



UNIVERSIDAD PERUANA  
**CAYETANO HEREDIA**

**EVALUACIÓN DE LA EXACTITUD DE  
LA GEOCODIFICACIÓN DE  
DIRECCIONES: ANÁLISIS ESPACIAL  
EN EL DISTRITO DE MI PERÚ,  
CALLAO**

TESIS PARA OPTAR EL GRADO DE  
MAESTRA EN INFORMÁTICA BIOMÉDICA  
EN SALUD GLOBAL CON MENCIÓN EN  
INFORMÁTICA EN SALUD

ROSARIO CAROLINA CUYA EVARISTO

LIMA - PERÚ

2022



## **ASESOR**

MD, MPH, PhD. CÉSAR PAUL EUGENIO CÁRCAMO CAVAGNARO

## **JURADO DE TESIS**

DRA. MABEL KAREL RAZA GARCÍA

PRESIDENTE

DR. ANDRES GUILLERMO LESCANO GUEVARA

VOCAL

MG. GABRIEL CARRASCO ESCOBAR

SECRETARIO

## **DEDICATORIA**

A mi padre José, mi ejemplo a seguir, por su amor, paciencia y consejos

A mi madre Felixa en el cielo, mi ángel, te extrañamos mucho

A mis hermanos y sobrinos, por su cariño y apoyo incondicional

## **AGRADECIMIENTOS**

A la Dra. Patricia García por brindarme la oportunidad de realizar este proyecto. Al equipo de HOPE, María, José, Jocelyn, por sus consejos y apoyo en la ejecución de la tesis. Agradezco especialmente a mi asesor el Dr. César Cárcamo por su constante guía y consejos, por su paciencia y dedicación durante el desarrollo de la investigación.

Al CONCYTEC – FONDECYT por permitirme realizar la maestría, por contribuir en la educación e investigación en nuestro país. A Daniel Córdor y Paola Pflucker, nuestros guías durante el desarrollo de la maestría, por su apoyo y consejos brindados. A mis compañeros de la maestría, grandes amigos, un grupo interdisciplinario que me ha permitido aprender importantes lecciones para mi formación personal y profesional.

Mi eterno agradecimiento a mi hermano César, quien me acompañó durante la fase de campo del estudio. A mi padre por sus palabras de aliento y sabiduría. A mis amigos Franklin, Kevin, Michelle y Gaby por los grandes momentos vividos en esta aventura, por sus consejos y su gran voluntad para apoyarme.

## **FUENTE DE FINANCIAMIENTO**

Esta investigación fue financiada por el Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC) a través del Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica del Perú (FONDECYT).

# ÍNDICE

## RESUMEN

## ABSTRACT

<b>1. INTRODUCCIÓN</b> .....	<b>1</b>
<b>2. PLANTEAMIENTO DEL PROBLEMA</b> .....	<b>3</b>
<b>3. MARCO TEÓRICO</b> .....	<b>5</b>
3.1. GEOGRAFÍA DE LA SALUD .....	5
3.2. SISTEMAS DE INFORMACIÓN GEOGRÁFICA .....	7
3.3. GEOCODIFICACIÓN .....	8
3.4. FUENTES DE INFORMACIÓN GEOESPACIAL .....	11
3.5. MODELO DE DATOS DE DIRECCIONES .....	13
3.6. GEORREFERENCIACIÓN.....	16
3.7. SISTEMA DE POSICIONAMIENTO POR SATÉLITE .....	17
3.8. PROYECTO HOPE .....	19
<b>4. JUSTIFICACIÓN DE LA INVESTIGACIÓN</b> .....	<b>21</b>
<b>5. OBJETIVOS</b> .....	<b>22</b>
5.1. OBJETIVO GENERAL .....	22
5.2. OBJETIVOS ESPECÍFICOS .....	22
<b>6. MATERIALES Y MÉTODOS</b> .....	<b>23</b>
6.1. DISEÑO DE ESTUDIO.....	23
6.2. POBLACIÓN .....	23
6.2.1. Criterios de inclusión .....	23
6.3. OPERACIONALIZACIÓN DE VARIABLES .....	24

6.4. PROCEDIMIENTOS Y TÉCNICAS.....	25
6.4.1. Base de datos de direcciones .....	25
6.4.2. Cartografía de referencia .....	25
6.4.3. Geocodificación manual .....	26
6.4.4. Georreferenciación.....	26
6.4.5. Integridad de la geocodificación a nivel manzana.....	27
6.4.6. Estimación de la exactitud de la geocodificación.....	28
6.4.7. Identificación de factores que conllevan a la inexactitud .....	28
6.4.8. Tiempo requerido para georreferenciar una vivienda y geocodificar su respectiva manzana .....	28
6.4.9. Costo de georreferenciación y geocodificación.....	29
6.4.10. Identificación de áreas desatendidas por las agentes comunitarias de HOPE .....	30
6.5. PLAN DE ANÁLISIS .....	30
6.5.1. Análisis Espacial.....	30
6.5.2. Análisis Estadístico.....	31
6.6. CONSIDERACIONES ÉTICAS.....	31
<b>7. RESULTADOS.....</b>	<b>32</b>
7.1. CALIDAD DE LOS DATOS DE DIRECCIONES RECIBIDOS.....	32
7.2. INTEGRIDAD DE LA GEOCODIFICACIÓN A NIVEL MANZANA .	33
7.3. EXACTITUD DE LA GEOCODIFICACIÓN .....	34
7.4. FACTORES QUE CONLLEVAN A LA INEXACTITUD EN LA GEOCODIFICACIÓN DE DIRECCIONES A NIVEL MANZANA.....	35

7.5. TIEMPO PROMEDIO REQUERIDO PARA GEORREFERENCIAR UNA VIVIENDA Y GEOCODIFICAR SU RESPECTIVA MANZANA.....	36
7.6. COSTO DE GEORREFERENCIACIÓN Y GEOCODIFICACIÓN .....	37
7.7. ÁREAS DESATENDIDAS POR LAS AGENTES COMUNITARIAS DE HOPE .....	39
<b>8. DISCUSIÓN.....</b>	<b>41</b>
<b>9. CONCLUSIONES .....</b>	<b>46</b>
<b>10. RECOMENDACIONES.....</b>	<b>47</b>
<b>REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>48</b>
<b>ANEXOS</b>	

## RESUMEN

La geocodificación o asignación de coordenadas geográficas a direcciones es requerida cada vez con más frecuencia en investigaciones en salud pública. Sin embargo, el uso del patrón de calles en los sistemas geocodificadores automatizados se dificulta cuando las direcciones son basadas en el nombre del asentamiento humano, manzana y lote, como ocurre con frecuencia en el Perú. Se realizó un estudio transversal descriptivo para estimar la exactitud de la geocodificación manual de direcciones, usando como referencia la georreferenciación por sistema de posicionamiento global. El proyecto HOPE facilitó las direcciones de 233 beneficiarias residentes del distrito de Mi Perú. La geocodificación manual se realizó empleando Google Maps, Geodir y el plano distrital 2016, lográndose geocodificar el 82.83% de las direcciones. Se encontró que una geometría incompleta de la cartografía; así como, un registro incorrecto de las direcciones, conllevan a la inexactitud de la geocodificación. El tiempo promedio para geocodificar una vivienda a nivel de manzana (1 minuto y 25 segundos) fue 17 veces menor que el requerido para georreferenciarla (25 minutos). La geocodificación manual es más tiempo- y costo-efectiva que la georreferenciación, y ofrece una exactitud adecuada en contextos como el de las zonas urbano-marginales del Perú.

**Palabras clave:** geocodificación, georreferenciación, exactitud, asentamientos humanos, manzanas.

## **ABSTRACT**

Geocoding or assignment of geographic coordinates to addresses is increasingly required in public health research. However, the use of the street pattern in automated geocoding systems is difficult when addresses are based on the name of the human settlement, block and lot, as is often the case in Peru. A descriptive cross-sectional study was carried out to estimate the accuracy of manual address geocoding, using georeferencing by global positioning system as a reference. The HOPE project provided the addresses of 233 beneficiaries residents of the Mi Perú district. The manual geocoding was carried out using Google Maps, Geodir and the 2016 district map, achieving geocoding in 82.83% of the addresses. We found that an incomplete geometry of the cartography; as well as, an incorrect registration of the addresses, lead to the inaccuracy of the geocoding. The average time to geocode an address at the block level (1 minute and 25 seconds) was 17 times shorter than that required to georeference it (25 minutes). Manual geocoding is more time- and cost-effective than georeferencing, and it offers an adequate accuracy in contexts such as that of urban-marginal areas of Peru.

**Keywords:** geocoding, georeferencing, accuracy, human settlements, blocks.

## 1. INTRODUCCIÓN

Los sistemas de información geográfica (SIG) cumplen un rol importante en la salud pública, ejerciendo acción en la vigilancia epidemiológica y la evaluación de las desigualdades de acceso a servicios de salud (1–3). La capacidad de vincular atributos espaciales con datos sociodemográficos y de salud, han convertido a los SIG en una herramienta para la predicción de patrones de enfermedades y asociaciones ecológicas (1,2). El análisis de riesgos a nivel de individuos en diversos estudios ha medido la proximidad residencial a una fuente de exposición (3,4). La ubicación específica del sujeto permite identificar con mayor exactitud los eventos de interés y brinda la posibilidad de realizar inferencias causales (5).

La geocodificación o codificación geográfica consiste en la asignación de coordenadas a direcciones residenciales o postales (6,7). Requiere de un conjunto de datos, algoritmos y cartografía georreferenciada (8). Este proceso implica analizar datos de entrada como direcciones, consultar con los datos espaciales de referencia mediante atributos como el nombre y numeración de calle, y determinar la ubicación del punto asignando sus respectivas coordenadas. En ocasiones el proceso requiere interpolación (9,10).

Las redes de calles son el modelo de datos de direcciones más empleado por los sistemas geocodificadores a nivel mundial (9,11). Diversos estudios han evaluado la calidad de la geocodificación de servicios comerciales, gratuitos o elaborados como recurso propio por un equipo de investigación (12–16). No obstante, estos sistemas, al emplear un patrón de calles como referencia, limitan la

ubicación de direcciones que no siguen esta nominación o no cuentan con una nomenclatura definida.

La calidad de los datos de direcciones y la concordancia con las unidades geográficas del patrón de referencia limitan la geocodificación (9). Se ha reportado que la exactitud posicional de diferentes métodos de geocodificación varía de acuerdo al área geográfica (urbana o rural), distribución espacial de ciudades y los asentamientos rurales (15). No existe un sistema geocodificador único que funcione para cada usuario o situación de interés (17).

La investigación realizada sobre el método más rentable para mejorar la calidad de los datos geocodificables es escasa. Se ha determinado que estimar el costo y/o tiempo requerido para realizar la corrección de la codificación geográfica a través de procesos manuales, permite identificar un recurso de bajo costo y fácil empleo para el proceso de geocodificación. Determinar cómo dicha corrección puede mejorar la calidad de la codificación geográfica requiere de la determinación de la exactitud posicional (7,18).

El presente estudio estima la exactitud posicional de la geocodificación de direcciones basadas en el nombre de asentamientos humanos y manzanas en el distrito de Mi Perú, empleando la base de datos del componente social del proyecto HOPE de la Universidad Peruana Cayetano Heredia. Esto permitió establecer el recurso más efectivo y rentable para la identificación de los domicilios en zonas urbano-marginales.

## **2. PLANTEAMIENTO DEL PROBLEMA**

El uso de la geocodificación de direcciones residenciales en estudios epidemiológicos se ha convertido en un componente esencial del análisis geoespacial, vinculando factores socioeconómicos con factores medio ambientales (2,3,6). La integridad de los registros geocodificados, la exactitud posicional y la concordancia con la cartografía de referencia permiten establecer la calidad del dato geocodificado (8). La exactitud es el valor que indica la proximidad a los datos reales de ubicación (19). Sin embargo, este factor depende de la calidad de los datos del registro de direcciones y los recursos cartográficos a emplearse (9,13–15).

El patrón de calles, parcelas y puntos de direcciones son empleados como referencia para la geocodificación (11). Las redes de calles, ampliamente utilizadas por los sistemas geocodificadores y herramientas de mapeo en línea como Google Maps, Google Earth, Waze, entre otros (14–16), no permiten la geocodificación en comunidades de reciente formación donde aún no se establece un nombre de calle y/o número de la puerta. El mismo problema ocurre cuando las direcciones están registradas bajo otra codificación como la combinación de asentamiento humano, manzana y lote.

Los modelos de parcelas de tierra y puntos de direcciones como referencia para la geocodificación presentan limitaciones en zonas donde las parcelas son grandes y contienen muchos domicilios a geocodificar, lo cual varía según criterio del investigador. No obstante, implica un mayor esfuerzo de trabajo y mayor cantidad de recursos cartográficos de diferentes fuentes; por lo cual, es poco empleada (6,8).

Los problemas de coincidencia de dirección se producen hasta en un sistema geocodificador de alta precisión. Los errores de digitación y diferentes formatos de direcciones generan puntos geocodificados incorrectos o inexistentes (6). Dichos errores deben ser caracterizados para el uso de los resultados que produce el sistema geocodificador. Los datos geocodificados que no asemejan a los valores reales, afectan negativamente a los resultados del análisis espacial (8).

La geocodificación en áreas urbanas recientes o zonas urbano marginales es compleja. El uso de esta metodología en mayor escala en el proyecto HOPE permitirá realizar un análisis espacial de la distribución de las participantes para mejorar la programación de actividades de campo, enfocándose en áreas de baja cobertura. Por lo tanto, se plantea estimar la exactitud de la ubicación geocodificada usando como referencia la georreferenciación e identificar los factores que conlleven a la inexactitud de la geocodificación.

### 3. MARCO TEÓRICO

#### 3.1. GEOGRAFÍA DE LA SALUD

La geografía de la salud estudia los efectos del medio ambiente en la salud, así como la distribución geográfica de la enfermedad. Conceptualiza la ubicación espacial como el contexto geográfico y el vínculo con otras áreas; su rol esencial en la salud, bienestar y enfermedad (20). La implicancia de esta disciplina se ha incrementado en las investigaciones epidemiológicas; explorando los contextos sociales, políticos y culturales para la salud. (21).

