UNIVERSITY OF SÃO PAULO

Faculty of Pharmaceutical Sciences

Post-Graduate Program in Pharmacy (Pathophysiology and Toxicology)

Pathophysiology Area

Using location history data from cell phones of infectious patients for disease surveillance

Jeevan Giddaluru

Thesis to obtain the title of Doctor of Science

Supervisor: Prof. Dr. Helder Takashi Imoto Nakaya

São Paulo

2023

UNIVERSITY OF SÃO PAULO

Faculty of Pharmaceutical Sciences

Post-Graduate Program in Pharmacy (Pathophysiology and Toxicology)

Pathophysiology Area

Using location history data from cell phones of infectious patients for disease surveillance

Jeevan Giddaluru

Corrected Version

Thesis to obtain the title of Doctor of Science

Supervisor: Prof. Dr. Helder Takashi Imoto Nakaya

São Paulo

2023

Ficha Catalográfica elaborada eletronicamente pelo autor, utilizando o programa desenvolvido pela Seção Técnica de Informática do ICMC/USP e adaptado para a Divisão de Biblioteca e Documentação do Conjunto das Químicas da USP

Bibliotecária responsável pela orientação de catalogação da publicação:
Marlene Aparecida Vieira - CRB - 8/5562

Giddaluru, Jeevan
G43u Usando dados do histórico de localização de telefones celulares de pacientes infecciosos para vigilância de doenças / Jeevan Giddaluru. - São Paulo, 2023.
64 p.

Tese (doutorado) - Faculdade de Ciências Farmacêuticas da Universidade de São Paulo. Departamento de Análises Clínicas e Toxicológicas.
Orientador: Nakaya, Helder Takashi Imoto

1. Vigilância de doenças. 2. Epidemiologia digital. 3. Sistema de Posicionamento Global. 4. Malária. 5. Tuberculose. I. T. II. Nakaya, Helder Takashi Imoto, orientador.

Jeevan Giddaluru

Using location history data from cell phones of infectious patients for disease surveillance

Judging Committee
of thesis to obtain the title of Doctor of Science

Prof. Dr. Helder Takashi Imoto Nakaya
Supervisor/president

_____
1st examiner

_____
2nd examiner

_____
3rd examiner

_____
4th examiner

São Paulo, _____ de _____ de 2023.

I dedicate this work to my family.
Your love, care, and belief have shaped this academic journey and crafted the very essence of who I am.

ACKNOWLEDGEMENTS

First and foremost, I extend my deepest gratitude to my supervisor, Prof. Dr. Helder Takashi Imoto Nakaya. His mentorship and expertise were foundational pillars of my Ph.D. journey. This research is a testament to innovation and discovery and has only been possible through his invaluable guidance. Prof. Nakaya's passion for science didn't just inspire me; it profoundly reshaped my academic perspective. His faith in my potential instilled a keen sense of purpose and a commitment to the health sciences that transcends the confines of this research. Without his mentorship, this journey would have remained unrealized. His influence on my personal growth and overall research trajectory has been immeasurable.

I reserve special thanks to my postdoc angels, Dr. Bernardo Oliveira and Dr. André Guilherme. Throughout my Ph.D. journey, your persistent support has been a lifeline. Our weekly meetings weren't just check-ins but instrumental in guiding and motivating me. Your constant encouragement and gentle pushes kept me focused. I am deeply grateful for your dedication and for being much more than just a mentor.

I sincerely thank my colleagues at the Computational Systems Biology Laboratory (CSBL). Your collaborative spirit, shared insights, and consistent support enriched my academic experience and made this journey rewarding and memorable. Working alongside such a dedicated and talented group has been an honor.

I extend my heartfelt gratitude to my friends Lucas Fleig, Alysson Urbanski, Vanessa Maso, Rodrigo Ogava, Ícaro Castro, Juan Carlos, André Cunha, and Peter Park. While our professional ties brought us together, the bonds we forged and the memories we created are what truly enriched my journey. The moments we shared went beyond our tasks in the lab. Thank you for the friendships and the unforgettable times we've experienced.

I convey my sincere appreciation to the friends and senior members of the lab: Thiago Hirata, Patrícia Gonzalez, Mauro Morais, and André Nicolau. Your wisdom, experience, and mentorship were invaluable to me. The insights you provided and the standards you upheld set a benchmark for my work. Your willingness to share your knowledge, assist with challenges, and provide constructive feedback greatly enriched my research experience. I am grateful for your guidance and the welcoming environment you fostered, making it an enriching space for growth and discovery.

I would like to take a moment to remember my late friend, Diógenes Lima. His mentorship was priceless, and his character stands as an inspiration for me. Even in his absence, the lessons he imparted and the legacy he left behind continue to guide me. His memory will forever remain etched in my heart and mind.

I am indebted to André Guilherme, Youvika Singh, and Manessh Singh. Living in Brazil, away from my native country, you became more than just friends—you became family. Your genuine care, warmth, and willingness to assist in times both big and small provided a sense of home and belonging. I genuinely believe that without you, my experience and life here would have felt incomplete. I am grateful for our countless laughter, understanding, and camaraderie moments. Thank you for making my time in Brazil fulfilling and complete.

My heartfelt thanks to Bárbara Almeida Penna Firme. While our paths have evolved, our friendship has been truly invaluable. The moments we've shared have added depth and color to every phase of this journey. This chapter in my life is more special with you by my side. Thank you for being a cherished part of my narrative and for the memories we continue to weave together.

I appreciate the Department of Clinical and Toxicological Analyses and the School of Pharmaceutical Sciences at the University of São Paulo for their consistent administrative and academic support. I sincerely thank Paola Minoprio, the Director of the Pasteur Institute of São Paulo (SPPU). Her generosity in providing essential facilities, space, and support was instrumental to continue my research during the pandemic. Finally, I express my appreciation to Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP, for generously funding this project (2019/16419-7). I am grateful for the resources they provided, which were pivotal in completing my doctoral studies.

To my dearest family, words can hardly encapsulate the depth of my gratitude. My mother, Krishnaveni Posa Giddaluru, has dedicated her entire life to ensuring my personal and professional growth. Every sacrifice she made, every challenge she overcame, was with the singular focus of my happiness and success. Her boundless love and relentless dedication have been the guiding lights of my journey. My sister, Sujana Giddaluru, has consistently reminded me of my responsibilities, pushing me to stand as a strong example. Her faith in my potential has been a source of unyielding resilience and motivation. Her dedication in difficult times is a continuous

reminder of how challenges ought to be tackled. Remembering my late father, Prasad Giddaluru, I am reminded of a man who never wavered in his commitment to my career growth despite the myriad obstacles he faced. His resilience in adversity has been a lesson in tenacity, perseverance, and devotion. I am eternally grateful to my family. I carry their teachings and affections with me, always!

ABSTRACT

GIDDALURU, J. **Using location history data from cell phones of infectious patients for disease surveillance**. 2023. Thesis (Doctorate) - Faculty of Pharmaceutical Sciences, University of São Paulo, São Paulo, 2023.

Infectious diseases significantly contribute to global morbidity and mortality, highlighting the critical need for robust disease surveillance systems. The rapid and accurate identification of infection hotspots is crucial for effective disease control and eliminating vector reservoirs. Traditional methods, reliant on patient-reported data, are vague, slow, and non-integrative, presenting substantial barriers to fully understanding the underlying causes of infection transmission. The widespread usage of smartphones presents a unique opportunity to access, analyze, and monitor digital data. Particularly, location data can offer potential insights into infectious disease dynamics, which has remained largely unexplored. Firstly, the present study leverages location history data from smartphones of malaria patients in Manaus, Amazonas region, to pinpoint mosquito-breeding sites. Upon quantifying the location data, the primary transmission hotspots were identified to be concentrated on the outskirts of the city of Manaus. Additionally, the quantification and hotspot validation confirmed that newly visited locations during the exposure period were potential sources of infection transmission. Secondly, the current study also employs a novel digital contact investigation method for a human-to-human transmission infection such as tuberculosis to measure the exposure risk between the active index cases and their close contacts. The digital contact investigation revealed varied exposure durations between the recruited paired index and close contact participants based on the outcome of close contact. To summarize, the present study determines distinct mobility patterns associated with both these infectious diseases, potentially aiding in drafting targeted public health strategies and policies for digital epidemiological surveillance.

Keywords: Disease surveillance. Digital epidemiology. Global positioning system. Malaria. Tuberculosis.

RESUMO

GIDDALURU, J. **Usando dados do histórico de localização de telefones celulares de pacientes infecciosos para vigilância de doenças**. 2023. Tese (Doutorado) - Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2023.

As doenças infecciosas são um dos principais contribuintes para a morbidade e a mortalidade globais, enfatizando a necessidade crítica de sistemas robustos de vigilância de doenças. A identificação rápida e precisa dos pontos críticos de infecção é fundamental para o controle eficaz de doenças e a eliminação de reservatórios de vetores. Os métodos tradicionais, que dependem de dados relatados por pacientes, são vagos, lentos e não integrativos, apresentando barreiras significativas para a compreensão total das causas subjacentes da transmissão de infecções. O uso generalizado de dispositivos móveis apresenta uma oportunidade única de acessar, analisar e monitorar dados digitais. Especialmente, dados de localização podem oferecer informações úteis sobre a dinâmica de doenças infecciosas, que permanecem em grande parte inexploradas. Primeiramente, o presente estudo utiliza dados de histórico de localização de smartphones de pacientes com malária em Manaus, na região do Amazonas, para identificar locais de reprodução de mosquitos. Ao quantificar os dados de localização, identificaram-se os principais pontos de transmissão concentrados nos arredores da cidade de Manaus. Além do mais, a quantificação e a validação em campo confirmaram que os locais recém-visitados durante o período de exposição eram potenciais fontes de transmissão da infecção. Em segundo lugar, o estudo atual também emprega um inovador método de investigação digital de contato para uma infecção por transmissão de humano para humano, como a tuberculose, a fim de medir o risco por exposição entre os casos índice ativos e seus contatos próximos. A investigação digital de contato revelou períodos de exposição variados entre os participantes recrutados em pares de casos índice e contatos próximos, com base no resultado do contato próximo. Em resumo, o presente estudo identifica padrões distintos de mobilidade associados a ambas essas doenças infecciosas, auxiliando potencialmente na elaboração de estratégias e políticas de saúde pública direcionadas para a vigilância epidemiológica digital.

Palavras-chave: Vigilância de doenças. Epidemiologia digital. Sistema de Posicionamento Global. Malária. Tuberculose.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

An. darlingi	-	*Anopheles darlingi*
DBSCAN	-	Density-based spatial clustering of applications with noise
FMT-HVD	-	Fundação de Medicina Tropical Doutor Heitor Vieira Dourado
GIS	-	Geographic Information Systems
GLH	-	Google location history
GPS	-	Global positioning system
RACD	-	Reactive Case Detection
HDBSCAN	-	Hierarchical density-based spatial clustering of applications with noise
IBIT	-	Instituto Brasileiro para Investigação da Tuberculose
IGRA	-	Interferon-gamma release assay
IRS	-	Indoor residual spraying
SiPoS	-	Sickness positioning system
SIVEP	-	Sistema de Informação de Vigilância Epidemiológica
SUS	-	Sistema Único de Saúde
TB	-	Tuberculosis

CONTENTS

CHAPTER I

1. BACKGROUND

1.1 Global impact of infectious diseases

Infectious diseases have long been the leading cause of morbidity and mortality worldwide, accounting for over 25% of the global disease burden (HIGH, 2004; HOTEZ et al., 2004). In 2019, infectious and parasitic diseases contributed to 5 million deaths worldwide (WHO, 2020a). Malaria, tuberculosis, and HIV/AIDS dominate the primary causes of death in low-income countries (WHO, 2020b). While the developed world has seen some success in mitigating the burden of infectious diseases (PINHEIRO; MATHERS; KRÄMER, 2009), the COVID-19 pandemic alone claimed 6.8 million lives by April 2023 (WHO, 2023), causing widespread health, societal, and economic disruption in both developed and developing countries (BIN ABDUL SATAR; HAKIMAH YAACOB, 2022; PADHAN; PRABHEESH, 2021). Many strategies have been introduced in recent decades to address the challenges of infectious diseases effectively. These strategies include accurate and affordable diagnostic testing (PANG; PEELING, 2007), global vaccination coverage (EDELSTEIN, 2017), proactive public health preventive measures, resilient health infrastructure, and robust disease surveillance systems (PALAGYI et al., 2019).

1.2 Evolution of infectious disease surveillance

Historically, disease surveillance played a pivotal role in eradicating infections such as smallpox (HEYMANN; BRILLIANT, 2011). In recent decades, disease surveillance has relied heavily on manual data collection, with local laboratories and hospitals reporting the data to public health agencies (M'IKANATHA et al., 2013). However, the era of advanced technology has uncovered a transformative phase in disease surveillance. Enhanced computing power, data-storage capabilities, and seamless data-sharing platforms have significantly improved the speed of information dissemination between medical care facilities and public health agencies, thereby enhancing the early detection capability (BANSAL et al., 2016). Concurrently, mobile technologies have facilitated expedited communication to implement outbreak intervention measures (NHAVOTO; GRÖNLUND, 2014).

Moreover, technological advancements in the Global Positioning System (GPS) encouraged experts to employ Geographic Information Systems (GIS) for

spatial epidemiological surveillance, replacing conventional geographical surveys (HIGHTOWER et al., 1998; SEVILLA-CASAS, 1993). With GIS technology being easily accessible, epidemiologists can delve deeper into the underlying geographical factors influencing disease transmission and trends (EISEN; EISEN, 2014) and subsequently support epidemiological modeling and prediction. This transition from manual data reporting to technological-driven surveillance has become a powerful tool for effectively monitoring and addressing infectious diseases.

1.3 Digital epidemiology

Though technological advancements enhanced the data acquisition methods, the primary sources of disease surveillance have been mainly confined to diagnostic and research laboratories, hospitals, and public agencies. However, the widespread usage of communication devices and internet availability enabled novel digital data sources that could provide real-time information on disease dynamics. This potential epidemiological footprint encompasses multifaceted digital data streams captured from smartphones, mobile applications, search engines, and wearable technologies (SALATHÉ et al., 2012). A recent COVID-19 epidemiological surveillance system has also incorporated unconventional data sources, such as financial transactions and QR code engagements (KIM, 2023).

Applying this rapidly expanding digital data to gain insights into epidemiology refers to Digital Epidemiology. Although relatively new, the first step towards digital epidemiology can be traced back to the pre-internet revolution. For example, rudimentary digital traces such as cell phone records, text messages, and mobile tower locations have provided insights into the transmission of diseases such as malaria and cholera (WESOLOWSKI et al., 2012; ZHOU et al., 2020). In the contemporary era, however, there has been a marked enhancement in the granularity of this data. Location data, in particular, is now being recorded with increased resolution. The wide-spread accessibility of such fine-grained location mobility data offers an understanding of human mobility patterns and their implications on disease transmission, one of the main recorded factors (PROTHERO, 1977). An intricate understanding of human mobility helps epidemiologists and health agencies pinpoint outbreak regions and implement targeted interventions. Although some studies report employing such high-granulated location data, the major

drawbacks were the limited number of participants and the high costs of distributing GPS-logger devices (VAZQUEZ-PROKOPEC et al., 2010, 2013).

Most of the population, including middle and low-income groups, have access to internet-enabled smartphones with an inbuilt GPS device (JAMES, 2020). Google LLC stores Android smartphone users' location data for commercial purposes and provides a personalized user experience (GOOGLE, [s.d.]). While some studies have quantified and evaluated the utility of Google Location History (GLH) data (COOLS et al., 2021; RUKTANONCHAI et al., 2018), a comprehensive exploration to understand the extent of its usefulness in studying transmission of infectious diseases has not yet been conducted. The present study investigates the potential application of GLH data obtained from patients' smartphones in understanding vector-borne and human-to-human disease transmission in a clinically monitored setting.

CHAPTER II

2. DIGITAL SURVEILLANCE OF VECTOR-BORNE TRANSMISSION

2.1 Introduction

Vector-borne transmission has long been a major concern in public health. Modern transport and globalization caused a resurgence in vector-borne infections from the 20th century (GUBLER, 1998). Mosquito-borne diseases share a significant contribution to the global vector-borne disease burden, causing over 627,000 deaths in 2020 (FRANKLINOS et al., 2019; ORGANIZATION, 2021). Human malaria remains the top vector-borne concern in tropical countries (SINGH et al., 2009). For example, Brazil registers over 40% of malaria cases in Latin America, and 99.6% of those are reported from the Amazonas region (OLIVEIRA-FERREIRA et al., 2010). The prevalence of other mosquito-borne diseases, such as chikungunya and dengue, has substantially increased (ARAÚJO et al., 2017, p. 200; COLLUCCI, 2016).

One of the most effective ways to reduce surging mosquito-borne transmission is by identifying transmission hotspots and implementing targeted intervention strategies (BOUSEMA et al., 2012). One strategy is "Reactive Case Detection (RACD)", which screens households or people in the neighborhood where malaria infection is confirmed. This case detection aims to identify additional infected symptomatic and asymptomatic cases (PERERA; CALDERA; WICKREMASINGHE, 2020). This information is then used to pinpoint places infected individuals have visited and locate potential mosquito breeding sites. Historically, such surveyed information has been coupled with GIS systems to find possible hotspot locations and interpret underlying spatial, environmental, and geographical factors (HIGHTOWER et al., 1998; KENGLUECHA et al., 2005; SHAFFER et al., 2020; WANJALA et al., 2011). However, this location information obtained through patient interviews is often inaccurate and incomplete because it depends on patients' memories, reducing the success rate of RACD. GLH data from the smartphones of the infected individuals could offer a more detailed exploration of places visited by the patients in the previous few days to several years. Moreover, the GLH data from smartphones can help quantify infected users' mobility patterns, which have been reported to correlate to mosquito-borne transmission (ABDUL-GHANI et al., 2020; PROTHERO, 1977; RUKTANONCHAI et al., 2016, 2018; WESOLOWSKI et al., 2012). Overall, the present chapter highlights the usability of this retrospective

location data to detect mosquito breeding sites and explore mobility patterns contributing to the infection.

2.2 Objective

To evaluate the efficacy of utilizing retrospective location data from smartphones of malaria patients to pinpoint potential mosquito-breeding hotspots and analyze relevant patient mobility patterns contributing to infection.

2.3 Methods

2.3.1 Data collection

2.3.1.1 Data collection platform

The participants in this study used the online platform Sickness Positioning System (SiPoS), accessible at https://sipos.fcf.usp.br/, to submit the location history data from their smartphones. This platform anonymously retrieves the GLH data from Android devices. It does not store personal information such as name, email, or phone number (CARDOZO, 2019).

