. Creating an API key requires billing and payment card information. Google offers a specified quantity of free pull requests on a daily basis, adequate for a social survey sample. Request-related quota limits and fees need to be well acknowledged. Python and R provide a range of packages for sending requests to Google APIs and storing the results. A number of documentations are available online to refer for the actual codes to be used. A team member with working knowledge in Python or R is required for this approach.

We computed the difference between the self-reported time that was collected from survey data (A) and the time taken data that was generated from Google API data (B) in order to evaluate the accuracy of the time that respondents said it took them to reach the closest facilities. This was done with the understanding that Google API data provides precise distances between two sites. The time difference (T) was calculated via the difference between Google API data (A) and self-reported data (B). The time difference data was subsequently classified into three categories: accurate, underestimation and overestimation.

T (Time Difference) = A (Google API data) – B (Self-Reported data)

Accurate reporting	Difference between API time and self-reported time is between -5 to 5 mins
Overestimating	Difference between API time and self-reported time is greater than 5 mins
Underestimating	Difference between API time and self-reported time is less than -5 mins

Ethics Approval and Consent of Participants

The study was approved by the Institutional Committee for Ethics of Institute of Financial Management and Research, Chennai, India. Informed consent was obtained from each agreeing participant before interviews, after explaining the details of the study in a language that they could understand.

Statistical Analysis

Descriptive analysis was carried out to understand the demographic and financial characteristics of respondents who reported accurately, underestimated and overestimated time taken to walk to nearby facilities. We further ran regression analysis to understand what entrepreneur and enterprise-related factors affect the accuracy of the time reported by the respondents. We used ordinal logit regression with self-reporting accuracy as the dependent variable (under-estimation and over-estimation was clubbed into the inaccurate category). STATA and R were used for all descriptive and correlating analyses.

Results and Discussion

The present analysis utilized data from 1238 enterprises belonging to the districts of Bangalore, Dhanbad, Garhwa and Mandya. Table 1 depicts the socio-demographic information of the entrepreneurs and enterprise characteristics. In our sample, 9% of the respondents are women entrepreneurs and 91% of the respondents are men. The proportion of women entrepreneurs in the sample is slightly less than the proportion at the national level which stands at 14% (NSSO, 2013-14). The average age of male respondents was 41 and female respondents was 39. In terms of years of operation, 7% of enterprises have been operating for less than three years, 63% between 3 to 10 years and 30% above 10 years. The bulk of enterprises hailed from the services sector (38%), followed by production (33%), and trading (29%). The surveyed enterprises operated mainly out of local markets (43%) followed by major markets (40%). On average, enterprises employed five full-time paid employees. Retail trade, manufacturing of food products, personal services, manufacturing of textiles and wholesale trade were major business activities carried out by the respondents.

Characteristic	Category	Accurate	
		Frequency	Percent
District	Bangalore	303	24.47%
	Dhanbad	316	25.53%
	Garhwa	313	25.28%
	Mandya	306	24.72%
Gender of the entrepreneur	Female	111	8.97%
	Male	1127	91.03%
Age of the entrepreneur	<= 30	188	15.19%
	31-40	200	16.16%
	41-50	494	39.90%
	> 50	356	28.76%
Age of the enterprise	Less than 3 years	86	6.95%
	3 to 5 years	395	31.91%
	6 to 10 years	385	31.10%
	11-20 years	224	18.09%
	More than 20 years	148	11.95%
Sector	Production	414	33.44%
	Services	467	37.72%
	Trading	357	28.84%

Characteristic	Category	Accurate	
		Frequency	Percent
Enterprise registered on UDYAM platform	Yes	841	67.93%
	No	397	32.07%
Location of the enterprise	Major market	489	39.50%
	Local market	527	42.57%
	Stand alone	63	5.09%
	Home-based	158	12.76%