Esta disciplina forma parte de la geografía humana; la cual estudia la interrelación entre las personas, lugares, entornos; así como, su variación en tiempo y espacio (21). Su importante rol en salud pública se centra en la ubicación de los eventos de interés en salud para facilitar la representación y el análisis espacial de la información de manera rápida y precisa (20).

El término “geografía de la salud” surgió en el Congreso de la Unión Geográfica Internacional (UGI) de 1976. Dicho congreso determinó realizar el cambio de nominación de geografía médica a geografía de la salud, y estableció dos campos de acción (22):

- a) La geografía de las enfermedades: Esta línea de aplicación abarca la identificación de factores de riesgo de enfermedades conjuntamente con el reconocimiento de patrones espaciales.
- b) La geografía de los servicios de salud: Este campo se encuentra vinculado a la planificación y accesibilidad a los servicios de salud.

Sin embargo, la geografía médica ha estado presente a lo largo de la historia. En el siglo V a.c., el tratado hipocrático “Del aire, las aguas y los lugares” buscó identificar los factores de riesgo asociados a enfermedades endémicas de un lugar. Este tratado fue considerado el primer libro escrito sobre geografía médica, climatología y antropología (20). En 1792, surgió el campo de estudio de la geografía médica en una obra dedicada a la topografía de las enfermedades (22).

John Snow en 1854 realizó un estudio cartográfico sobre la incidencia de casos de cólera con epicentro en una bomba de agua. Este hallazgo resaltó la importancia de la cartografía y las correlaciones espaciales en la salud (22,23). Diversos conceptos fueron acuñados a este campo de estudio, siendo definida en 1950 como el estudio sistémico de las enfermedades de la tierra y de la población. La introducción de métodos cualitativos y cuantitativos, la creación de pautas de distribución espacial y una visión ecológica conllevaron a la especialización de la investigación geográfica (22).

La dinámica actual de la geografía de la salud no se encuentra limitada a la geografía espacial del evento o carácter de interés; abarca el entendimiento del problema de salud y las acciones que conlleven a su solución. El principal desafío es comprender individualmente cada problema de salud; así como, su relación con el entorno como la pobreza, densidad poblacional, urbanización, el deterioro ambiental, entre otros (20,22).

### **3.2. SISTEMAS DE INFORMACIÓN GEOGRÁFICA**

Los mapas han sido empleados para expresar información de manera rápida y entendible, a través del uso de lenguaje cartográfico de símbolos y colores. El desarrollo de la tecnología permitió que los mapas sean más detallados y exactos. La cartografía se convirtió en una herramienta capaz de generar mayor información. El uso de la estadística espacial en conjunto con la cartografía digital originó el surgimiento de los sistemas de información geográfica (20).

Los SIG constituyen un sistema computarizado que permite recopilar, almacenar, procesar, analizar y modelar datos espaciales (24). El ingreso de información se realiza importando tablas de datos georreferenciadas o con características que permitan ser vinculadas a la cartografía digital. Las funciones analíticas permiten procesar los datos de mapas de referencia y datos espaciales de interés (25). Los productos obtenidos son la presentación de los datos en forma de mapas, planos o cartas (24).

Los usos de los SIG varían desde la localización de un punto de interés (individuo, centro de salud, entre otros) mediante la geocodificación o la georreferenciación (3,6,26), hasta la integración de datos de vigilancia epidemiológica (27). La ubicación de los datos debe ser confiable y precisa para que los SIG representen con exactitud los eventos de interés en la superficie terrestre. Los datos, la escala y la exactitud son elementos relevantes para las aplicaciones de SIG (25).

Los modelos ráster y vectorial son las representaciones de espacio empleadas en los SIG. Estos modelos van a depender del tipo de datos espaciales de entrada. En el primero, las celdas de la cuadrícula son unidades de análisis como son los píxeles de las imágenes satelitales y fotografías aéreas georreferenciadas. Con respecto al segundo, el modelo vectorial emplea puntos, líneas y polígonos según la geometría del espacio como representación de la información de interés (25).

El modelo vectorial emplea como puntos a las ubicaciones específicas de un evento o lugar de interés, como un establecimiento de salud o una vivienda. Las líneas hacen referencia a un segmento georreferenciado como las redes de calles, mientras que un polígono está vinculado a un área georreferenciada como las manzanas. Los datos espaciales de líneas o polígonos son esenciales para el proceso de geocodificación de direcciones o códigos postales (5,6).

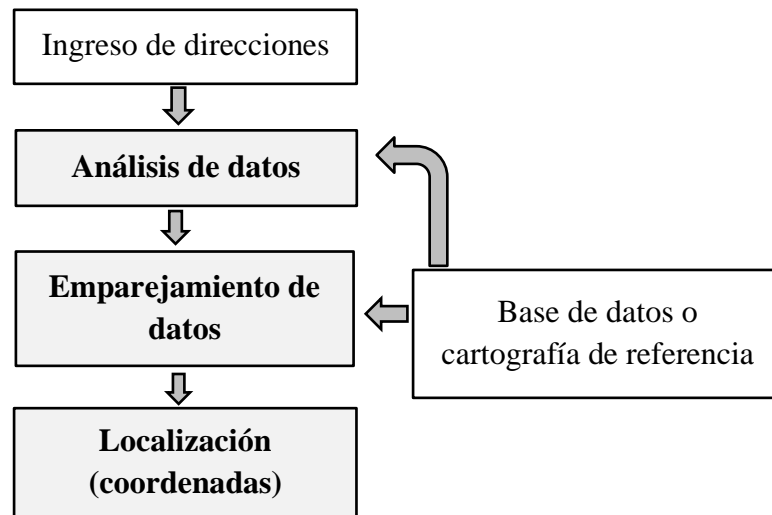
### **3.3. GEOCODIFICACIÓN**

La geocodificación es el proceso de asignar coordenadas a una dirección (6,12). Las primeras investigaciones realizadas en geocodificación se desarrollaron en áreas distintas al sector salud. Sin embargo, su uso en las ciencias de salud fue incrementándose, siendo considerada actualmente como una herramienta esencial y práctica para el análisis espacial de la salud (6).

Este recurso es un complejo conjunto de datos, algoritmos y cartografía de referencia (11). Existe una gran variedad de algoritmos que permiten vincular una dirección a su respectiva ubicación geográfica en una cartografía. Estos

algoritmos pueden conllevar a errores de los resultados geocodificados (18,28–30). Es por ello, que no existe un sistema geocodificador único, dependen del área de estudio, temporalidad, calidad de datos y las variables a analizar por el investigador (17,24).

### **Proceso de geocodificación**



El proceso de geocodificación incluye pasos fundamentales como el análisis, el emparejamiento de datos y la localización como valores de coordenadas. Los datos de direcciones requieren de un análisis que permita estructurarlos. Estos deben seguir un modelo de direcciones concordante con la cartografía de referencia como las redes de calles (nombre y número de calle). El emparejamiento de datos compara y vincula las direcciones estructuradas a una cartografía de referencia, donde se establecen los puntos de geocodificación. La localización establece las coordenadas de la dirección geocodificada (30).

La geocodificación requiere una cartografía de referencia para establecer los puntos de geocodificación resultantes en la localización. Esta cartografía de referencia es un recurso cartográfico como un plano, un mapa o una base de datos georreferenciados como Google Maps (14,31,32). Los recursos cartográficos son fuentes de información geoespacial que permiten identificar la ubicación geográfica de la dirección. La inclusión de más recursos cartográficos para optimizar la integridad de la geocodificación va a depender del área geográfica y el modelo de direcciones (31–33).

La calidad de la geocodificación depende de la integridad y el nivel de exactitud posicional. La **integridad o completitud** es definida como la proporción de direcciones geocodificadas del total de direcciones destinadas a ese proceso. Esta se ve afectada por los errores previos y durante el proceso de geocodificación (6,8,14). La **exactitud** se define como el nivel de proximidad con respecto a la realidad en relación a los atributos, temporalidad y posición (9,34). Si bien existen varios tipos de exactitud, el presente estudio se circunscribe a exactitud posicional, es decir, la distancia entre el punto georreferenciado y el punto más cercano en la manzana geocodificada (6,12,15,34).

Los errores correspondientes a la información geográfica pueden reducirse mediante el mapeo en campo del lugar, empleando información registrada por sistemas de posicionamiento global (GPS) y adquiriendo cartografía de referencia elaborada por centros de investigación especializados en información geoespacial (6,8).

### 3.4. FUENTES DE INFORMACIÓN GEOESPACIAL

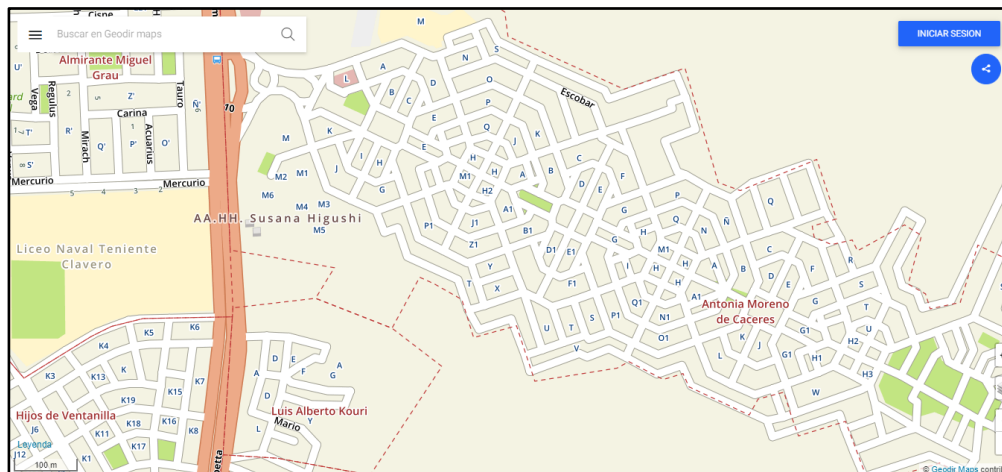
Un dato espacial contiene valores posicionales únicos, es decir presenta coordenadas u otro identificador espacial que permita ubicarlo en un espacio geográfico. Hace referencia a objetos geográficos como edificios, casas, ríos o calles. Estos datos están asociados a información no espacial o atributos que le brinda características específicas como el nombre, la dirección, dimensiones u otros elementos (24,35,36).

La disponibilidad de datos espaciales confiables es un requisito importante para garantizar una geocodificación exacta y precisa (35). Los sistemas geocodificadores emplean recursos cartográficos para realizar el vínculo entre la dirección y sus coordenadas. La selección de estos recursos o fuentes de información geoespacial es un desafío para el investigador.

Los geocodificadores de software SIG como ArcGIS (comercial) y QGIS (gratuito) emplean repositorios de datos espaciales de entidades gubernamentales, repositorios de código abierto como OpenStreetMap (OSM) y herramientas de mapeo en línea. ArcGIS también cuenta con repositorios privados del proveedor *Environmental Systems Research Institute Inc* (ESRI) (35,36).

Empresas privadas han ido desarrollando software SIG basado en *cloud computing* que permite almacenar la información en servidores de Internet y luego ser empleada en servidores físicos. ESRI ArcGIS Online brinda este servicio de manera comercial (35). Geodir es una empresa privada que cuenta con una plataforma en la nube que provee mapeo en línea gratuito (Geodir

Maps) y servicios de geocodificación (Geodir Geocoder) de manera comercial (37). A diferencia de un software SIG, esta plataforma nacional cuenta con mapas online ajustados a la realidad peruana, emplea las redes de calles y zonas urbanas como urbanizaciones, asentamientos humanos y manzanas (Figura N°1) (38).



**Figura N°1. Captura de pantalla de la plataforma de mapas en línea de Geodir en el distrito de Ventanilla, Provincia Constitucional del Callao**

Los servicios de mapas en líneas o cartografía web de empresas como Google o Microsoft han mejorado sus características y capacidades de mapeo desde su creación en el 2005. Herramientas como Google Earth, Google Maps y Bing Maps (Microsoft) están transformando la industria del software SIG al realizar aportes significativos de datos espaciales de la web. No son consideradas SIG porque aún no realizan todas las funciones de análisis espacial de un SIG. Sin embargo, en los próximos años, pueden desarrollar herramientas sofisticadas de análisis comparadas a los SIG (35,36).

Los mapas y planos georreferenciados físicos o digitalizados brindan datos espaciales confiables cuando son realizados por entidades gubernamentales o entidades privadas de prestigio. El Perú cuenta con plataformas web del gobierno que brindan cartografía georreferenciada del territorio nacional. Una de las instituciones que brinda dicha información es el Instituto Nacional de Estadística e Informática (INEI), otorga previa solicitud, cartografía actualizada a la fecha del último censo nacional, en formato compatible con software SIG. La cartografía disponible se encuentra limitada a nivel de manzanas en los asentamientos humanos de zonas urbano-marginales y áreas urbanas de reciente formación.

### **3.5. MODELO DE DATOS DE DIRECCIONES**

La geocodificación vincula una matriz de direcciones con una cartografía de referencia que cuenta con datos espaciales y atributos. Dichos atributos como el nombre de calle y número de casa asignados a la coordenada, deben ser concordantes con la nomenclatura presentada en la matriz de direcciones a geocodificar. Requiere seguir un modelo de datos de direcciones. Este permite organizar los atributos de forma lógica y específica para coincidir con la nomenclatura de una dirección a geocodificar (11,19).

