DSSR

Discussion Paper No. 149

Alternative approaches to monitoring working hours: Use of location data from cell phones

Yuta Kuroda and Shinpei Sano

November, 2025

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

Alternative approaches to monitoring working hours: Use of location data from cell phones[*]

Yuta Kuroda[1] and Shinpei Sano[2]

[1] Graduate School of Economics and Management, Tohoku University, Sendai, Japan

[2] Graduate School of Economics, Kobe University, Kobe, Japan

Abstract

This study examines the potential of mobile phone location data for measuring working hours. We compare hourly population estimates from Mobile Spatial Statistics with card-reader attendance records of teachers at Japanese public high schools. The analysis shows a strong correlation between location-based indicators and actual working hours. Mobile data provide more accurate proxies in areas with few residents, where background noise is limited, and in schools with many employees. These results suggest that large-scale mobility data can serve as a valuable resource for labor research when direct administrative records are not accessible.

JEL classifications: C81, C55, J22

Keywords: Working hours, Mobility data, Big data, Attendance records, Mobile spatial statistics

[*] Corresponding author: Yuta Kuroda (kuroyu0725@gmail.com).

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI

During the preparation of this work the authors used DeepL and Grammarly in order to improve its language. After using these tools, the authors reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Data availability statement

The demographic data used to construct the mobility indices were provided by NTT Docomo, and the raw data cannot be shared without the provider's permission, but can be purchased at https://mobaku.jp/. Teacher attendance records also cannot be shared without authorization from the data holder, as they contain personal information. All other datasets are publicly available. The dataset to reproduce our results can be provided upon request.

1. Introduction

Working hours are one of the central indicators in economics, and their accurate measurement is crucial for estimating labor productivity and hourly wages (Nekoei, 2022). Excessive working hours may also reduce productivity (Pencavel, 2015), harm physical and mental health (Cygan-Rehm et al., 2018), and lower overall life satisfaction (Hamermesh et al., 2017), highlighting the importance of proper work-time management. Nevertheless, most contemporary studies rely on self-reported survey data, which are costly to collect, infrequent, and potentially subject to systematic bias (Baum-Snow & Neal, 2009; Nekoei, 2022). To enable more precise empirical analyses and the design of effective policies, it is therefore essential to develop high-frequency, accurate, objective, and broadly applicable methods for measuring working hours that allow for comparisons across sectors.

This study examines the applicability and limitations of mobile phone location data for measuring working hours. Since the COVID-19 pandemic, location data from mobile devices have been increasingly used to capture human activity (e.g., Alexander & Karger, 2023; Couture et al., 2022; Monte, 2020). Much of this research focuses on visits to specific places or social interactions (e.g., Liang et al., 2022; Mas & Koshimura, 2024). More recently, several studies have gone further by using location data as proxies for economic activities such as consumption in commercial districts (Cepparulo, 2025) or workplace attendance (Arai et al., 2023; Sakuma et al., 2024). Nevertheless, as a relatively new data source, the scope and validity of mobile location data remain insufficiently understood.

To address this gap, we use card-reader records of teacher attendance at public high schools in Osaka and Sakai as a benchmark. Specifically, we compare the number of teachers on school premises by time of day, based on attendance records, with population estimates derived from mobile phone location data around each school. This comparison allows us to assess under what conditions location data can reliably approximate actual working hours.

2. Data and settings

The mobile phone location data used in this study are obtained from NTT DOCOMO, Inc., Japan's largest mobile network operator, and consist of hourly population estimates known as Mobile Spatial Statistics (MSS). These estimates are generated from the continuous and automatic aggregation of location information from approximately 85 million subscriber devices operating on the NTT DOCOMO network. Unlike GPS-based systems, MSS rely on cellular base station signals. As a result, data collection does not depend on installing or activating specific applications; as long as a device is powered on, its presence can be captured. Japan is divided into grids of 500 × 500 meters, and for each grid the estimated population is reported on an hourly basis. In major metropolitan areas, population estimates are also available at the finer resolution of 250 × 250 meters. MSS has been widely used in studies on Japan because of its high temporal and spatial granularity, accuracy, and representativeness (e.g., Arai et al., 2023; Kuroda et al., 2025; Sakuma et al., 2024). Further details and technical descriptions of the MSS can be found in the NTT DOCOMO Technical Journal (NTT DOCOMO, Inc., 2013).