Table 1: Entrepreneur and enterprise characteristics of the sample

We compared the time taken to reach to the nearest infrastructure facilities based on self-reporting and Google API. The average time taken (in minutes) to walk to the nearest bus stop from the enterprise location based on self-reporting was 22.87 (S.D = 35.03) compared to 27.69 (S.D = 34.09) estimated from Google API. The time taken to reach the nearest railway station based on self-reporting was 56.98 (S.D = 59.51) compared to 35.76 (S.D = 45.12) estimated from Google API. The time taken to reach the nearest bank branch based on self-reporting was 12.70 (S.D = 15.64) compared to 24.56 (S.D = 37.82) estimated from Google API. Similarly, the time taken to reach the nearest hospital based on self-reporting was 22.06 (S.D = 30.22) compared to 21.83 (S.D = 35.61) estimated from Google API. Based on the Google API data, the average distance to the nearest bus stop was 1.7 kilometres, followed by 1.5 kilometres to the bank, 2.2 kilometres to the railway station and 1.3 kilometres to the hospital. The difference was observed to be higher for the railway station compared to the other three infrastructure facilities. The density plots to compare of the distribution of travel time between self-reported and API data is presented in Figure 1.

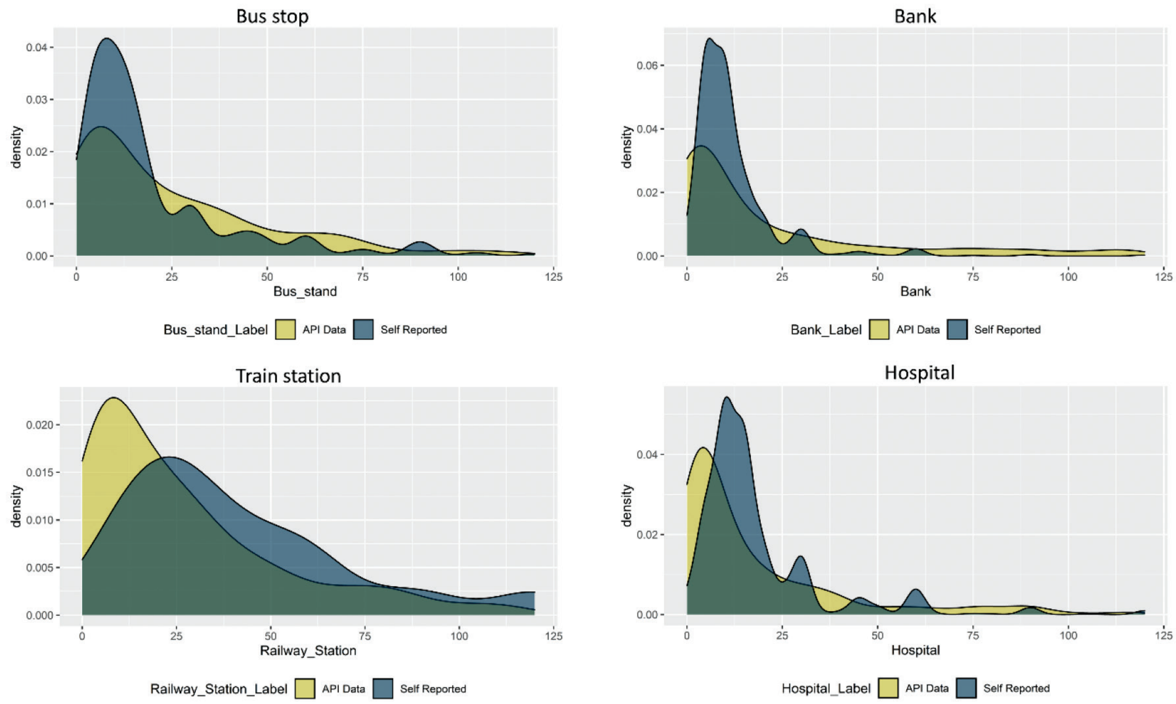

Figure 1: Comparison of travel time - self-reported vs API data

For all four infrastructure facilities, the proportion of respondents that over-estimated the time taken were higher. This is consistent with prior research results analysed in this systematic review (Kelly et al., 2013). At the overall level (mean difference in time taken for all four infrastructure facilities), 25% of respondents reported accurately, 45% over-estimated, and 30% under-estimated; there was a variance in accuracy between infrastructure points. As shown in Figure 2, 20% respondents accurately estimated the time taken to reach the closest bus stop, whereas 49% over-estimated and 31% under-estimated. For the nearest train station, 13% of respondents reported accurately, while 71% over-estimated and 16% under-estimated. For nearest bank branch, 36% respondents reported accurately while 39% over-estimated and 25% under-estimated. For the time taken to reach the nearest hospital, 21% reported accurately while 55% over-estimated and 24% under-estimated. Entrepreneurs are likely to visit banks frequently for business purposes and less likely to visit railway station more often. Out of the four facilities, higher proportion of respondents were able to accurately report the time taken for reaching the nearest bank and the proportion was lowest for time taken to reach the nearest railway station.