El modelo de datos de direcciones más empleado en el mundo se basa en las redes de calles digitalizadas como segmentos de líneas de calle. Dicho segmento contiene nombre de calle y rango de números de casa en cada lado de la calle. La codificación geográfica de direcciones se realiza vinculando el nombre de calle, luego el número de la casa y finaliza al colocar un punto en

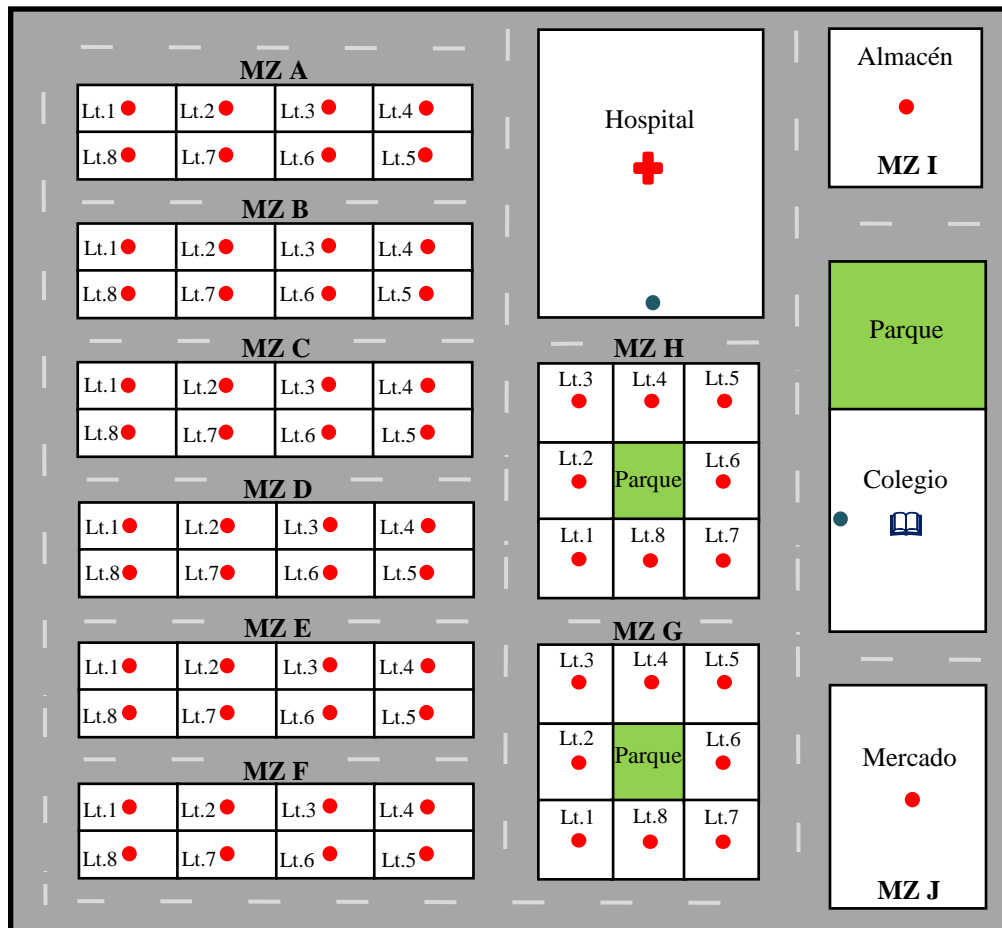
dicha dirección. Sin embargo, no todas las direcciones logran ser geocodificadas de forma directa. La ausencia del número de casa a buscar en la cartografía de referencia, conduce al geocodificador a emplear la interpolación para ubicar la dirección. Este modelo de datos es empleado por los geocodificadores de la mayoría de software SIG (8,9,11).

El modelo de dirección de parcelas permite coincidir parcelas de tierra individuales con las direcciones, en lugar de emplear la línea de calle. A cada parcela o polígono se le asigna una coordenada específica ubicada en su punto central o centroide (Figura N°2), vinculado con la dirección a geocodificar (6,9,11). Este método brinda una geocodificación más exacta que la ubicación establecida en la línea de calle cuando la parcela representa una dirección. Sin embargo, en parcelas grandes con varios domicilios como las manzanas, la exactitud es menor. Las coordenadas del centroide son asumidas como únicas para todos los lotes que se ubican dentro de la manzana (6).

La geocodificación basada en puntos de direcciones surge como un modelo de direcciones para subsanar las limitaciones de las parcelas. Estos puntos representan las ubicaciones de todas estructuras de interés que cuentan con una dirección definida (Figura N°2). Se obtienen a partir de la georreferenciación en campo o la verificación de la ubicación mediante imágenes satelitales (8,9,11).

La geocodificación automatizada realizada por un software SIG o geocodificadores en línea requiere de bases de datos espaciales. Las redes de calles y mapas de parcelas son un repositorio de dichos datos. Este repositorio

es una matriz de puntos de direcciones georreferenciados. Las bases de datos espaciales más empleadas y accesibles son Google Maps, Google Earth, Bing Maps y OSM (35,36). Estos recursos de mapeo en línea son empleados también como referencia en la geocodificación manual.



**Figura N°2. Geocodificación de parcelas (puntos rojos) y puntos de direcciones (azules)**

### 3.6. GEORREFERENCIACIÓN

La georreferenciación brinda la capacidad de ubicar un sujeto, objeto o evento de interés con exactitud en un espacio geográfico. Se realiza estableciendo las coordenadas del lugar de interés mediante un estudio de campo con receptores de constelaciones de los sistemas de navegación por satélite (GNSS) o a través de imágenes satelitales (19,35,39).

El proceso de identificar el objeto o lugar y la asignación de sus correspondientes coordenadas se denomina emparejamiento. Los datos vectoriales necesitan ubicar el punto o línea de interés y asignar sus coordenadas. Las imágenes ráster requieren identificar puntos de control terrestre. Éstos son puntos de referencia con coordenadas conocidas, como cruces de calles o establecimientos. La imagen se divide en grupos de píxeles con características similares. Estos grupos se correlacionan con los puntos de control y se genera una capa ráster georreferenciada (39).

La georreferenciación generalmente se realiza empleando un sistema de referencia geodésico como WGS84 (*World Geodetic System 84*). Este comprende un marco de coordenadas geográficas para la Tierra, un elipsoide referencial y un geoide que define el nivel del mar. Existen otros sistemas de referencia, el uso de cada uno de ellos va a depender del campo de aplicación específico. Sin embargo, el WGS84 es ampliamente usado a nivel mundial y con mayor frecuencia (39).

La captación de coordenadas en la georreferenciación in situ se realiza con mediciones GNSS (por su nombre en inglés *Global Navigation Satellite*

*System*) de dispositivos especializados. Estos incluyen receptores de constelaciones GNSS en su matriz interna o adheridos al exterior. Los receptores GNSS de alta precisión y exactitud (errores de posición inferiores a 1 mm) de instrumentos topográficos empleados en el ejército, estudios de ingeniería y cartografía; son costosos y requieren personal especializado para su uso. Los estudios epidemiológicos suelen emplear dispositivos más accesibles y económicos, con un error posicional que varía de 10 a 20 metros aproximadamente (19). Sin embargo, este margen de error está decreciendo con la optimización de los dispositivos con sistema de posicionamiento.

### **3.7. SISTEMA DE POSICIONAMIENTO POR SATÉLITE**

El GNSS agrupa una red de satélites que proporciona servicios de navegación, posicionamiento y cronometraje en el mundo. El desarrollo y uso múltiple de constelaciones GNSS ha mejorado la visibilidad satelital, la precisión posicional y la confiabilidad de los datos espaciales obtenidos. Dichas constelaciones son el Sistema de Posicionamiento Global (GPS) de origen norteamericano, GLONASS proveniente de Rusia, Galileo del continente europeo y BeiDou de origen chino, como las resaltantes.(40–42).

El GPS ha sido el principal GNSS, surgió para su uso militar, pero en el año 1983 se autorizó que civiles tengan acceso a ellas para usos de posicionamiento y navegación, con un error intencional a nivel posicional. La disponibilidad selectiva (SA, *selective availability*) fue una degradación premeditada de las señales GPS públicas, cuyo objetivo era evitar su uso por enemigos (43). Recién en el año 2000 se suspende la disponibilidad selectiva,

eliminando este error. La estimación sobre la exactitud del GPS con y sin SA muestra que, en una dispersión posicional de puntos, el 95% se ubica en un radio de 45 metros con SA; mientras que sin SA se ubican en un radio de 6.3 metros (44).

El GPS ha sido ampliamente empleado en receptores GNSS, unidades portátiles y dispositivos móviles inteligentes. Estudios realizados sobre el uso de múltiples constelaciones GNSS en conjunto evidenciaron su potencial en el posicionamiento espacial de precisión (41,45,46). Seo (45) reportaron que la integración de GLONASS en el sistema de posicionamiento GPS redujo el error máximo de 9.5 metros a 6.4 m.

Las mediciones GNSS fueron incluidas en teléfonos inteligentes y tabletas con sistema operativo Android N (*Nougat* o versión 7) en 2016. Broadcom, empresa líder en tecnología de infraestructura, anunció al mundo el primer chip GNSS de doble frecuencia en 2017. Otros productores empezaron a lanzar sus equipos con doble frecuencia, revolucionando los métodos de posicionamiento GNSS de los teléfonos inteligentes (47). La inclusión del sistema GLONASS, Galileo y BeiDou a los receptores GNSS ha sido denominado como un servicio multi-GNSS, el cual optimiza la geometría espacial y el posicionamiento (40,46).

La Oficina Nacional de Coordinación para el Posicionamiento, Navegación y Cronometraje de la Fuerza Espacial de EEUU menciona que los teléfonos inteligentes con GPS suelen tener una exactitud de 4.9 metros de radio en condiciones normales bajo cielo (48). Un estudio realizado sobre la

exactitud posicional de un dispositivo GPS estándar y un receptor GPS integrado a teléfonos móviles reportó un rango de 3.4 a 4.7 m para el primero, y de 3 a 5 m para el segundo. Sin embargo, los receptores multi-GNSS han demostrado una mayor exactitud que receptores más antiguos. Se ha reportado que unidades portátiles como el GPSMap 66 y teléfonos inteligentes con Android 8, reportaron valores de 1.15 m y 2.26 m, respectivamente, bajo cielo despejado (42).

### **3.8. PROYECTO HOPE**

El proyecto HOPE surge como una estrategia de emprendimiento social no lucrativo en salud pública de la Universidad Peruana Cayetano Heredia. Tiene como principal objetivo la prevención y la optimización del manejo del Cáncer de cuello uterino (CaCu). Esta innovación social busca concientizar, informar y poner a disposición la prueba rápida molecular por autotoma para la detección del virus del papiloma humano (VPH) (49,50).

El componente social del proyecto HOPE ofrece la prueba rápida en comunidades con desventajas socioeconómicas. Mujeres voluntarias (nominadas Ladies HOPE) son capacitadas como agentes comunitarios de salud para promover la detección del cáncer de cuello uterino a través del auto muestreo de VPH. Ellas venden un kit de autotoma subsidiado y reciben un incentivo por cada prueba vendida. Cada Lady HOPE ofrece el kit de autotoma e informa sobre la importancia del despistaje de VPH a mujeres de su comunidad casa por casa o en centros públicos como mercados (49).

La ubicación de los domicilios de las beneficiarias es esencial para un adecuado seguimiento de beneficiarias e implementación de estrategias para la prevención del VPH. Sin embargo, la cartografía disponible para la zona y la nomenclatura de las direcciones limita este proceso y aún no se ha implementado un recurso eficaz que permita localizar las direcciones.

#### **4. JUSTIFICACIÓN DE LA INVESTIGACIÓN**

La geocodificación es una herramienta esencial en el análisis espacial de factores sociodemográficos en la salud (6,12). Los sistemas geocodificadores comerciales y gratuitos emplean las redes de calle (5,6,9,11). Las direcciones basadas en AAHH, manzanas y lotes requieren métodos basados en parcelas y puntos de direcciones para su geocodificación. Los sistemas dedicados a la geocodificación de estas direcciones son escasos y disponibles de manera comercial.

La presencia de múltiples comunidades donde la nomenclatura de las direcciones es basada en AA. HH, urbanizaciones, sectores, agrupaciones poblacionales, zonas, manzanas y lotes, exhibe la necesidad de un proceso que permita la geocodificación de dichas zonas geográficas de manera rápida y eficiente.

La geocodificación de las direcciones de zonas urbano-marginales y de reciente formación brinda la oportunidad de realizar análisis espaciales necesarios para la investigación y el establecimiento de estrategias eficaces para la población, tanto en el sector salud como educación, seguridad ciudadana y otras áreas.

En vista de la necesidad encontrada, se justifica identificar un recurso rápido y eficiente para la ubicación de las direcciones basadas en AAHH y manzanas. Estimar la exactitud posicional que brinda la geocodificación de estas direcciones. Además, identificar los factores que conlleven a la inexactitud de la geocodificación en zonas urbano-marginales y áreas de reciente formación.

## **5. OBJETIVOS**

### **5.1. OBJETIVO GENERAL**

Estimar la exactitud posicional de la ubicación geocodificada a nivel de manzana del registro de direcciones en el marco del proyecto HOPE, usando la georreferenciación por GPS como gold standard.

### **5.2. OBJETIVOS ESPECÍFICOS**

- Estimar la integridad de la geocodificación, medida como la proporción de viviendas geocodificadas.
- Identificar los problemas que resultan en inexactitud en la geocodificación basada en registro de direcciones.
- Comparar el tiempo promedio requerido para geocodificar y georreferenciar una dirección a nivel manzana.
- Comparar los costos requeridos para realizar la geocodificación y georreferenciación de direcciones.
- Determinar la existencia de áreas desatendidas por las agentes comunitarias del proyecto HOPE.

## **6. MATERIALES Y MÉTODOS**

### **6.1. DISEÑO DE ESTUDIO**

El presente es un estudio transversal descriptivo

### **6.2. POBLACIÓN**

La población fue conformada por las direcciones de las mujeres beneficiarias del componente social del proyecto HOPE, registradas entre marzo del 2019 y febrero del 2020. Las direcciones corresponden a viviendas del distrito de Mi Perú, provincia constitucional del Callao. El estudio abarcando la totalidad de 243 registros de mujeres tamizadas según los criterios de inclusión. El proyecto HOPE siguió un método de muestreo no probabilístico en el que las promotoras eligen por conveniencia a las beneficiarias.

#### **6.2.1. Criterios de inclusión**

- ✓ Mujeres que residen en el distrito de Mi Perú.
- ✓ Mujeres tamizadas del proyecto HOPE desde marzo 2019 hasta febrero 2020.