Our analysis focuses on Osaka and Sakai, the core cities of Osaka Prefecture, the second largest metropolitan area in Japan after Tokyo. We use MSS data at 500-meter and 250-meter resolutions for 2019–2021. The 500-meter data are available for all dates and hours, whereas the 250-meter data are restricted to July–August and 13 hours per day

(6:00–10:00 and 15:00–23:00). We analyze three representative periods—pre-COVID-19, early pandemic, and mid-pandemic—to capture changes in workplace attendance behavior. Since population levels remain stable during daytime (10:00–15:00) and late night (23:00–6:00), working patterns can be effectively observed by focusing on morning (6:00–10:00) and evening (15:00–23:00) hours.

In addition, we obtained attendance data for 46 public high schools from the Osaka Prefectural Board of Education. These records are based on card-reader entries and exits, and in principle cover all teachers and staff. We aggregate them to construct hourly measures of the number of teachers present at each school. It should be noted that card-reader attendance records do not capture work performed outside school premises, such as tasks at home or business trips, and therefore do not represent teachers' total working hours. However, the focus of this study is to assess how effectively mobile location data can approximate time spent at the workplace. Thus, this limitation is not critical for our analysis.

Fig. 1 shows the 500-meter grid boundaries and the locations of public high school premises in Osaka and Sakai, the study area. As the figure illustrates, school premises sometimes span multiple grids. To address this, we allocate the MSS-based population estimates to each school according to the proportion of its area within each grid, thereby constructing an indicator of the surrounding population. We then assess how well this indicator reflects the actual number of teachers present at school, thereby examining the potential applicability of mobile location data in studies of working hours.

