

Comparación de las herramientas de geocodificación ArcGIS e Ideca: uso y recomendaciones en salud pública a partir de las experiencias en la pandemia

Juliana María Damelines¹;
Kimberly Piñeros Chía¹;
Sofía Ríos Oliveros¹

¹Subsecretaría de Salud Pública -
Secretaría Distrital de Salud.

Resumen

La geocodificación es un proceso que permite localizar una acción, un evento o situación; muestra gran utilidad para el análisis espacial y es ampliamente usada en diferentes contextos. En el campo de la salud pública se ha observado desde hace varios años la necesidad de geocodificar los eventos de interés en el espacio geográfico, con el fin de analizar características geoespaciales que sirvan para orientar la toma de decisiones. A partir de la emergencia sanitaria por COVID-19, esta necesidad tuvo grandes retos relacionados con la calidad del registro de la información y la capacidad de los geocodificadores para dar respuesta diaria de la situación. En este análisis se mostrarán las bondades de dos herramientas para geocodificar, a las cuales tiene acceso la Secretaría Distrital de Salud, se destaca en cada una sus virtudes y potencialidades para el análisis y el aporte en la toma de decisiones frente a los retos planteados. A manera de conclusión, se puede identificar que existe una coincidencia, entre las localidades, del 60,06%. El porcentaje coincidente de datos entre la localidad reportada en el seguimiento por COVID-19 y las localidades obtenidas entre una u otra herramienta de geocodificación

permiten complementar la localización de los registros de seguimiento. Cuando los datos no son coincidentes entre la localidad reportada y la geocodificada, se puede identificar si se debe ajustar la localidad reportada en el registro o si se debe buscar la dirección a través de diferentes mecanismos de vigilancia en salud pública.

Palabras clave: localización, cartografía, geocodificación, eventos en salud.

1. Introducción

La cartografía ha sido una herramienta útil para la presentación de datos. Desde la epidemiología ha sido utilizada hace varias décadas para identificar o tener un panorama de la distribución de las enfermedades, condiciones o eventos en salud, mortalidades, así como para hacer análisis comparativos o de asociación que se logran a partir de las facilidades de visualización que genera la representación gráfica (1). De manera progresiva se han implementado los sistemas de información geográfica (SIG) para mejorar los procesos cartográficos que permiten la visualización de los eventos (2). A raíz de la pandemia a nivel mundial estos sistemas se hicieron más visibles gracias a la necesidad de evidenciar cómo se desarrollaba el avance de la epidemia y cómo traspasaba fronteras en un corto periodo de tiempo (3).

Esta situación puso a las naciones a generar reportes de información diaria de casos detectados y mortalidad, contrastando la capacidad de otros sistemas de información, como los de vigilancia epidemiológica, al brindar información de la captación y procesamiento de muestras para detección de positivos o muertes a causa del virus. Esto conllevó a los países a migrar información muy rápidamente, mejorar sus procesos de producción de información en los laboratorios, así como los de vigilancia epidemiológica de campo —tratan de responder

a la velocidad exigida—, para mostrar, en gran medida, el desarrollo y avance de los sistemas de información de cada uno de los países y, específicamente, dentro del país, las brechas en gestión de la información con las que cuentan los municipios.

En Colombia, esta necesidad puso en la agenda pública la capacidad de cada municipio de obtener información detallada, por lo que generó diversos tableros de control que exigían cada día más y mejor información de la vigilancia en salud pública. La forma de representación fue una cartografía que refleja la situación en términos de casos y muertes, y permite visualizar el comportamiento diario de cada municipio, fue una fuente de información ágil para la comunidad en general. Adicionalmente, de manera rápida, el país logró comparar la situación de los municipios a partir de técnicas agregadas de datos o de generación de índices, los cuales se han ido consolidando y mejorando a medida que transcurre la epidemia (4).

En Bogotá la situación no fue diferente, dada la complejidad y volumen de población fue necesario tener tableros de información propios que permitieran visualizar con más detalle la situación. A través de algunos alistamientos previos con herramientas de geocodificación, la Secretaría Distrital de Salud de Bogotá (SDS) avanzó en el registro estandarizado de direcciones de residencia en los sistemas de información, para así localizar y entender la configuración de los eventos en el espacio geográfico de la capital.

La geocodificación de la residencia de los eventos en salud es un elemento fundamental en el desarrollo de estadísticas, no solo de acceso al sistema de salud, sino también del seguimiento y control de las estrategias que buscan las condiciones de salud de la población (5). Geocodificar es asignar coordenadas geográficas a los eventos y estrategias en el espacio, a partir de las diversas direcciones

postales (6). Se entiende por *procesos de geocodificación* el conjunto de herramientas y métodos para transformar un dato, como la dirección de residencia, en información geográfica detallada, que permite localizar los lugares donde suceden los eventos en salud, con el fin de avanzar a un análisis de las acumulaciones históricas-actuales, ambientales y políticas, que sustentan el contexto donde se producen los problemas de salud (7).