### 6.3. OPERACIONALIZACIÓN DE VARIABLES

**Tabla N°1. Operacionalización de variables**

<b>Variable</b>	<b>Definición</b>	<b>Naturaleza y escala de medición</b>	<b>Fuente de información</b>	<b>Unidad de medida</b>
<b>Exactitud</b>	Distancia entre el punto georreferenciado y el punto más cercano en la manzana geocodificada	Cuantitativa Continua De razón	Coordenadas UTM	Metros (m)
<b>Número de viviendas geocodificadas</b>	Cantidad de direcciones que se geocodificaron a nivel manzana	Cuantitativa Discreta	Direcciones geocodificadas manualmente	
<b>Número de viviendas georreferenciadas</b>	Cantidad de direcciones que se georreferenciaron	Cuantitativa Discreta	Direcciones georreferenciadas	
<b>Tiempo total de geocodificación a nivel de manzana</b>	Tiempo total para el proceso de geocodificación a nivel manzana	Cuantitativa Continua De razón	Medición directa	segundos
<b>Tiempo total de georreferenciación de direcciones</b>	Tiempo total para el proceso de georreferenciación de las viviendas	Cuantitativa Continua De razón	Medición directa	segundos
<b>Costo total de geocodificación de direcciones</b>	Costo total para el proceso de geocodificación de las direcciones a nivel manzana	Cuantitativa Continua De razón	Medición directa	soles
<b>Costo total de georreferenciación de direcciones</b>	Costo total para el proceso de georreferenciación de las viviendas	Cuantitativa Continua De razón	Medición directa	soles

## **6.4. PROCEDIMIENTOS Y TÉCNICAS**

Los pasos realizados para la colección de datos se detallan a continuación.

### **6.4.1. Base de datos de direcciones**

El estudio HOPE cuenta con un registro digitalizado y físico de datos sociodemográficos de las beneficiarias, las cuales incluyen las direcciones residenciales. Se realizó una revisión de los formatos físicos para detectar errores de digitación.

Se elaboró una tabla donde se ingresaron los siguientes criterios: dirección completa, distrito, sector, asentamiento humano y manzana. A cada dirección se asignó un código identificador, empleado como código único para el proceso de geocodificación y georreferenciación.

### **6.4.2. Cartografía de referencia**

Para el presente estudio se utilizó como cartografía de referencia el plano digital actualizado en formato *shapefile* (.shp) para el distrito de Mi Perú, Callao elaborado por el Instituto Nacional de Estadística e Informática (INEI). El plano incluye: polígonos de áreas verdes, ubigeo, sectores, poligonal urbana y manzanas censales (Anexo N°1). Cada manzana del distrito se encuentra poligonizada, lo cual facilitó la selección e identificación de cada una de ellas. No obstante, las manzanas se encuentran bajo una nominación propia de la institución. Este plano no contiene información sobre lotes.

### **6.4.3. Geocodificación manual**

La geocodificación se realizó empleando el modelo de direcciones de parcelas. Este proceso se realiza vinculando la dirección con el centroide de las parcelas o polígonos (en este caso manzanas); en lugar de una línea central de calle (5,6,8). Asignando un único punto geocodificado para todos los lotes que se ubiquen dentro de cada manzana.

La nomenclatura particular de las manzanas en la cartografía de referencia requirió de recursos cartográficos complementarios que permitan nominar a los polígonos por nombre de asentamiento humano y manzana. La geocodificación manual empleó los siguientes recursos: Google Maps, Geodir Maps y el plano distrital digitalizado del 2016 realizado por la Gerencia de Desarrollo Urbano del Municipio Distrital (Anexo N°2).

Se realizó la geocodificación de las viviendas empleando de manera individual cada recurso cartográfico y luego los tres de manera conjunta.

### **6.4.4. Georreferenciación**

Por su elevada exactitud posicional, para este estudio se empleó la georreferenciación como gold standard para evaluar la exactitud de la geocodificación.

El proceso de georreferenciación requirió un celular con Android 9.0 al cual se instaló el aplicativo móvil del software Magpi. Se empleó un

formato simple de registro para ingresar coordenadas geográficas de cada dirección.

El punto a georreferenciar se determinó al situarse en la puerta del domicilio, asegurándose estar bajo cielo y esperar los segundos necesarios para el GPS se ubique adecuadamente (Figura N°3).



**Figura N°3. Georreferenciación de vivienda en el distrito de Mi Perú**

#### **6.4.5. Integridad de la geocodificación a nivel manzana**

La integridad de la geocodificación se estimó como la proporción de direcciones geocodificadas empleando cada recurso cartográfico individualmente y al emplear los 3 recursos en conjunto.

#### **6.4.6. Estimación de la exactitud de la geocodificación**

Las coordenadas geográficas obtenidas por georreferenciación y geocodificación de las direcciones fueron registradas como puntos rojos y amarillos, respectivamente, en la cartografía de referencia (Anexo N°3). Se estimó la exactitud de la geocodificación como la distancia entre el punto georreferenciado y el punto más cercano en la manzana geocodificada. En el Anexo N°4 se fundamenta esta decisión.

#### **6.4.7. Identificación de factores que conllevan a la inexactitud**

Se identificó los factores que conllevan a la inexactitud durante el proceso de geocodificación. Los estudios realizados sobre geocodificación mediante redes de calles reportan los siguientes factores (6,11).

**Tabla N°2. Factores asociados a la inexactitud de la geocodificación mediante redes de calles.**

<b>Registro de direcciones</b>	<b>Cartografía de referencia</b>
- Errores ortográficos	- Geometría inexacta
- Formatos incorrectos en la nomenclatura.	- Calles faltantes
	- Calles que no existen en la realidad.

#### **6.4.8. Tiempo requerido para georreferenciar una vivienda y geocodificar su respectiva manzana**

Se estimó el tiempo requerido para georreferenciar una vivienda y para geocodificarla hasta el nivel de manzana. Para esto se calculó el

tiempo total requerido para realizar el procedimiento, y se dividió entre el número de viviendas georreferenciadas o geocodificadas.

El tiempo total de georreferenciación se determinó registrando la hora inicio en el punto de partida hacia el distrito de Mi Perú y la hora término en el punto de partida igualmente, por día de trabajo realizado, estimando la sumatoria de horas realizadas para la georreferenciación de las direcciones.

El tiempo total de geocodificación se estimó registrando la hora de inicio y término del proceso de geocodificación manual por cada recurso cartográfico empleado Geodir, Google Maps, plano distrital. Por último, registrando el tiempo total empleando los 3 recursos.

#### **6.4.9. Costo de georreferenciación y geocodificación**

Se establecieron los costos fijos y costos variables para ambos procesos. Los costos fijos consistieron en los gastos generales como el servidor, software y celular. Los costos variables fueron los costos de internet, telefonía, cartografía de referencia, movilidad y costo por hora del operador, tanto para la revisión de la base de datos, la georreferenciación y la geocodificación.

Los costos de personal se fijaron empleando como referencia los salarios expedidos al personal del INEI, correspondiente a cargos profesionales equivalentes para el proceso de revisión de calidad de datos, georreferenciación y geocodificación. Se estimó 8 horas de trabajo por 5

días a la semana, un total de 160 horas de trabajo por mes aproximadamente. El salario fue dividido entre la cantidad de horas trabajadas para estimar un costo/hora de trabajo por operario.

Los costos de personal estimados para la revisión de calidad de calidad de la base de datos de direcciones y del operador responsable se obtuvo del salario de un encuestador de campo de la Encuesta Nacional de Hogares (ENAH) 2022 (51). El costo de personal para la geocodificación se estimó del salario de un especialista en revisión cartográfica y de datos del proyecto Censo Nacional Económico (CENEC) 2022 (52).

#### **6.4.10. Identificación de áreas desatendidas por las agentes comunitarias de HOPE**

Las áreas desatendidas por las Ladies HOPE fueron identificadas al obtener la distribución espacial de las manzanas geocodificadas en el distrito (Anexo N°3).

### **6.5. PLAN DE ANÁLISIS**

#### **6.5.1. Análisis Espacial**

La geocodificación manual se realizó mediante el software libre QGIS 3.10.2 “A Coruña” y los recursos cartográficos de Geodir, Google Maps y el plano distrital del 2016. Se presentó la distribución espacial de los puntos geocodificados y georreferenciados, para estimar la exactitud posicional de la geocodificación

### **6.5.2. Análisis Estadístico**

El análisis estadístico se realizó mediante la determinación de frecuencias y porcentajes de direcciones geocodificadas y no geocodificadas. Para describir la *Exactitud* se usó la mediana, rango intercuartílico y el rango de la variable. Se calculó el tiempo promedio requerido para geocodificar y geocodificar una vivienda. Además de la diferencia de costos requeridos para cada procedimiento.

### **6.6. CONSIDERACIONES ÉTICAS**

El presente estudio fue aprobado por el Comité de Ética de la Universidad Peruana Cayetano Heredia (código SIDISI 201566). Se solicitó la autorización del proyecto HOPE para obtener la base de datos de direcciones de las beneficiarias en el distrito de Mi Perú. No se tomó contacto con ninguna de las beneficiarias. La protección y confiabilidad de los datos se garantizó solicitando únicamente el registro de direcciones de las beneficiarias, manteniendo el anonimato de las participantes. Los datos fueron almacenados en una computadora con contraseña donde solo tuvo acceso el investigador.

## 7. RESULTADOS

### 7.1. CALIDAD DE LOS DATOS DE DIRECCIONES RECIBIDOS

Se recibió una base de datos con 243 registros, correspondientes al mismo número de beneficiarias del proyecto HOPE, residentes en el distrito. Se excluyó 10 (4.12%) registros que presentaron direcciones duplicadas (correspondientes a pares de beneficiarias que compartían la misma dirección), quedando una base de datos con 233 direcciones a geocodificar y georreferenciar.

El registro digitalizado evidenció que el 6.01% de direcciones presentaba errores ortográficos como ausencia de tilde y exceso o ausencia de letras en el nombre de AAHH. El 47.64% no registraban nombre del AAHH, sector o agrupación poblacional. Además, se observó que el nombre del distrito fue registrado bajo una codificación numérica en el 9.44% de las direcciones. Lo cual, obstaculiza el proceso de geocodificación sin su previa corrección (Tabla N°3).

**Tabla N°3. Direcciones corregidas para el proceso de geocodificación**

<b>Características corregidas</b>	<b>Cantidad de direcciones</b>	<b>Porcentaje</b>
Errores ortográficos en el nombre de AAHH	14/233	6.01%
Ausencia del nombre de AAHH, agrupación o sector poblacional	111/233	47.64%
Formato incorrecto de nombre del distrito	22/233	9.44%

Se realizó la corrección de la nomenclatura de direcciones, corroborando la información brindada con los registros físicos de direcciones, conjuntamente con el listado de AAHH, sectores o agrupaciones poblacionales existentes en el distrito de Mi Perú, según el “Plan estratégico de desarrollo concentrado del distrito de Mi Perú 2014 – 2024” (53). Sin embargo, al 7.30% (17/233) de las direcciones no se logró identificar el nombre de AAHH o sector poblacional correspondiente.

## 7.2. INTEGRIDAD DE LA GEOCODIFICACIÓN A NIVEL MANZANA

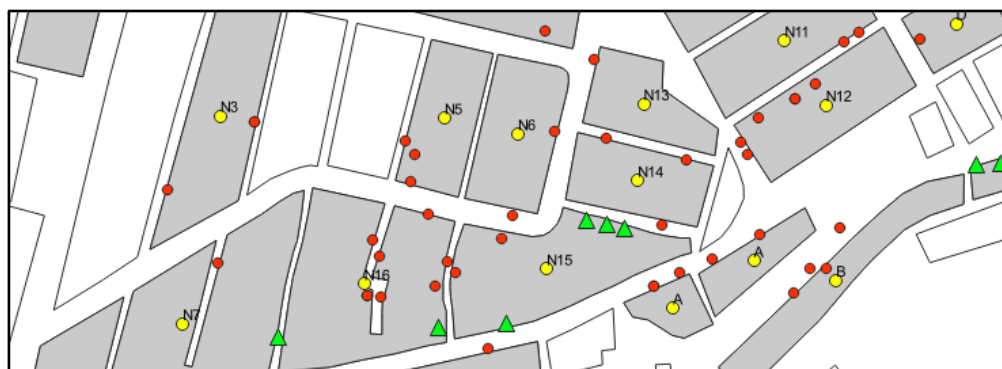
El proceso de geocodificación manual se realizó empleando recursos cartográficos de manera individual y luego de manera conjunta. El 73.82% de las direcciones se geocodificaron empleando el aplicativo web de mapas Geodir. El 78.54% del registro se logró geocodificar mediante el uso de Google Maps y el 77.68% se geocodificó empleando el plano distrital digitalizado del 2016. Se determinó que, empleando los 3 recursos cartográficos, se pudo geocodificar el 82.83% de las direcciones (Tabla N°4).

**Tabla N°4. Direcciones geocodificadas y no geocodificadas**

<b>Recurso cartográfico</b>	<b>Direcciones geocodificadas</b>	<b>Direcciones no geocodificadas</b>
Geodir	73.82% (172/233)	26.18% (61/233)
Google Maps	78.54% (183/233)	21.46% (50/233)
Plano distrital	77.68% (181/233)	22.32% (52/233)
Los 3 recursos	82.83% (193/233)	17.17% (40/233)

### 7.3. EXACTITUD DE LA GEOCODIFICACIÓN

Se observó que en general las direcciones georreferenciadas se ubicaron al margen de sus respectivas manzanas geocodificadas (Figura N°4). Se calculó la distancia entre el punto georreferenciado y el punto más cercano en la manzana geocodificada. Solo en 5 direcciones, esta distancia fue mayor de cero (Tabla N°5). Por tanto, la distancia tiene una mediana de cero, un rango intercuartílico de cero a cero y un rango de cero a 147 metros.



**Figura N°4. Puntos georreferenciados (rojo) ubicados a la periferia de sus respectivas manzanas con sus puntos geocodificados (amarillo).**

**Tabla N°5. Errores encontrados en el mapa de Geodir**

Errores	Exactitud posicional (metros)
Intercambio de nombre entre las Manzanas J5 y J4	23.07
	35.33
	58.95
Cambio de nombre de la Mz E4 por E3	47.41
AAHH Los Girasoles no concuerda con su ubicación real	147.23

#### 7.4. FACTORES QUE CONLLEVAN A LA INEXACTITUD EN LA GEOCODIFICACIÓN DE DIRECCIONES A NIVEL MANZANA

Se identificó las razones por las cuales algunas direcciones fueron inexactas o no se lograron geocodificar según cada recurso cartográfico y empleando los 3 juntos.

Se presentaron discrepancias en la información obtenida de los diferentes recursos cartográficos; en Geodir, se observó geometría inexacta en un AAHH y error de nomenclatura en dos manzanas (Tabla N°5). Tanto Geodir, Google maps y el plano distrital presentaron la ausencia de geometría de algunos AAHH y manzanas. Además, concordaron que la única agrupación poblacional en el distrito no se encuentra definida geográficamente (Tabla N°6). Con respecto al registro de direcciones, se observó que el 6.44% (15/233) de las direcciones no presentaron nombre de AAHH y el 0.86% (2/233) se ubican fuera del distrito.