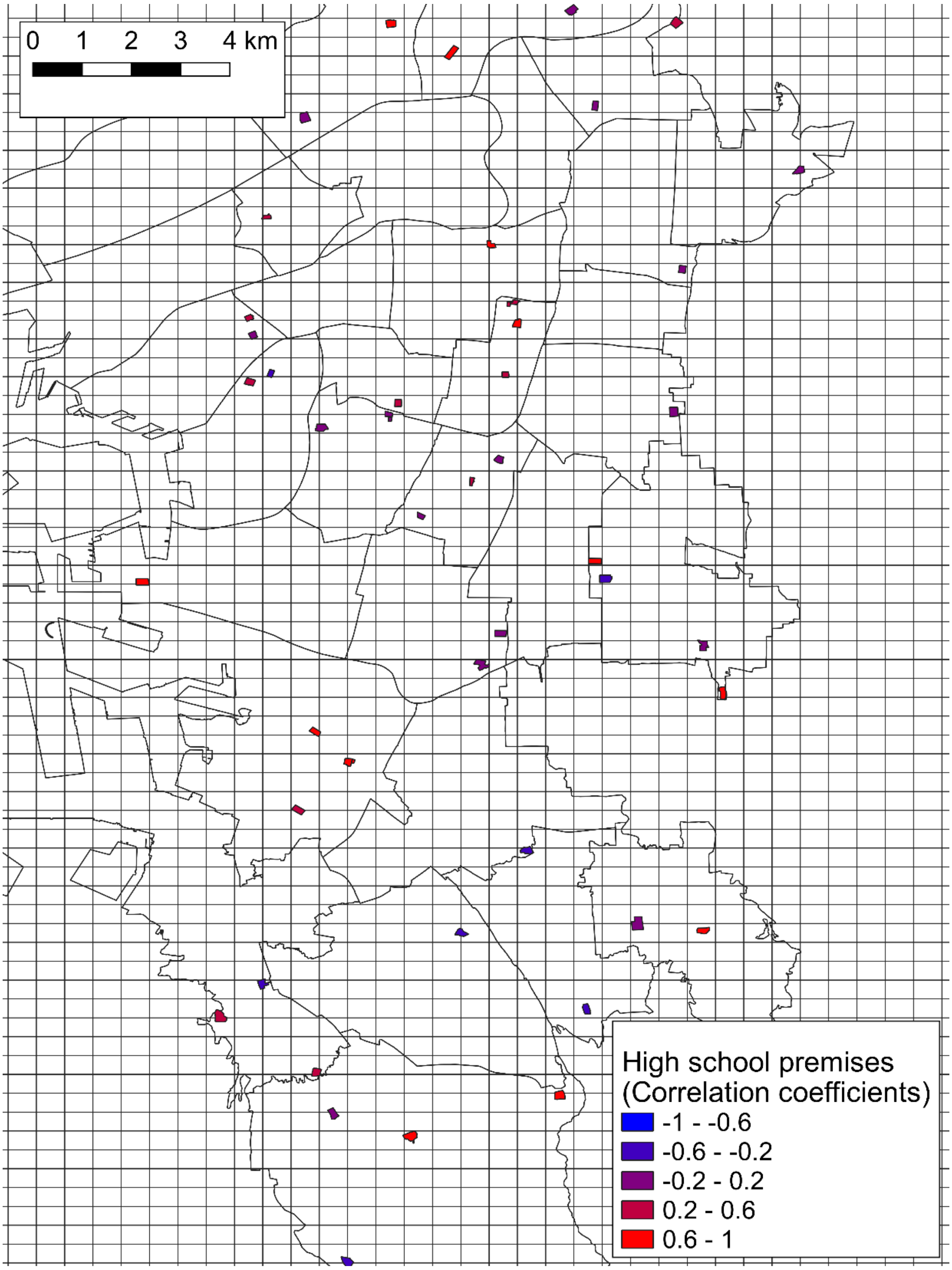


Fig. 1. Public high school premises and 500-meter grid boundaries in the study area

Note: This figure shows the 500-meter grid boundaries and the locations of public high school premises in Osaka and Sakai. Colors indicate the correlation coefficients between MSS-based population indicators and teacher attendance records.

3. Methods

We begin with a simple comparison between the surrounding population, estimated from MSS, and the number of teachers present at school, measured by card-reader attendance records. Although the original data are available at an hourly level, we mainly use daily averages to facilitate visualization.[1] To make the two measures comparable, both variables are standardized using the full sample period. Plotting the standardized values allows for a visual inspection of their correlation. In addition, we calculate school-level correlation coefficients and use them for the analysis.

To examine which geographic, neighborhood, and school characteristics are associated with the applicability of location data, we estimate the following specification.

$$Cor_i = \alpha + \beta G_i + \gamma SN_i + \varepsilon_i \tag{1}$$

where the dependent variable, Cor_i, is the correlation coefficient between MSS-based population estimates and attendance-based teacher counts at the school–day level. We use three versions of this measure: (i) the 24-hour average at the 500-meter resolution, (ii) the peak-time average (6:00–10:00 and 15:00–23:00) at the 500-meter resolution, and (iii) the peak-time average at the 250-meter resolution.

The vector G_i captures the spatial relationship between school premises and MSS grids, including the number of overlapping grids, the share of the dominant grid, and the share of area outside the premises. These variables allow us to assess how technical

[1] Results are nearly identical when using hourly instead of daily data.

features of the data affect the measurement of workplace attendance. The vector SN_i represents school and neighborhood characteristics, such as school size (number of teachers), daytime–nighttime population ratio, population density, and the number of establishments. This specification allows us to assess the extent to which technical constraints—such as the relationship between grids and school premises or data resolution—and school and neighborhood characteristics—such as school size or local population density—affect the reliability of location data in measuring working hours. Descriptive statistics of the variables used in the analysis are reported in Appendix Table A1.

4. Results

Fig. 2 presents standardized daily indices of teacher attendance and surrounding population based on mobile phone location data for July–August in 2019, 2020, and 2021. The indices are constructed from 500-meter all-day averages at the school level and then standardized across schools. Both series exhibit clear declines during weekends and the mid-August summer holiday, confirming the plausibility of the data. Although the two indicators diverge at times, the visual evidence suggests that location-based measures closely track attendance records overall. Appendix Figure A1 shows similar results using hourly data, indicating that the findings do not depend on the level of temporal aggregation and that mobile indices can meaningfully capture working hours.