Rounding-off is a regular occurrence in self-reported data. Literature acknowledges the rounding-off dilemma, but largely disregards it as a chance occurrence. (Serfling, 2006) found that rounding in survey data does not occur at random but is correlated with interview situation characteristics and the persons involved. It is commonly noted that people think and report both times and distances in rounded-off numbers (Witlox, 2007). We analysed the self-reported figures to check for signs of rounding off. When compared against the API data, we find that the self-reported numbers tend to be multiples of 5, 10 and 15, which are typically 'anchor points' (Rietveld, 2001) represented

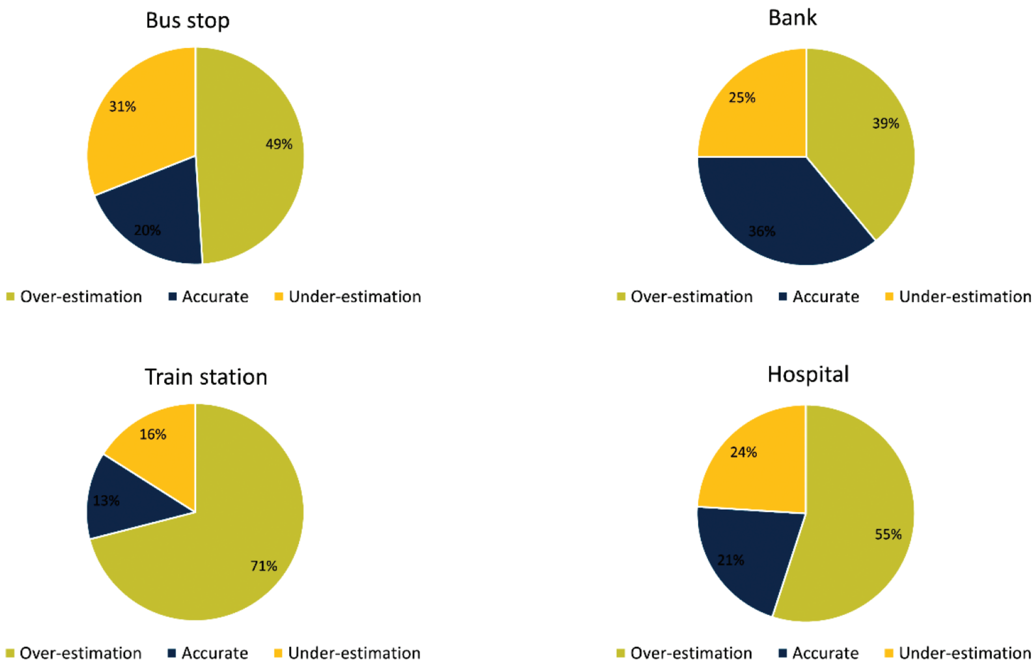

Figure 2: Accuracy of self-reporting for the four infrastructure facilities

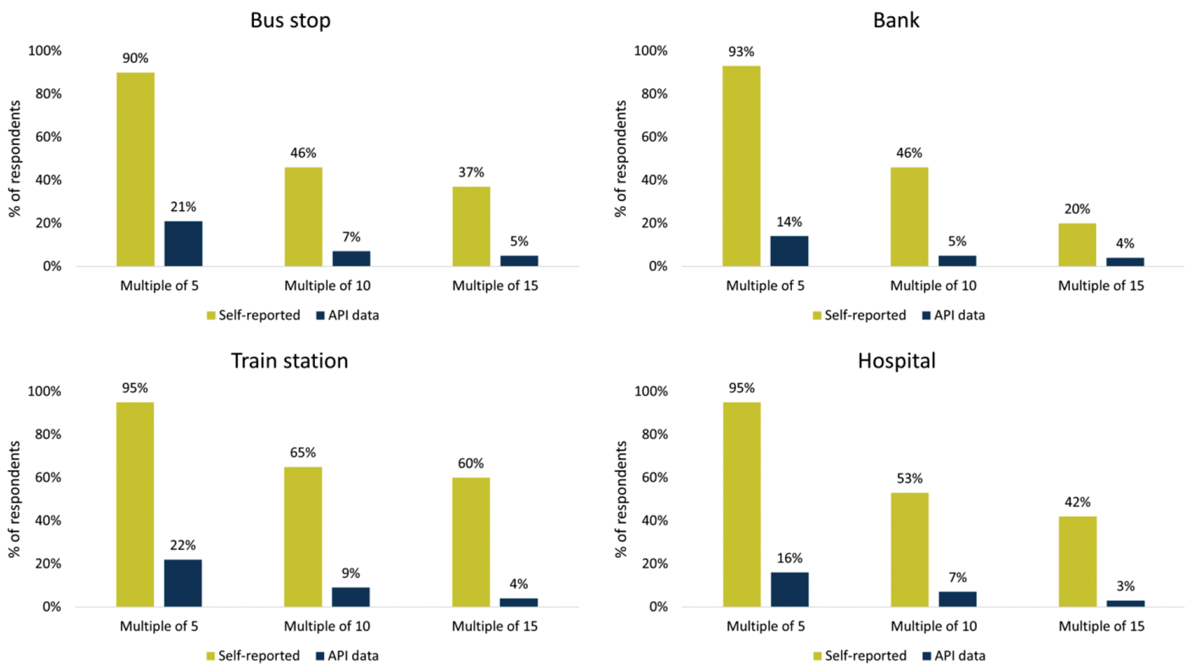

Figure 3: Comparison of travel time in multiples of 5,10 and 15

on modern clocks. More than 90% of responses are rounded off to multiples of 5. Note that these round-offs may co-exist with the general inaccuracy of distances as well.

The proportion of respondents who indicated travel times in multiples of 5, 10, and 15 has been compared to the API data in Figure 3.

For all the four infrastructure facilities, 90% or more respondents reported the time it took to reach them in multiples of five, whereas the percentage according to the Google API data was approximately 20% (21% for bus stop, 22% for railway station, 14% for bank branch, and 16% for hospitals). When examining the time taken in multiples of 10 and 15, similar patterns may be noticed for all infrastructure facilities. Rounding-off in surveys may reflect the interest and motivation of respondents to answer distance and time questions. This motivation concern must be addressed, while alternative approaches for estimating travel distance must be investigated.