**Tabla N°6. Razones por las cuales algunas direcciones no se geocodificaron según recurso cartográfico**

Hallazgos	Geodir	Google Maps	Plano distrital	Los 3 recursos
- Manzana (Mz) ausente	17	5	15	4
- AAHH ausente	16	18	13	13
- Nombre de Mz o AAHH erróneo	5	4	1	0
- Agrupación poblacional ausente	6	6	6	6
<b>Total</b>	44	33	35	23
<b>Porcentaje que no se geocodificó*</b>	<b>18.88%</b>	<b>14.16%</b>	<b>15.02%</b>	<b>9.87%</b>

\*Sobre 233 direcciones

**7.5. TIEMPO PROMEDIO REQUERIDO PARA  
GEORREFERENCIAR UNA VIVIENDA Y GEOCODIFICAR SU  
RESPECTIVA MANZANA**

El proceso de limpieza y ordenamiento de las 233 direcciones requirió un tiempo de 2 horas y 49 minutos. Por tanto, se estima un tiempo de 43.52 segundos por dirección en la revisión de la calidad de dato. Este primer proceso sirvió tanto para la geocodificación como para la georreferenciación, por lo que se incluye en los costos y tiempos para ambos procesos.

Se determinó que el tiempo promedio requerido para georreferenciar una vivienda es 25 minutos. Mientras que, para geocodificar una vivienda al nivel manzana, el tiempo varío por cada recurso cartográfico, 1'27" con Geodir, 1'56" con Google Maps, 1'22" empleando el plano distrital y 1'25" con el uso de 3 recursos (Tabla N°7).

**Tabla N°7. Tiempo promedio requerido para georreferenciar y geocodificar una vivienda y su respectiva manzana**

	Georreferenciación de viviendas	Geocodificación de manzanas			
		Geodir	Google Maps	Plano distrital	Los 3 recursos
<b>Tiempo total requerido para cada proceso (horas, minutos y segundos)</b>	86:16:00	04:08:51	05:52:58	04:06:36	04:34:22
<b>Número de viviendas realizadas</b>	207	172	183	181	193
<b>Tiempo promedio requerido para cada proceso (minutos y segundos) por dirección</b>	00:25:00	00:01:27	00:01:56	00:01:22	00:01:25

## **7.6. COSTO DE GEORREFERENCIACIÓN Y GEOCODIFICACIÓN**

Los costos fijos de servidor y celular fueron aportados por el investigador. Los costos variables de internet y telefonía fueron recursos propios del investigador. Sin embargo, deben ser estimados para determinar la implementación y viabilidad de estas metodologías en un estudio de mayor cobertura geográfica y/o inversión comercial externa al grupo de investigación. Los recursos cartográficos y el software empleado fueron de acceso libre y gratuito (Tabla N°8).

El costo/hora de trabajo por operario fue S/. 10.63 para la revisión de la calidad de datos de direcciones y la georreferenciación en campo. El costo/hora de trabajo por especialista fue S/. 21.88 para la geocodificación manual. La cantidad de tiempo se aproximó a la hora de trabajo cumplida. (Tabla N°9).

El proceso de georreferenciación implicó un costo de movilización desde el punto de partida hasta el distrito de Mi Perú, dentro del cual se requirió una movilidad interna para desplazarse desde cada AAHH. La geocodificación no requirió gastos de movilidad. Sin embargo, requirió de una cartografía de referencia para la establecer los puntos geocodificados y estimar la exactitud posicional.

**Tabla N°8. Costo total para la geocodificación y georreferenciación**

		<b>Cantidad</b>	<b>Georreferenciación</b>	<b>Geocodificación</b>
<b>Costos Fijos</b>	Servidor	01	S/. 00.00	S/. 00.00
	Software	01	-----	S/. 00.00
	Celular	01	S/. 00.00	-----
<b>Costos variables</b>	Costo total Operador	----	S/. 839.77	S/. 43.76
	Costo total revisión de calidad de datos	----	S/. 31.89	S/. 31.89
	Servicio Telefonía	01	S/. 00.00	-----
	Internet	-----	S/. 00.00	S/. 00.00
	Cartografía de referencia	01	S/. 00.00	S/. 80.00
	Movilidad	-----	S/. 144.00	S/. 00.00
Total			S/. 1015.66	S/. 155.65

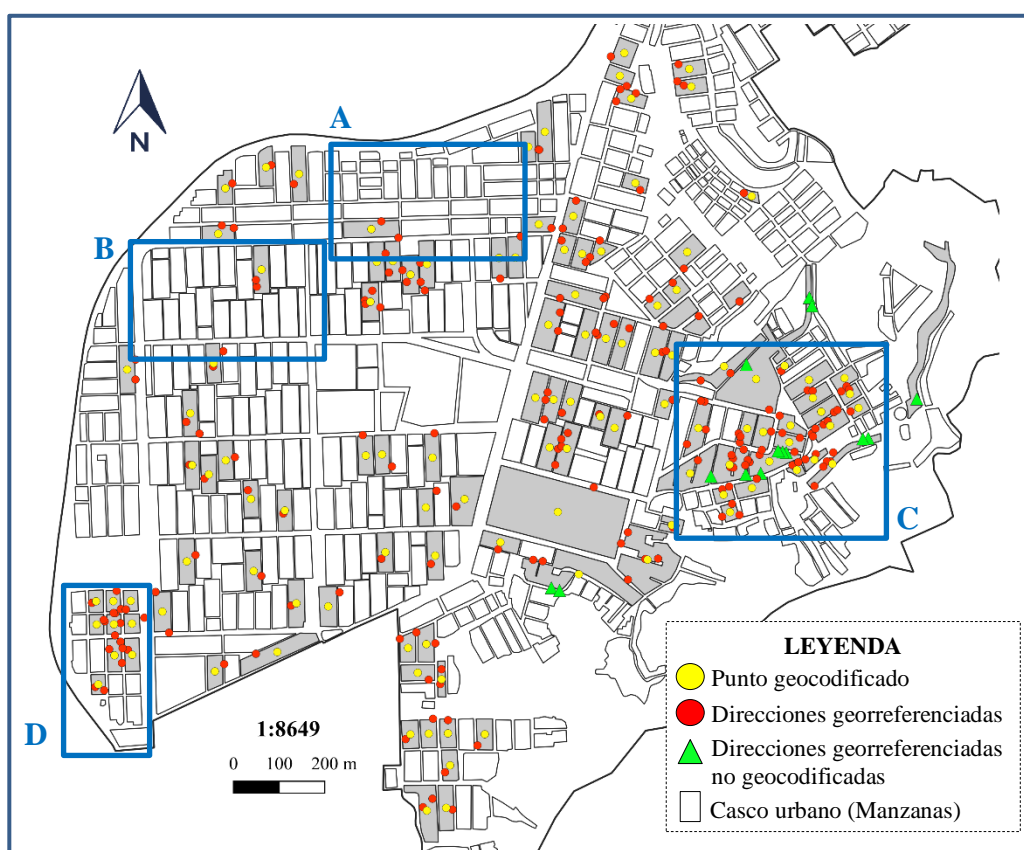
**Tabla N°9. Costo/hora de trabajo por operario para la geocodificación y georreferenciación**

	<b>Georreferenciación</b>	<b>Geocodificación</b>
Costo/ hora revisión de calidad de datos	S/. 10.63	S/. 10.63
Horas invertidas en la revisión de datos	3 horas	3 horas
Costo/hora operador	S/. 10.63	S/. 21.88
Cantidad de horas invertidas	79 horas	2 horas
Cantidad de direcciones realizadas por proceso	207	193

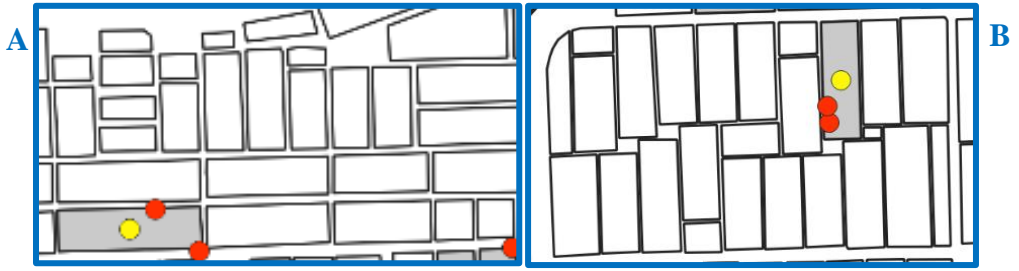
El costo promedio para georreferenciar una vivienda fue S/. 4.91 y para geocodificarla hasta el nivel de manzana fue de S/. 0.81.

## 7.7. ÁREAS DESATENDIDAS POR LAS AGENTES COMUNITARIAS DE HOPE

La distribución espacial de las manzanas geocodificadas identificó zonas densas de beneficiarias del proyecto HOPE y áreas geográficas con ausencia de ellas. Se determinaron áreas atendidas y desatendidas por las agentes comunitarias de HOPE (Figura N°5).



**Figura N°5. Conglomerados atendidos y desatendidos por las agentes comunitarias de HOPE**



A.B. Áreas de menor densidad de mujeres tamizadas



C.D. Áreas de mayor densidad de mujeres tamizadas

## 8. DISCUSIÓN

El presente estudio evidenció que la geocodificación manual ofrece una exactitud adecuada para la ubicación de la residencia de beneficiarias de un programa de salud implementado en zonas urbano-marginales del Perú. El proceso es más tiempo- y costo-efectivo que la georreferenciación. Nuestro estudio es el primero en reportar resultados sobre exactitud de la geocodificación de direcciones basados en asentamientos humanos y manzanas, ajustados a la realidad peruana.

La integridad de la geocodificación del presente estudio concuerda con la hallada en otros estudios sobre geocodificación automatizada en EE.UU. Comparaciones entre geocodificadores comerciales han reportado cifras que oscilan entre 44% y 84% según Krieger (34) y del 30% al 98% según Whitsel (54). Investigaciones posteriores encontraron cifras aún mayores (98% a 100%) usando un servidor comercial (55) o herramientas gratuitas (32). Resultados similares al encontrado se observaron en un estudio francés, el cual emplea un geocodificador en línea gratuito (81.4%) y (84.4%) ArcGIS (14).

El uso de múltiples cartografías de referencia ha optimizado la integridad de la geocodificación (82.83%). Este estudio logró un ligero aumento de la cantidad de direcciones geocodificadas al emplear tres recursos cartográficos en conjunto: Google Maps, Geodir maps y el plano distrital. Este resultado favorable en la integridad coincide con lo reportado por McElroy (31), quien alcanzó el 97% de direcciones geocodificadas usando 3 mapas de calles como referencia. Alcanzó el 76% de mediante geocodificación automatizada en ArcView 3.2 de ESRI. El 21% restante se geocodificó mediante mapeo web, ubicando la mayoría de direcciones

rurales dejadas por el geocodificador anterior. Una investigación sobre geocodificación para el registro de pacientes con cáncer de Nebraska reportó un incremento de la integridad hasta 95% empleando Google maps para ubicar a las direcciones no geocodificadas por ArcGIS. (56). Un estudio sobre la aplicación de un SIG a la gestión informativa de la tuberculosis pulmonar (33) muestra también resultados favorables en la integridad al emplear dos métodos de geocodificación, mediante un servidor comercial (92.26%) y la geocodificación manual para completar el 100% de direcciones geocodificadas de una base de datos.

Nuestro estudio obtuvo una geocodificación perfecta de las direcciones, lo cual supera a lo obtenido por otros estudios sobre exactitud posicional (12–15,31–34,54,55,57). Sin embargo, estas investigaciones geocodificaron a nivel de vivienda, a diferencia de este estudio donde la geocodificación es a nivel de manzana. Si bien, no ofrece el mismo nivel de exactitud, la geocodificación de direcciones por asentamientos humanos y manzanas es aceptable. Este modelo de direcciones por parcelas es empleado en otros estudios para geocodificar las direcciones a nivel de código postal, pueblo o ciudad (9,14,54). El resultado obvio es que se está haciendo una comparación injusta entre dos sistemas que intrínsecamente tienen exactitudes diferentes. Para un estudio donde se requiere ubicar a la vivienda como un seguimiento de casos, la georreferenciación brinda la exactitud requerida. No obstante, un estudio donde se analiza la distribución espacial no requiere el mismo nivel de exactitud. Otros ejemplos del uso de áreas mayores a la vivienda para determinar áreas desatendidas son los estudios en cáncer de mama (58) y VIH (59) que geocodifican a nivel de áreas censales.

Una cifra alta de integridad no siempre coincide con una mayor exactitud posicional. Este estudio encontró que el 2.91% de las direcciones geocodificadas con Geodir maps fueron inexactas debido a errores en el nombre de las manzanas o a la ubicación incorrecta de AAHH. Se geocodificaron en un rango de 23.07 a 147.23 m de la ubicación real de la manzana donde se ubica el domicilio. Estas distancias son coherentes con las reportadas en otros estudios sobre exactitud posicional a nivel de vivienda. El 72% a 86% de las direcciones geocodificadas se ubicaron con errores de posición  $\leq 100$  m. Estos estudios han estimado rangos de distancia superiores a los 800 m (12,15,60). Faure (14) y Singh (32) encontraron direcciones geocodificadas (0.8% y 4%) a distancias superiores a 3 km de su ubicación real.

La revisión de la calidad de los datos de direcciones permitió identificar que el 63.09% de las direcciones presentaban errores. La corrección de estos errores ha sido un paso previo a la geocodificación en otras investigaciones (13,31,54). Un estudio sobre la exactitud de la geocodificación para la evaluación de la exposición basada en SIG en estudios epidemiológicos (14) encontró una mayor cantidad de direcciones (28.8%) con errores ortográficos durante la limpieza de datos. Además, identificó que las direcciones incompletas y con grafía incorrecta del municipio resultaron inexactas, ubicándolas a más de 30 km. Wu (61) logró ubicar correctamente a las direcciones que oscilaban entre 70 a 110 m de su ubicación real, tras corregir la ortografía en el nombre de carreteras. Singh (32) encontró que las direcciones inexactas ( $< 3$  km) presentaban un símbolo “*et*” (&) no conocido por el geocodificador. Goldberg (18) optimizó en un 16% la integridad de la

geocodificación a nivel de centroides, tras la corrección de errores en la base de datos.