To examine whether spatial resolution matters, Fig. 3 compares attendance records with MSS-based indices at 500-meter and 250-meter grids, focusing on peak-time averages (6:00–10:00 and 15:00–23:00). The 250-meter series appears marginally closer to actual attendance, but the difference is small, implying that even 500-meter grids can adequately capture work attendance behavior.

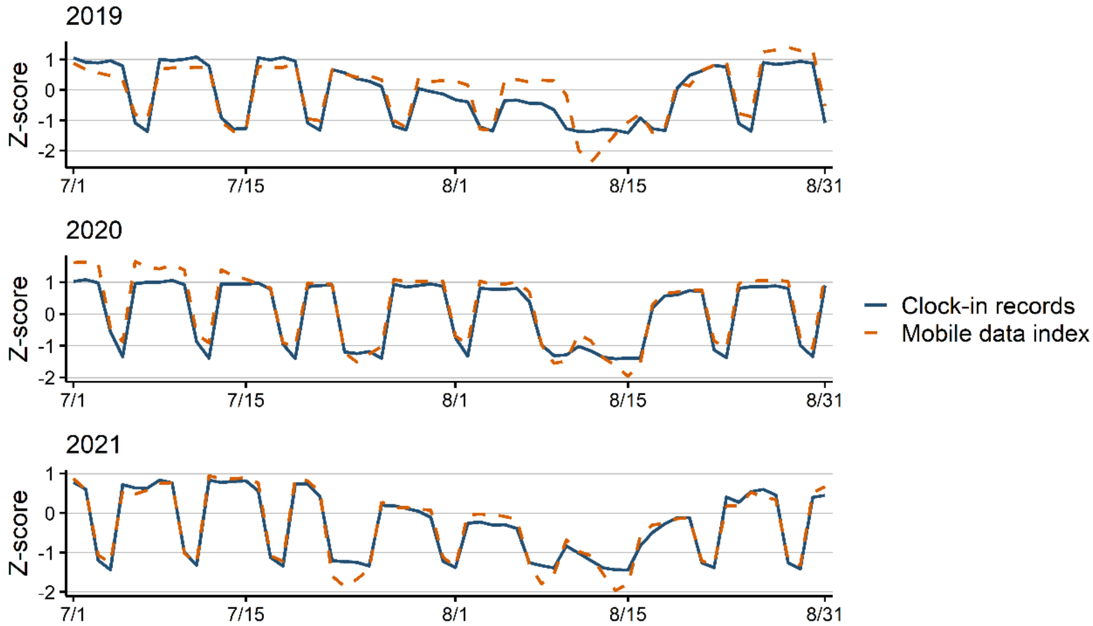


Fig. 2. Standardized daily school attendance and surrounding population

Note: This figure plots standardized daily values of teacher attendance (clock-in records) and surrounding population (mobile data index) from July 1 to August 31 for 2019, 2020, and 2021. The vertical axis represents standardized values (Z-scores), and the horizontal axis shows calendar dates. The blue solid line represents attendance records, and the orange dashed line represents the 500-meter MSS population index.