2.3.1.2 Data collection from malaria patients

Fundação de Medicina Tropical Doutor Heitor Vieira Dourado (FMT-HVD) in Manaus, who actively engages in malaria research and diagnostics, collaborated to collect location history data from the smartphones of malaria patients. Those individuals diagnosed with malaria at the FMT-HVD were invited to participate in this research study. With the participants' consent and agreement, they were guided to submit the data. The medical collaborator assigned each participating individual a unique code, allowing the data in the storage server to remain anonymous.

Considering the sensitivity of the data, ethical committee approval was obtained (CAAE: 68428917.0.0000.0005, Parecer: 2.135.257).

2.3.2 Data processing

2.3.2.1 Filtering noise/outlier points

GPS trajectories often include coordinates that individuals have never actually visited. The primary reason for this data discrepancy is attributed to residual errors such as clock-related, signal propagation, and system errors (KARAIM; ELSHEIKH;

NOURELDIN, 2018). One effective method to identify these erroneous coordinates within a trajectory is to calculate the speed between the consecutive GPS points and apply a speed threshold. Points exceeding this threshold are considered noise and excluded from the trajectory. A 500 km/hr speed threshold was applied for the outlier detection and exclusion.

Moreover, every GPS point within the trajectory is accompanied by an estimated margin of error known as the accuracy radius error. This measurement indicates a possible dispersion of the recorded coordinates from the actual location, i.e., the actual coordinates are likely to be within the specified radius error measurement. GPS points with more than 50 meters of accuracy radius error were discarded from the trajectories.

2.3.2.2 Stop location detection

A stop location of an individual refers to a specific location where the individual remained stationary for a particular duration before resuming the journey. Identifying a stop location within a GPS trajectory requires a distance and time threshold.

The stop algorithm checks for two conditions: first, if a sequence of consecutive points in a GPS trajectory falls within the specified distance threshold (e.g., 100 meters), and second, if the time elapsed between the first and last points of the same sequence exceeds the time threshold (e.g., 15 minutes) (ZHENG, 2015). The sequence of points satisfying the above conditions represents a stop location. The mean coordinates of these recorded points are considered as the stop location coordinates. Coordinate points falling outside these stop sequence points are labeled moving points captured between the stop locations. The stop location detection process is repeated until the end of the trajectory (Table 1).

A distance threshold of 100 meters and a time threshold of 15 minutes was applied.

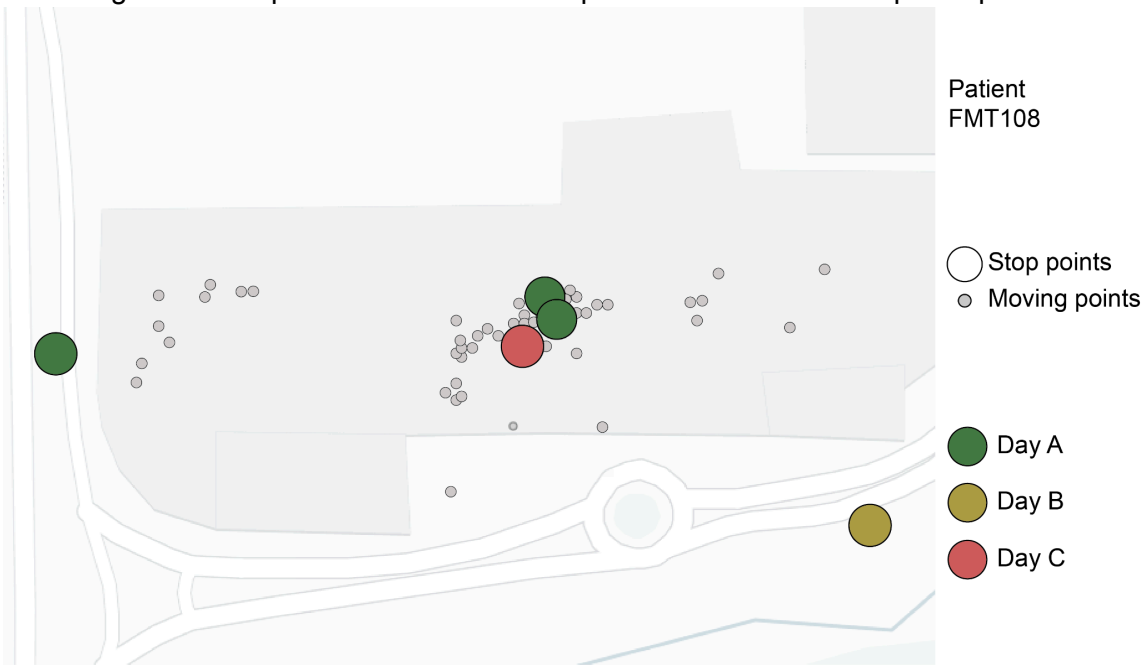
Table 1 - Stop point location detection algorithm

Stop Point Algorithm:
1. Initialize an empty list of stop points S
2. For each anchor point p_i in a chronologically ordered Trajectory T
a. Find the successors of p_i, i.e., p_{i+1}, p_{i+2}, p_{i+3},...,p_n
b. Calculate the distance between p_i and each successor point p_j

c. If there exists a successor point p_k (where i < k <= n) such that the distance between p_i and p_k is greater than $d_{threshold}$
   i. Calculate the time difference between p_i and the last successor point p_l (where i < l <= k) within the $d_{threshold}$
   ii. If the time difference is greater than $t_{threshold}$, a stay point is detected
   iii. If the stay point is detected, assign the mean of coordinates p_i, p_{i+1}, p_{i+2},...,p_l as the stop location coordinates. Add the mean coordinates to the list of stop points
   iv. Set the next anchor point as p_{i+1}
   v. Continue the next iteration of stop detection from step 2a.
3. Return the list of stop points.

Source: Modified version of (ZHENG, 2015)

Figure 1 – Stop locations of a malaria patient identified near a public place.



Source: The author, 2023.

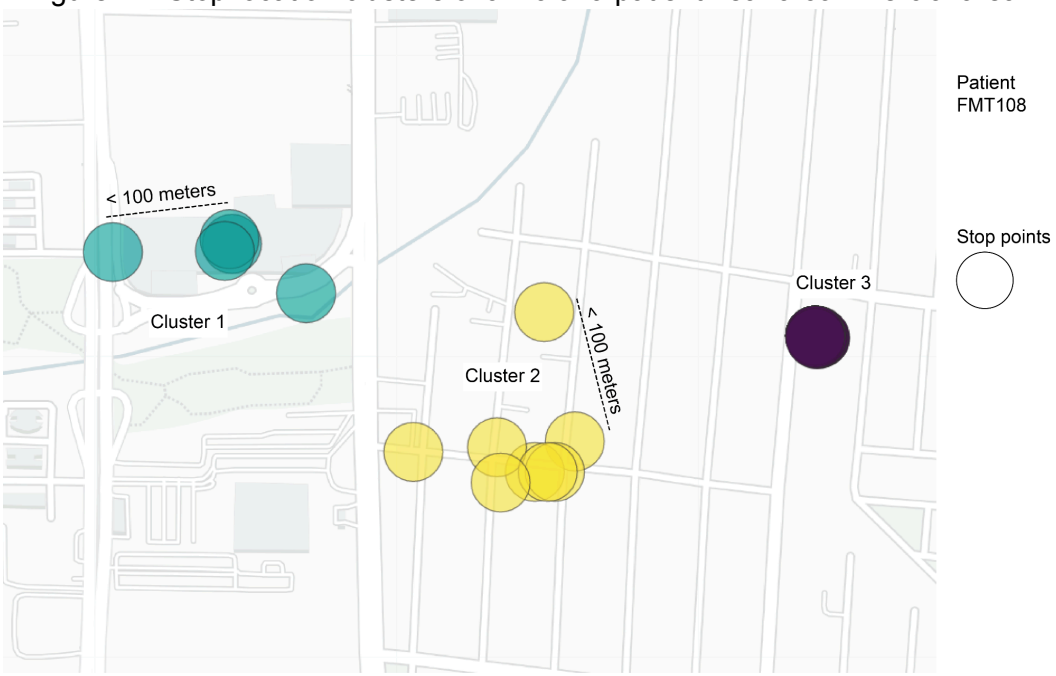
2.3.2.3 Clustering stop locations

The generated list of stop locations commonly includes revisiting locations. A revisiting location corresponds to a visit to the same location but at a different time, such as home and workplace. It is necessary to group and label such occurrences to identify and analyze patterns associated with those places. A density-based clustering algorithm, Density-based spatial clustering of applications with noise

(DBSCAN), was utilized to categorize these places (PEDREGOSA, 2011). The algorithm mainly uses two parameters, a clustering radius and a minimum number of stop points, to be considered for cluster formation.

A clustering radius of 100 meters with a minimum stop point of one is used to cluster and label the stop point locations.

Figure 2 – Stop location clusters of a malaria patient near a commercial area.



Source: The author, 2023.

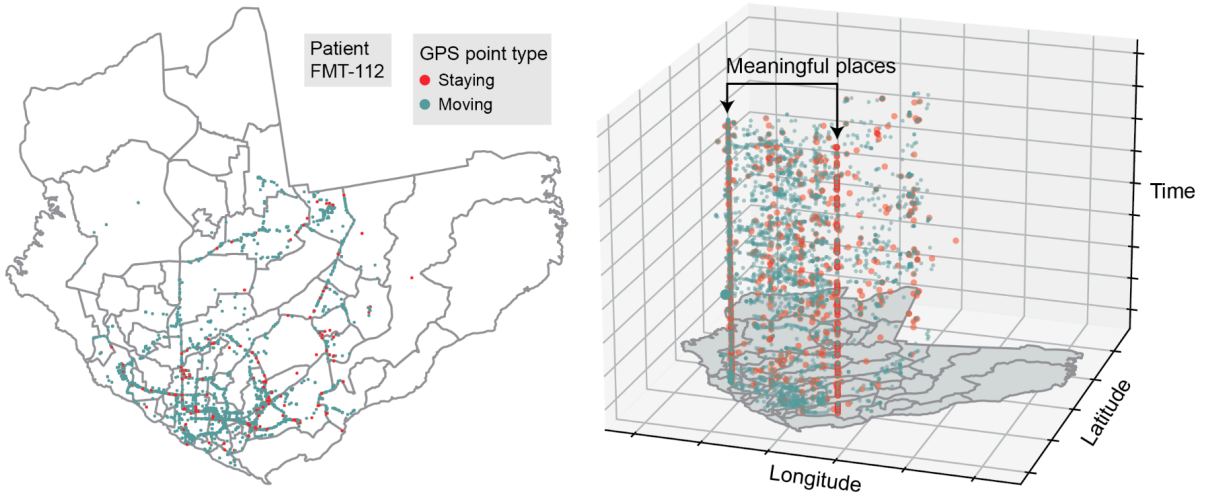
2.4 Results and discussion

Over two hundred and fifty malaria patients participated in the submission of location history data for this study. First, participants with no GLH data (i.e. an absent or empty JSON file) were excluded from the study. Second, location history recorded only from the day of diagnosis to 90 days before the day of diagnosis was used for further processing and the rest of the data was discarded. At last, only a hundred and nine patients' GLH data had usable location data for further analysis.

2.4.1 Exploring locations visited by malaria patients

Each patient's data was processed separately. After filtering/deleting outlier GPS points, locations visited by each patient where the patient stayed at least for 15 minutes were identified (refer to methods 2.3.2). Figure 3 shows locations the malaria patient FMT-112 visited within the 90-day period in the municipality of Manaus.

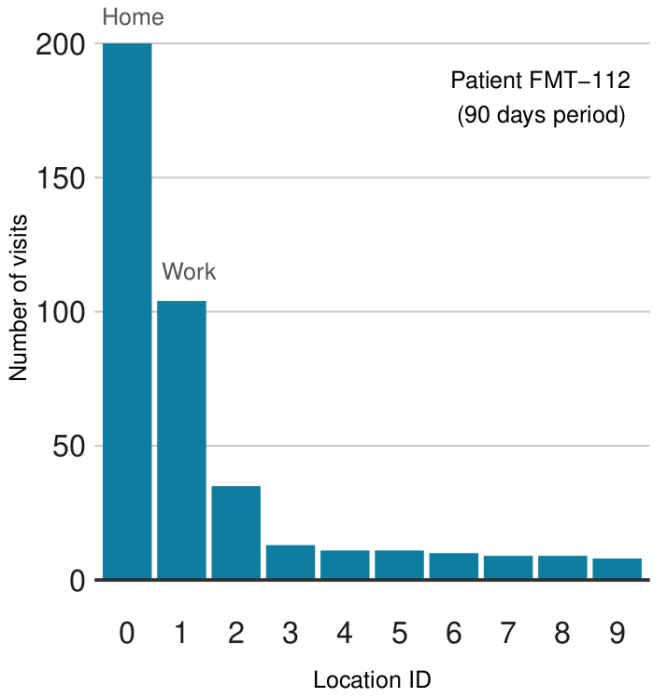
Figure 3 – Stop point locations and moving points of the malaria patient FMT-112 during the 90-day period.



2D and 3D maps showing the identified stop points (red) and moving points (blue), i.e., GPS points captured while moving from one place to another. Meaningful places marked on the 3D map represent the home and work location of the patient revisited during the 90 days. Source: The author, 2023.

These identified visited locations generally include revisited locations, such as home or work. Such locations were determined through clustering (refer to methods 3.2.3). The top ten unique visitations of the patient FMT-112 in the ninety-day period are shown in Figure 4.

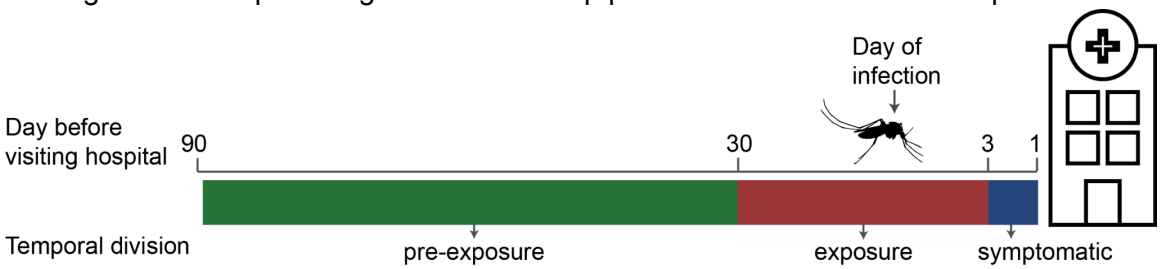
Figure 4 – Frequency of top 10 unique locations visited by the malaria patient FMT-112.



Source: The author, 2023.

All the visited locations by the patients were temporally classified based on the number of days relative to their diagnosis date. This temporal segmentation included three periods: ‘symptomatic’ if the visited location was within three days before the diagnosis date; ‘exposure’ if the visited location was between 3 and 30 days before the diagnosis date; ‘pre-exposure’ if the visit was 60 and 90 days before diagnosis date (Figure 5). This temporal segmentation was mainly determined by the reported mean time between the exposure and the onset of symptoms (NISHIURA et al., 2007).

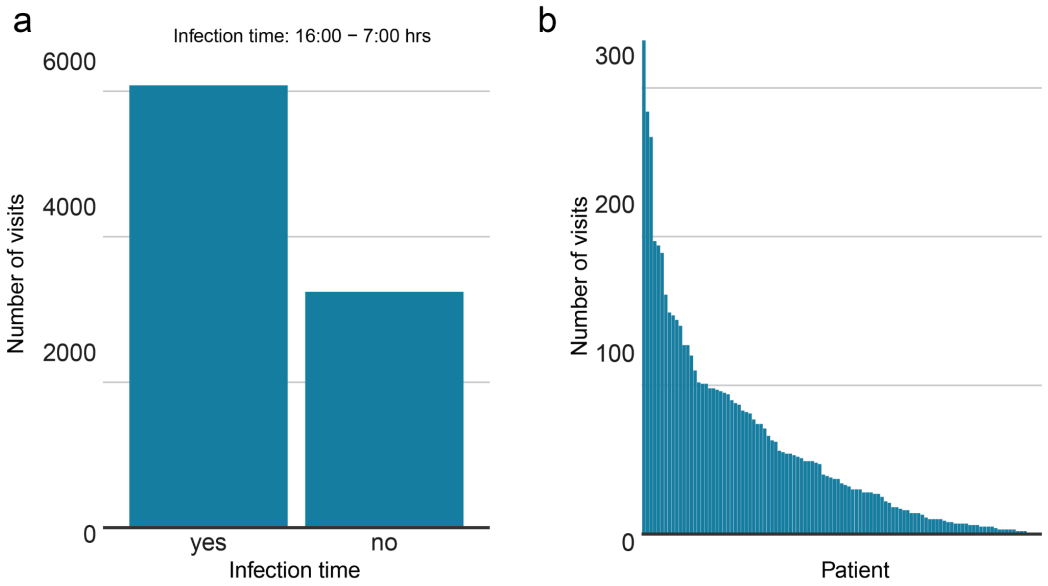
Figure 5 – Temporal segmentation of stop point location data of malaria patients.



Source: The author, 2023.

Since the study aimed to locate hotspot transmission locations, stop locations visited only during the exposure period were used for further analysis. Also, female anopheline mosquitoes have an active feeding time between 17:00 and 6:00 hours (ROZENDAAL, 1989). A buffer hour was added to this range, and the stop locations of the 109 patients that fell between 16:00 and 7:00 hours during the exposure period were extracted. About 1/3rd of the stop locations recorded within the exposure period were outside the above-defined time range (Figure 6a). It is important to note that most of these potential infected stop locations belonged to only a small proportion of overall patients, as shown in Figure 6b. This observation has been reported previously (GONZÁLEZ; HIDALGO; BARABÁSI, 2008).

Figure 6 – Frequency of visits of all malaria patients.