El tiempo promedio para geocodificar una vivienda fue entre 82 a 116 segundos dependiendo de la cartografía de referencia empleada (Geodir maps, Google maps y el plano distrital). Sin embargo, en todos los casos el tiempo requerido para la geocodificación fue inferior al requerido para georreferenciar una vivienda (1500 segundos). Estudios que incluyeron la corrección de errores en el registro de direcciones para el proceso de geocodificación automatizada reportaron tiempo de procesamiento similares al encontrado en nuestro estudio, 69 y 194 segundos por registro (7,18). No obstante, estos tiempos son superiores a los reportados en estudios sobre geocodificación automatizada por geocodificadores comerciales ( $\leq$  1 segundo por dirección) (13,31,34,54).

Los costos requeridos para la geocodificación en este estudio fueron menor al costo de la georreferenciación. Esto concuerda con lo reportado por estudios donde se ha realizado la geocodificación con recursos gratuitos (7,14). Aquellas investigaciones que recurrieron a la geocodificación por servidores comerciales incluyeron pagos por el uso del geocodificador (31). Sin embargo, este costo se vuelve menor cuando los volúmenes de datos a geocodificar son enormes. Los geocodificadores automatizados procesan lotes de 2500 direcciones a más, dependiendo del geocodificador. Por ende, en estos estudios, los costos (menores a un dólar por dirección) son menores al requerido para la georreferenciación *in situ* de cada una de las direcciones (31,34).

La distribución espacial de las manzanas geocodificadas permitió identificar áreas desatendidas por las agentes comunitarias. Estos hallazgos fueron entregados a la investigadora principal del proyecto HOPE. Las áreas menos densas de beneficiarias reflejan brechas en los esfuerzos de difusión de la prueba rápida molecular por autotoma para la detección del VPH. El análisis de la distribución espacial para determinar áreas de baja participación en estrategias de salud ya ha sido utilizado para enfermedades de transmisión sexual, tuberculosis (62,63), cáncer de mama (58), melanoma maligno cutáneo (64), VIH (59), VPH (65), enfermedades cardiacas y malaria (66),

El uso de la cartografía de referencia hasta el nivel de manzana es la principal limitante del estudio. Emplear una cartografía hasta el nivel de lotes permitirá establecer un mayor nivel de exactitud de la geocodificación. No obstante, esta cartografía está limitada a geocodificadores comerciales o entidades públicas como COFOPRI e INEI.

## 9. CONCLUSIONES

- La exactitud posicional de la geocodificación a nivel de manzanas es adecuada en zonas urbanas donde las direcciones son basadas en asentamientos humanos y manzanas.
- Este estudio logró una geocodificación perfecta a nivel de manzana en el 82.83% de las direcciones evaluadas. El uso de múltiples cartografías optimiza la integridad de la geocodificación; sin embargo, requiere mayor tiempo de procesamiento.
- Errores en la cartografía de referencia como geometría incompleta o nomenclatura incorrecta conllevan a la inexactitud de la geocodificación. Además, un registro incompleto de las direcciones no permite la geocodificación exacta de las mismas.
- La geocodificación reduce el tiempo requerido para ubicar geográficamente una vivienda a nivel de manzana con respecto a la georreferenciación.
- El costo total para realizar la geocodificación es menor al requerido para la georreferenciación, al emplear recursos de acceso libre y no requerir gastos de movilización.
- Se identificaron áreas que han sido desatendidas por las agentes comunitarias de HOPE en el distrito Mi Perú.

## **10.RECOMENDACIONES**

- Se debe fomentar la geocodificación como un proceso rápido y accesible para el análisis espacial de zonas geográficas donde las redes de calles no son efectivas.
- La inclusión de la geocodificación de direcciones con AAHH y manzana, en el sector salud, permitirá un adecuado seguimiento de casos y la implementación de estrategias en salud pública en dichas zonas urbanas.
- Se recomienda la creación de sistemas geocodificadores automatizados de libre acceso que permitan reducir el tiempo y costo de la geocodificación de direcciones basadas en asentamientos humanos y manzanas congruentes con la realidad peruana.
- Se debe mantener una cartografía actualizada, particularmente en áreas de reciente urbanización, de manera que se faciliten evaluaciones e intervenciones comunitarias.

## REFERENCIAS BIBLIOGRÁFICAS

1. Lewis-Michl EL, Melius JM, Kallenbach LR, Ju CL, Talbot TO, Orr MF. Breast Cancer Risk and Residence near Industry or Traffic in Nassau and Suffolk Counties, Long Island, New York. *Arch Environ Health Int J* [Internet]. 1 de agosto de 1996 [citado 2 de febrero de 2020];51(4):255-65. Disponible en: <https://doi.org/10.1080/00039896.1996.9936024>
2. English P, Neutra R, Scalf R, Sullivan M, Waller L, Zhu L. Examining associations between childhood asthma and traffic flow using a geographic information system. *Environ Health Perspect* [Internet]. 1 de septiembre de 1999 [citado 2 de febrero de 2020];107(9):761-7. Disponible en: <https://ehp.niehs.nih.gov/doi/10.1289/ehp.99107761>
3. Ganguly R, Batterman S, Isakov V, Snyder M, Breen M, Brakefield-Caldwell W. Effect of geocoding errors on traffic-related air pollutant exposure and concentration estimates. *J Expo Sci Environ Epidemiol* [Internet]. septiembre de 2015 [citado 2 de febrero de 2020];25(5):490-8. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4532655/>
4. Jacquemin B, Siroux V, Sanchez M, Carsin AE, Schikowski T, Adam M, et al. Ambient Air Pollution and Adult Asthma Incidence in Six European Cohorts (ESCAPE). *Environ Health Perspect* [Internet]. junio de 2015 [citado 2 de febrero de 2020];123(6):613-21. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4455584/>
5. Dearwent SM, Jacobs RR, Halbert JB. Locational uncertainty in georeferencing public health datasets. *J Expo Sci Environ Epidemiol* [Internet]. septiembre de 2001 [citado 2 de febrero de 2020];11(4):329-34. Disponible en: <https://www.nature.com/articles/7500173>
6. Rushton G, Armstrong MP, Gittler J, Greene BR, Pavlik CE, West MM, et al. Geocoding in Cancer Research: A Review. *Am J Prev Med* [Internet]. 1 de

- febrero de 2006 [citado 2 de febrero de 2020];30(2):S16-24. Disponible en: [https://www.ajpmonline.org/article/S0749-3797\(05\)00362-4/abstract](https://www.ajpmonline.org/article/S0749-3797(05)00362-4/abstract)
7. McDonald YJ, Schwind M, Goldberg DW, Lampley A, Wheeler CM. An analysis of the process and results of manual geocode correction. *Geospatial Health* [Internet]. 11 de mayo de 2017 [citado 2 de febrero de 2020];12(1):526. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5978681/>
  8. Zandbergen PA, Hart TC, Lenzer KE, Camponovo ME. Error propagation models to examine the effects of geocoding quality on spatial analysis of individual-level datasets. *Spat Spatio-Temporal Epidemiol* [Internet]. abril de 2012 [citado 2 de febrero de 2020];3(1):69-82. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3319659/>
  9. Zandbergen PA. Geocoding Quality and Implications for Spatial Analysis. *Geogr Compass* [Internet]. 2009 [citado 2 de febrero de 2020];3(2):647-80. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-8198.2008.00205.x>
  10. Chow TE, Dede-Bamfo N, Dahal KR. Geographic disparity of positional errors and matching rate of residential addresses among geocoding solutions. *Ann GIS* [Internet]. 2 de enero de 2016 [citado 26 de julio de 2021];22(1):29-42. Disponible en: <https://doi.org/10.1080/19475683.2015.1085437>
  11. Zandbergen PA. A comparison of address point, parcel and street geocoding techniques. *Comput Environ Urban Syst* [Internet]. 1 de mayo de 2008 [citado 2 de febrero de 2020];32(3):214-32. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0198971507000890>
  12. Ward MH, Nuckols JR, Giglierano J, Bonner MR, Wolter C, Airola M, et al. Positional accuracy of two methods of geocoding. *Epidemiol Camb Mass* [Internet]. julio de 2005;16(4):542-7. Disponible en: [https://journals.lww.com/epidem/Fulltext/2005/07000/Positional\\_Accuracy\\_of\\_Two\\_Methods\\_of\\_Geocoding.17.aspx](https://journals.lww.com/epidem/Fulltext/2005/07000/Positional_Accuracy_of_Two_Methods_of_Geocoding.17.aspx)

13. Duncan DT, Castro MC, Blossom JC, Bennett GG, Gortmaker SL. Evaluation of the positional difference between two common geocoding methods. *Geospatial Health* [Internet]. mayo de 2011;5(2):265-73. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/21590677/>
14. Faure E, Danjou AMN, Clavel-Chapelon F, Boutron-Ruault MC, Dossus L, Fervers B. Accuracy of two geocoding methods for geographic information system-based exposure assessment in epidemiological studies. *Environ Health Glob Access Sci Source* [Internet]. febrero de 2017;16(1). Disponible en: <https://ehjournal.biomedcentral.com/articles/10.1186/s12940-017-0217-5>
15. Bonner MR, Han D, Nie J, Rogerson P, Vena JE, Freudenheim JL. Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology* [Internet]. julio de 2003;14(4):408-12. Disponible en: [https://journals.lww.com/epidem/fulltext/2003/07000/positional\\_accuracy\\_of\\_geocoded\\_addresses\\_in.6.aspx](https://journals.lww.com/epidem/fulltext/2003/07000/positional_accuracy_of_geocoded_addresses_in.6.aspx)
16. Whitsel EA, Rose KM, Wood JL, Henley AC, Liao D, Heiss G. Accuracy and repeatability of commercial geocoding. *Am J Epidemiol* [Internet]. noviembre de 2004;160(10):1023-9. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/15522859/>
17. Jacquez GM, Rommel R. Local indicators of geocoding accuracy (LIGA): theory and application. *Int J Health Geogr* [Internet]. 28 de octubre de 2009 [citado 3 de febrero de 2020];8:60. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2774310/>
18. Goldberg DW, Wilson JP, Knoblock CA, Ritz B, Cockburn MG. An effective and efficient approach for manually improving geocoded data. *Int J Health Geogr* [Internet]. 26 de noviembre de 2008;7:60. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2612650/>
19. Ribeiro AI, Olhero A, Teixeira H, Magalhães A, Pina MF. Tools for Address Georeferencing – Limitations and Opportunities Every Public Health Professional Should Be Aware Of. *PLOS ONE* [Internet]. 3 de diciembre de

- 2014 [citado 3 de febrero de 2020];9(12):e114130. Disponible en: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0114130>
20. Cuéllar Luna L, Gutiérrez Soto T. Desarrollo de la geografía médica o de la salud en Cuba. *Rev Cuba Hig Epidemiol* [Internet]. diciembre de 2014 [citado 24 de septiembre de 2021];52(3):388-401. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_abstract&pid=S1561-30032014000300011&lng=es&nrm=iso&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1561-30032014000300011&lng=es&nrm=iso&tlng=es)
21. Dummer TJB. Health geography: supporting public health policy and planning. *CMAJ Can Med Assoc J*. 22 de abril de 2008;178(9):1177-80.
22. Barcellos C, Buzai G, Santana P. Geografía de la salud: bases y actualidad. *Salud Colect* [Internet]. 2018 [citado 10 de septiembre de 2021];14(1). Disponible en: [http://www.scielo.org.ar/scielo.php?script=sci\\_arttext&pid=S1851-82652018000100001](http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1851-82652018000100001)
23. Berke EM. Geographic Information Systems (GIS): Recognizing the Importance of Place in Primary Care Research and Practice. *J Am Board Fam Med* [Internet]. 1 de enero de 2010 [citado 24 de septiembre de 2021];23(1):9-12. Disponible en: <https://www.jabfm.org/content/23/1/9>
24. Huisman O, de By RA. Principles of geographic information systems: an introductory textbook. Enschede: The International Institute for Geo-Information Science and Earth Observation (ITC); 2009. 540 p.
25. Nuckols JR, Ward MH, Jarup L. Using Geographic Information Systems for Exposure Assessment in Environmental Epidemiology Studies. *Environ Health Perspect* [Internet]. junio de 2004 [citado 3 de febrero de 2020];112(9):1007-15. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1247194/>
26. Baldovin T, Zangrando D, Casale P, Ferrarese F, Bertoncetto C, Buja A, et al. Geocoding health data with Geographic Information Systems: a pilot study in northeast Italy for developing a standardized data-acquiring format. *J Prev Med*

- Hyg [Internet]. junio de 2015 [citado 3 de febrero de 2020];56(2):E88-94. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4718356/>
27. Reynolds P, Von Behren J, Gunier RB, Goldberg DE, Hertz A, Harnly ME. Childhood cancer and agricultural pesticide use: an ecologic study in California. *Environ Health Perspect* [Internet]. marzo de 2002 [citado 3 de febrero de 2020];110(3):319-24. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1240773/>
28. Ratcliffe JH. On the accuracy of TIGER-type geocoded address data in relation to cadastral and census areal units. *Int J Geogr Inf Sci* [Internet]. 1 de julio de 2001 [citado 3 de febrero de 2020];15(5):473-85. Disponible en: <https://doi.org/10.1080/13658810110047221>
29. Goldberg DW, Ballard M, Boyd JH, Mullan N, Garfield C, Rosman D, et al. An evaluation framework for comparing geocoding systems. *Int J Health Geogr* [Internet]. 8 de noviembre de 2013 [citado 3 de febrero de 2020];12:50. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3834528/>
30. Lee K, Claridades ARC, Lee J. Improving a Street-Based Geocoding Algorithm Using Machine Learning Techniques. *Appl Sci* [Internet]. enero de 2020 [citado 2 de mayo de 2022];10(16):5628. Disponible en: <https://www.mdpi.com/2076-3417/10/16/5628>
31. McElroy JA, Remington PL, Trentham-Dietz A, Robert SA, Newcomb PA. Geocoding addresses from a large population-based study: lessons learned. *Epidemiol Camb Mass* [Internet]. julio de 2003;14(4):399-407. Disponible en: [https://journals.lww.com/epidem/Fulltext/2003/07000/Geocoding\\_Addresses\\_from\\_a\\_Large\\_Population\\_based.5.aspx](https://journals.lww.com/epidem/Fulltext/2003/07000/Geocoding_Addresses_from_a_Large_Population_based.5.aspx)
32. Singh SK. Evaluating two freely available geocoding tools for geographical inconsistencies and geocoding errors. *Open Geospatial Data Softw Stand* [Internet]. 1 de mayo de 2017 [citado 23 de noviembre de 2021];2(1):11. Disponible en: <https://doi.org/10.1186/s40965-017-0026-3>