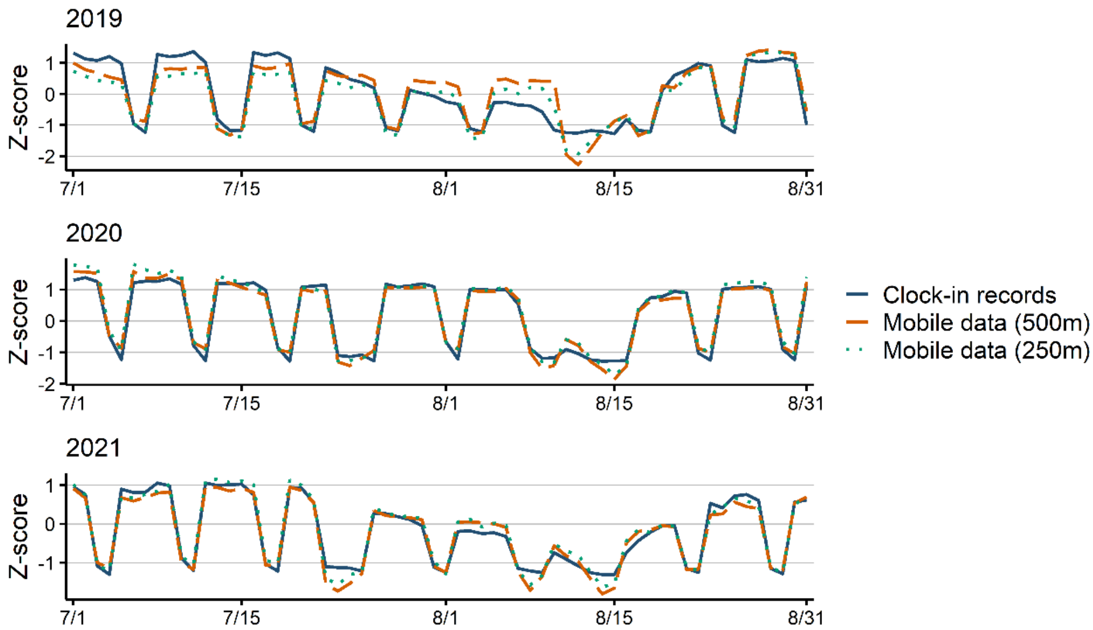


Fig. 3. Standardized daily attendance and surrounding population at peak times (500m vs. 250m grids)

Note: This figure plots standardized daily values of teacher attendance (clock-in records) and surrounding population indices derived from MSS at the 500-meter and 250-meter grid resolutions. The data cover July 1 to August 31 for 2019, 2020, and 2021. The vertical axis represents standardized values (Z-scores), and the horizontal axis shows calendar dates. Unlike Figure 2, which uses all-day averages, this figure focuses on peak-time periods (6:00–10:00 and 15:00–23:00) to compare the performance of different spatial resolutions. The blue solid line represents attendance records, the orange dashed line represents the 500-meter MSS population index, and the light blue dotted line represents the 250-meter MSS population index.

Fig. 2 and 3 report averages across all schools. However, the reliability of mobile indices may vary depending on school and neighborhood characteristics. Fig. 4 shows the distribution of school-level correlation coefficients between attendance records and MSS-based indices. Most correlations are positive, yet there is considerable heterogeneity across schools. While resolution and aggregation time do not substantially alter the average fit, the distribution suggests that finer spatial units, such as 250 meters, may yield stronger correlations with actual attendance.

Table 1 explores the determinants of these school-level correlations. Geographic factors related to the alignment between school premises and grid boundaries show little explanatory power. By contrast, the day–night population ratio around schools has a significant effect: mobile indices perform better in areas with fewer non-working residents. Because mobile data capture all individuals regardless of employment status, high residential density introduces noise and weakens the signal. Consequently, schools located in areas with limited unrelated population activity are more accurately represented by location-based measures.