Table 2 depicts the variation in the difference in reporting across key entrepreneur and enterprise characteristics. About 9% of the respondents in the study were female entrepreneurs while their proportion was significantly less among respondents who overestimated the time. Very few studies have studied the effect of gender of the respondent on estimation accuracy and found a positive yet non-significant correlation with accurate reporting (Fillekes et al, 2019). The respondents' mean age was 41 years. Existing literature (Kelly et al., 2013) suggests age of the person to be a significant factor in accurate reporting but we did not find any significant variation in age groups across the accuracy categories. This might be because the respondents are a homogenous group (enterprise owners) compared to the heterogenous nature of people in the general population.

Characteristic	Category	Accurate		Over-estimation		Under-estimation		p-Value
		Frequency	Percent (%)	Frequency	Percent (%)	Frequency	Percent (%)	
Gender	Male	35	11.51%	36	6.44%	40	10.67%	**0.0173***
	Female	269	88.49%	523	93.56%	335	89.33%	
Age of the entrepreneur	<= 30	39	12.83%	92	16.46%	57	15.20%	0.7411
	31-40	55	18.09%	89	15.92%	56	14.93%	
	41-50	119	39.14%	225	40.25%	150	40.00%	
	> 50	91	29.93%	153	27.37%	112	29.87%	
Sector	Production	94	30.82%	133	23.79%	187	49.87%	**<0.0001***
	Services	141	46.23%	182	32.56%	144	38.40%	
	Trading	69	22.62%	244	43.65%	44	11.73%	
Location of the enterprise	Major market	111	36.51%	301	53.94%	77	20.53%	**<0.0001***
	Local market	147	48.36%	175	31.36%	205	54.67%	
	Stand alone	9	2.96%	8	1.43%	46	12.27%	
	Home-based	37	12.17%	74	13.26%	47	12.53%	

