

Detección de Zonas de Alta Prevalencia Potencial de Chagas en Argentina

Una aplicación de big data para la identificación de potenciales focos
de una enfermedad infecciosa endémica.

Documento
de Trabajo
2018-02

Detección de Zonas de Alta Prevalencia Potencial de Chagas en Argentina

Documento de Trabajo

2018-02

**Detección de Zonas de Alta Prevalencia
Potencial de Chagas en Argentina**

Coordinación

Brenda Walter

Coordinadora Senior de la Fundación Bunge y Born

Colaboradores

Guillermo Bozzoli

Tomás Olego

Germán Rosati

Antonio Vazquez Brust

Fellows de la Fundación Bunge y Born

Participaron en el *working paper* original:

Roberto Chuit

Fellow de la Fundación Mundo Sano

Martín Andrés Minnoni

Founder CTO, Grandata

Carlos Sarraute

Director de Research, Grandata

Diego Weinberg

Subgerente de Implementación de Proyectos, Fundación Mundo Sano

Carolina Lang

Fellow de la Fundación Bunge y Born

Agradecemos a todos los colegas que nos han acompañado a lo largo de este trabajo. A todo el equipo de la Fundación Bunge y Born, especialmente a Alejandra Candia y Gerardo della Paolera por sus valiosos aportes. También recibimos observaciones y comentarios muy útiles por parte de Marcelo Abril de la Fundación Mundo Sano. El equipo de Grandata Labs aportó los Call Detail Records (CDRs), así como su conocimiento y experiencia en el procesamiento de datos. Todos los hallazgos, interpretaciones y opiniones expresadas en esta publicación son de los autores y no representan necesariamente el punto de vista de sus respectivas instituciones.

Índice

Introducción.....	4
Antecedentes / Trabajo relacionado.....	5
Generación del Mapa de Alta Precisión de zonas de riesgo de Chagas en Argentina	8
Índice de Receptividad al Vector de Transmisión.....	9
Condiciones de la vivienda.....	11
Índice de Afinidad con la Zona Endémica de Chagas.....	13
Call Detail Records (CDRs).....	14
Metodología.....	14
Índice de Vulnerabilidad Sanitaria.....	18
Índice de Prevalencia Potencial de Chagas (IPPECh).....	20
Resultados.....	21
Conclusiones.....	26
Referencias.....	28
Anexo A: Acceso a los datos	29
Anexo B: Tabla extendida. Tres localidades fuera del área endémica de Chagas con mayor Prevalencia Potencial, por provincia.....	28

Introducción

En 2018 la Fundación Bunge y Born, con la colaboración de la Fundación Mundo Sano y Grandata, se propuso ahondar en la investigación sobre la difusión territorial de la enfermedad de Chagas en la Argentina, donde es considerada una enfermedad desatendida. El impulso del proyecto fue la posibilidad de acceder a una fuente de datos de inmenso potencial, cuyo análisis resulta innovador en el contexto de la salud pública: los registros de llamados de telefonía celular de millones de abonados en todo el país, registrados durante un lapso de cinco meses.

Como hipótesis inicial se asumió que los grupos poblacionales que tienen un alto nivel de comunicación con zonas endémicas como el Gran Chaco, sea por haber vivido allí o por otros vínculos, tienen mayor probabilidad de estar infectadas. La enfermedad puede portarse, sin síntomas, durante décadas¹. Esto posibilita que población migrante, con el correr de los años, se nuclea en regiones alejadas del área endémica, donde el sistema de salud pública local no está preparado para diagnosticar la enfermedad.

A partir del análisis de miles de millones de registros anónimos de llamadas fue posible cuantificar las comunicaciones que se realizan conectando usuarios de telefonía intra y extra-área endémica. Con ello se identificaron localidades fuera de la zona endémica de la enfermedad que muestran una alta afinidad con la región del Gran Chaco, permitiendo inferir la existencia de patrones migratorios.