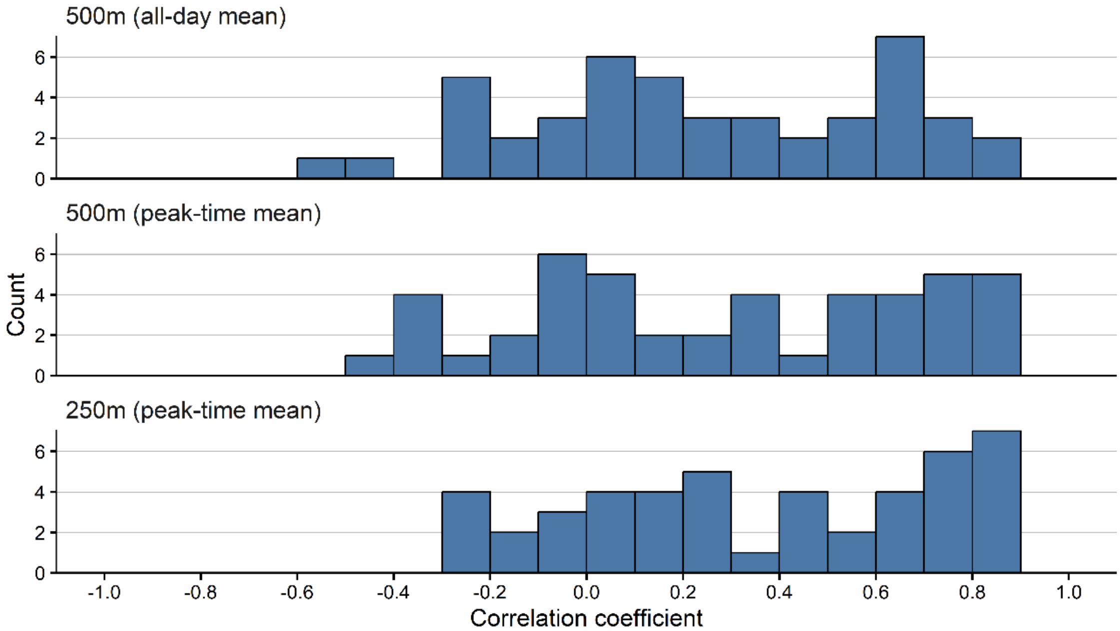


Fig. 4. Distribution of school-level correlations between attendance records and MSS population indices

Note: This figure shows histograms of school-level correlation coefficients between MSS-based population estimates and teacher attendance records. Three variants of the MSS population index are used: the 500-meter all-day mean, the 500-meter peak-time mean (6:00−10:00 and 15:00–23:00), and the 250-meter peak-time mean. The vertical axis indicates the number of schools, and the horizontal axis represents correlation coefficients.

	500m All day (1)	500m Peak hours (2)	250m Peak hours (3)
Geographic characteristics			
School site area	-44.2	-31.3	-45.3
	(112.8)	(123.7)	(329.9)
Number of overlapping meshes	10.90	7.27	2.76
	(29.6)	(32.5)	(21.7)
Dominant mesh share	0.252	0.585	0.129
	(0.354)	(0.389)	(0.298)
Residual mesh area	-40.9	-27.2	-41.6
	(111.7)	(122.6)	(328.2)
School and neighborhood characteristics			
Number of teachers	0.007	0.010	0.008
	(0.005)	(0.005)	(0.005)
Day/night population ratio	0.246*	0.241*	0.210*
	(0.096)	(0.106)	(0.099)
Population	-0.009	-0.003	0.003
	(0.061)	(0.066)	(0.064)
Number of Businesses	-0.011	-0.0004	0.004
	(0.048)	(0.053)	(0.052)
Observations	46	46	46
Adjusted R-squared	0.2808	0.2846	0.1919

Table 1. Regression results on school-level correlation coefficients

Note: This table reports the estimation results of equation (1), examining how geographic, school, and neighborhood characteristics affect the correlation between MSS-based population indices and teacher attendance records. The dependent variable is the school-level correlation coefficient, calculated using three variants of MSS data: the 500-meter all-day mean (column 1), the 500-meter peak-time mean (column 2), and the 250-meter peak-time mean (column 3). Standard errors are reported in parentheses. An asterisk (*) indicates statistical significance at the 5% level.

5. Conclusion

This study contributes to the economic literature by evaluating the applicability of mobile phone location data for measuring working hours, using high-frequency and accurate administrative attendance records as a benchmark. Because such data are objective, temporally detailed, and collected without imposing additional costs on workers, they can partly overcome long-standing limitations of conventional labor statistics, such as missing records and self-reporting biases, whether intentional or unintentional. At the same time, the usefulness of location data is bounded by the nature of work. It is particularly suitable for industries where physical presence at the workplace is essential—such as education, healthcare, agriculture, retail, and construction—but it cannot adequately capture remote work, which has become more prevalent since COVID-19. From a policy perspective, these findings suggest that mobile location data could serve as a complementary source of evidence for monitoring workplace conditions in sectors where administrative records are limited, thereby supporting the design of more effective labor market policies.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 23K17556, 23K25529, and 24K16385. We appreciate the cooperation of NTT DOCOMO, Inc. and the Osaka Prefectural Board of Education in providing us with valuable data. Any errors that remain are our own. The views expressed are those of the authors and do not necessarily reflect those of any organizations with which the authors are affiliated.