Table 2: Entrepreneur and enterprise characteristics and accuracy of reporting
*Significant codes: '***' - 0.001 '**' - 0.01 '*' 0.05 '.' 0.1*

A higher proportion of entrepreneurs running production-related enterprises tend to underestimate the time taken to reach various infrastructure facilities while a higher proportion of entrepreneurs running trading enterprises tend to overestimate the time. The relationship between the enterprise sector and location of the enterprise is an important factor in understanding this variation. As depicted in Figure 4, trading enterprises are likely to be located in main market areas, while production-related enterprises are usually situated in local markets that are specialised for their products. This fact is further augmented in Figure 5 where we depict API distance to different infrastructure facilities based on nature of the enterprise. Production and service-related enterprises had a flatter bell curve compared to trading-related enterprises.

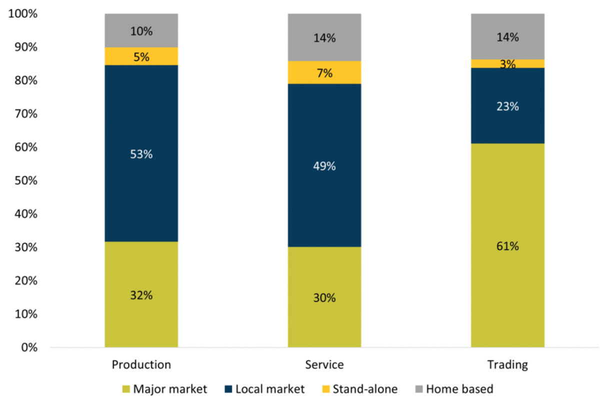

Figure 4: Location vs Sector of the enterprise

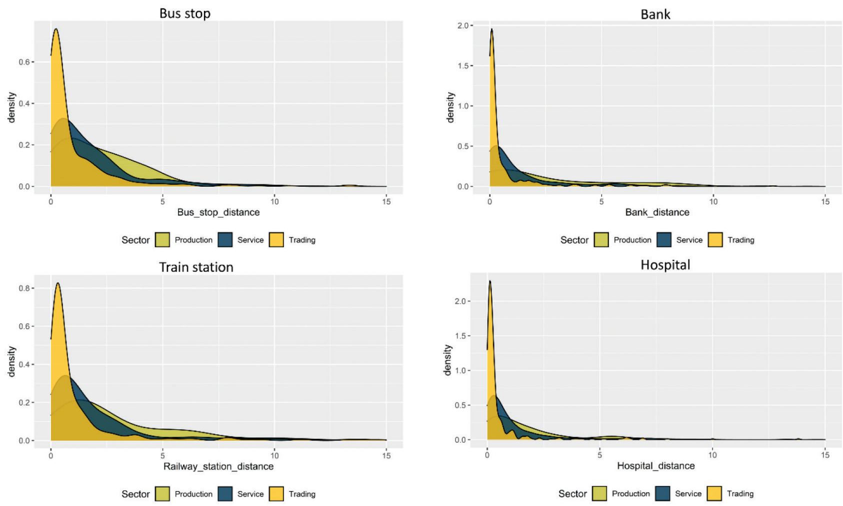

Figure 5: Comparison of API distance to various infrastructure points for different sectors

We compared the actual distance of the enterprises to different infrastructure facilities obtained through Google API to the reporting accuracy of the respondents. For all infrastructure-related facilities, it seems that entrepreneurs are more likely to report either accurately or overestimate the time taken if the infrastructure is closer to their enterprise location. They are more likely to underestimate the time taken if the infrastructure is located farther away from their enterprise. The scatter plots in Figure 6 depict the relationship between the Google API distance to each infrastructure facility and the difference between self-reported and API time taken to reach them. This is an important takeaway especially for surveys conducted in rural areas where most of the infrastructure points are located at a greater distance. Respondents may underestimate the distance and time taken which have to be accounted for and corrected during analysis. We tried to estimate the significance of this relationship by conducting associational analysis. (Witlox, 2007) suggests that the mode of transport used by the respondents for daily commute is an important factor associated with the accuracy of the self-reported data. We will explore this relationship in our future studies as we did not capture this information in our study.