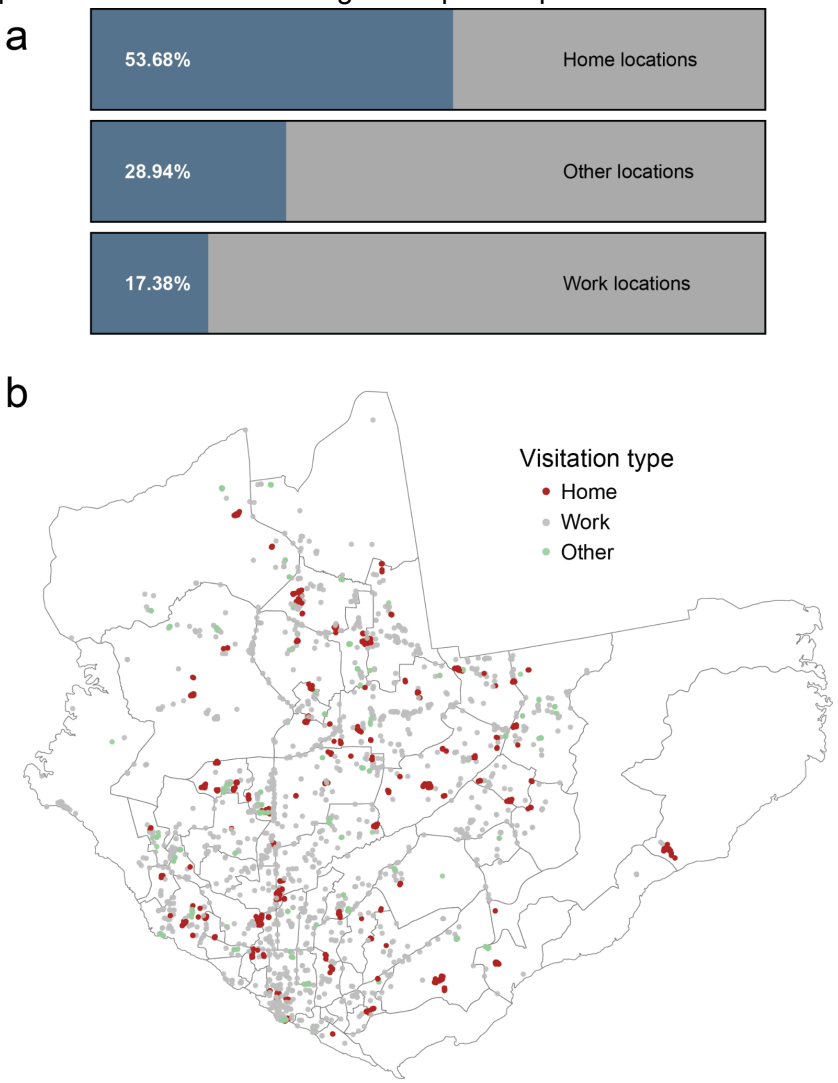
(a) Frequency of visits of all patients separated by potential infection time. (b) Frequency of visits by each patient. Source: The author, 2023.

Only the stop locations of the exposure period identified within the infectious time range were used in the further analysis. More than 50% of these stop locations included home visits. Home and work location coordinates were determined based on the clustering, frequency, and time of visitation (22:00 - 7:00 hours for home and 7:00 - 22:00 hours for work). Figure 7a represents the percentage of total visits categorized by stop location type. Figure 7b shows the Manaus map plotted with these stop locations categorized based on the location type.

Home and work location coordinates are the most used geographical data to map mosquito-borne transmission hotspots. (HIGHTOWER et al., 1998) reported the first implementation of GPS technology to record and spatially map locations such as homes, hospitals, and known mosquito breeding sites for malaria transmission. Other malaria studies have employed similar methods to identify the transmission clusters (ALEMU et al., 2014; WANJALA et al., 2011). Disease control measures such as indoor residual spraying (IRS) and other intervention methods are implemented upon identifying the home and work clusters. Spatial mapping of home and workplaces is also standard practice in chikungunya disease (NSOESIE et al., 2015), dengue (CHANSANG; KITTAYAPONG, 2007), and other vector-borne disease surveillance (WASHINGTON et al., 2004). This crucial epidemiological and surveillance information was generated from the GLH data of the patients without the involvement of manual data collection methods. Hence, utilizing GLH data could be cost-effective instead of distributing handheld GPS devices for field agents to record patients' home and work locations.

In addition to home and work locations, GLH data consists of other visited places by the patients, as shown in Figure 7b. These other visited locations also represent potential zones where mosquito-breeding sites could exist. One way to narrow down these additional visited locations for potential hotspots is to overlay this visitation information with regions where malaria cases are frequently reported. This overlayed information could be favorable in detecting new hotspot locations near the previously reported infected zones with similar landscape characteristics (CARDO; VEZZANI; CARBAJO, 2013).

Figure 7 – Stop locations identified during the exposure period and within the infectious time.



(a) Proportion of stop locations categorized by type of stop location (home, work, and other). (b) Manaus map with stop locations visited by all 109 patients. The stop locations are colored by the type. Source: The author, 2023.

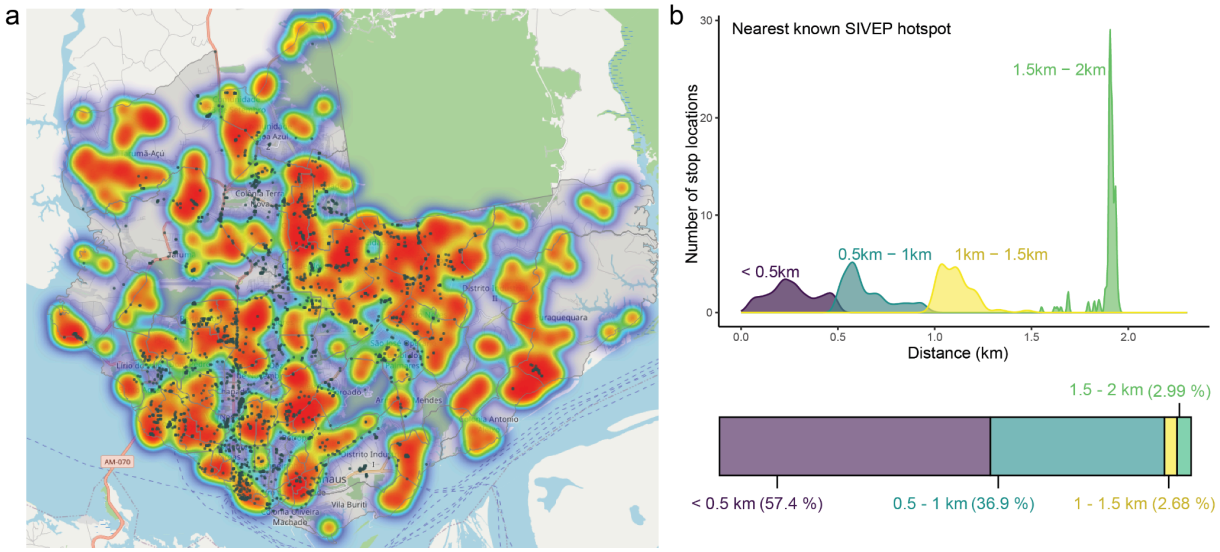
2.4.2 Overlapping patient visits with government-defined hotspots

Sistema de Informação de Vigilância Epidemiológica (SIVEP) - Malaria is an epidemiological surveillance database created by Sistema Único de Saúde (SUS), a Brazilian government-run healthcare system, where diagnosis and surveyed information of reported malaria patients is stored. The database includes the list of approximate location coordinates where malaria cases were previously reported.

These SIVEP location coordinates were overlapped with the identified stop locations (home, work, and others) of all patients. Figure 8a shows the overlapped stop locations and SIVEP regions. Further, the distance from each stop location to all the SIVEP hotspots was calculated, and the nearest SIVEP hotspot from each stop

location was identified. Over half of the stop locations were within 0.5 km from a SIVEP hotspot, while around 94% were under 1 km from a reported SIVEP hotspot. Notably, all the stop locations were within 2 km from the nearest hotspot (Figure 8b). The flight range of the Anopheles darlingi (*An. darlingi*) mosquito species, a major malaria-causing species in the Amazonas region, generally flies up to 2 km in its lifetime (CDC, [s.d.]), highlighting that all stop locations in the exposure period within the infectious time-range fall under potential infection zones.

Figure 8 – Overlapping stop location data with SIVEP-Malaria hotspot data.



(a) Manaus map with SIVEP hotspots data (heatmap) overlapped with the identified stop locations of patients (points). b) Density of distance measurements between stop locations to the nearest SIVEP hotspot. Source: The author, 2023.

2.4.3 Implementing clustering on potential transmission locations

Another way to exploit the stop location information is by employing clustering methods. Although spatial clustering has previously been employed to identify the transmission hotspots, the location history information in such studies was primarily determined through patient interviews (COLEMAN et al., 2009).

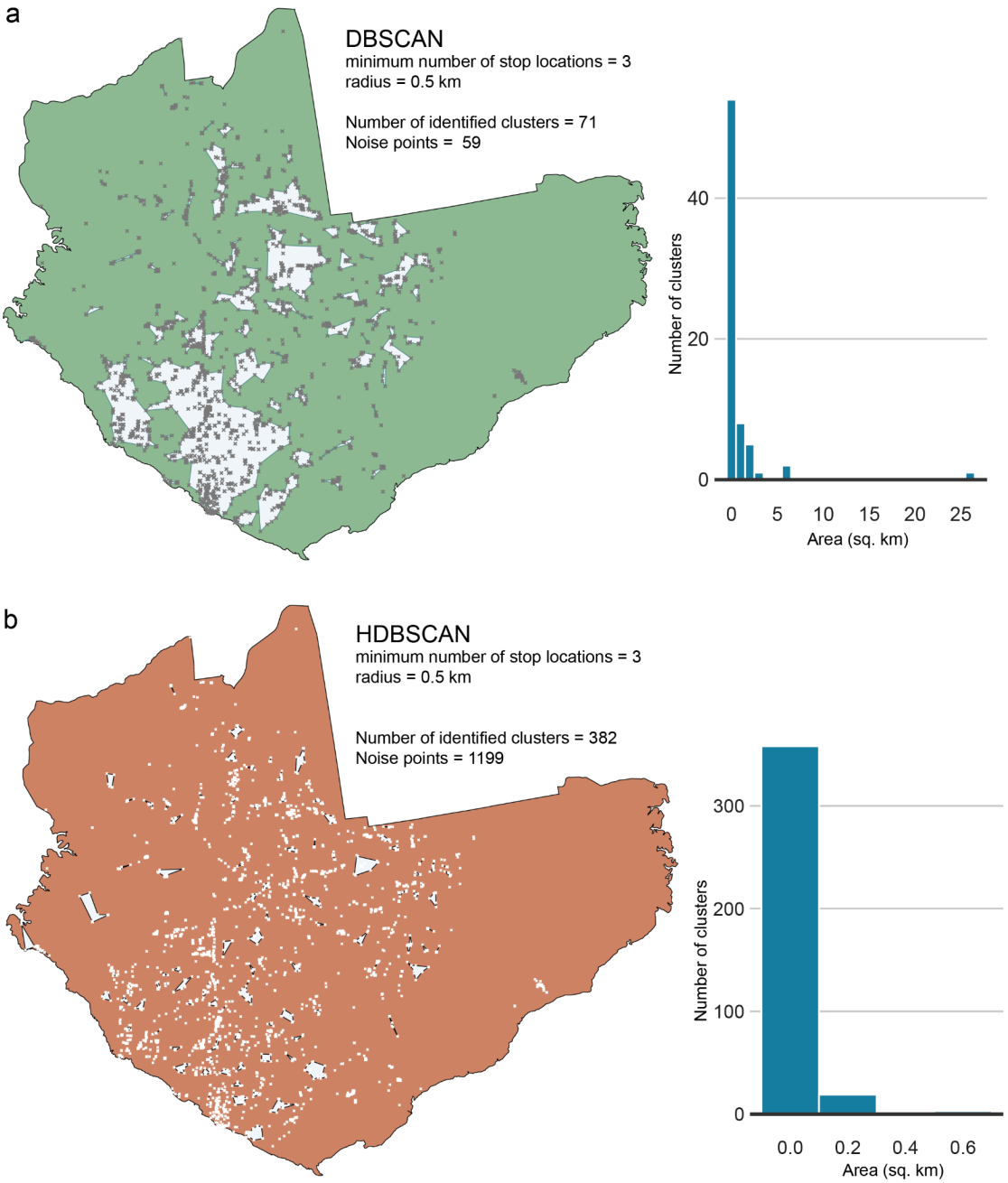
Density-based clustering methods help identify clusters based on proximity and separation of points (BHADANE; SHAH, 2020). Unlike K-means clustering, the density-based algorithms do not require defining the number of clusters and can detect clusters of varying shapes. Implementing density-based clustering such as DBSCAN for the stop locations of the patients requires two parameters. A minimum number of points (stop locations) to form a cluster, and a specified distance within which neighboring points must lie to be considered part of the same cluster. The

DBSCAN algorithm was applied by establishing a minimum of 3 stop locations and a cluster radius of 0.5 km (epsilon = cluster radius km / 6371.0088). Other parameters, the *'haversine'* distance metric and the *'ball_tree'* Nearest Neighbors algorithm, were utilized (PEDREGOSA, 2011).

Within the Manaus municipality, 71 clusters were identified, each exhibiting various shapes. The areas of these clusters spanned between 0.1 sq. km and 25 sq. km (Figure 9a). As expected, clusters with large sizes, i.e, high density of visits, were found in the Manaus city center (Centro) and the highest population density areas, such as Compensa and Cidade Nova (SEPLAN-CTI, 2015).

Although epidemiologically useful, in practice, finding mosquito-breeding sites and implementing intervention strategies in large irregular areas could be resource and time-intensive. Therefore, a modified version of the DBSCAN method called Hierarchical density-based spatial clustering of applications with noise (HDBSCAN) (CAMPELLO; MOULAVI; SANDER, 2013; PEDREGOSA, 2011) was implemented on the stop location information. In addition to the density and sparsity aspect, HDBSCAN constructs a hierarchy of clusters, splits the clusters based on varying point densities, and finds the most stable clusters. Using the same parameters as applied to the DBSCAN method, HDBSCAN resulted in 382 clusters, with the largest cluster covering an area of mere 0.6 sq. km (Figure 9b), thus identifying significantly smaller clusters.

Figure 9 – Density-based clustering of stop locations of all malaria patients.



Clusters generated by DBSCAN (a) and HDBSCAN (b) plotted on the Manaus map, along with their respective distribution of areas (sq. km) Source: The author, 2023.

2.4.4 Quantifying malaria hotspot clusters

While the smaller clusters have been identified, it is necessary to determine the importance of each cluster to prioritize them during the field survey to implement mosquito-control intervention strategies. One factor that could ascertain a cluster's importance is finding ground truth labels, such as SIVEP hotspots inside and adjacent to the boundaries of these clusters. This distribution of SIVEP hotspots and the number of previously reported malaria cases at these hotspots could provide

robust ground truth knowledge. Further, the number of current reported patients that visited these clusters could help understand the severity of existing disease spread within the cluster.

Hence, incorporating these variables, a score was calculated for each cluster. Based on the importance of the variable metric, weights w_1, w_2 and w_3 were assigned to SIVEP hotspot density ($SIVEP_{Density}$), average number of historical cases ($Avg_{Historical\ Cases}$) and current patient density of cluster ($Patient_{Density}$), respectively. The cluster score formula was defined as follows:

$$Cluster\ Score = (w_1 \times SIVEP_{Density}) + (w_2 \times Avg_{Historical\ Cases}) + (w_3 \times Patient_{Density})$$

where, $w_1 = 0.5$, $w_2 = 0.3$, $w_3 = 0.2$

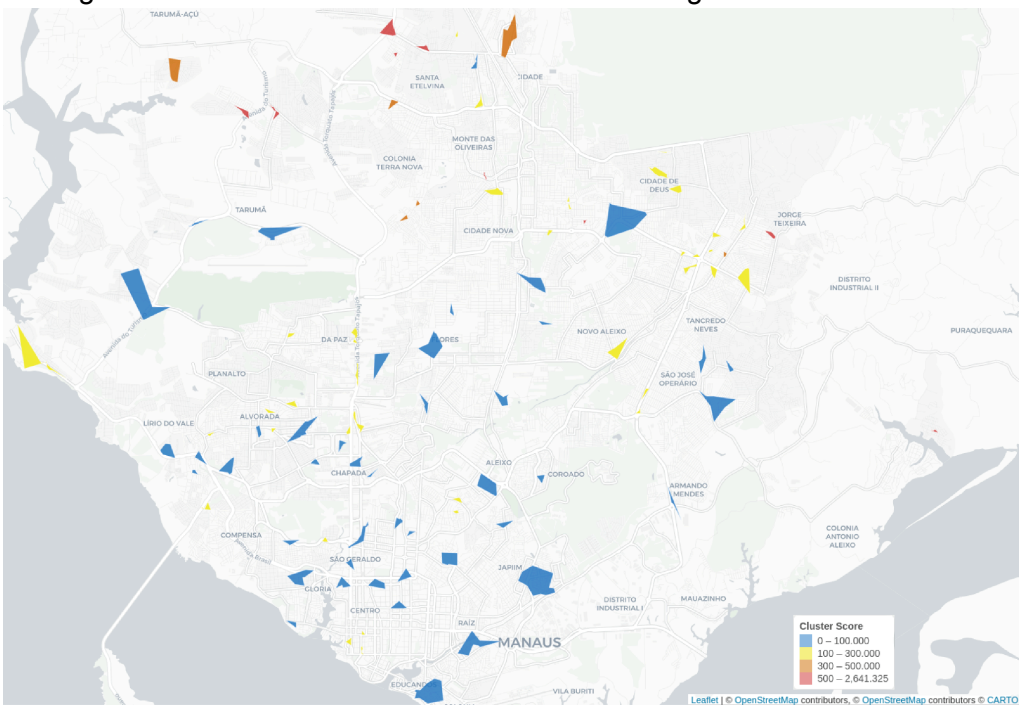
$$SIVEP_{Density} = \frac{No.\ of\ SIVEP\ hotspots\ inside\ cluster + No.\ of\ SIVEP\ hotspots\ within\ 0.5\ km\ outside\ cluster}{Area\ of\ the\ cluster}$$

$$Avg_{Historical\ Cases} = \frac{Total\ no.\ of\ historical\ cases\ at\ the\ SIVEP\ Hotspots}{No.\ of\ SIVEP\ hotspots\ inside\ cluster + No.\ of\ SIVEP\ hotspots\ within\ 0.5\ km\ outside\ cluster}$$

$$Patient_{Density} = \frac{Number\ of\ presently\ identified\ patients\ within\ the\ cluster}{Area\ of\ the\ cluster}$$

Almost all clusters in the city center and densely populated regions had the lowest ranking score (Figure 10). Most clusters with the highest score were observed in the outskirts of Manaus. This observation is apparent because Amazonas forests border the city's outskirts, where larval breeding in the water collections at the forest fringes is often reported (BARROS; HONÓRIO, 2015). Since Manaus has a sizable irregular forest cover scattered inside the city, the bordering regions of these forest patches within the Manaus city could also present potential mosquito-breeding sites. Thus, the resulting clustering scores provide valuable insight for prioritizing certain clusters and offer a strategic focus for targeting potential mosquito-breeding sites within and around the city.