Reference

Alexander, D., & Karger, E. (2023). Do Stay-at-Home Orders Cause People to Stay at Home? Effects of Stay-at-Home Orders on Consumer Behavior. *Review of Economics and Statistics*, *105*(4), 1017–1027. https://doi.org/10.1162/rest_a_01108

Arai, N., Hamano, M., Katayama, M., Murakami, Y., & Yamada, K. (2023). Nightless city: Impacts of policymakers' questions on overtime work of government officials. *ISER DP*, No. 1206. https://doi.org/10.2139/ssrn.4413072

Baum-Snow, N., & Neal, D. (2009). Mismeasurement of usual hours worked in the census and ACS. *Economics Letters*, *102*(1), 39–41. https://doi.org/10.1016/j.econlet.2008.11.005

Cepparulo, B. (2025). Is mobility a good proxy for consumption? *Economics Letters*, *255*, 112454. https://doi.org/10.1016/j.econlet.2025.112454

Couture, V., Dingel, J. I., Green, A., Handbury, J., & Williams, K. R. (2022). JUE Insight: Measuring movement and social contact with smartphone data: a real-time application to COVID-19. *Journal of Urban Economics*, *127*, 103328. https://doi.org/10.1016/j.jue.2021.103328

Cygan-Rehm, K., & Wunder, C. (2018). Do working hours affect health? Evidence from statutory workweek regulations in Germany. *Labour Economics*, *53*, 162–171. https://doi.org/10.1016/j.labeco.2018.05.003

Hamermesh, D. S., Kawaguchi, D., & Lee, J. (2017). Does labor legislation benefit workers? Well-being after an hours reduction. *Journal of the Japanese and International Economies*, *44*, 1–12. https://doi.org/10.1016/j.jjie.2017.02.003

Kuroda, Y., Sato, T., & Matsuda, Y. (2025). How long do voluntary lockdowns keep people at home? The role of social capital during the COVID-19 pandemic. *Japanese Economic Review*, in Press. https://doi.org/10.1007/s42973-025-00206-8

Liang, Y., Yin, J., Pan, B., Lin, M. S., Miller, L., Taff, B. D., & Chi, G. (2022). Assessing the validity of mobile device data for estimating visitor demographics and visitation patterns in Yellowstone National Park. *Journal of Environmental Management*, *317*, 115410. https://doi.org/10.1016/j.jenvman.2022.115410

Mas, E., & Koshimura, S. (2024). Feasibility of anomalous event detection based on Mobile Spatial Statistics: A study of six cases in Japan. *International Journal of Disaster Risk Reduction: IJDRR*, *110*, 104625. https://doi.org/10.1016/j.ijdrr.2024.104625

Monte, F. (2020). Mobility Zones. *Economics Letters*, *194*, 109425. https://doi.org/10.1016/j.econlet.2020.109425

Nekoei, A. (2022). The measurement of labor supply using March CPS: A cautionary tale. *Economics Letters*, *216*, 110512. https://doi.org/10.1016/j.econlet.2022.110512

NTT DOCOMO, Inc., Technical Journal, 14:3 (2013) https://www.docomo.ne.jp/english/corporate/technology/rd/technical_journal/bn/vol14_3/

Pencavel, J. (2015). The productivity of working hours. *Economic Journal*, *125*(589), 2052–2076. https://doi.org/10.1111/ecoj.12166

Sakuma, M., Matsuo, K., Tsutsumi, M., & Imazeki, T. (2024). Measuring office attendance during the COVID-19 pandemic with mobility data to quantify local trends and characteristics. *Asia-Pacific Journal of Regional Science*, *8*(1), 185–237. https://doi.org/10.1007/s41685-023-00324-4