A partir de lo anterior, este documento se centrará en presentar los resultados obtenidos mediante la geocodificación, realizada a la base de datos de casos positivos COVID-19, usando el geocodificador de direcciones de la Infraestructura de Datos Espaciales para el Distrito Capital (Ideca) y el geocodificador de StreetMap Premium for ArcGIS for Desktop Map —en adelante ArcGis—, con el fin de identificar sus coincidencias y diferencias a la hora de localizar los eventos en salud, que para efectos de este escrito se suscribe a COVID-19.

2. Materiales y métodos

Proceso exploratorio para comparar dos herramientas de geocodificación de direcciones. La información analizada fue el registro preliminar de dirección de personas positivas por COVID-19 del 19 de septiembre del 2020. La comparación se realizó por medio de una matriz de datos, en la cual los casos por residencia se organizaron como filas y las variables de localización, como la localidad, en columnas. Se identificó que del total de 250490 personas registradas de la base por geocodificar, 1917 personas estaban registradas como fuera de Bogotá, 14584 no tenían dirección y 116 no poseían localidad de registro en la base de información, pero sí dirección.

Seguido, se organizó el registro de dirección conforme al estándar de nomenclatura

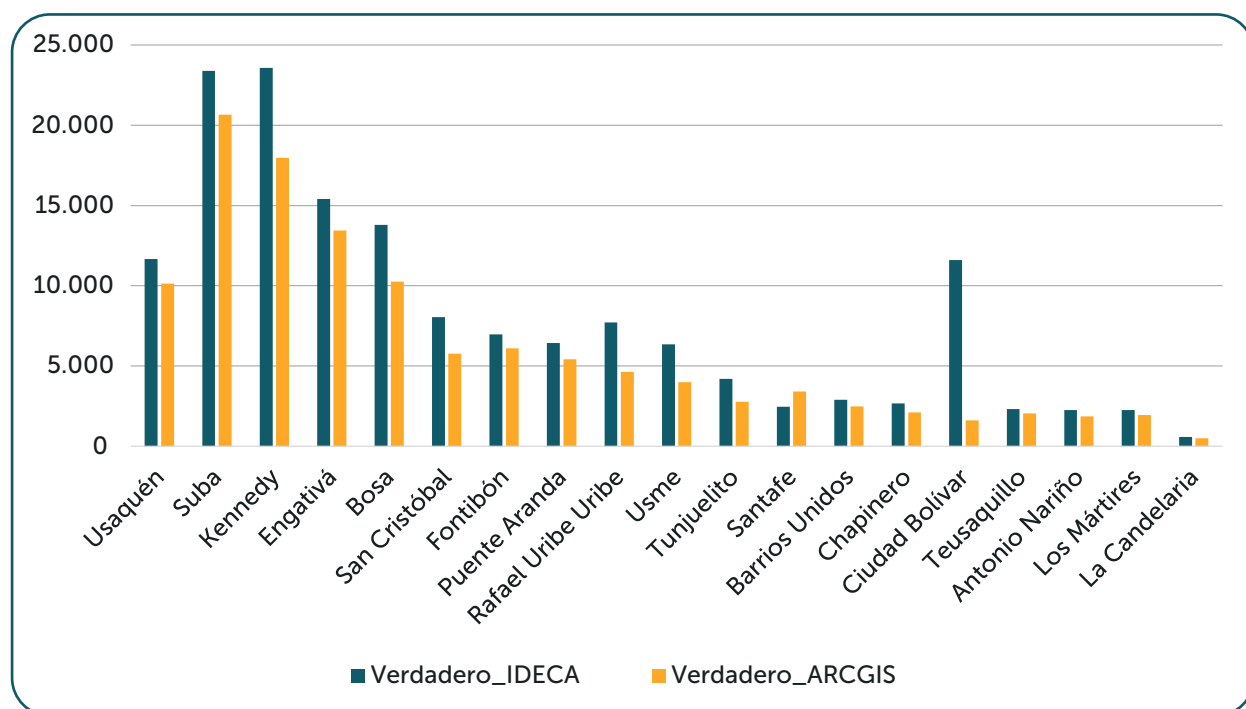
urbana de Bogotá adoptado por la SDS, en donde calles se registran como CL, carreras como KR; transversales, TV; diagonales, DG; las avenidas calle, AC, y las avenidas carrera, AK. Igualmente, se excluyeron del registro: (a) complementos de direcciones, (b) caracteres no alfanuméricos y (c) diversas escrituras de la palabra número. Posteriormente, se generaron tablas Excel acordes con los formatos de los geocodificadores por emplear. Para el geocodificador Ideca, los campos fueron: identificador único, dirección estandarizada y ciudad. También se extrajeron caracteres y tildes, y se cambiaron las palabras *sur* y *este* por las letras *S* y *E*. Para el geocodificador ArcGIS se generó una tabla Excel con los campos de localidad, dirección y ciudad, también se redactó en el campo de localidad específicamente Rafael Uribe y Santa Fe. Paralelamente, no fue necesario cambiar las expresiones *este* y *sur* por las letras *E* y *S*, lo que sí fue necesario para Ideca.

Realizada la geocodificación, se comparó, por un lado, la localidad identificada por cada geocodificador con la registrada en la base de positivos de COVID-19 y, por el otro, la localidad geocodificada por las dos herramientas. El geocodificador ArcGIS no arroja datos de UPZ ni localidad, por lo que se realizó una unión espacial de los registros para identificarlas.