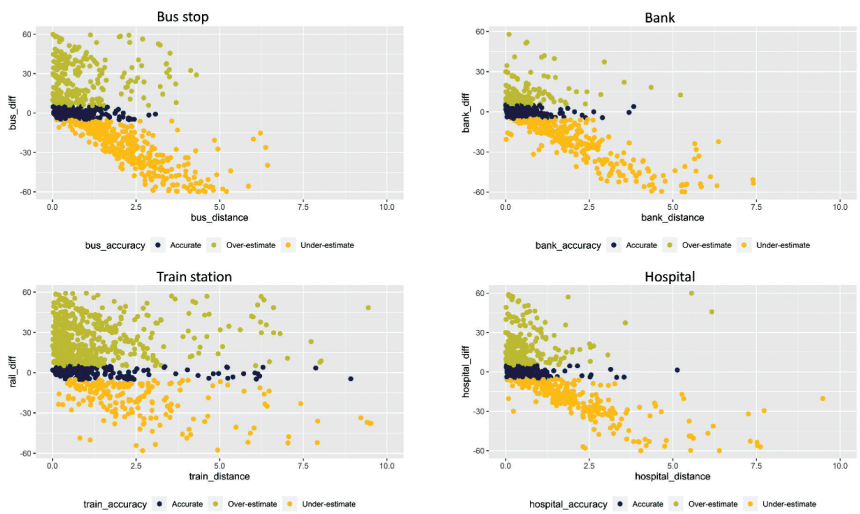

Figure 6: API distance to infrastructure facilities vs accuracy of reporting

We conducted ordinal logit regression and estimated the Odds Ratio (OR) to understand what factors are associated with respondents accurately reporting the time taken to reach the nearest facilities. Respondents owning service-related enterprises are more likely to accurately report time taken compared to respondents owning production related enterprises at 90% confidence level. On the contrary, respondents owning trading-related enterprises are less likely to accurately report time taken than respondents owning production-related enterprises at 90% confidence level. The higher the distance to the nearest bank branch from the enterprise location, the respondents are less likely to accurately report the time taken at 95% confidence level. Similarly, higher the distance to the nearest hospital from the enterprise location, the respondents are less likely to accurately report the time taken at 90% confidence level.

Estimating accurate distance and time taken data is critical for policy-level decision-making on transportation models and other infrastructure planning. Developing trip specific models for carpooling with accurate distance data can help in developing carpooling promotion policies and strategies (Liu et al., 2019). Understanding community mobility through accurate estimation of different types of destinations and time spent out of the house can help in improving physical health, mental health, and overall quality of life (Hallo et al., 2004). There are not many studies conducted in India to understand the accuracy of self-reporting of respondents and exploring other accurate methods to estimate distance and time taken data. Capturing accurate access of infrastructure data is critical for developing policy inputs to address development challenges in the areas of employment, entrepreneurship, and women's economic empowerment.

Variables	Categories	Estimate	*p* Value	Accurate (Ref = Inaccurate)	
				OR	95% CI
Age of the entrepreneur		0.007	0.293	1.007	0.994 – 1.021
Gender of the entrepreneur (Ref = Female)	Male	-0.342	0.132	0.710	0.455 – 1.108
Age of the enterprise		0.007	0.324	1.007	0.993 – 1.021
Industry type (Ref = Production)	Services	0.298	0.070 .	1.347	0.976 – 1.858
	Trading	-0.322	0.099 .	0.725	0.495 – 1.062
UDYAM registration (Ref = No)	Yes	0.010	0.946	1.010	0.757 – 1.348
Location (Ref = Home based)	Major market	-0.102	0.652	0.903	0.579 – 1.407
	Local market	0.240	0.272	1.272	0.828 – 1.954
	Stand-alone shop	-0.588	0.156	0.556	0.247 – 1.407
API distance to nearest bus stop		-0.015	0.783	0.985	0.885 – 1.096
API distance to nearest train station		0.043	0.322	1.044	0.959 – 1.137
API distance to nearest bank branch		-0.125	0.014 *	0.883	0.799 – 0.975
API distance to nearest hospital		-0.111	0.054 .	0.895	0.799 – 1.002

Table 3: Factors associated with over-estimation and under-estimation of time taken

*Significant codes: '***' - 0.001 '**' - 0.01 '*' 0.05 '.' 0.1*

Conclusion

Our study contributes to the literature by looking at self-reporting discrepancies in the context of India, a geography which is largely understudied in this aspect. We have proposed a methodology for accurately estimating distance and travel time using Google API, whose applications extend well beyond the realm of social research. In addition, we have investigated the associations between self-reporting accuracies and the characteristics of entrepreneurs and their enterprises.