La información de los registros de llamados fue combinada con bases de datos epidemiológicos, sociodemográficos y sanitarios para realizar un índice de riesgo que identifica potenciales nichos de infección por Chagas, diseñado para apoyar políticas públicas que optimicen la asignación de recursos para diagnóstico, tratamiento y educación para la prevención.

1. La enfermedad de Chagas evoluciona en dos fases: aguda y crónica. La fase aguda se inicia en el momento de adquirir el parásito por cualquiera de sus vías, se caracteriza por presentar alta parasitemia, es detectable por métodos parasitológicos directos y dura entre 30 y 90 días. Generalmente es asintomática pero alrededor del 5% presenta manifestaciones clínicas generales inespecíficas (síndrome febril prolongado, hepatoesplenomegalia, edema generalizado, adenomegalia, irritabilidad, somnolencia, anorexia, diarrea). El chagoma de inoculación es una lesión característica del período agudo de esta enfermedad, y se produce en el sitio donde penetra el protozoario. Si se produce en el ojo se denomina complejo oftalmoganglionar (signo de Romaña), aunque puede encontrarse en cualquier región de la piel. Las expresiones clínicas graves de esta fase son: miocarditis y meningoencefalitis.

En la fase crónica la parasitemia suele ser baja por lo que el diagnóstico se realiza por medio de las técnicas serológicas. En Argentina se divide en fase crónica sin patología demostrable (el ECG y la radiografía de tórax son normales), o con patología demostrable.

Fuente: Curso virtual "Chagas, del conocimiento a la acción" de la Fundación Bunge y Born.

Antecedentes / Trabajo relacionado

La llamada era de la Big Data se caracteriza por la constante generación por parte de la sociedad de datos en gran volumen, al mismo tiempo que se incrementa la potencia de las computadoras y se inventan nuevas técnicas de análisis.

En este contexto surgen fuentes de datos que resultan novedosas para la investigación científica. Entre ellas las bases de registros de llamados de telefonía móvil, que son de especial interés para el estudio de las migraciones humanas². El uso de este nuevo recurso ofrece notables ventajas respecto de opciones tradicionales como las encuestas "origen-destino", que requieren entrevistar a un gran número de personas y pedirles que reconstruyan, de memoria, el motivo y trayecto de sus viajes durante las semanas recientes.

Los registros de llamadas ofrecen una cobertura geográfica y poblacional enormes en comparación con las encuestas (por ejemplo, cubriendo millones de personas en todo el territorio de un país), y con un costo de adquisición mucho menor a los significativos recursos necesarios para llevar a cabo encuestas de movilidad a miles de personas.

Diversos estudios científicos han realizado el procesamiento de millones de registros de llamados móviles para determinar la forma en que la gente se mueve, aprovechando que el gran nivel de detalle el análisis a distintas escalas geográficas (tanto viajes dentro de un barrio como entre países). A pesar de sus ventajas, cabe recordar la naturaleza aún novedosa de esta fuente ya que su aprovechamiento como recurso de investigación se encuentra en su período inicial y exploratorio. Algunos de los desafíos actuales provienen de la naturaleza de los datos. Al ser recolectados como parte de la provisión de un servicio comercial, y no para su uso en investigación científica, es esperable la presencia de sesgos, ruido y omisiones no documentadas. Por otra parte, -en contraste con las encuestas de movilidad tradicionales- los registros de llamadas son muy detallados en sus atributos espaciales y temporales, pero mucho más superficiales en los sociodemográficos.

Con incomparable área de cobertura y volumen, las bases de datos de llamadas celulares se están siendo utilizando cada vez más por diversas disciplinas, entre ellas la epidemiología. En este campo existen diversos antecedentes³ en el análisis de miles de millones de registros para modelar el patrón de difusión de endemias.