3. Resultados

Se identificó que, del total de 250490 registros, 243746 (97,3 %) contaban con la variable de *dirección* y, a su vez, con la referencia de localidad acorde con el autorregistro de información. De estos, 154530 (63,39 %) fueron encontrados coincidentemente con el registro de localidad de residencia en la base a partir del geocodificador Ideca, y 117101 (48,04 %) con ArcGIS.

Figura 1. Coincidencia entre localidad registrada y geocodificada.

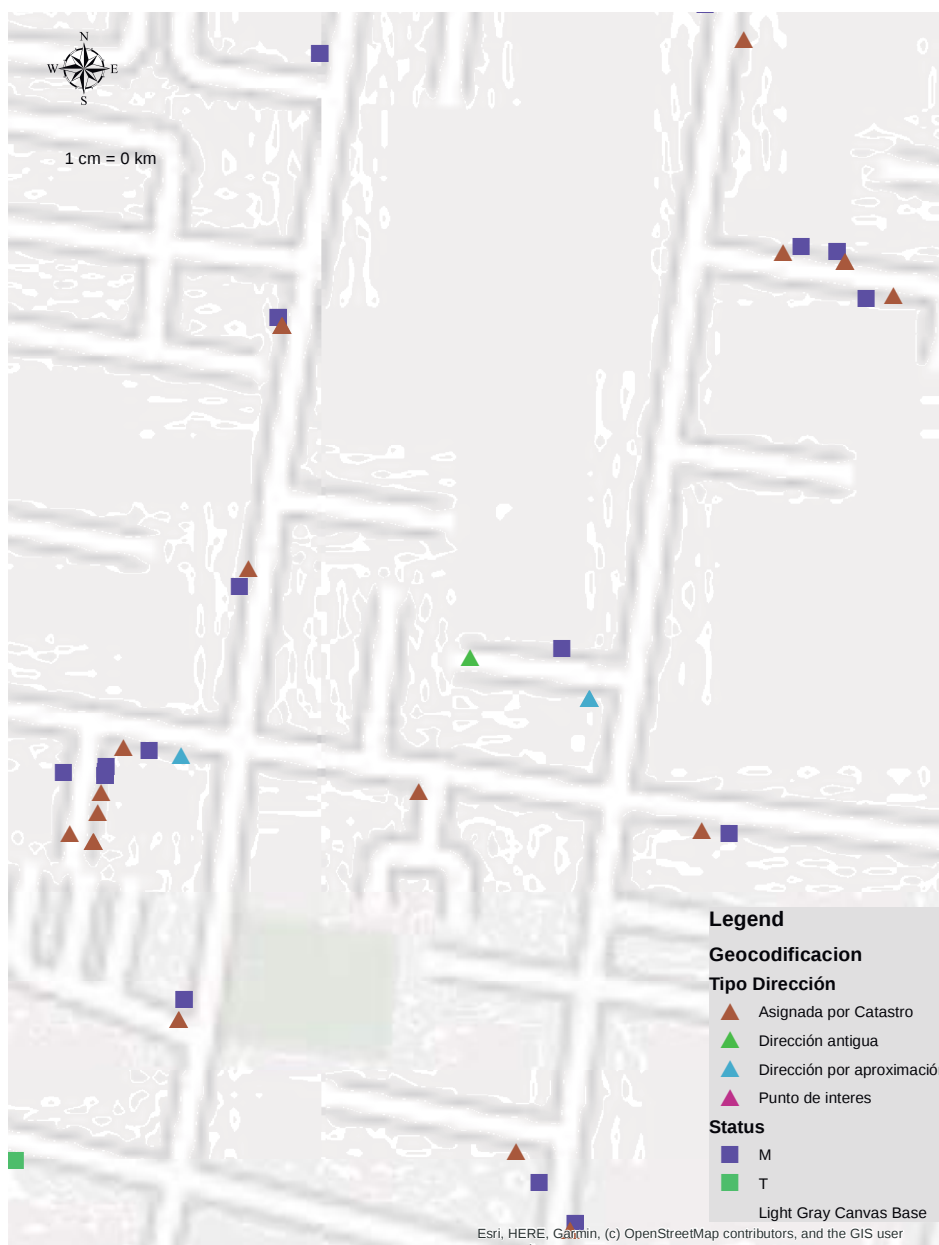


Fuente: base casos positivos por COVID-19 preliminar.

Al comparar cada localidad, se identificó que con las dos herramientas de geocodificación 146 414 (60,06 %) registros fueron coincidentes. Las localidades que obtuvieron mejor coincidencia para los dos geocodificadores, con respecto a la localidad registrada, fueron Barrios Unidos, La Candelaria, Los Mártires y Teusaquillo.

El mayor número de registros no coincidentes lo aporta el geocodificador ArcGIS, con 128 205 (51,18 %) registros, frente a 89 216 (35,61 %) registros de IDECA. 3090 casos no localizados por geocodificador de IDECA fueron localizados por el de ARCGIS, de los cuales 1783 fueron coincidentes entre la localidad registrada y la geocodificada.