Figure 10 – HDBSCAN clusters colored according to the cluster score.



High scored HDBSCAN clusters identified in the outskirts of Manaus, whereas low scored clusters in the high population density regions. Source: The author, 2023.

It is also crucial to acknowledge that the clustering scores may not represent the completed picture. For instance, SIVEP hotspots, rather than pinpointing the exact location of each infected case, multiple cases are often grouped under a single hotspot location, resulting in potential oversimplification of the actual distribution and concentration of cases. Furthermore, ecological factors such as intermediate forest cover (LAPORTA et al., 2021) and stagnant water bodies (CATRY et al., 2018) around the clusters can be incorporated into the calculation of clustering score to provide a more comprehensive assessment.

Finally, the potential mosquito-breeding sites previously validated by the field agents using only the patients' stop location information were overlapped with the identified HDBSCAN clusters to validate the clustering findings. Figure 11 shows the presence of suspected and confirmed *An. darlingi* larvae breeding sites at the marked locations on the Manaus satellite maps. Suspected locations signify vast locations that could not be fully assessed but had a history of larvae presence. All the clusters near the suspected breeding sites have the highest cluster score. The validated positive larvae breeding site near cluster 8 and the microscopic image of collected *An. darlingi* larvae are shown in Figure 11.

Moreover, the stop locations of the patients inside these clusters near the suspected and positive breeding sites were checked for the frequency of visitation to these locations by the patients, i.e., how frequently the patients visited these locations during the exposure period and within the infectious time range. Notably, the majority of these stop locations had minimal visitation frequency, as depicted in the violin plot of Figure 11. This implies that these visits were not part of the patients' regular visits and were likely visited for the first time during the exposure period, highlighting the potential role of newly visited places in infection before the onset of symptoms. This quantification of critical mobility patterns associated with the infection showcases potential utility of location history data from smartphones of the patients.

Figure 11 – HDBSCAN clusters and nearby validated Anopheles breeding sites by field agents.

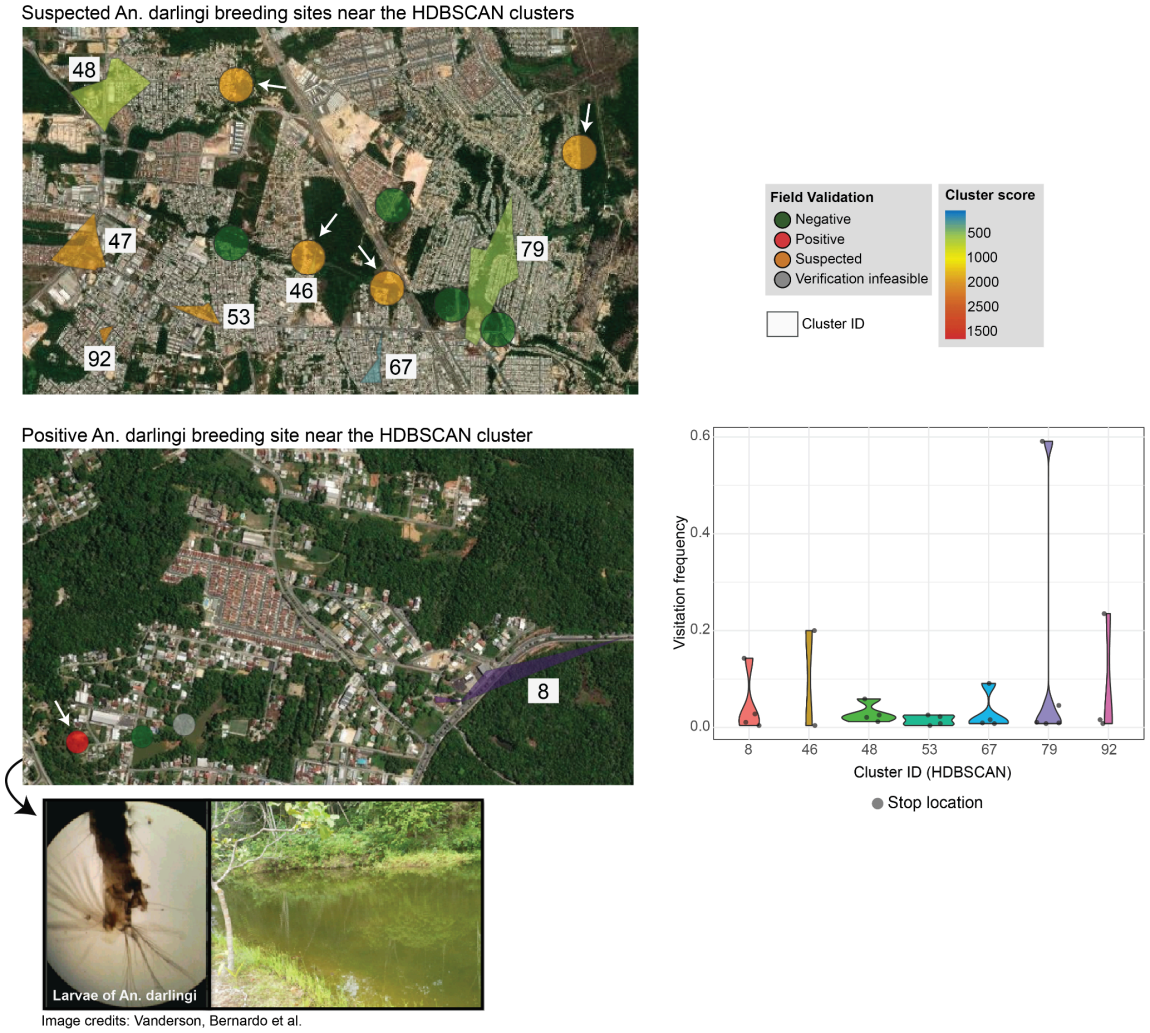


Image credits: Vanderson, Bernardo et al.

Zoomed satellite maps showing suspected and confirmed positive Anopheles breeding sites identified near HDBSCAN clusters derived from stop locations of patients. The violin plot shows visitation frequency of stop locations at each of these clusters. Source: The author, 2023.

2.5 Conclusion

In summary, this study has successfully leveraged the location history data from the smartphones of malaria patients to uncover potential transmission hotspots within the Manaus municipality. The step-by-step processing and analysis of the GPS history data of the patients helped identify key locations visited by the patients before the onset of symptoms where the patients potentially were infected. These potential transmission visits significantly overlap with government-defined (SIVEP) hotspots, validating the reliability of using GLH data. Furthermore, implementing density-based clustering methods on this data facilitated identifying and isolating smaller and more specific potential transmission zones within the Manaus region. The study employed a scoring system to rank the identified clusters, revealing those in the outskirts of Manaus, bordered by the Amazonas forests, as high-priority regions for further investigation and intervention. The study additionally emphasized the significance of certain visited locations based on their visitation frequency. Locations less frequently visited, i.e., non-regular visits of the patients, were pinpointed near the validated suspected and confirmed *An. darlingi* larvae breeding sites, thus underscoring the potential role of such locations in malaria transmission. This innovative use of GLH data presents an efficient and cost-effective approach for enhancing malaria surveillance, aiding in the timely implementation of control and intervention measures.

CHAPTER III

3. DIGITAL SURVEILLANCE OF HUMAN-TO-HUMAN TRANSMISSION

3.1 Introduction

Infectious diseases that spread from one individual to another represent a considerable portion of the worldwide disease prevalence. In 2019, tuberculosis was responsible for 1.2 million fatalities (WHO, 2020a), making up 20% of all deaths from infectious diseases. Besides, emerging communicable infections such as COVID-19 surpassed it in the number of deaths between 2020 and 2023 (WHO, 2023). Respiratory tract infections such as influenza also cause substantial yearly hospitalizations worldwide (LAFOND et al., 2021). Despite the eradication attempts of other respiratory diseases, such as measles (DE QUADROS et al., 2004), which transmits through airborne droplets, it has seen a resurgence in Manaus, the Amazonas region of Brazil, in 2018 (ELIDIO et al., 2019). Urbanization, human density, and global mobility have been the contributing factors in amplifying the risk of this human-to-human transmission and emerging infections (ALAM, 2021; NEIDERUD, 2015; SANTIAGO-ALARCON; MACGREGOR-FORS, 2020; TARWATER; MARTIN, 2001).

Contact tracing and contact investigation evolved as crucial strategies in mitigating the spread of human-to-human transmission. The primary objective of contact tracing involves identifying and monitoring individuals who have been in close proximity to a network of potentially infected contacts (EAMES; KEELING, 2003). The history of contact tracing traces back to the 1600s during the Great Plague (NAKAYAMA, 2022) and in 18th-century Europe, where potentially infected people such as travelers were identified and quarantined to stop the spread of the infection (TOGNOTTI, 2013). On the other hand, contact investigation focuses on identifying possible infection sources and understanding the conditions under which the transmission occurred. In the contemporary era, contact tracing and contact investigation effectively reduce the spread of infections like tuberculosis (ERKENS et al., 2010; SHRIVASTAVA; SHRIVASTAVA; RAMASAMY, 2014). These methods were also crucial in successfully containing the MERS outbreak 2013 (MAILLES et al., 2013).

In the recent COVID-19 pandemic, digital contact tracing systems have been introduced to identify potential transmission risks, offering an efficient alternative to

traditional manual contact tracing techniques (ELMESALAWY; SALAMA; ANANY, 2020; HEO et al., 2020; RAHMAN et al., 2020; SINGAPORE, [s.d.]). These digital systems are primarily based on bluetooth and location technologies, which notify potential exposure risks and facilitate health bodies to craft disease surveillance and health policies dynamically. Furthermore, human mobility datasets from private enterprises enabled us to understand the impact of population movement patterns preceding, during, and post-pandemic (HUANG et al., 2020; LEE et al., 2020). However, it is essential to note that these digital disease surveillance systems and human mobility reports were rapidly deployed in response to the urgency of the pandemic, leading to potential oversights of disease dynamics. To date, there has been a lack of studies conducted in a clinically monitored environment to quantify and explore this crucial digital data gathered from patients for monitoring and investigating the underlying causes of transmission.

The present chapter explores the usability of location history data from the smartphones of index cases and their close contacts exposed to a human-to-human transmissible infection such as tuberculosis (TB).

3.2 Objective

To assess the effectiveness of employing location data from smartphones of tuberculosis patients and their close contacts for contact investigation and discerning critical mobility patterns related to infection transmission.

3.3 Methods

3.3.1 Data collection

3.3.1.1 Data collection platform

The study participants provided their location history (GLH) data from their Android smartphones via the SiPoS platform, accessed at https://sipos.fcf.usp.br/ (CARDOZO, 2019) or through the REDCap platform. The GLH obtained through the REDCap platform was made anonymous before undergoing data storage and processing. The anonymizing process involved removing metadata from the acquired Google Takeout location history file.

3.3.1.2 Data collection from TB patients and their close contacts

Instituto Brasileiro para Investigação da Tuberculose (IBIT), specializing in international TB monitoring, collaborated to collect location data from the individuals presenting TB symptoms at the Fundação José Silveira Hospital in Salvador, Bahia.

Once diagnosed with TB, the patients and their immediate close contacts were invited to join the study. Participants were categorized into two cohorts: Cohort A, which included diagnosed active TB patients, and Cohort B, which comprised their close contacts (Table 2). Every participant underwent an Interferon-Gamma Release Assay (IGRA) test on the day of enrollment and at the end of the month six after enrollment. IGRA test checks for a person's immune reactivity to *M. Tuberculosis* and aids in the detection of previous infection to TB and latent TB infection. Based on Cohort B participants' IGRA test results on the enrollment day and six months after enrollment, Cohort B participants were categorized into four groups (Table 3).

Considering the sensitivity of the data, ethical committee approval was obtained (CAAE number 87022618.4.0000.5543, Parecer: 2.614.181).

Table 2 - Tuberculosis cohorts

Cohort	Description
A	Patients diagnosed with active tuberculosis (index cases)
B	Participants in close contact with Cohort A patients

Source: The author, 2023.

Table 3 - Outcome of participants in Cohort B

Outcome of Cohort B	
1.	IGRA-Negative at Baseline and IGRA-Positive at Month 6
2.	IGRA-Negative at Baseline and at Month 6
3.	IGRA-Positive at Baseline
4.	IGRA-Negative at Baseline and progressed to active TB at month 6

Source: The author, 2023.

3.3.2 Data processing

The processing of the acquired location history data from the participants was carried out as detailed in the methods section of chapter one (2.3.2), and the steps include:

1. Filtering noise/outlier points (section 2.3.2.1)
2. Stop location detection (section 2.3.2.2)
3. Clustering stop locations (section 2.3.2.3)

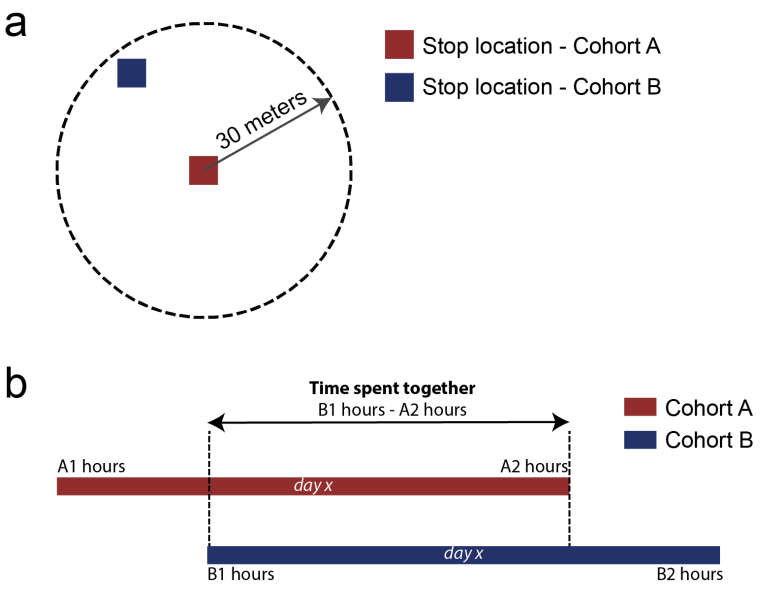
3.3.3 Quantifying spatial-temporal overlap between participants

The criteria for establishing a spatial-temporal overlap between an index case and a close contact i.e, if the index case and close contact were in the same place at the same time, was determined by two specific conditions, as outlined below:

1. If a stop location exists where both the index case and the corresponding close contact are within a maximum distance of 30 meters from each other (Figure 12a).
2. If there is a time interval during which both index case and close contact are present at the above-mentioned overlapping stop location (Figure 12b).

When criteria mentioned above are satisfied, the number of times the index case and the close contact have encountered each other, and the duration of time spent together at these encounters are retrieved.

Figure 12 - Spatial-temporal overlap algorithm



Source: The author, 2023.

3.4 Results and discussion

In this pilot study, one hundred participants were enrolled at IBIT. Of these, forty-six were part of Cohort A (active TB), while the remaining participants belonged to Cohort B (close contacts). Some Cohort A participants had more than one corresponding close contact participating in Cohort B. Any participants from either cohort with missing GLH data, i.e., either having no GLH JSON file or an empty GLH file, were excluded from the study.

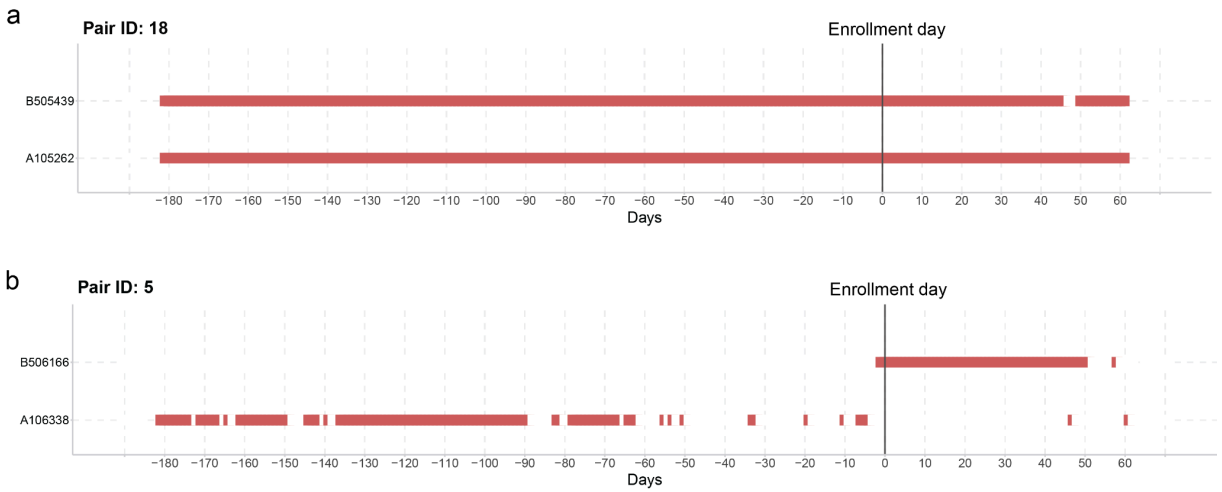
The primary objective of this study was to evaluate the feasibility of employing contact investigation using GPS data to identify and quantify the exposure between individuals in Cohort A and Cohort B. Considering the TB incubation occurs within ten weeks after the initial exposure (BEHR; EDELSTEIN; RAMAKRISHNAN, 2018), the study timeline encompassed location data from six months before, and two months after the participants' enrollment in the study. Since post-enrollment data was needed, data was retrieved at the end of the study, thereby GLH file consisting of location history data both before and after enrollment date.

3.4.1 Evaluating TB exposure risk through spatial-temporal overlap analysis

Given that TB transmission occurs through inhalation of respiratory droplets from an infected person (DONALD et al., 2018), the exposure from the index case to close contact can potentially be determined by examining their GPS data for the availability of spatial-temporal overlaps. In other words, exposure is likely if both individuals are simultaneously found at the same location.

Initially, out of forty-three recruited cohort pairs, only thirteen had requisite location history data available during the study period. At first, each participant's data was processed separately to detect stop locations (refer to methods 3.3.2). Once the data is processed, each cohort pair, i.e., an index case from Cohort A and the corresponding close contact from Cohort B, was accessed for availability of stop locations during the study period. To accurately quantify the spatial-temporal overlap between paired participants, the stop location data for a specific day must be available for both individuals. Finally, only six pairs had intersecting stop location data among the aforementioned thirteen cohort pairs. Figure 13 illustrates intersecting stop location data availability for two example cohort pairs.