Future research is required to study the self-reporting accuracies among the general population in India to understand what specific socio-demographic factors like age, education and health status of the individual influence the perceptions and assumptions of time and distance. Google APIs have wide range of applications and features beyond estimating travel time and distance. These applications need to be explored to leverage their use in the social sector.

Easy access to infrastructure and its use is key to economic development in both urban and rural areas. Policymakers require precise, real-time infrastructure data in order to build optimal solutions while working with constrained resources. Incorporating alternative data sources to supplement primary research is essential for achieving widespread social impact by leveraging the latest technological advancements.

Acknowledgement

The authors thank the entrepreneurs who consented and participated in the study. The authors acknowledge the tireless contribution of enumerators, field executives and field managers who completed the data collection in a timely manner during the COVID-19 pandemic. We also thank Mridulya Narasimhan, Debdulal Thakur, and Sijo John for their inputs on the study and the manuscript.

Declaration of conflicting interest

The authors have declared that no competing interests exist.

References

Briceno-Garmendia, C., Estache, A., & Shafik, N. (2004). Infrastructure Services in Developing Countries: Access, Quality, Costs and Policy Reform. World Bank Policy Research Working Paper 3468.

Bricka, S. G., Paleti, R., & Bhat, C. (2012). An analysis of the factors influencing differences in self-reported and GPS-recorded trips. Transportation Research Part C: Emerging Technologies, 67-88.

Curl, A., Nelson, J. D., & Anable, J. (2015). Same question, different answer: A comparison of GIS-based journey time accessibility with self-reported measures from the National Travel Survey in England. Computers, Environment and Urban Systems, 86-97.

Fillekes, M. P., Rocke, C., Katana, M., & Weibel, R. (2019). Self-reported versus GPS-derived indicators of daily mobility in a sample of healthy older adults. Social Science and Medicine, 193-202.

Forrest TL, P. D. (2005). Comparison of Trip Determination Methods in Household Travel Surveys Enhanced by a Global Positioning System. Transportation Research Record, 63-71.

Hallo, J. C. (2005). A Case Study Comparison of Visitor Self-reported Travel Routes and GPS Recorded Travel Routes. Northeastern Recreation Research Symposium. U.D Department of Agriculture.

Hallo, J. C., Manning, R. C., Valliere, W., & Budruk, M. (2004). A case study comparison of visitor self-reported and GPS recorded travel routes. Northeastern Recreation Research Symposium. USDA.

Janz, K. F. (2006). Physical activity in epidemiology: moving from questionnaire to objective measurement. British Journal of Sports Medicine.

Kelly, P., Krenn, P., Titze, S., Stopher, P., & Foster, C. (2013). Quantifying the Difference Between Self-Reported and Global Positioning Systems-Measured Journey Durations: A Systematic Review. Transport Reviews, 33(4), 443-459. From https://www.tandfonline.com/actionshow CitFormats?doi=10.1080%2F01441647.2013.815288

Kumari, A., & Sharma, A. K. (2017). Physical & social infrastructure in India & its relationship with economic development. World Development Perspectives, 30-33.

Lei, L., Desai, S., & Vanneman, R. (2019). The Impact of Transportation Infrastructure on Women's Employment in India. *Fem Econ*.

Liu, X., Yan, X., Liu, F., Wang, R., & Leng, Y. (2019). A trip-specific model for fuel saving estimation and subsidy policy making of carpooling based on empirical data. Applied Energy, 295-311.

Macintyre, S., Macdonald, L., & Ellaway, A. (2008). Lack of agreement between measured and self-reported distance from public green parks in Glasgow, Scotland. International Journal of Behavioral Nutrition and Physical Activity, 5 (26). https://ijbnpa.biomedcentral.com/articles/10.1186/1479-5868-5-26

Marcelle D, R. e. (2014). Influence of GPS and self-reported data in travel demand models. Procedia-Social and Behavioral Sciences, 467- 476.

NSSO, M. (2013-14). Economic Census.