Figura 2. Localización en paralelo de los dos geocodificadores.

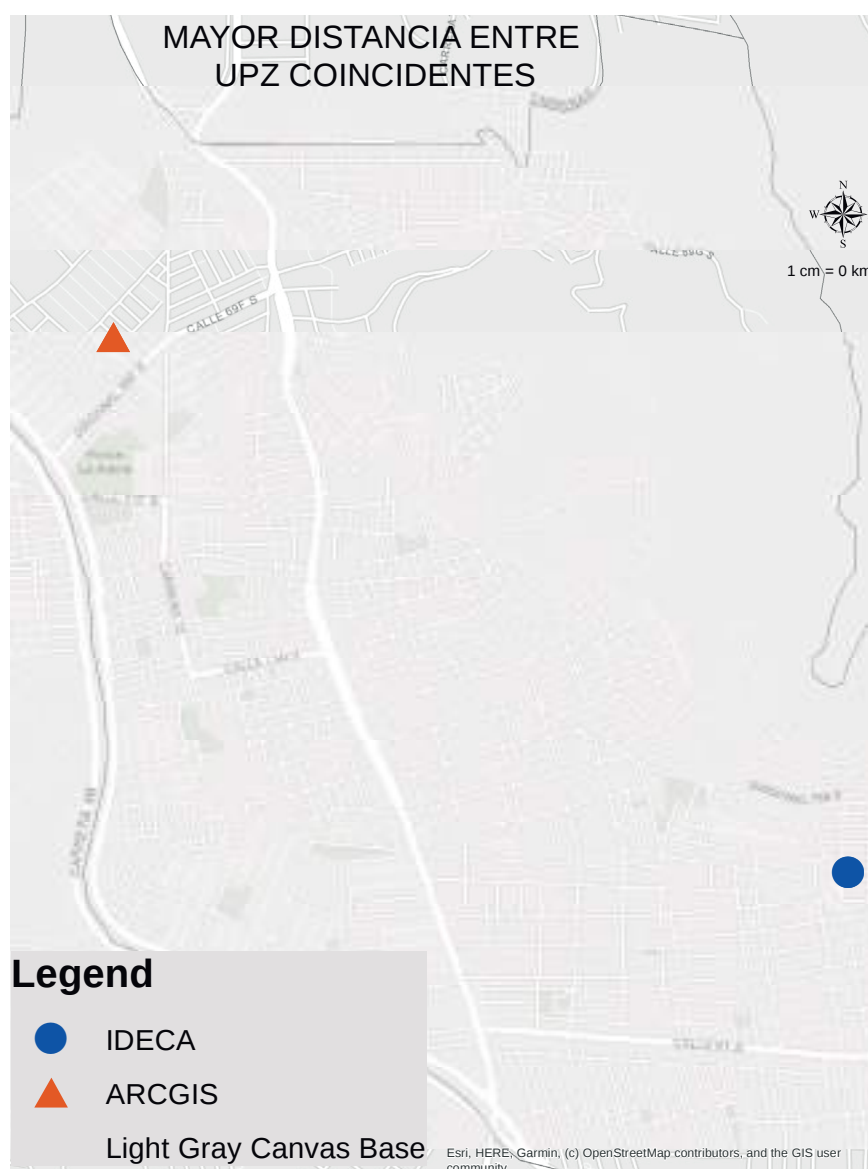


Fuente: base de casos positivos por COVID-19 preliminar.

La menor coincidencia entre la localidad reportada y la geocodificada del geocodificador Ideca fue en la localidad de Santa Fe y donde paralelamente ArcGIS tuvo una mayor coincidencia entre la geocodificación y el reporte. La diferencia más amplia se observó entre la localidad registrada y la geocodificada, las cuales fueron: Ciudad Bolívar, Bosa, Kennedy, Rafael Uribe y San Cristóbal.

El geocodificador ArcGIS asignó a 87972 datos no localizados la coordenada X (−74,069411), Y (4,61496) u 0. El de Ideca, por su parte, catalogó a 36217 datos como coordenada 0. Usaquén concentró ampliamente los datos coincidentes entre las dos geocodificaciones, pero no coincidentes con la localidad registrada, con valores totales de 44652 para Ideca y 46183 para ArcGIS.

Figura 3. No coincidencia entre localidad registrada y geocodificada.