Figure 13 - Stop location data of example cohort pairs (pairs 18 and 5)



A timeline graph shows the presence (red) and absence (white) of stop location information between the cohorts each day before and after enrollment during the study period. Cohort pair 18 (a) shows the availability of necessary stop location information to perform a spatial-temporal overlap analysis, whereas pair 5 (b) shows almost no intersecting stop location data between the participants. Source: The author, 2023.

The final six cohort pairs included two pairs with the Cohort B participants converting to IGRA-positive status at the six-month mark, while one of these Cohort B participants also progressed to active TB. Cohort B participants of three other pairs remained IGRA-negative at the end of the six months, whereas the last pair's close contact tested IGRA-positive at the baseline, i.e., on the day of enrollment. Table 4 provides the outcome information of Cohort B participants with their respective cohort pairs IDs.

Table 4 - Pair ID with their corresponding Cohort B outcomes

Pair ID	Cohort A	Cohort B	Outcome of Cohort B
18	A105262	B505439	IGRA-Negative at Baseline; IGRA-positive at Month 6
32	A104180	B504249	IGRA-Negative at Baseline; Active TB at Month 6
20	A105161	B505289	IGRA-Negative at Baseline and at Month 6
22	A105192	B505339	IGRA-Negative at Baseline and at Month 6
23	A105268	B505452	IGRA-Negative at Baseline and at Month 6
7	A105198	B505350	IGRA-Positive at Baseline

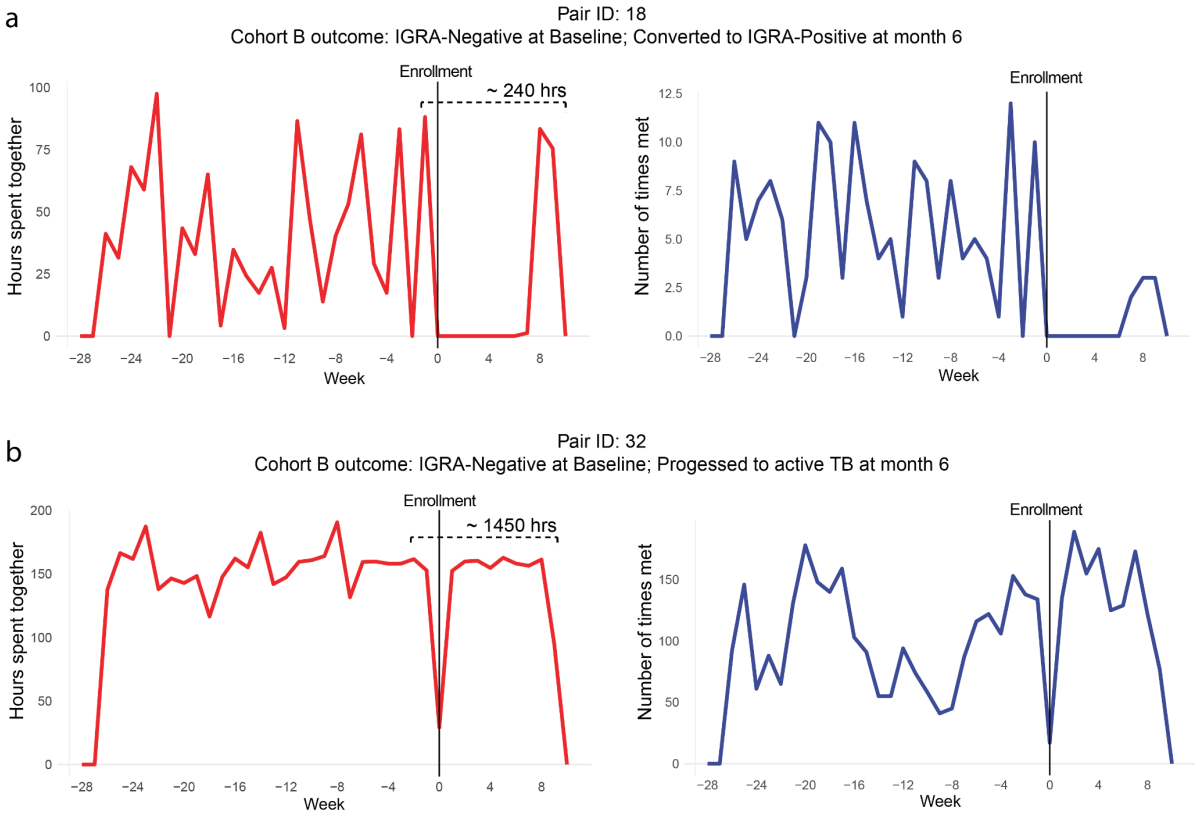
Source: The author, 2023.

The algorithm outlined in the methods section (3.3.3) was used to analyze the spatial-temporal overlap between the index cases and their close contact of all six pairs, hence retrieving the frequency of encounters and the time spent together between the participants in each cohort pair.

The close contact from cohort pair 18, initially IGRA-negative at enrollment, converted to IGRA-positive by the end of the sixth month. Figure 14a shows this Cohort B individual spent over 150 hours with the index case during the two months after enrollment. Notably, the potential for transmission could have existed before the enrollment, as the weekly time spent together between the participants was as high as 90 hours just before the enrollment began. This total duration of exposure observed has previously been identified as a significant predictive factor for the onset of latent tuberculosis infection (REICHLER et al., 2020). Interestingly, there was a decline in hours spent together immediately after the enrollment, suggesting that the active TB participant may have been isolated from the close contact during this period.

The close contact of pair 32, who developed active TB, spent an average of 150 hours every week with the index case throughout the study period (Figure 14b). Firstly, this high spatial-temporal overlap between the participants demonstrates an elevated level of exposure risk from the active TB patient. Such extended exposure to frequent aerosolization of infected respiratory secretions (2005), would intensify the degree of exposure, making the progression to active TB infection more likely (ACUÑA-VILLAORDUÑA et al., 2018). Also, the presence of these respiratory secretions in indoor settings, along with the poor quality of ventilation at the encountered locations directly contribute to the transmission of TB infection (HOUK et al., 1968; KENYON et al., 1996; MOORE et al., 1999; RILEY, 1957). It is also important to recognize factors such as the age and immunosuppression status of the close contacts, which further increase this risk during a potential exposure (CARAUX-PAZ et al., 2021; HASAN et al., 2018).

Figure 14 – Weekly time spent and frequency of encounters of converter cohorts



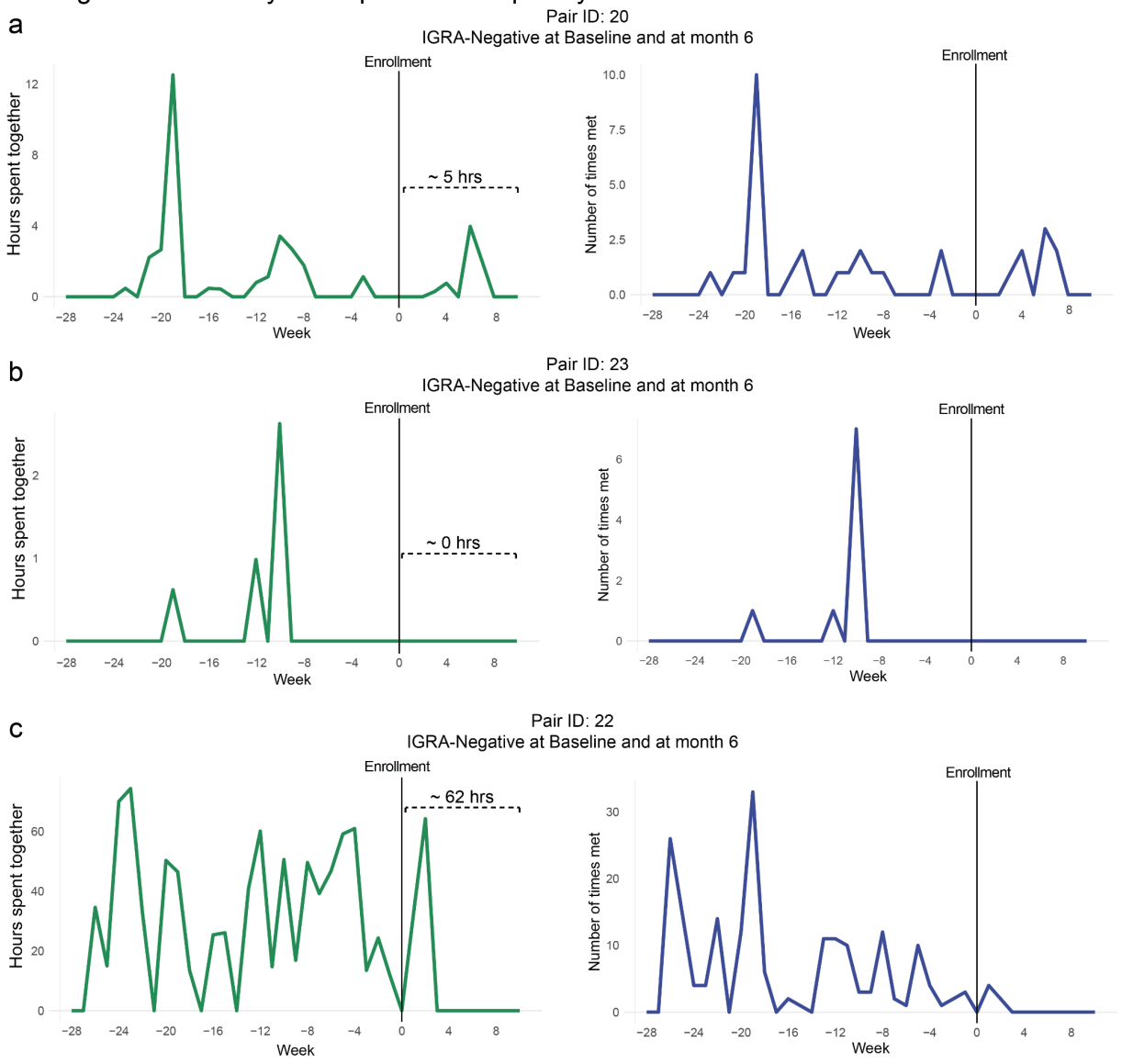
The graphs illustrate high hours of exposure before and after enrollment between the index cases and close contacts of pairs 18 and 32, who converted to IGRA-positive status (one of them also progressing to active TB) at the end of study period. Source: The author, 2023.

Close contacts in three other specific cohort pairs (20, 22, and 23) remained in IGRA-negative status after the sixth month. As illustrated in Figure 15, the time spent between the participants in these pairs was significantly less, ranging from no contact to a maximum of 62 hours in total throughout the two months following enrollment. Conversely, the close contact of pair 7, who tested IGRA-positive on the day of enrollment, had spent over 420 hours with the corresponding index case in the month leading up to the enrollment (Figure 16).

By using location history data from smartphones, this analysis facilitates a digitally-based spatial-temporal contact investigation, providing a comprehensive and quantifiable assessment of potential exposure between index cases and their close contacts. Conventionally, this crucial contact assessment is carried out via direct interviews of index cases or designated proxies, a method susceptible to recall bias (2005). Furthermore, location data objectively records interactions between infected and susceptible individuals. The speed and consistency of automated data collection

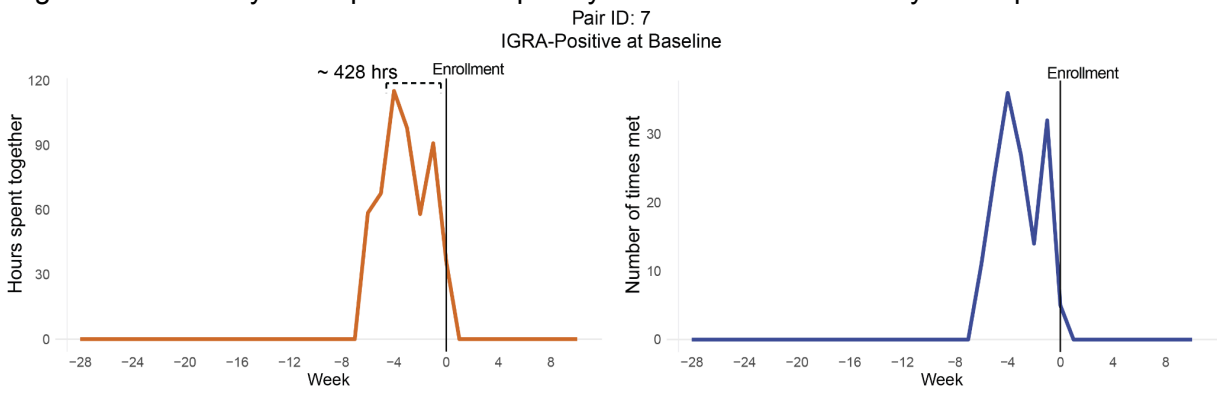
and processing provide additional advantages, enabling a more time-efficient and reliable method for contact investigation. In large outbreak settings, where transmission is prevalent even among casual contacts (DUTHIE et al., 2008), this digital approach proves to be a more effective method for contact investigation.

Figure 15 - Weekly time spent and frequency of encounters of non-converter cohorts



The line graphs illustrate comparatively low hours of exposure before and after enrollment between the index cases and close contacts of pairs 20, 23 and 22, who maintained IGRA-negative status at the end of study period. Source: The author, 2023.

Figure 16 - Weekly time spent and frequency of encounters of already IGRA-positive cohort



The line graphs illustrate high hours of exposure a month before the enrollment between the index case and close contact of pair 7, who tested IGRA-positive on the day of enrollment. Source: The author, 2023.

3.4.2 Movement patterns associated with active TB patients and close contacts

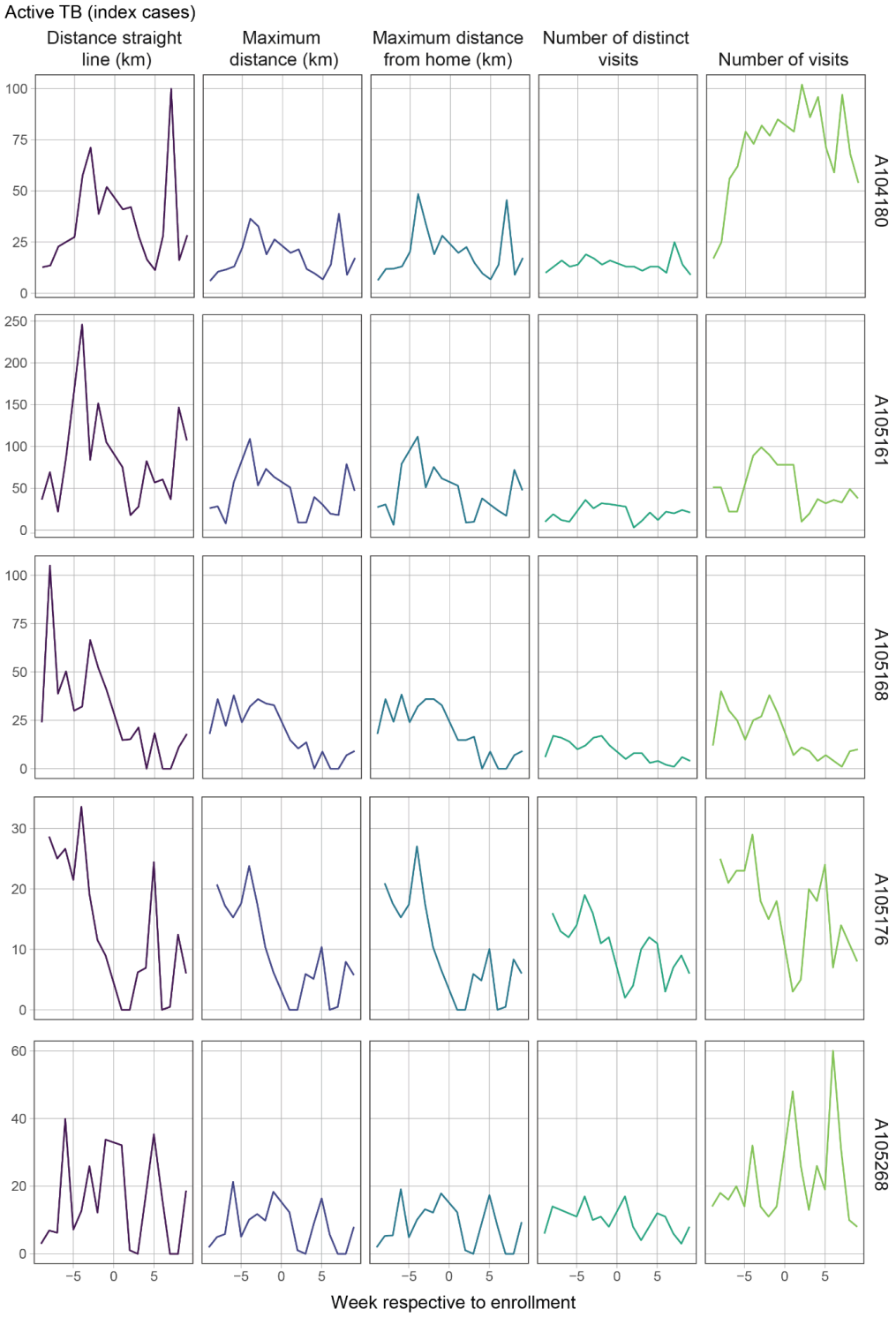
A particularly salient aspect of understanding the transmission dynamics of a human-to-human transmissible infection involves looking at the movement patterns of index cases and their immediate contacts. Some features that precisely describe the movement patterns relevant to TB transmission include daily commuting distance and visits to non-home locations (BROWN et al., 2022).

At first, distance variables such as total and maximum distance traveled on each day by the participants were calculated. Subsequently, home locations were determined based on the clustering, frequency, and time of visitation (22:00 - 7:00 hrs) to the stop locations. The daily maximum distance traveled from home was calculated using the identified home coordinates and other recognized stop locations in the trajectory. Furthermore, the number of locations and unique locations visited daily by the participants were identified.