33. Pacheco González J, Fabregat Rodríguez M, Morffi García I, González Hermida A. Aplicación de un sistema de información geográfica a la gestión informativa de la tuberculosis pulmonar. *MediSur* [Internet]. diciembre de 2010 [citado 24 de noviembre de 2021];8(6):398-406. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1727-897X2010000600002](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1727-897X2010000600002)
  
34. Krieger N, Waterman P, Lemieux K, Zierler S, Hogan JW. On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. *Am J Public Health* [Internet]. julio de 2001 [citado 3 de febrero de 2020];91(7):1114-6. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1446703/>
  
35. Jensen JR, Jensen RR. *Introductory geographic information systems*. Boston: Pearson; 2013. 437 p.
  
36. Campbell J, Shin M. *Essentials of Geographic Information Systems* [Internet]. Lynchburg: Saylor Foundation; 2011. 260 p. Disponible en: <https://digitalcommons.liberty.edu/textbooks/2/>
  
37. Geodir. ¿Cuáles son las ventajas de Geodir maps vs Google maps? [Internet]. Blog GeoDir. 2018 [citado 10 de noviembre de 2021]. Disponible en: <https://geodir.wordpress.com/2018/01/06/buscador-comparativo-de-google-maps-y-geodir-maps/>
  
38. Geodir. API Geocodificación Inversa (búsqueda de dirección) [Internet]. Blog GeoDir. 2019 [citado 28 de junio de 2021]. Disponible en: <https://geodir.wordpress.com/2019/08/29/api-geocodificacion-inversa-busqueda-de-direccion/>
  
39. Hackeloeer A, Klasing K, Krisp JM, Meng L. Georeferencing: a review of methods and applications. *Ann GIS* [Internet]. 2 de enero de 2014 [citado 1 de octubre de 2021];20(1):61-9. Disponible en: <https://doi.org/10.1080/19475683.2013.868826>

40. Li X, Zhang X, Ren X, Fritsche M, Wickert J, Schuh H. Precise positioning with current multi-constellation Global Navigation Satellite Systems: GPS, GLONASS, Galileo and BeiDou. *Sci Rep* [Internet]. 9 de febrero de 2015 [citado 11 de noviembre de 2021];5:8328. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4321187/>
41. Yoon D, Kee C, Seo J, Park B. Position Accuracy Improvement by Implementing the DGNSS-CP Algorithm in Smartphones. *Sensors* [Internet]. 18 de junio de 2016 [citado 28 de octubre de 2021];16(6):910. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4934336/>
42. Lachapelle G, Gratton P, Horrejt J, Lemieux E, Broumandan A. Evaluation of a Low Cost Hand Held Unit with GNSS Raw Data Capability and Comparison with an Android Smartphone. *Sensors* [Internet]. 29 de noviembre de 2018 [citado 11 de noviembre de 2021];18(12):4185. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6308615/>
43. National Coordination Office for Space-Based Positioning, Navigation and Timing. Selective Availability [Internet]. *GPS.gov*. 2021 [citado 25 de noviembre de 2021]. Disponible en: <https://www.gps.gov/systems/gps/modernization/sa/>
44. National Coordination Office for Space-Based Positioning, Navigation and Timing. Data from the First Week without selective availability [Internet]. *GPS.gov*. 2018 [citado 24 de noviembre de 2021]. Disponible en: <https://www.gps.gov/systems/gps/modernization/sa/data/>
45. Seo KY, Park SH, Jang WS. Performance Analysis of Software-Based GPS/GLONASS Augmentation System for Maritime DGNSS Service. *En 2011* [citado 23 de noviembre de 2021]. p. 1845-50. Disponible en: <http://www.ion.org/publications/abstract.cfm?jp=p&articleID=9732>
46. Angrisano A, Petovello M, Pugliano G. Benefits of Combined GPS/GLONASS with Low-Cost MEMS IMUs for Vehicular Urban Navigation. *Sensors*

- [Internet]. 19 de abril de 2012 [citado 11 de noviembre de 2021];12(4):5134-58. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3355462/>
47. Zhu H, Xia L, Wu D, Xia J, Li Q. Study on Multi-GNSS Precise Point Positioning Performance with Adverse Effects of Satellite Signals on Android Smartphone. *Sensors* [Internet]. 11 de noviembre de 2020 [citado 10 de noviembre de 2021];20(22):6447. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7697570/>
48. National Coordination Office for Space-Based Positioning, Navigation and Timing. GPS Accuracy [Internet]. GPS.gov. 2021 [citado 28 de octubre de 2021]. Disponible en: <https://www.gps.gov/systems/gps/performance/accuracy/>
49. Shin MB, Garcia PJ, Saldarriaga EM, Fiestas JL, Ásbjörnsdóttir KH, Iribarren SJ, et al. Cost of community-based human papillomavirus self-sampling in Peru: A micro-costing study. *Lancet Reg Health - Am* [Internet]. 1 de abril de 2022 [citado 2 de mayo de 2022];8:100160. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2667193X21001563>
50. Morán F, Cárcamo C, Valderrama M, García PJ. Preferencias y satisfacción hacia un programa de tamizaje con pruebas autoadministradas de detección del virus de papiloma humano. *Rev Peru Med Exp Salud Pública* [Internet]. 30 de junio de 2017 [citado 3 de febrero de 2020];34(2):228-32. Disponible en: <https://rpmpesp.ins.gob.pe/index.php/rpmpesp/article/view/2453>
51. INEI [Instituto Nacional de Estadística e Informática). Sistema de consecución de recursos humanos [Internet]. Convocatoria bajo locación de servicios ENAHO 2022. [citado 22 de abril de 2022]. Disponible en: [https://consecucion.inei.gob.pe/convocatorias/2014/convoca/default\\_n.asp?id=22903](https://consecucion.inei.gob.pe/convocatorias/2014/convoca/default_n.asp?id=22903)
52. INEI [Instituto Nacional de Estadística e Informática). Sistema de Consecución de recursos humanos [Internet]. Convocatoria bajo locación de servicios CENEC 2022. [citado 2 de mayo de 2022]. Disponible en:

[https://consecucion.inei.gob.pe/convocatorias/2014/convoca/default\\_n.asp?id=23039](https://consecucion.inei.gob.pe/convocatorias/2014/convoca/default_n.asp?id=23039)

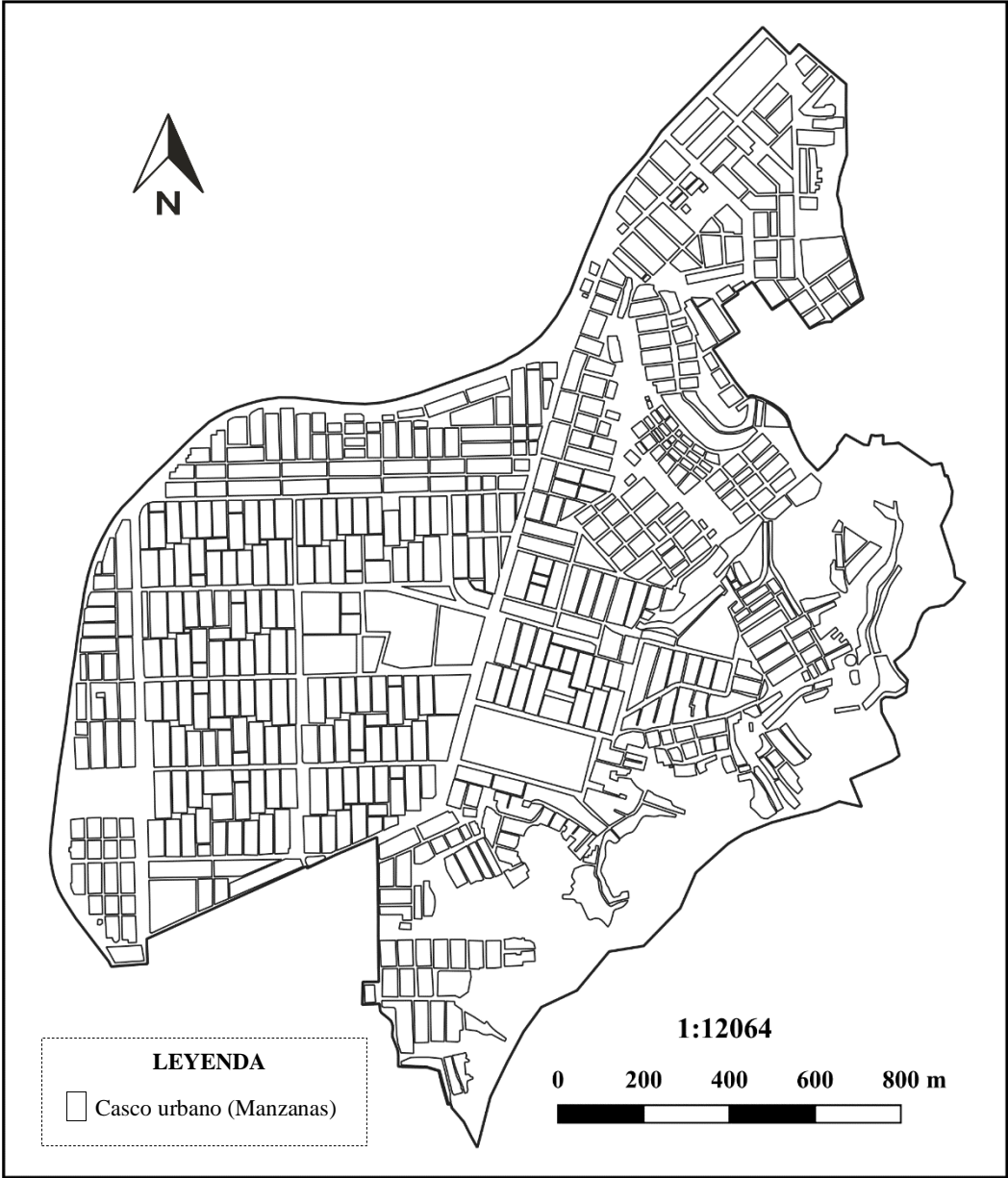
53. Véliz C J. Plan Estratégico de Desarrollo Concertado del distrito de Mi Perú 2014 - 2024 [Internet]. Municipalidad de Centro Poblado Nuestra Señora de las Mercedes de Mi Perú; 2014. Disponible en: <https://es.scribd.com/doc/254222901/PLAN-ESTRATEGICO-DE-DESARROLLO-CONCERTADO-DEL-DISTRITO-DE-MI-PERU-2014-2024>
54. Whitsel EA, Quibrera PM, Smith RL, Catellier DJ, Liao D, Henley AC, et al. Accuracy of commercial geocoding: assessment and implications. *Epidemiol Perspect Innov* [Internet]. 20 de julio de 2006 [citado 26 de noviembre de 2021];3:8. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1557664/>
55. Schootman M, Sterling DA, Struthers J, Yan Y, Laboube T, Emo B, et al. Positional accuracy and geographic bias of four methods of geocoding in epidemiologic research. *Ann Epidemiol* [Internet]. junio de 2007;17(6):464-70. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/17448683/>
56. Lin G, Gray J, Qu M. Improving geocoding outcomes for the Nebraska Cancer Registry: learning from proven practices. *J Regist Manag* [Internet]. 2010;37(2):49-56. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/21086822/>
57. Kravets N, Hadden WC. The accuracy of address coding and the effects of coding errors. *Health Place* [Internet]. marzo de 2007;13(1):293-8. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S1353829205000468?via%3Dihub>
58. Padilla CM, Painblanc F, Soler-Michel P, Vieira VM. Mapping Variation in Breast Cancer Screening: Where to Intervene? *Int J Environ Res Public Health* [Internet]. julio de 2019 [citado 5 de mayo de 2022];16(13):2274. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6651541/>

59. Edmonds A, Breskin A, Cole SR, Westreich D, Ramirez C, Cocohoba J, et al. Poverty, Deprivation, and Mortality Risk Among Women With HIV in the United States. *Epidemiol Camb Mass* [Internet]. 1 de noviembre de 2021;32(6):877-85. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/34347686/>
60. Zhan FB, Brender JD, De Lima I, Suarez L, Langlois PH. Match rate and positional accuracy of two geocoding methods for epidemiologic research. *Ann Epidemiol* [Internet]. noviembre de 2006;16(11):842-9. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/17027286/>
61. Wu J, Funk TH, Lurmann FW, Winer AM. Improving Spatial Accuracy of Roadway Networks and Geocoded Addresses. *Trans GIS* [Internet]. 2005 [citado 23 de noviembre de 2021];9(4):585-601. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9671.2005.00236.x>
62. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/Ethnicity, Gender, and Monitoring Socioeconomic Gradients in Health: A Comparison of Area-Based Socioeconomic Measures—The Public Health Disparities Geocoding Project. *Am J Public Health* [Internet]. octubre de 2003 [citado 16 de junio de 2022];93(10):1655-71. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1448030/>
63. Krieger N, Waterman PD, Chen JT, Soobader MJ, Subramanian SV. Monitoring socioeconomic inequalities in sexually transmitted infections, tuberculosis, and violence: geocoding and choice of area-based socioeconomic measures--the public health disparities geocoding project (US). *Public Health Rep* [Internet]. 2003 [citado 16 de junio de 2022];118(3):240-60. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1497538/>
64. Strömberg U, Parkes BL, Holmén A, Peterson S, Holmberg E, Baigi A, et al. Disease mapping of early- and late-stage cancer to monitor inequalities in early detection: a study of cutaneous malignant melanoma. *Eur J Epidemiol*

[Internet]. 2020 [citado 16 de junio de 2022];35(6):537-47. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7320924/>