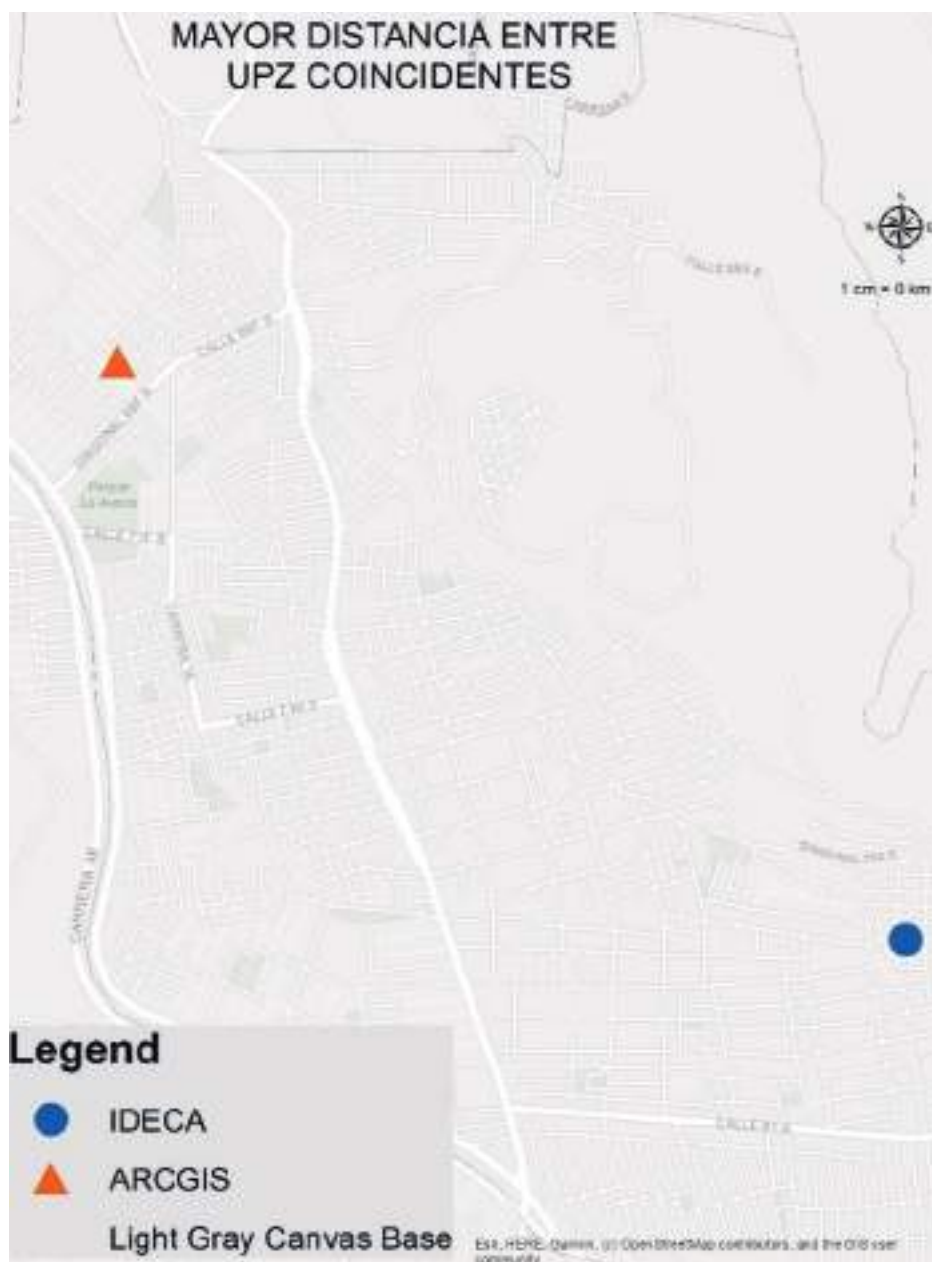


Fuente: base casos positivos por COVID-19 preliminar.

Ninguna coordenada es igual entre los geocodificadores, sin embargo, pueden encontrarse a menos de un metro de distancia. Si por ejemplo tomamos aquellos casos donde la UPZ es coincidente en las dos geocodificaciones, los cuales suman 152179 registros del total (60,75 %), vemos que dentro de estas UPZ las distancias mayores entre una coordenada y

otra del mismo registro puede estar a más de 800 metros y menos de dos kilómetros; en este rango encontramos 1103 registros. A modo de ejemplo, la figura 4 representa las 2 coordenadas obtenidas de un registro con la mayor distancia entre las coordenadas de UPZ coincidentes, localizadas en la localidad de Usme, UPZ Gran Yomasa con 1957,16 metros.

Figura 4. Mayor distancia entre UPZ coincidentes.



Fuente: base casos positivos por COVID-19 preliminar.

A partir de la anterior caracterización de los datos de geocodificación, se identifica que, para realizar cartografías o análisis epidémicos por localidad, UPZ u otras desagregaciones espaciales, es necesario comparar la localidad identificada en la geocodificación con la localidad reportada en el registro, y así validar los datos que podrán ser usados en diferentes análisis definidos para el seguimiento de eventos en salud y que para este escrito se orienta al comportamiento de COVID-19.

Para hacer una cartografía de seguimiento es importante partir de los datos de localización validados y obtenidos con la herramienta que arrojó un mayor resultado coincidente entre la localidad geocodificada con la localidad de registro de la base de positivos COVID-19. Posteriormente, es necesario complementar la información de localización faltante con los datos obtenidos coincidentemente a partir de la segunda herramienta de geocodificación, con el objetivo de aumentar el número de registros localizados y de no sobrestimar los datos preliminares que se van construyendo y siguiendo diariamente desde vigilancia en salud pública. Los mapas, a continuación, son ejemplos del resultado del ejercicio de organizar, geocodificar y analizar los datos para producir cartografía sobre posibles configuraciones del evento en el espacio geográfico de Bogotá. (ver mapas páginas 11 y 12).