Ribeiro, M. D., Larranaga, A. M., Arellana, J., & Cybis, H. B. (2014). Influence of GPS and self-reported data in travel demand models. Procedia - Social and Behavioral Sciences, 467-476.

Rietveld, P., & Universiteit, V. (2001). Rounding of Arrival and Departure Times in Travel Surveys: An Interpretation in Terms of Scheduled Activities. Tinbergen Institute Discussion Paper, No. 01-110/3, 01-110/3(..). https://www.econstor.eu/handle/10419/85888

Serfling, O. (2006). Rounding behavior of respondents in household surveys - Evidence from Swiss data. University of Basel.

Stopher P, F. C. (2009). National Academis. https://trid.trb.org/view/809805

Stopher P, S. L. (2007). In-Depth Comparison of Global Positioning System and Diary Records. Transportation Research Record.

Stopher, P. F. (2011). Assessing the accuracy of the Sydney Household Travel Survey with GPS. Transportation, 723–741.

Stopher, P., & Bullock, P. (2002). Exploring the Use of Passive GPS Devices to Measure Travel. International Conference on Applications of Advanced Technologies in Transportation Engineering. Institute of Transport Studies. https://www.researchgate.net/publication/238740813_Exploring_the_Use_of_Passive_GPS_Devices_to_Measure_Travel

Vanwolleghem, G. S. (2016). Children's GPS-determined versus self-reported transport in leisure time and associations with parental perceptions of the neighborhood environment. Int J Health Geogr, 15-16.

Wang, F., & Xu, Y. (2011). Estimating O–D travel time matrix by Google Maps API: implementation, advantages, and implications. Annals of GIS, 199-209.

Witlox, F. (2007, May). Evaluating the reliability of reported distance data in urban travel behaviour analysis. Journal of Transport Geography, 15(3), 172-183. https://www.sciencedirect.com/science/article/abs/pii/S0966692306000305

Appendix 1

Sample python scripts from the Google API exercise for extracting details of bus stops within 20 kilometres radius of a block headquarters

```
“source”: [
   “import requests\n”,
   “import json\n”,
   “import pandas as pd\n”,
   “assembly = pd.read_excel(r\”C:\\Users\\Shruti_Kakade\\Downloads\\data\\assembly.xlsx-
\”)\n”,
   “l = ‘bus_station’\n”,
   “a = assembly[‘Assemblyname’].values[1]\n”,
   “\n”,
   “radius = ‘20000’\n”,
   “api_key = ‘AIzaSyBt66keM717RRNBbnch47JiWqYPiPSm_vE’\n”,
   “#api_key = ‘AIzaSyBkc15b88TuRItOZfwC6pbqL4SrbXCnn2E’ #insert your Places API\n”,
   “location = str(assembly[assembly[‘Assemblyname’]==a].values[0][1])+ str(‘,’) + str(assem-
bly[assembly[‘Assemblyname’]==a].values[0][2])\n”,
   “types = [l]\n”,
   “URL = \”https://maps.googleapis.com/maps/api/place/nearbysearch/json?location=%s&radi-
us=%s&types=%s&key=AIzaSyBkc15b88TuRItOZfwC6pbqL4SrbXCnn2E\” % (location,radi-
us,types)\n”,
   “def run_loop(location,radius,types, URL):\n”,
   “    final_data=[]\n”,
   “    while True:\n”,
   “        response = requests.request(\”POST\”, URL)\n”,
   “        response = json.loads(response.text)\n”,
   “        results = response[‘results’]\n”,
   “\n”,
   “        for result in results:\n”,
   “            final_data.append(result)\n”,
   “\n”,
   “        if ‘next_page_token’ not in response:\n”,
   “            break\n”,
   “        else:\n”,
   “            next_page_token = response[‘next_page_token’]\n”,
   “\n”,
   “        next_page_token = ‘&pagetoken=%s’ % str(next_page_token)\n”,
   “        URL = \”https://maps.googleapis.com/maps/api/place/nearbysearch/json?loca-
tion=%s&radius=%s&types=%s&key=AIzaSyBkc15b88TuRItOZfwC6pbqL4SrbXCnn2E%s\”
% (location,radius,types,next_page_token)\n”,
   “\n”,
   “        return final_data\n”,
    ]
```