Figures 17 and 18 show the computed mobility measurements of active TB patients during the study period. The mobility measurements of active TB patients illustrated in Figure 17 show a significant drop after a positive diagnosis. However, most patients quickly rebound to previous mobility patterns in a couple of weeks. In practice, patients diagnosed with active TB should be isolated for two weeks after starting the relevant medication (D. AHMAD; W.K.C. MORGAN, 2000) for the patients to become non-infectious. Contrarily, the active TB patients shown in Figure 18 recorded increased or no change in distance and visitation measurements after

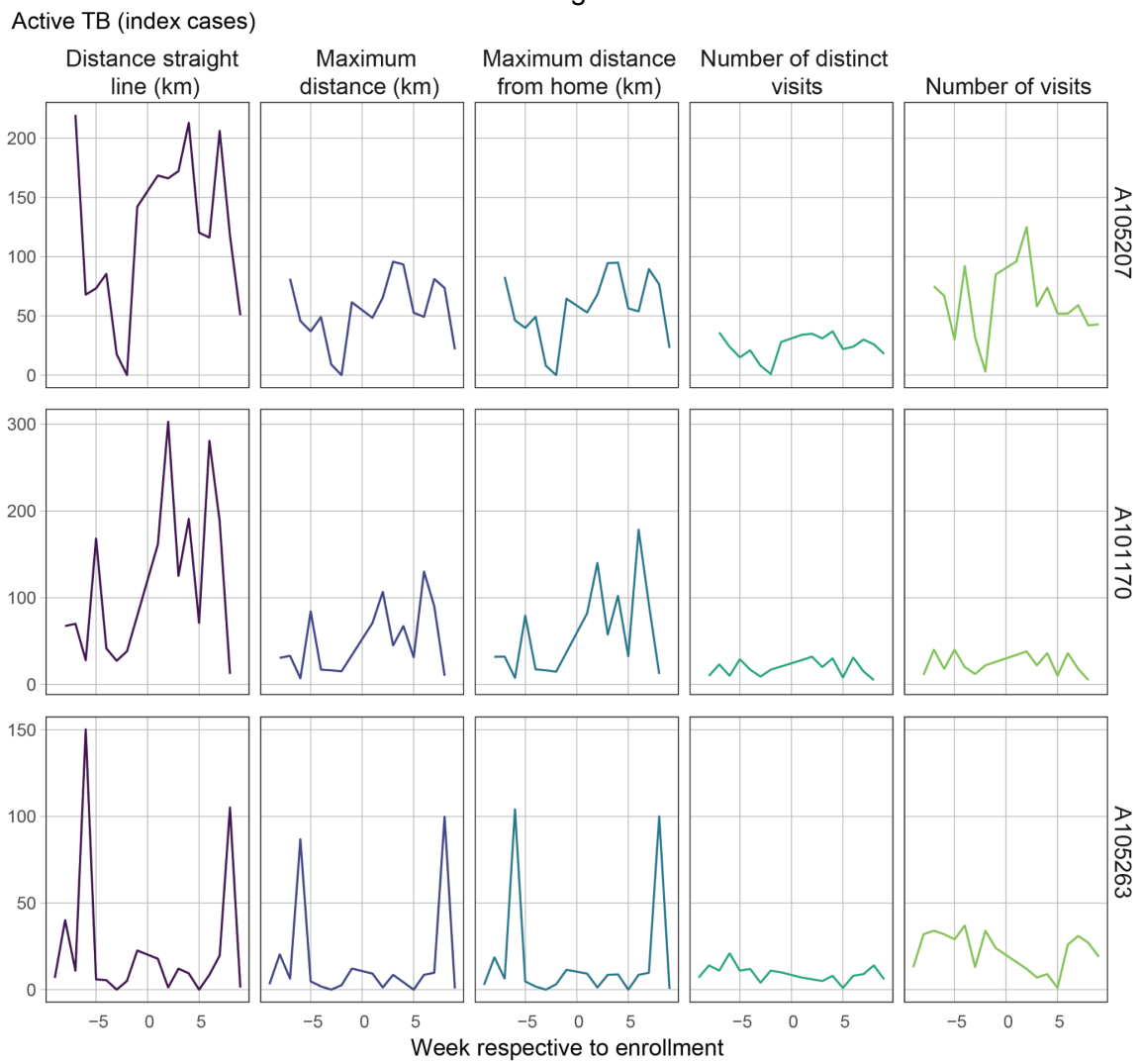
enrollment. This suggests that these patients might act as infection reservoirs and represent potential transmission of TB.

Figure 17 – Active TB patients with decrease in mobility measures after active TB diagnosis



Drop in distance traveled and number of visitations observed each week two months before and after active TB diagnosis. Week 0 represents the enrollment week. Source: The author, 2023.

Figure 18 – Active TB patients with increased or no change in mobility measures after active TB diagnosis



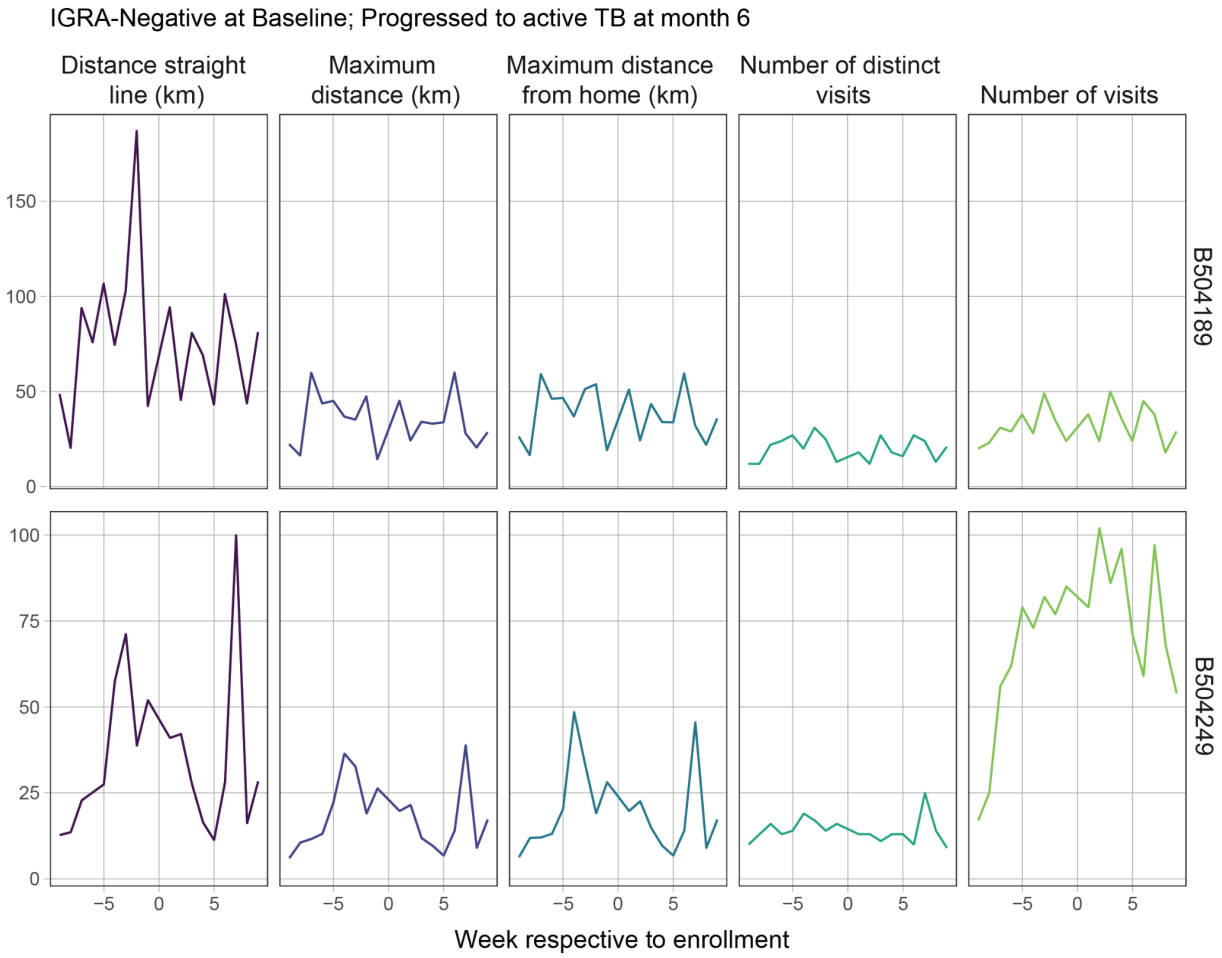
Increase or no change in distance traveled and number of visitations observed each week two months before and after active TB diagnosis. Week 0 represents the enrollment week. Source: The author, 2023.

Such drops or increases in mobility patterns have been reported during infection outbreaks by using anonymized population mobility data (SHIBAMOTO; HAYAKI; OGISU, 2022). Although these population-level anonymized studies help design outbreak intervention strategies (BUCHEL et al., 2021), they are inadequate to examine the underlying relationship between mobility and the infectious disease. Hence, examining and correlating the mobility patterns of individual patients, along with their diagnosis information on a large scale can offer an opportunity to predict

crucial outcomes such as intensity of illness, TB treatment (MULHOLLAND et al., 2023), and death by sickness (ROBSKY et al., 2020).

Two cohort B participants who progressed to active TB later during the study period were tested IGRA-negative on enrollment day; hence, latent TB treatment not being prescribed. As observed in Figure 19, these individuals continued their regular mobility patterns and potentially played a primary role as superspreaders. (DU et al., 2023) report that such latent TB-negative close contacts were more likely to be the superspreaders during an outbreak.

Figure 19 – Mobility patterns of cohort B participants who progressed to active TB



Cohort B participants who initially tested IGRA-negative but progressed to active TB at the end of the study period show regular moving patterns, thus likely to be superspreaders. Source: The author, 2023.

3.5 Conclusion

In conclusion, the present chapter provides insights into quantifying the exposure risk between active TB patients and their immediate close contacts using location data retrieved from smartphones. The recruited paired index and close contact participants observed varied exposure durations based on the outcome of close contact. Considering direct transmission as the primary risk, the time spent together by active TB patients and close contacts, derived from location history data, can directly correlate with the likelihood of TB transmission. Furthermore, the computed movement patterns, such as daily commuting distance and visits, showcase a promising approach to understanding the mobility patterns associated with both active TB patients and their outcomes. These findings emphasize the potential of smartphone location data in augmenting the contact investigation process of human-to-human transmission.

CHAPTER IV

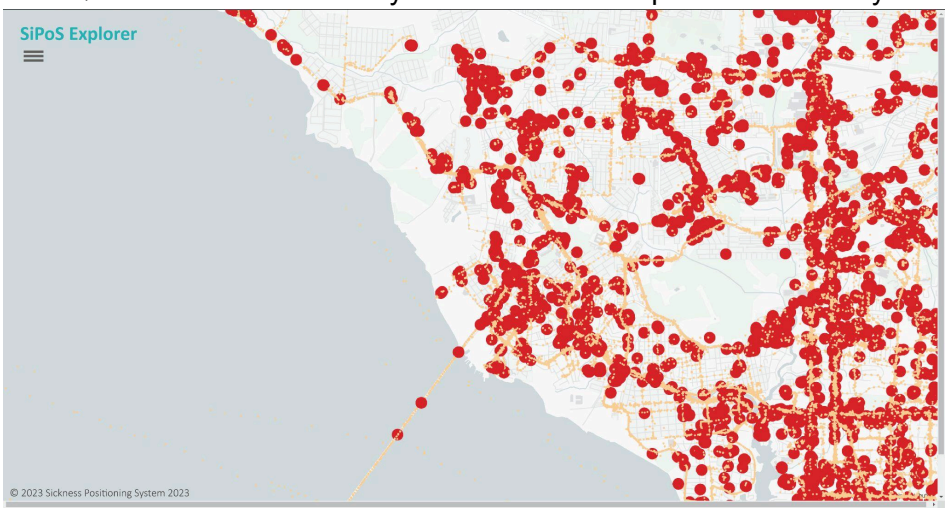
4. SUMMARY

4.1 Overview

The present study demonstrates the overall applicability of location history data from patients' smartphones for improving surveillance of infectious diseases. This study mainly used the location history data digitally sourced from Google LLC, which is widely available, focusing on malaria and tuberculosis transmission. The methods developed, and analyses performed in this study quantify the crucial retrospective location history of malaria patients, thus allowing the identification of hotspot clusters in malaria-endemic regions like Manaus. This innovative approach to pinpointing vector reservoirs extends its applicability to diseases like Schistosomiasis and Chagas. Furthermore, the study also develops methods to implement digital contact investigation by identifying exposure risk between active TB patients and their close contacts. This digitally-based contact investigation method can be applied to other human-to-human contact infections like measles and COVID-19.

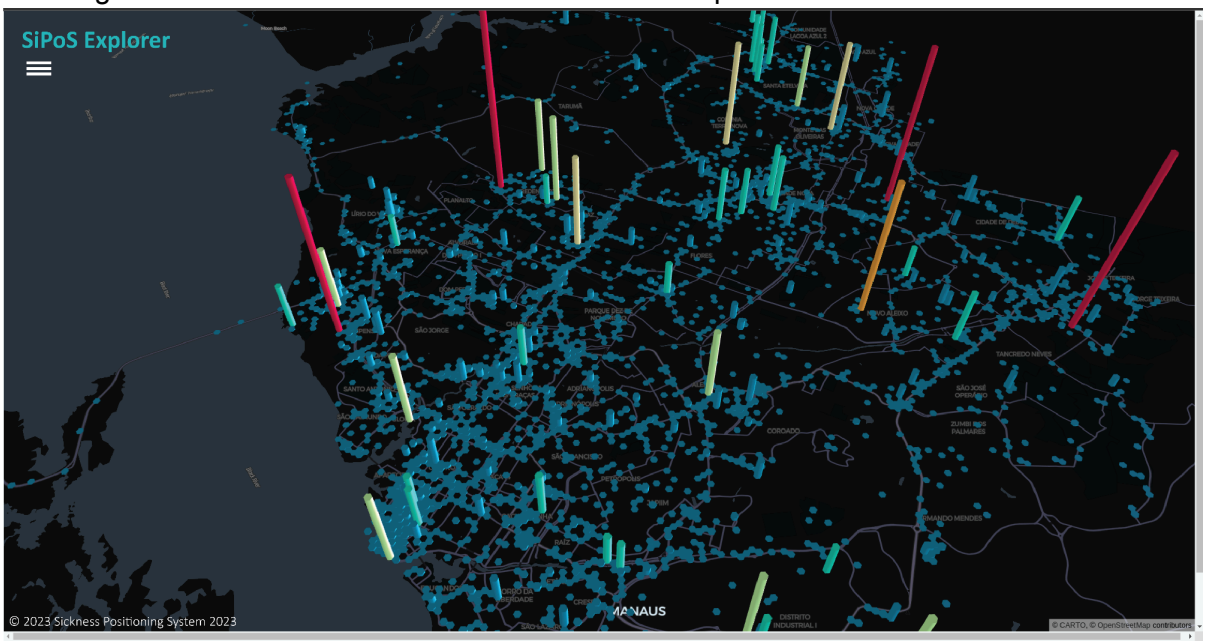
Once quantified into meaningful information, this transformed location history data retrieved from the smartphones can be effectively used with visualization tools for real-time exploration and examination of infection transmission hotspots. Visual exploration of the analyzed malaria patients' data is shown in Figures 20 and 21. OUTBREAK, an open-source disease surveillance tool, was made available at https://outbreak.sysbio.tools/ to explore such data (Figure 22).

Figure 20 - Quantified location history data of all malaria patients in the city of Manaus



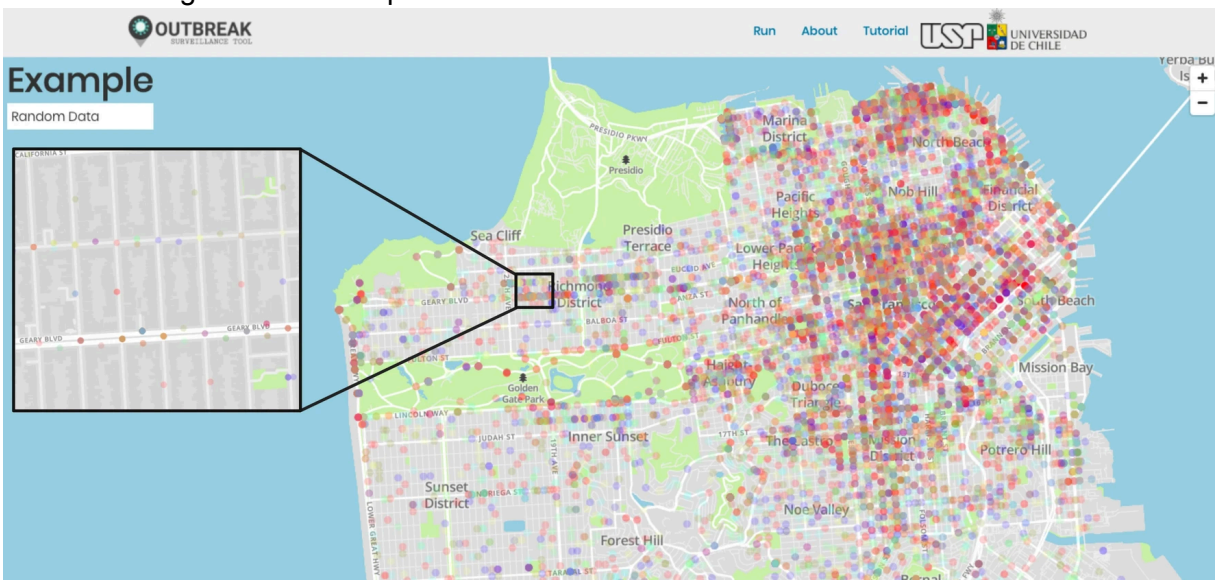
Exploration of staying point locations (red color) and moving points (orange color) of all malaria patients on an interactive map. Source: The author, 2023.

Figure 21 - Potential infected locations of malaria patients based on their visitation



Locations visited by all malaria patients are clustered and sized (height of the bar) based on the visitation frequency. Source: The author, 2023.

Figure 22 – Example use of OUTBREAK tool for disease surveillance



Example GPS data uploaded and explored on OUTBREAK online tool, with street-level zoom facility. Each GPS point can be assigned an HTML color. Source: The author, 2023.

To conclude, this crucial digital location history data from the patients' smartphones demonstrates a robust potential use by local health and municipal agencies for efficient real-time disease surveillance, rapid infection containment, and implementation of infection prevention programs.

4.2 Limitations

4.2.1 Data privacy

Ensuring the privacy of patients is a critical aspect of studies utilizing GPS data for the surveillance of infectious diseases. While the present study has taken steps to anonymize patients' location data, it is crucial to acknowledge that extracting home location information introduces a potential risk of patient identification. Nevertheless, it is essential to note that gathering such information is a standard practice during patient interactions at hospitals. Employing GLH data in the surveillance process complements the routine data collection process by providing a more reliable and timely strategy for analyzing transmission. However, this sensitive information must be restricted to a specific group of experts directly engaged in infection control efforts and, thus, carefully monitored. This restrictive access is vital in balancing the advantages of data-driven insights and the imperative to safeguard patient privacy. Furthermore, stringent confidentiality measures must be rigorously implemented, examined, and approved by the ethical committee, including secure data storage.