4. Discusión

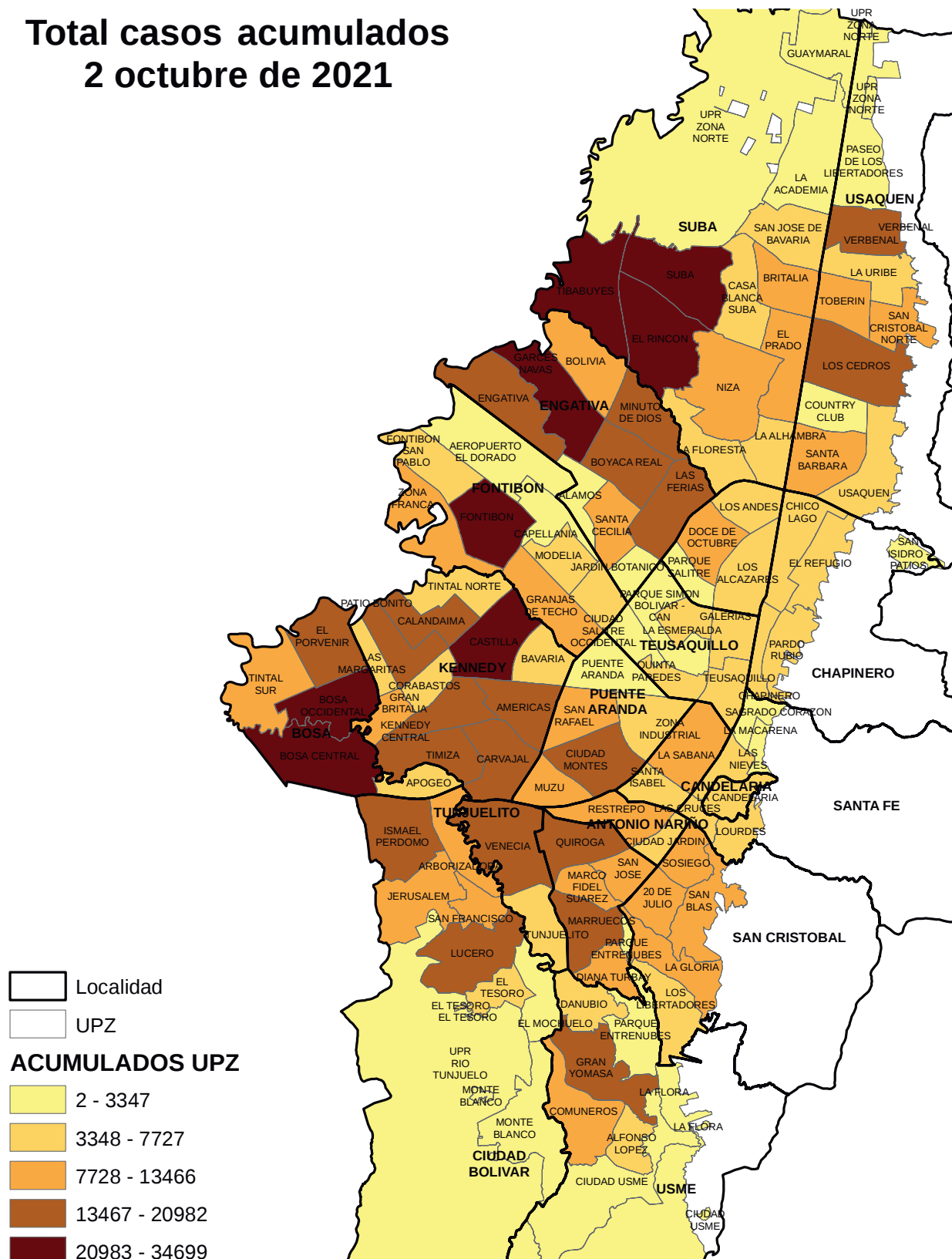
Los análisis geoespaciales han mostrado ser muy útiles para la toma de decisiones en salud pública (8,9). La pandemia por COVID-19 ha mostrado las bondades de tener salidas cartográficas con datos muy actualizados y con la posibilidad de acceso a la comunidad en general. Sin embargo, es necesario ampliar los estudios y la forma en la que se están obteniendo los datos, ya que esto permitirá tener información más

precisa para la toma de decisiones. Un estudio de comparación empírico realizado en Canadá indica una tasa de coincidencia por encima del 90% para dos geocodificadores analizados, uno de ellos ArcGIS online y el otro es un servicio de geocodificación construido manualmente que utiliza conjuntos de datos de referencia DMTI Spatial (DMTI). Las diferencias en porcentajes con nuestro análisis radican, por un lado, en la diferencia en la calidad de los datos registrados, que son la fuente fundamental para geocodificar y, por otro, en las herramientas empleadas pues, como se mencionó, usamos el geocodificador de Ideca y StreetMap Premium for ArcGIS for Desktop Map. Sin embargo, en ambos análisis se coincide en la necesidad de integrar procedimientos de codificación geográfica para aumentar las coincidencias y disminuir la incertidumbre posicional (10).