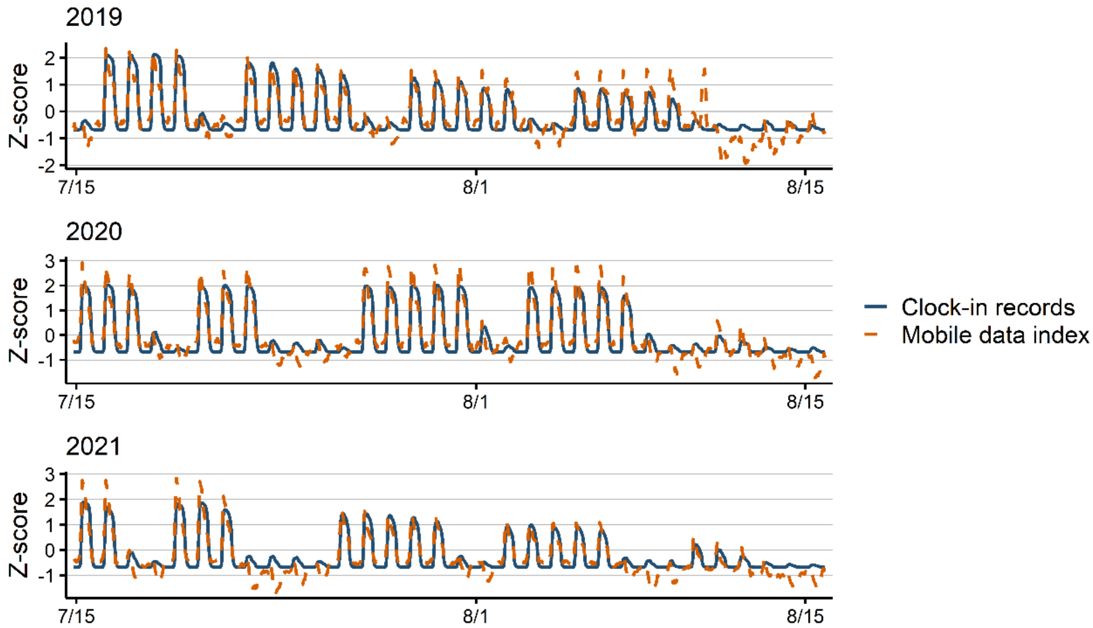
Appendix

Table A1. Descriptive statistics of school-level correlation, geographic characteristics, and neighborhood attributes

	Mean	Std. Dev.	Min.	Max.
Work records–mobile data correlation				
500m (all-day mean)	0.247	0.368	-0.566	0.843
500m (peak-time mean)	0.288	0.404	-0.478	0.898
250m (peak-time mean)	0.369	0.371	-0.266	0.890
Geographic characteristics				
School site area (km^2)	0.032	0.010	0.014	0.063
Number of overlapping meshes (500m)	2.109	0.994	1.000	4.000
Number of overlapping meshes (250m)	3.348	1.037	1.000	5.000
Dominant mesh share (500m)	0.817	0.193	0.378	1.000
Dominant mesh share (250m)	0.657	0.193	0.334	1.000
Residual mesh area (km^2, 500m)	0.527	0.263	0.230	1.038
Residual mesh area (km^2, 250m)	0.190	0.068	0.032	0.296
School and neighborhood characteristics				
Number of teachers	59.348	11.445	27.000	79.000
Day/night population ratio	1.248	0.877	0.670	4.877
Population (×1000)	2.977	1.688	0.003	6.081
Number of Businesses (×100)	1.486	1.642	0.112	9.270
Total number of high schools		46		

Note: This table reports descriptive statistics for the 46 public high schools in the study area. The first panel shows correlations between teacher attendance records and MSS-based population indicators at different spatial and temporal resolutions. The second panel summarizes school-level geographic characteristics, including site area, the number of overlapping meshes, and the share of the dominant mesh. The third panel presents school and neighborhood attributes, such as the number of teachers, local population, and the number of businesses.

Figure A1. Hourly patterns of school attendance and surrounding population



Note: This figure plots standardized hourly values of teacher attendance (clock-in records) and surrounding population (mobile data index) from July 15 to August 15 for 2019, 2020, and 2021. The vertical axis shows standardized values (Z-scores), and the horizontal axis represents dates (hours). While the main analysis in this study is based on daily variables, this figure uses hourly data to illustrate the dynamics of workplace attendance and location-based population measures.