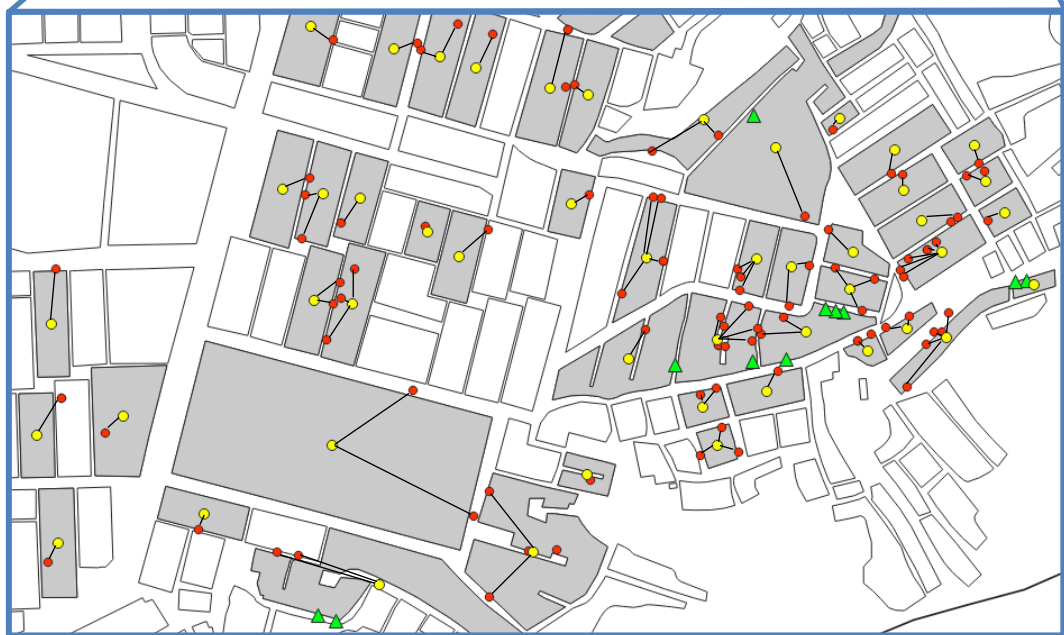
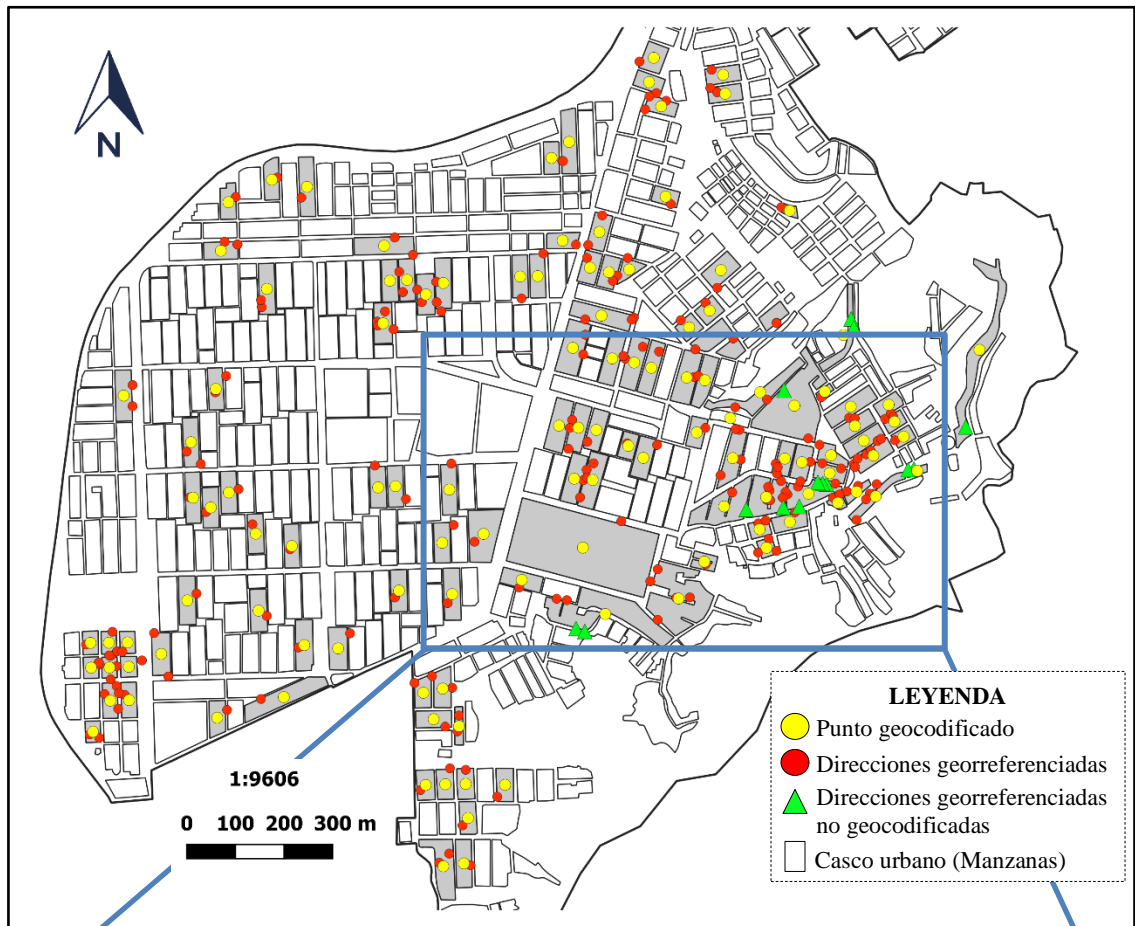
65. Terán-Figueroa Y, Muñoz-Carreón P, Moya MF, Galán-Cuevas S, Noyola-Range N, Gutiérrez-Enríquez SO, et al. The burden of cervical cancer in patients with limited access to health services. *Ginecol Obstet Mex.* 2015;83(3):162-72.
66. Cho J, You SC, Lee S, Park D, Park B, Hripcsak G, et al. Application of Epidemiological Geographic Information System: An Open-Source Spatial Analysis Tool Based on the OMOP Common Data Model. *Int J Environ Res Public Health* [Internet]. noviembre de 2020 [citado 16 de junio de 2022];17(21):7824. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7663469/>

**ANEXO N°1: Plano urbano del distrito de Mi Perú 2017, otorgado por INEI**



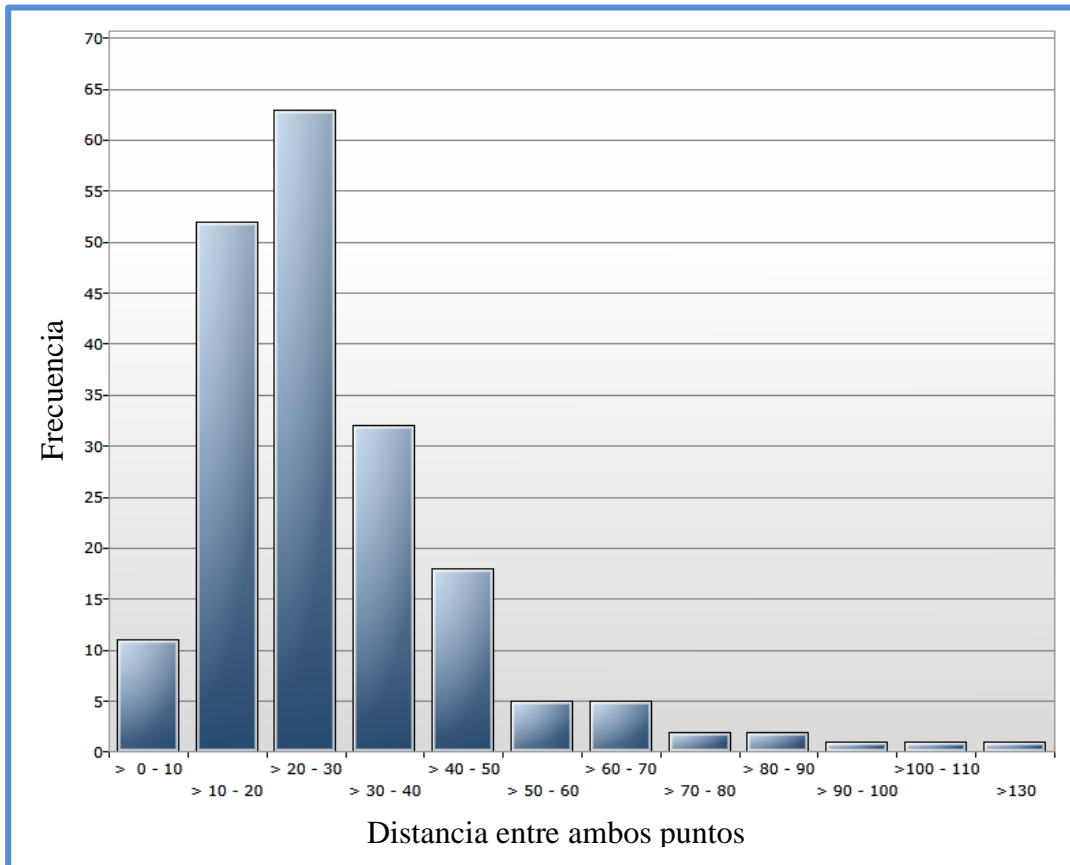


**ANEXO N°3: Distribución espacial de las direcciones georreferenciadas y puntos geocodificados en el distrito de Mi Perú**



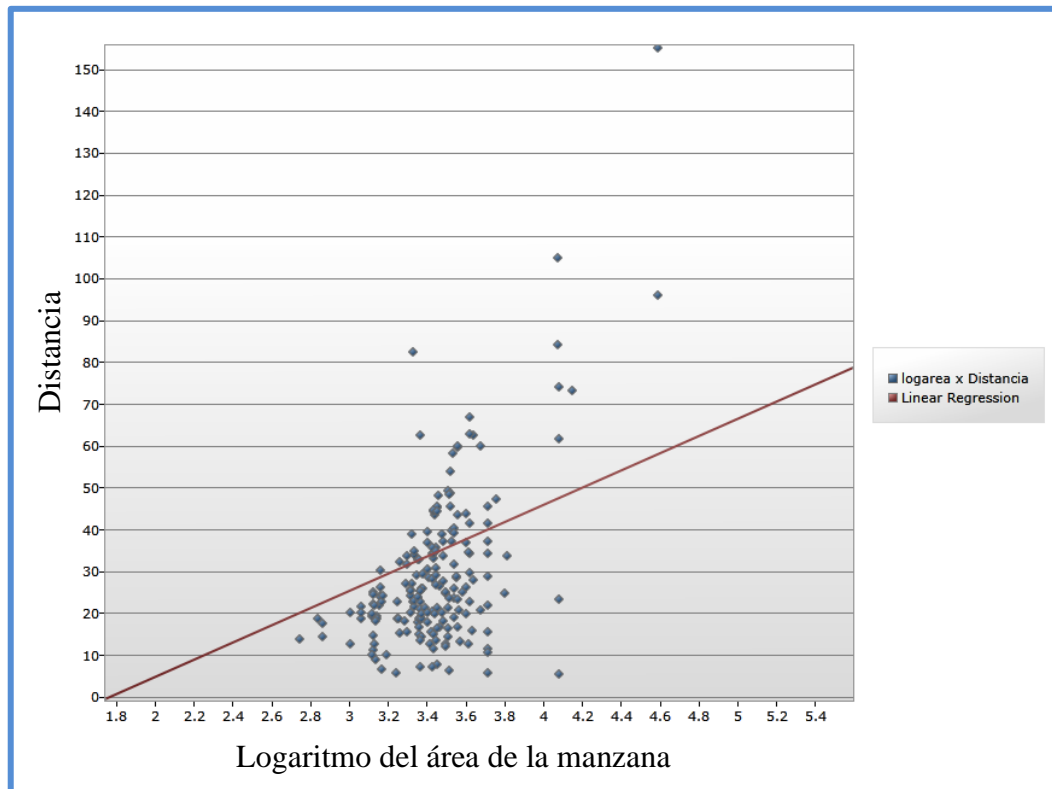
**ANEXO N°4: Distancia entre el punto georreferenciado y el centroide de la manzana (punto geocodificado)**

La mediana de la distancia entre estos dos puntos fue de 24.40 m, con un rango de 5.31 m a 154.78 m, y con una distribución asimétrica (Fig. A4-1)



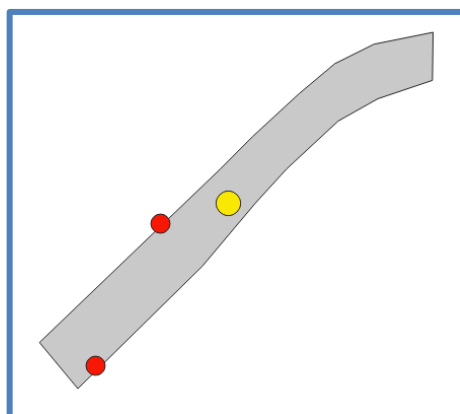
**Figura A4-1. Distribución de la distancia entre el punto georreferenciado y el centroide de la manzana.**

Existe cierta correlación entre el logaritmo del área de la manzana y la distancia entre los dos puntos ( $r=0.34$ ,  $p<0.0001$ ) (Figura A4-2).



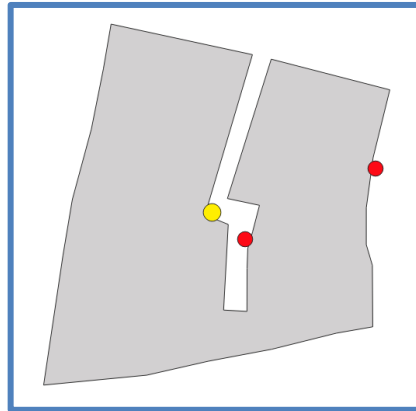
**Figura A4-2. Gráfico de dispersión con línea de regresión**

Adicionalmente esta distancia también se ve afectada por la forma de la manzana. En comparación con las cuadradas, manzanas alargadas pueden tener distancias más cortas o largas dependiendo de la ubicación de la vivienda (Figura A4-3).



**Figura A4-3. Manzana alargada con puntos rojos (georreferenciados) cercano y alejado del punto central (geocodificado)**

Finalmente, en el caso de algunas manzanas de forma no rectangular podremos tener una distancia muy pequeña para viviendas ubicadas cerca al centroide (Figura A4-4).



**Figura A4-4. Manzana de forma no rectangular con puntos rojos (georreferenciados) cercano y alejado del punto central (geocodificado).**

Por estos motivos se prefirió no usar la distancia entre el punto georreferenciado y el centroide de la manzana (punto geocodificado) para la estimación de la exactitud de la geocodificación.