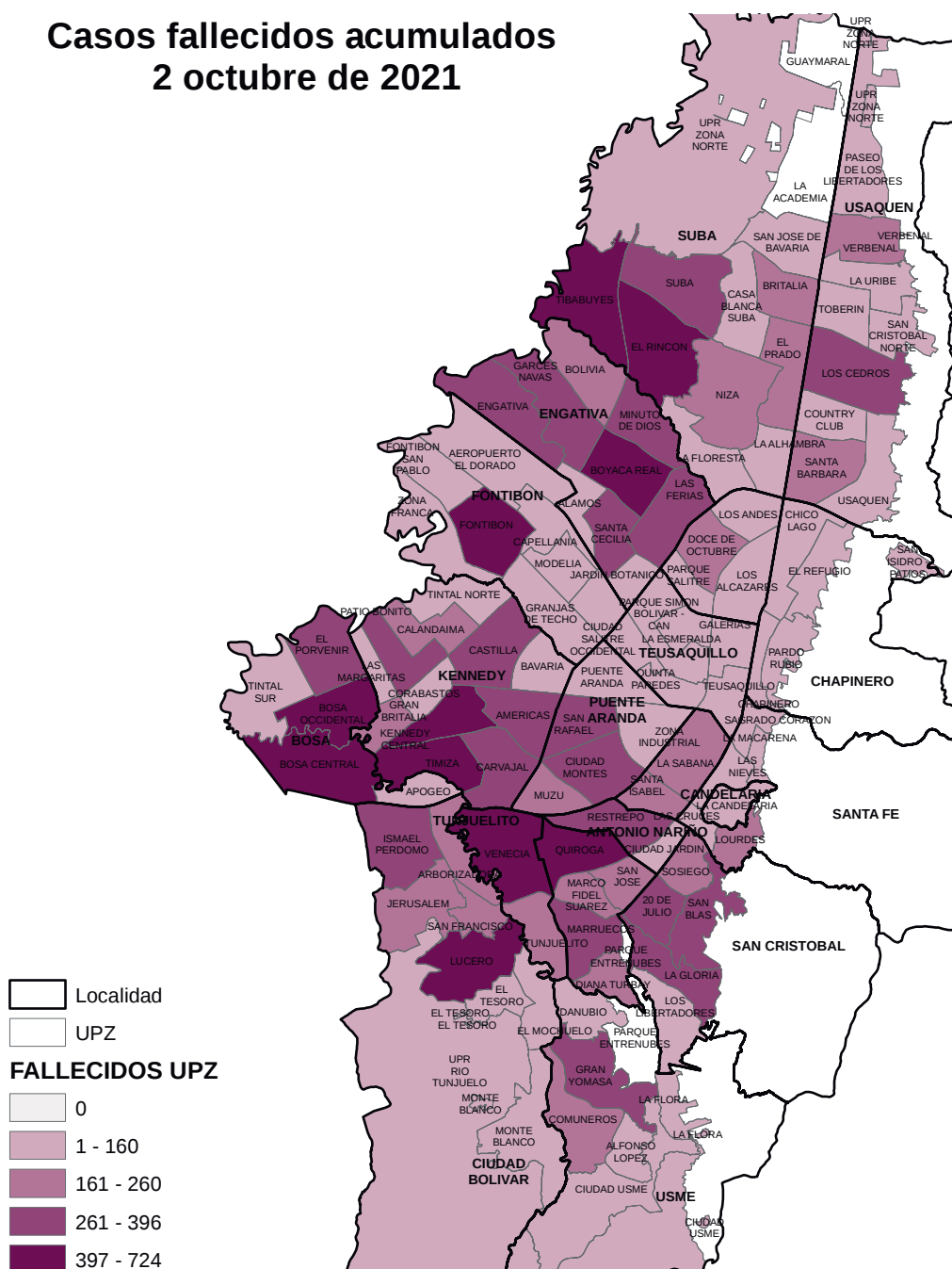
Otros estudios revelan la importancia de tener coordenadas conocidas para probar los datos resultado de la codificación, que ofrecen una mejora en la coincidencia de las direcciones, haciendo énfasis en que la clasificación errónea de la exposición resultante de un error posicional depende de la base de datos geográfica, el método de codificación geográfica y la prevalencia de la exposición (11). De igual forma, se exponen las posibles subestimaciones o sobrestimaciones que se pueden generar en un área geográfica por las dificultades e imprecisiones de los geocodificadores, que en su mayoría están muy relacionados con la importancia de la calidad del dato que se ingresan (12).

Figura 5. Cartografías de seguimiento semanal COVID-19 por UPZ.

Total casos acumulados
2 octubre de 2021



Casos fallecidos acumulados 2 octubre de 2021



Fuente: base casos positivos por COVID-19 preliminar.

Conclusiones

Comparar la geocodificación realizada por las herramientas de Ideca y ArcGIS permitió identificar que existe una coincidencia entre las herramientas del 60,06 %. Para responder a la pregunta ¿cuál geocodificador nos permite

una mejor aproximación espacial?, existe una mayor coincidencia con la herramienta de Ideca (63,39 %). Esto no quiere decir que una herramienta sea más perfecta que la otra, pues en algunos casos la geocodificación de ArcGIS puede encontrar información que no fue obtenida por la primera herramienta.