4.2.2 Data size, parameter selection, and analysis

The small data size was one of the major limiting factors in the present study. Out of eight hundred individuals reporting to FMT-HVD with malaria symptoms, only two hundred and fifty eventually participated in the study. One primary reason was that the patient lacked a mobile phone with a GPS facility. Other reasons include technical issues during the data acquisition; for example, the participant needed to remember their Google login password for authentication. Moreover, the retrieved GLH file often lacked the necessary data, for instance, data falling within the exposure and symptomatic timeline and GPS points with reasonable accuracy. Since the patient's data was excluded in these cases, retrieving adequate data for identifying mosquito breeding hotspots can remain challenging. In the tuberculosis pilot study, out of forty-three recruited pairs, only six had the necessary data to perform the digital contact investigation. While the derived hours of exposure risk between these close contacts and index cases across all categories conform to the previous epidemiological studies, more data pair participants are required to validate these findings at the clinical level. Despite the current limitations, the existing data

represents a step forward in transmission hotspot identification and contact investigation.

Parameter selection, such as the maximum distance to be considered as clustering radius to cluster stop points and the number of patients in a cluster to be considered a hotspot, play a crucial role in determining the effectiveness of the methods and findings of the present study. The parameters chosen in the current study were determined based on the present literature, practical feasibility, and alignment with the objectives. When calculating the clustering score to prioritize the hotspot clusters, the weights were arbitrarily chosen based on the importance of the variable determined from the previous studies. As more patient data becomes available, future studies may implement iterative methods for selecting optimal parameters, thus maximizing the utility of the collected location history data.

In addition, more variables, such as the presence of stagnant water bodies and forest cover around the identified clusters, can be incorporated into the hotspot cluster scoring. Given the coordinates of the clusters, the presence of these environment features near the clusters can be retrieved from the Open Street Maps or Google Maps APIs. However, this information may only sometimes be available through the APIs and may lead to inaccuracies in cluster scoring. Hence, domain experts should develop more advanced methods to identify these environmental features by pattern recognition from satellite map images. In the case of digital contact investigation, the patients' and their close contacts' diagnostic and hospital data can be integrated into the analysis to estimate health outcomes. Nevertheless, such data integration is practical and definitive only in full-length, comprehensive studies with more participants.

In summary, despite the existing constraints, this study marks progress in pinpointing transmission hotspots and recognizing mobility patterns contributing to infection and conducting digital contact investigations. With ongoing technological advancements and expanding location mobility data sources, forthcoming research should prioritize refining methodologies, addressing patients' data privacy issues, and broadening the spectrum of variables considered.

REFERENCES[1]

ABDUL-GHANI, Rashad; FOUQUE, Florence; MAHDY, Mohammed A. K.; ZHONG, Qingxia; AL-ERYANI, Samira M. A.; ALKWRI, Abdulsamad; BEIER, John C. Multisectoral Approach to Address Chikungunya Outbreaks Driven by Human Mobility: A Systematic Review and Meta-Analysis. **The Journal of Infectious Diseases**, *[S. l.]*, v. 222, n. Supplement_8, p. S709–S716, 2020. DOI: 10.1093/infdis/jiaa500.

ACUÑA-VILLAORDUÑA, Carlos et al. Intensity of exposure to pulmonary tuberculosis determines risk of tuberculosis infection and disease. **European Respiratory Journal**, *[S. l.]*, v. 51, n. 1, p. 1701578, 2018. DOI: 10.1183/13993003.01578-2017.

ALAM, Md. Zakiul. Is Population Density a Risk Factor for Communicable Diseases Like COVID-19? A Case of Bangladesh. **Asia Pacific Journal of Public Health**, *[S. l.]*, v. 33, n. 8, p. 949–950, 2021. DOI: 10.1177/1010539521998858.

ALEMU, Kassahun; WORKU, Alemayehu; BERHANE, Yemane; KUMIE, Abera. Spatiotemporal clusters of malaria cases at village level, northwest Ethiopia. **Malaria Journal**, *[S. l.]*, v. 13, n. 1, p. 223, 2014. DOI: 10.1186/1475-2875-13-223.

ARAÚJO, Valdelaine Etelvina Miranda De; BEZERRA, Juliana Maria Trindade; AMÂNCIO, Frederico Figueiredo; PASSOS, Valéria Maria De Azeredo; CARNEIRO, Mariângela. Aumento da carga de dengue no Brasil e unidades federadas, 2000 e 2015: análise do Global Burden of Disease Study 2015. **Revista Brasileira de Epidemiologia**, *[S. l.]*, v. 20, n. suppl 1, p. 205–216, 2017. DOI: 10.1590/1980-5497201700050017.

BANSAL, Shweta; CHOWELL, Gerardo; SIMONSEN, Lone; VESPIGNANI, Alessandro; VIBOUD, Cécile. Big Data for Infectious Disease Surveillance and Modeling. **Journal of Infectious Diseases**, *[S. l.]*, v. 214, n. suppl 4, p. S375–S379, 2016. DOI: 10.1093/infdis/jiw400.

BARROS, Fábio S. M.; HONÓRIO, Nildimar A. Deforestation and Malaria on the Amazon Frontier: Larval Clustering of Anopheles darlingi (Diptera: Culicidae) Determines Focal Distribution of Malaria. **The American Journal of Tropical Medicine and Hygiene**, *[S. l.]*, v. 93, n. 5, p. 939–953, 2015. DOI: 10.4269/ajtmh.15-0042.

BEHR, Marcel A.; EDELSTEIN, Paul H.; RAMAKRISHNAN, Lalita. Revisiting the timetable of tuberculosis. **BMJ**, *[S. l.]*, p. k2738, 2018. DOI: 10.1136/bmj.k2738.

BHADANE, Chetashri; SHAH, Ketan. Clustering Algorithms for Spatial Data Mining. *Em*: PROCEEDINGS OF THE 2020 3RD INTERNATIONAL CONFERENCE ON GEOINFORMATICS AND DATA ANALYSIS 2020, Marseille France. **Anais** [...]. . *Em*: ICGDA 2020: 2020 3RD INTERNATIONAL CONFERENCE ON GEOINFORMATICS AND DATA ANALYSIS. Marseille France: ACM, 2020. p. 5–9. DOI: 10.1145/3397056.3397068. Disponível em: https://dl.acm.org/doi/10.1145/3397056.3397068. Acesso em: 17 set. 2023.

BIN ABDUL SATAR, Abdul Hadi; HAKIMAH YAACOB. Mapping economic and social

effects of Covid-19 in ASEAN economies. **Journal of Management Info**, *[S. l.]*, v. 9, n. 1, p. 102–120, 2022. DOI: 10.31580/jmi.v9i1.2592.

BOUSEMA, Teun; GRIFFIN, Jamie T.; SAUERWEIN, Robert W.; SMITH, David L.; CHURCHER, Thomas S.; TAKKEN, Willem; GHANI, Azra; DRAKELEY, Chris; GOSLING, Roly. Hitting Hotspots: Spatial Targeting of Malaria for Control and Elimination. **PLoS Medicine**, *[S. l.]*, v. 9, n. 1, p. e1001165, 2012. DOI: 10.1371/journal.pmed.1001165.

BROWN, Tyler S.; ROBINSON, D. Ashley; BUCKEE, Caroline O.; MATHEMA, Barun. Connecting the dots: understanding how human mobility shapes TB epidemics. **Trends in Microbiology**, *[S. l.]*, v. 30, n. 11, p. 1036–1044, 2022. DOI: 10.1016/j.tim.2022.04.005.

BUCHEL, Olha; NINKOV, Anton; CATHEL, Danise; BAR-YAM, Yaneer; HEDAYATIFAR, Leila. Strategizing COVID-19 lockdowns using mobility patterns. **Royal Society Open Science**, *[S. l.]*, v. 8, n. 12, p. 210865, 2021. DOI: 10.1098/rsos.210865.

CAMPELLO, Ricardo J. G. B.; MOULAVI, Davoud; SANDER, Joerg. Density-Based Clustering Based on Hierarchical Density Estimates. *Em*: PEI, Jian; TSENG, Vincent S.; CAO, Longbing; MOTODA, Hiroshi; XU, Guandong (org.). **Advances in Knowledge Discovery and Data Mining**. Lecture Notes in Computer ScienceBerlin, Heidelberg: Springer Berlin Heidelberg, 2013. v. 7819p. 160–172. DOI: 10.1007/978-3-642-37456-2_14. Disponível em: http://link.springer.com/10.1007/978-3-642-37456-2_14. Acesso em: 17 set. 2023.

CARAUX-PAZ, Pauline; DIAMANTIS, Sylvain; DE WAZIÈRES, Benoit; GALLIEN, Sébastien. Tuberculosis in the Elderly. **Journal of Clinical Medicine**, *[S. l.]*, v. 10, n. 24, p. 5888, 2021. DOI: 10.3390/jcm10245888.

CARDO, María Victoria; VEZZANI, Darío; CARBAJO, Aníbal Eduardo. The role of the landscape in structuring immature mosquito assemblages in wetlands. **Wetlands Ecology and Management**, *[S. l.]*, v. 21, n. 1, p. 55–70, 2013. DOI: 10.1007/s11273-012-9279-x.

CARDOZO, Lucas Esteves. **Mapeamento de *hotspots* de transmissão de malária utilizando geolocalização de pacientes**. 2019. Mestrado em Fisiopatologia - Universidade de São Paulo, São Paulo, 2019. DOI: 10.11606/D.9.2019.tde-04022019-170221. Disponível em: http://www.teses.usp.br/teses/disponiveis/9/9142/tde-04022019-170221/. Acesso em: 23 set. 2023.

CATRY, Thibault; LI, Zhichao; ROUX, Emmanuel; HERBRETEAU, Vincent; GURGEL, Helen; MANGEAS, Morgan; SEYLER, Frédérique; DESSAY, Nadine. Wetlands and Malaria in the Amazon: Guidelines for the Use of Synthetic Aperture Radar Remote-Sensing. **International Journal of Environmental Research and Public Health**, *[S. l.]*, v. 15, n. 3, p. 468, 2018. DOI: 10.3390/ijerph15030468.

CDC. **Guidelines for the investigation of contacts of persons with infectious tuberculosis; recommendations from the National Tuberculosis Controllers Association and CDC, and Guidelines for using the QuantiFERON®-TB Gold**

test for detecting Mycobacterium tuberculosis infection. MMWR, , 2005. . Acesso em: 20 ago. 2023.

CDC, CDC. **Anopheles Mosquito Life Cycle**. [s.l: s.n.]. Disponível em: https://www.cdc.gov/mosquitoes/about/life-cycles/anopheles.html.

CHANSANG, Chitti; KITTAYAPONG, Pattamaporn. Application of Mosquito Sampling Count and Geospatial Methods to Improve Dengue Vector Surveillance. **The American Journal of Tropical Medicine and Hygiene**, *[S. l.]*, v. 77, n. 5, p. 897–902, 2007. DOI: 10.4269/ajtmh.2007.77.897.

COLEMAN, Marlize; COLEMAN, Michael; MABUZA, Aaron M.; KOK, Gerdalize; COETZEE, Maureen; DURRHEIM, David N. Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. **Malaria Journal**, *[S. l.]*, v. 8, n. 1, p. 68, 2009. DOI: 10.1186/1475-2875-8-68.

COLLUCCI, Cláudia. Brazil sees sharp rise in chikungunya cases. **BMJ**, *[S. l.]*, p. i4560, 2016. DOI: 10.1136/bmj.i4560.

COOLS, Dillan; MCCALLUM, Scott Christian; RAINHAM, Daniel; TAYLOR, Nathan; PATTERSON, Zachary. Understanding Google Location History as a Tool for Travel Diary Data Acquisition. **Transportation Research Record: Journal of the Transportation Research Board**, *[S. l.]*, v. 2675, n. 5, p. 238–251, 2021. DOI: 10.1177/0361198120986169.

D. AHMAD; W.K.C. MORGAN. How long are TB patients infectious? **Canadian Medical Association Journal**, *[S. l.]*, v. 163, n. 2, p. 157, 2000.

DE QUADROS, Ciro A.; IZURIETA, Hector; VENCZEL, Linda; CARRASCO, Peter. Measles Eradication in the Americas: Progress to Date. **The Journal of Infectious Diseases**, *[S. l.]*, v. 189, n. Supplement_1, p. S227–S235, 2004. DOI: 10.1086/377741.

DONALD, P. R.; DIACON, A. H.; LANGE, C.; DEMERS, A. M.; VON GROOTE-BIDDLINGMEIER, F.; NARDELL, E. Droplets, dust and guinea pigs: an historical review of tuberculosis transmission research, 1878–1940. **The International Journal of Tuberculosis and Lung Disease**, *[S. l.]*, v. 22, n. 9, p. 972–982, 2018. DOI: 10.5588/ijtld.18.0173.

DU, Jingli; SU, Yue; DONG, Enjun; QIAO, Juan; WANG, Ruilan; ZHAO, Wenjuan; LIANG, Jianqin; GONG, Wenping. LTBI-negative close contacts of tuberculosis are more likely to develop the disease: enlightenment and lessons from a cluster outbreak. **Frontiers in Public Health**, *[S. l.]*, v. 11, p. 1136355, 2023. DOI: 10.3389/fpubh.2023.1136355.

DUTHIE, S.; BLACK, C.; DOUGLAS, G.; JACKSON, A. D.; WEBSTER, D. Tuberculosis outbreak associated with a mosque: challenges of large scale contact tracing. **Eurosurveillance**, *[S. l.]*, v. 13, n. 51, 2008. DOI: 10.2807/ese.13.51.19069-en. Disponível em: https://www.eurosurveillance.org/content/10.2807/ese.13.51.19069-en. Acesso em: 23 set. 2023.

EAMES, K. T. D.; KEELING, M. J. Contact tracing and disease control. **Proceedings**

of the Royal Society of London. Series B: Biological Sciences, *[S. l.]*, v. 270, n. 1533, p. 2565–2571, 2003. DOI: 10.1098/rspb.2003.2554.

EDELSTEIN, Michael. Measuring vaccination coverage better will help achieve disease control. **International Health**, *[S. l.]*, v. 9, n. 3, p. 142–144, 2017. DOI: 10.1093/inthealth/ihx013.

EISEN, Rebecca J.; EISEN, Lars. Use of Geographic Information Systems in Infectious Disease Surveillance. *Em*: M'IKANATHA, Nkuchia M.; ISKANDER, John K. (org.). **Concepts and Methods in Infectious Disease Surveillance**. Oxford, UK: John Wiley & Sons, Ltd, 2014. p. 219–229. DOI: 10.1002/9781118928646.ch20. Disponível em: https://onlinelibrary.wiley.com/doi/10.1002/9781118928646.ch20. Acesso em: 23 ago. 2023.

ELIDIO, Guilherme Almeida et al. Measles outbreak: preliminary report on a case series of the first 8,070 suspected cases, Manaus, Amazonas state, Brazil, February to November 2018. **Eurosurveillance**, *[S. l.]*, v. 24, n. 2, 2019. DOI: 10.2807/1560-7917.ES.2019.24.2.1800663. Disponível em: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2019.24.2.1800663. Acesso em: 22 set. 2023.

ELMESALAWY, Mahmoud M.; SALAMA, Abdullah I.; ANANY, M. G. Tracy: Smartphone-based Contact Tracing Solution that Supports Self-investigation to Limit the Spread of COVID-19. *Em*: 2020 2ND NOVEL INTELLIGENT AND LEADING EMERGING SCIENCES CONFERENCE (NILES) 2020, Giza, Egypt. **Anais** [...]. . *Em*: 2020 2ND NOVEL INTELLIGENT AND LEADING EMERGING SCIENCES CONFERENCE (NILES). Giza, Egypt: IEEE, 2020. p. 623–628. DOI: 10.1109/NILES50944.2020.9257915. Disponível em: https://ieeexplore.ieee.org/document/9257915/. Acesso em: 22 set. 2023.

ERKENS, C. G. M. et al. Tuberculosis contact investigation in low prevalence countries: a European consensus. **European Respiratory Journal**, *[S. l.]*, v. 36, n. 4, p. 925–949, 2010. DOI: 10.1183/09031936.00201609.

FRANKLINOS, Lydia H. V.; JONES, Kate E.; REDDING, David W.; ABUBAKAR, Ibrahim. The effect of global change on mosquito-borne disease. **The Lancet Infectious Diseases**, *[S. l.]*, v. 19, n. 9, p. e302–e312, 2019. DOI: 10.1016/S1473-3099(19)30161-6.

GONZÁLEZ, Marta C.; HIDALGO, César A.; BARABÁSI, Albert-László. Understanding individual human mobility patterns. **Nature**, *[S. l.]*, v. 453, n. 7196, p. 779–782, 2008. DOI: 10.1038/nature06958.

GOOGLE, LLC. **Google Location History**. , [s.d.]. Disponível em: https://policies.google.com/technologies/location-data.

GUBLER, Duane. Resurgent Vector-Borne Diseases as a Global Health Problem. **Emerging Infectious Diseases**, *[S. l.]*, v. 4, n. 3, p. 442–450, 1998. DOI: 10.3201/eid0403.980326.

HASAN, Tasnim; AU, Eric; CHEN, Sharon; TONG, Allison; WONG, Germaine. Screening and prevention for latent tuberculosis in immunosuppressed patients at risk for tuberculosis: a systematic review of clinical practice guidelines. **BMJ Open**,

[S. l.], v. 8, n. 9, p. e022445, 2018. DOI: 10.1136/bmjopen-2018-022445.

HEO, JoonNyung et al. A Patient Self-Checkup App for COVID-19: Development and Usage Pattern Analysis. **Journal of Medical Internet Research**, *[S. l.]*, v. 22, n. 11, p. e19665, 2020. DOI: 10.2196/19665.

HEYMANN, David L.; BRILLIANT, Larry. Surveillance in eradication and elimination of infectious diseases: A progression through the years. **Vaccine**, *[S. l.]*, v. 29, p. D141–D144, 2011. DOI: 10.1016/j.vaccine.2011.12.135.

HIGH, Kevin P. Infection as a cause of age-related morbidity and mortality. **Ageing Research Reviews**, *[S. l.]*, v. 3, n. 1, p. 1–14, 2004. DOI: 10.1016/j.arr.2003.08.001.

HIGHTOWER, A. W.; ODHIAMBO, R.; NAHLEN, B. L.; OMBOK, M.; OLOO, A. J.; HAWLEY, W. A.; LAL, A. A.; OTIENO, R. A geographic information system applied to a malaria field study in western Kenya. **The American Journal of Tropical Medicine and Hygiene**, *[S. l.]*, v. 58, n. 3, p. 266–272, 1998. DOI: 10.4269/ajtmh.1998.58.266.