Se recomienda usar ambas herramientas con el fin de mejorar la calidad de los datos en el sistema de información, en los indicadores y en el diseño de estrategias institucionales acordes con las necesidades y demandas en salud del territorio.

Específicamente y dada la emergencia sanitaria, se logró evidenciar la importancia de contar con el registro de la dirección de residencia para la mayoría de las personas identificadas como caso positivo para COVID-19 y que este registro cumpliera con unos criterios mínimos de calidad del dato, lo cual se debe potencializar en otros eventos en salud pública. Se requiere adoptar y establecer procedimientos técnicos para localizar eventos de interés en salud, de manera que referentes distritales y locales involucren la geocodificación a la hora de analizar y diseñar estrategias en salud pública.

Es necesario mejorar continuamente la calidad del dato de dirección de las personas implicadas por los diversos eventos y estrategias en salud para así obtener con mayor precisión la configuración de las problemáticas de la población y poder mejorar la toma de decisiones en salud. En este punto, es muy importante que instituciones territoriales y prestadoras de servicios, encargadas de registrar el estado de salud de las personas, implementen estándares de captura y registro del dato de dirección de residencia.

La geocodificación de las direcciones de residencia permitirá avanzar en el análisis histórico del contexto donde se producen las problemáticas en salud, si es orientada a localizar los elementos y relaciones producidos en el espacio geográfico que han constituido afectaciones o mejoras en el estado de salud de las personas. Paralelamente, es necesario fortalecer procesos, herramientas, aptitudes e incluso actitudes con respecto a localizar los eventos por parte de profesionales e instituciones relacionadas con los datos en salud.

Es importante recordar el llamado que diversas investigaciones en salud hacen para usar y fortalecer el manejo de herramientas alusivas a los SIG a la hora de analizar el estado de salud en el territorio, puesto que estas permiten un mejor conocimiento de la relación existente entre la distribución de los eventos de interés en salud y de las diversas características del contexto geográfico próximo a las personas, lo cual es un aporte a la planificación de servicios y medidas en salud.

Referencias

1. Narro Robles JR, Ponce de León R. La representación cartográfica: una herramienta de la epidemiología. *Salud Publica Mex.* 1986; 28(3):283-91.
2. Iñiguez Rojas LB, Barcellos C. La cartografía en salud pública: viejos problemas y nuevas oportunidades. *Rev. Dep. Geogr.* 2014;(spe cartogeo):390-412.
3. John Hopkins Coronavirus Resource Center [Internet]. Baltimore: CSSE and JHU; 2020 [actualizado 2022 ene 31]. p. 1-6. Disponible en: <https://coronavirus.jhu.edu/map.html>
4. Ministerio de Salud y Protección Social [Internet]. Bogotá: Sistema Integrado de Información de la Protección Social. s. f. Tableros de control - Municipios de Colombia según su afectación por COVID-19; Minsalud. [consultado 202 agosto 14]. Disponible en: <https://bit.ly/3KZOiRN>
5. Puerto Segura E, Rodríguez Díaz V, Rodríguez Romero E. Procesos de geocodificación en el ámbito sanitario. Una comparativa de resultados. En: Ojeda J, Pita MF y Vallejo I, editores. *Tecnol la Inf Geográfica la Inf geográfica al Serv los Ciudad Secr publicaciones la Univ Sevilla.* 2010;337-47.
6. Quesada J, Nolasco A, Moncho J. Comparación de las aplicaciones de Google y Yahoo para la geocodificación de direcciones postales con fines epidemiológicos. *Rev Esp Salud Publica.* 2012;87(2):201-6. Disponible en: <https://bit.ly/3uj9Dj3>

7. Íñiguez Rojas L, Barcellos C. Geografía y salud en América Latina: evolución y tendencias. *Rev Cuba Salud Pública*. 2003;29(4):330-43.
8. Fisher JA, Spaur M, Buller ID, Flory AR, Freeman LEB, Hofmann JN, et al. Spatial heterogeneity in positional errors: A comparison of two residential geocoding efforts in the agricultural health study. *Int J Environ Res Public Health*. 2021;18(4):1-13.
9. De Cos O, Castillo V, Cantarero D. Facing a second wave from a regional view: Spatial patterns of covid-19 as a key determinant for public health and geoprevention plans. *Int J Environ Res Public Health*. 2020;17(22):1-18.
10. Shah TI, Bell S, Wilson K. Geocoding for public health research: Empirical comparison of two geocoding services applied to Canadian cities. *Can Geogr*. 2014;58(4):400-17.
11. Jones RR, DellaValle CT, Flory AR, Nordan A, Hoppin JA, Hofmann JN, et al. Accuracy of residential geocoding in the Agricultural Health Study. *Int J Health Geogr*. 2014;13(1):1-10.
12. Kinnee EJ, Tripathy S, Schinasi L, Shmool JLC, Sheffield PE, Holguin F, et al. Geocoding error, spatial uncertainty, and implications for exposure assessment and environmental epidemiology. *Int J Environ Res Public Health*. 2020;17(16):1-23.