HOTEZ, Peter J.; REMME, Jan H. F.; BUSS, Paulo; ALLEYNE, George; MOREL, Carlos; BREMAN, Joel G. Combating Tropical Infectious Diseases: Report of the Disease Control Priorities in Developing Countries Project. **Clinical Infectious Diseases**, *[S. l.]*, v. 38, n. 6, p. 871–878, 2004. DOI: 10.1086/382077.

HOUK, Vernon N.; KENT, Donald C.; BAKER, John H.; SORENSEN, Kurt; HANZEL, George D. The Byrd Study: In-depth Analysis of a Micro-outbreak of Tuberculosis in a Closed Environment. **Archives of Environmental Health: An International Journal**, *[S. l.]*, v. 16, n. 1, p. 4–6, 1968. DOI: 10.1080/00039896.1968.10665007.

HUANG, Xiao; LI, Zhenlong; JIANG, Yuqin; LI, Xiaoming; PORTER, Dwayne. Twitter reveals human mobility dynamics during the COVID-19 pandemic. **PLOS ONE**, *[S. l.]*, v. 15, n. 11, p. e0241957, 2020. DOI: 10.1371/journal.pone.0241957.

JAMES, Jeffrey. The smart feature phone revolution in developing countries: Bringing the internet to the bottom of the pyramid. **The Information Society**, *[S. l.]*, v. 36, n. 4, p. 226–235, 2020. DOI: 10.1080/01972243.2020.1761497.

KARAIM, Malek; ELSHEIKH, Mohamed; NOURELDIN, Aboelmagd. GNSS Error Sources. *Em*: RUSTAMOV, Rustam B.; HASHIMOV, Arif M. (org.). **Multifunctional Operation and Application of GPS**. [s.l.] : InTech, 2018. DOI: 10.5772/intechopen.75493. Disponível em: http://www.intechopen.com/books/multifunctional-operation-and-application-of-gps/gnss-error-sources. Acesso em: 23 set. 2023.

KENGLUECHA, Ampornpan; SINGHASIVANON, Pratap; TIENSUWAN, Montip; JONES, James W.; SITHIPRASASNA, Ratana. Water quality and breeding habitats of anopheline mosquito in northwestern Thailand. **The Southeast Asian Journal of Tropical Medicine and Public Health**, *[S. l.]*, v. 36, n. 1, p. 46–53, 2005.

KENYON, Thomas A.; VALWAY, Sarah E.; IHLE, Walter W.; ONORATO, Ida M.; CASTRO, Kenneth G. Transmission of Multidrug-Resistant *Mycobacterium tuberculosis* during a Long Airplane Flight. **New England Journal of Medicine**, *[S. l.]*, v. 334, n. 15, p. 933–938, 1996. DOI: 10.1056/NEJM199604113341501.

KIM, Eun-Sung. The Birth of Digital Epidemiology in South Korea. **East Asian Science, Technology and Society: An International Journal**, *[S. l.]*, p. 1–25, 2023. DOI: 10.1080/18752160.2023.2177405.

LAFOND, Kathryn E. et al. Global burden of influenza-associated lower respiratory tract infections and hospitalizations among adults: A systematic review and meta-analysis. **PLOS Medicine**, *[S. l.]*, v. 18, n. 3, p. e1003550, 2021. DOI: 10.1371/journal.pmed.1003550.

LAPORTA, Gabriel Z. et al. Malaria transmission in landscapes with varying deforestation levels and timelines in the Amazon: a longitudinal spatiotemporal study. **Scientific Reports**, *[S. l.]*, v. 11, n. 1, p. 6477, 2021. DOI: 10.1038/s41598-021-85890-3.

LEE, Minha; ZHAO, Jun; SUN, Qianqian; PAN, Yixuan; ZHOU, Weiyi; XIONG, Chenfeng; ZHANG, Lei. Human mobility trends during the early stage of the COVID-19 pandemic in the United States. **PLOS ONE**, *[S. l.]*, v. 15, n. 11, p. e0241468, 2020. DOI: 10.1371/journal.pone.0241468.

MAILLES, A. et al. First cases of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) infections in France, investigations and implications for the prevention of human-to-human transmission, France, May 2013. **Eurosurveillance**, *[S. l.]*, v. 18, n. 24, 2013. DOI: 10.2807/ese.18.24.20502-en. Disponível em: https://www.eurosurveillance.org/content/10.2807/ese.18.24.20502-en. Acesso em: 22 set. 2023.

M'IKANATHA, Nkuchia M.; LYNFIELD, Ruth; JULIAN, Kathleen G.; VAN BENEDEN, Chris A.; VALK, Henriette De. Infectious disease surveillance: a cornerstone for prevention and control. *Em*: M'IKANATHA, Nkuchia M.; LYNFIELD, Ruth; VAN BENEDEN, Chris A.; DE VALK, Henriette (org.). **Infectious Disease Surveillance**. Oxford, UK: John Wiley & Sons Ltd, 2013. p. 1–20. DOI: 10.1002/9781118543504.ch1. Disponível em: https://onlinelibrary.wiley.com/doi/10.1002/9781118543504.ch1. Acesso em: 23 ago. 2023.

MOORE, Marisa; VALWAY, Sarah E.; IHLE, Walter; ONORATO, Ida M. A Train Passenger with Pulmonary Tuberculosis: Evidence of Limited Transmission During Travel. **Clinical Infectious Diseases**, *[S. l.]*, v. 28, n. 1, p. 52–56, 1999. DOI: 10.1086/515089.

MULHOLLAND, Grace E. et al. Geographic mobility and treatment outcomes among people in care for tuberculosis in the Lake Victoria region of East Africa: A multi-site prospective cohort study. **PLOS Global Public Health**, *[S. l.]*, v. 3, n. 6, p. e0001992, 2023. DOI: 10.1371/journal.pgph.0001992.

NAKAYAMA, Don K. Social Distancing and Contact Tracing During the Great Plague of 1665. **The American Surgeon**, *[S. l.]*, v. 88, n. 2, p. 165–166, 2022. DOI: 10.1177/00031348211065094.

NEIDERUD, Carl-Johan. How urbanization affects the epidemiology of emerging infectious diseases. **Infection Ecology & Epidemiology**, *[S. l.]*, v. 5, n. 1, p. 27060, 2015. DOI: 10.3402/iee.v5.27060.

NHAVOTO, José António; GRÖNLUND, Åke. Mobile Technologies and Geographic Information Systems to Improve Health Care Systems: A Literature Review. **JMIR mHealth and uHealth**, *[S. l.]*, v. 2, n. 2, p. e21, 2014. DOI: 10.2196/mhealth.3216.

NISHIURA, Hiroshi; LEE, Hyeong-Woo; CHO, Shin-Hyeong; LEE, Wook-Gyo; IN, Tae-Suk; MOON, Sung-Ung; CHUNG, Gyung Tae; KIM, Tong-Soo. Estimates of short- and long-term incubation periods of Plasmodium vivax malaria in the Republic of Korea. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, *[S. l.]*, v. 101, n. 4, p. 338–343, 2007. DOI: 10.1016/j.trstmh.2006.11.002.

NSOESIE, Elaine O. et al. Spatial and Temporal Clustering of Chikungunya Virus Transmission in Dominica. **PLOS Neglected Tropical Diseases**, *[S. l.]*, v. 9, n. 8, p. e0003977, 2015. DOI: 10.1371/journal.pntd.0003977.

OLIVEIRA-FERREIRA, Joseli; LACERDA, Marcus Vg; BRASIL, Patrícia; LADISLAU, José Lb; TAUIL, Pedro L.; DANIEL-RIBEIRO, Cláudio Tadeu. Malaria in Brazil: an overview. **Malaria Journal**, *[S. l.]*, v. 9, n. 1, p. 115, 2010. DOI: 10.1186/1475-2875-9-115.

ORGANIZATION, World Health. **World Malaria Report 2021**. Geneva: World Health Organization, 2021.

PADHAN, Rakesh; PRABHEESH, K. P. The economics of COVID-19 pandemic: A survey. **Economic Analysis and Policy**, *[S. l.]*, v. 70, p. 220–237, 2021. DOI: 10.1016/j.eap.2021.02.012.

PALAGYI, Anna; MARAIS, Ben J.; ABIMBOLA, Seye; TOPP, Stephanie M.; MCBRYDE, Emma S.; NEGIN, Joel. Health system preparedness for emerging infectious diseases: A synthesis of the literature. **Global Public Health**, *[S. l.]*, v. 14, n. 12, p. 1847–1868, 2019. DOI: 10.1080/17441692.2019.1614645.

PANG, Tikki; PEELING, Rosanna W. Diagnostic tests for infectious diseases in the developing world: two sides of the coin. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, *[S. l.]*, v. 101, n. 9, p. 856–857, 2007. DOI: 10.1016/j.trstmh.2007.04.014.

PEDREGOSA, Fabian. Scikit-learn: Machine Learning in Python. **The Journal of Machine Learning Research**, *[S. l.]*, v. 12, p. 2825–2830, 2011. DOI: 10.5555/1953048.2078195.

PERERA, Ruwanthi; CALDERA, Amandhi; WICKREMASINGHE, A. Rajitha. Reactive Case Detection (RACD) and foci investigation strategies in malaria control and elimination: a review. **Malaria Journal**, *[S. l.]*, v. 19, n. 1, p. 401, 2020. DOI: 10.1186/s12936-020-03478-0.

PINHEIRO, Paulo; MATHERS, Colin D.; KRÄMER, Alexander. The Global Burden of Infectious Diseases. *Em*: KRÄMER, Alexander; KRETZSCHMAR, Mirjam; KRICKEBERG, Klaus (org.). **Modern Infectious Disease Epidemiology**. New York, NY: Springer New York, 2009. p. 3–21. DOI: 10.1007/978-0-387-93835-6_1. Disponível em: http://link.springer.com/10.1007/978-0-387-93835-6_1. Acesso em: 23 ago. 2023.

PROTHERO, R. Mansell. Disease and Mobility: A Neglected Factor in Epidemiology.

International Journal of Epidemiology, *[S. l.]*, v. 6, n. 3, p. 259–267, 1977. DOI: 10.1093/ije/6.3.259.

RAHMAN, Md. Tanvir; KHAN, Risala T.; KHANDAKER, Muhammad R. A.; SELLATHURAI, Mathini; SALAN, Md. Sifat A. An Automated Contact Tracing Approach for Controlling Covid-19 Spread Based on Geolocation Data From Mobile Cellular Networks. **IEEE Access**, *[S. l.]*, v. 8, p. 213554–213565, 2020. DOI: 10.1109/ACCESS.2020.3040198.

REICHLER, Mary R. et al. Duration of Exposure Among Close Contacts of Patients With Infectious Tuberculosis and Risk of Latent Tuberculosis Infection. **Clinical Infectious Diseases**, *[S. l.]*, v. 71, n. 7, p. 1627–1634, 2020. DOI: 10.1093/cid/ciz1044.

RILEY, Richard L. The J. Burns Amberson Lecture Aerial Dissemination of Pulmonary Tuberculosis. **American Review of Tuberculosis and Pulmonary Diseases**, *[S. l.]*, v. 76, n. 6, p. 931–941, 1957.

ROBSKY, Katherine O.; HUGHES, Seamus; KITYAMUWESI, Alex; KENDALL, Emily A.; KITONSA, Peter James; DOWDY, David W.; KATAMBA, Achilles. Is distance associated with tuberculosis treatment outcomes? A retrospective cohort study in Kampala, Uganda. **BMC Infectious Diseases**, *[S. l.]*, v. 20, n. 1, p. 406, 2020. DOI: 10.1186/s12879-020-05099-z.

ROZENDAAL, J. A. Biting and resting behavior of Anopheles darlingi in the Suriname rainforest. **Journal of the American Mosquito Control Association**, *[S. l.]*, v. 5, n. 3, p. 351–358, 1989.

RUKTANONCHAI, Nick W.; DELEENHEER, Patrick; TATEM, Andrew J.; ALEGANA, Victor A.; CAUGHLIN, T. Trevor; ZU ERBACH-SCHOENBERG, Elisabeth; LOURENÇO, Christopher; RUKTANONCHAI, Corrine W.; SMITH, David L. Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data. **PLOS Computational Biology**, *[S. l.]*, v. 12, n. 4, p. e1004846, 2016. DOI: 10.1371/journal.pcbi.1004846.

RUKTANONCHAI, Nick Warren; RUKTANONCHAI, Corrine Warren; FLOYD, Jessica Rhona; TATEM, Andrew J. Using Google Location History data to quantify fine-scale human mobility. **International Journal of Health Geographics**, *[S. l.]*, v. 17, n. 1, p. 28, 2018. DOI: 10.1186/s12942-018-0150-z.

SALATHÉ, Marcel et al. Digital Epidemiology. **PLoS Computational Biology**, *[S. l.]*, v. 8, n. 7, p. e1002616, 2012. DOI: 10.1371/journal.pcbi.1002616.

SANTIAGO-ALARCON, Diego; MACGREGOR-FORS, Ian. Cities and pandemics: urban areas are ground zero for the transmission of emerging human infectious diseases. **Journal of Urban Ecology**, *[S. l.]*, v. 6, n. 1, p. juaa012, 2020. DOI: 10.1093/jue/juaa012.

SEPLAN-CTI. **Demographic Density in Manaus**, 2015.

SEVILLA-CASAS, Elías. Human mobility and malaria risk in the Naya river basin of Colombia. **Social Science & Medicine**, *[S. l.]*, v. 37, n. 9, p. 1155–1167, 1993. DOI: 10.1016/0277-9536(93)90255-3.

SHAFFER, Jeffrey G. et al. Clustering of asymptomatic Plasmodium falciparum infection and the effectiveness of targeted malaria control measures. **Malaria Journal**, *[S. l.]*, v. 19, n. 1, p. 33, 2020. DOI: 10.1186/s12936-019-3063-9.

SHIBAMOTO, Masahiko; HAYAKI, Shoka; OGISU, Yoshitaka. COVID-19 infection spread and human mobility. **Journal of the Japanese and International Economies**, *[S. l.]*, v. 64, p. 101195, 2022. DOI: 10.1016/j.jjie.2022.101195.

SHRIVASTAVA, Saurabh R.; SHRIVASTAVA, Prateek S.; RAMASAMY, Jegadeesh. Assessing the utility of contact tracing in reducing the magnitude of tuberculosis. **Infection Ecology & Epidemiology**, *[S. l.]*, v. 4, n. 1, p. 25265, 2014. DOI: 10.3402/iee.v4.25265.

SINGAPORE, Ministry of Health. **Trace Together**. , [s.d.]. Disponível em: https://www.tracetogether.gov.sg/.

SINGH, Vineeta; MISHRA, Neelima; AWASTHI, Gauri; DASH, Aditya P.; DAS, Aparup. Why is it important to study malaria epidemiology in India? **Trends in Parasitology**, *[S. l.]*, v. 25, n. 10, p. 452–457, 2009. DOI: 10.1016/j.pt.2009.06.004.

TARWATER, Patrick M.; MARTIN, Clyde F. Effects of population density on the spread of disease. **Complexity**, *[S. l.]*, v. 6, n. 6, p. 29–36, 2001. DOI: 10.1002/cplx.10003.

TOGNOTTI, Eugenia. Lessons from the History of Quarantine, from Plague to Influenza A. **Emerging Infectious Diseases**, *[S. l.]*, v. 19, n. 2, p. 254–259, 2013. DOI: 10.3201/eid1902.120312.

VAZQUEZ-PROKOPEC, Gonzalo M. et al. Using GPS Technology to Quantify Human Mobility, Dynamic Contacts and Infectious Disease Dynamics in a Resource-Poor Urban Environment. **PLoS ONE**, *[S. l.]*, v. 8, n. 4, p. e58802, 2013. DOI: 10.1371/journal.pone.0058802.

VAZQUEZ-PROKOPEC, Gonzalo; STODDARD, Steven T.; PAZ-SOLDAN, Valerie A.; ELDER, John P.; MORRISON, Amy C.; KOCHEL, Tadeusz J.; KITRON, Uriel; SCOTT, Thomas W. Assessing and Maximizing the Acceptability of Global Positioning System Device Use for Studying the Role of Human Movement in Dengue Virus Transmission in Iquitos, Peru. **The American Journal of Tropical Medicine and Hygiene**, *[S. l.]*, v. 82, n. 4, p. 723–730, 2010. DOI: 10.4269/ajtmh.2010.09-0496.

WANJALA, Christine L.; WAITUMBI, John; ZHOU, Guofa; GITHEKO, Andrew K. Identification of malaria transmission and epidemic hotspots in the western Kenya highlands: its application to malaria epidemic prediction. **Parasites & Vectors**, *[S. l.]*, v. 4, n. 1, p. 81, 2011. DOI: 10.1186/1756-3305-4-81.

WASHINGTON, Charles H. et al. Spatial clustering of filarial transmission before and after a Mass Drug Administration in a setting of low infection prevalence. **Filaria Journal**, *[S. l.]*, v. 3, n. 1, p. 3, 2004. DOI: 10.1186/1475-2883-3-3.

WESOLOWSKI, Amy; EAGLE, Nathan; TATEM, Andrew J.; SMITH, David L.; NOOR, Abdisalan M.; SNOW, Robert W.; BUCKEE, Caroline O. Quantifying the Impact of Human Mobility on Malaria. **Science**, *[S. l.]*, v. 338, n. 6104, p. 267–270, 2012. DOI:

10.1126/science.1223467.

WHO. **Global health estimates**. 2020a. Disponível em: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death. Acesso em: 23 ago. 2023.

WHO. **The top 10 causes of death**. [s.l: s.n.]. Disponível em: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. Acesso em: 23 ago. 2023b.

WHO, Geneva. **WHO COVID-19 Dashboard**. 2023. Disponível em: https://covid19.who.int. Acesso em: 23 ago. 2023.

ZHENG, Yu. Trajectory Data Mining: An Overview. **ACM Transactions on Intelligent Systems and Technology**, *[S. l.]*, v. 6, n. 3, p. 1–41, 2015. DOI: 10.1145/2743025.

ZHOU, Ying; XU, Renzhe; HU, Dongsheng; YUE, Yang; LI, Qingquan; XIA, Jizhe. Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: a modelling study using mobile phone data. **The Lancet Digital Health**, *[S. l.]*, v. 2, n. 8, p. e417–e424, 2020. DOI: 10.1016/S2589-7500(20)30165-5.

[1] De acordo com a Associação Brasileira de Normas Técnicas (ABNT NBR 6023).