En línea con esto, en 2015, la Fundación Mundo Sano y Grandata llevaron a cabo el análisis de 18.000 millones de comunicaciones para detectar posibles zonas de riesgo de Enfermedad de Chagas en dos países de América Latina, Argentina y México. El trabajo fue presentado en el *conference paper* "Uncovering the Spread of an Infectious Disease with Mobile Phone Data"⁴, expuesto en el 2° *Simposio Argentino de Grandes Datos, AGRANDA 2016*, y en "Analyzing the Spread of

2. J. Candia, M. C. González, P. Wang, T. Schoenharl, G. Madey, y A.-L. Barabási. *Uncovering individual and collective human dynamics from mobile phone records*. *Journal of physics A: mathematical and theoretical*, 41(22):224015, 2008.

3. E. Frias-Martinez, G. Williamson, y V. Frias-Martinez. *An agent-based model of epidemic spread using human mobility and social network information*. En *Privacy, Security, Risk and Trust (PASSAT) y 2011 IEEE Third International Conference on Social Computing (SocialCom)*, páginas 57–64. IEEE, 2011.

4. Monasterio, Salles, Lang, Weinberg, Minnoni, Travizano, Sarraute. (2016). *Uncovering the Spread of an Infectious Disease with Mobile Phone Data*. Disponible en <https://www.researchgate.net/publication/319730478>. *Uncovering the Spread of an Infectious Disease with Mobile Phone Data*. Último acceso: 03/02/2020

5. Monasterio, Salles, Lang, Weinberg, Minnoni, Travizano, Sarraute. (2016). *Analyzing the Spread of Chagas Disease with Mobile Phone Data*. 607-612. 10.1109/ASONAM.2016.7752298. Disponible en: https://www.researchgate.net/publication/310809125_Analyzing_the_Spread_of_Chagas_Disease_with_Mobile_Phone_Data. Último acceso: 03/02/2020

*Chagas Disease with Mobile Phone Data*⁵, en ocasión de *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2016*. Ese trabajo mostró que los registros de llamadas geolocalizados (con información precisa sobre su lugar de su origen y destino) son ricos en información social e individual, y pueden utilizarse para inferir si un individuo ha vivido en un área endémica en algún momento de su vida. Así se estableció que distintos patrones de llamadas pueden ser inferidos, estableciéndose su relación con fenómenos sociales como las migraciones estacionales o de largo plazo. Más aún, de existir estadísticas a nivel detallado (esto es, al máximo nivel de desagregación posible), se podría inferir, bajo ciertos supuestos, los patrones migratorios de individuos infectados por parásitos, virus o bacterias.

Con esos antecedentes, en 2018 la Fundación Bunge y Born convoca una vez más a la Fundación Mundo Sano y a Grandata para desarrollar un proyecto sobre Detección de Zonas de Alta Prevalencia Potencial de Chagas en Argentina. Para esta investigación se dispuso de registros de unas 50 millones de llamadas diarias entre celulares de la Argentina, a lo largo de cinco meses entre 2011 y 2012. El estudio avanzó en dos direcciones fundamentales respecto a los trabajos mencionados:

- Se mejoró el modelo que cuantifica la interacción entre zonas endémicas y no endémicas previamente utilizado. Mientras que el modelo anterior tomó toda la zona endémica como un área sin diferenciar, con las mismas probabilidades de exposición al Chagas en toda su superficie, en esta ocasión se utilizó información complementaria acerca de las viviendas radicadas en la zona endémica para poder identificar un gradiente de regiones con distintas probabilidades de transmisión. La diferencia es determinada por el grado en que las condiciones locales de vivienda facilitan la cohabitación de personas con el vector de la enfermedad, es decir, la vinchuca (*Triatoma Infestans*).
- Se incorporó una dimensión que afecta las probabilidades de contagio de la enfermedad, llamada vulnerabilidad sanitaria, que refleja el grado en que la población local encuentra dificultades para acceder a la salud pública. Este indicador combina variables socioeconómicas de la población junto a métricas de accesibilidad a centros de salud pública.

En el transcurso del proyecto se analizaron los registros de llamados y combinaron con información complementaria en dimensiones como salud pública, características sociodemográficas y calidad de las viviendas. La combinación de las fuentes permitió calcular un Índice de Prevalencia Potencial de Enfermedad de Chagas para cada coordenada del territorio argentino. El término prevalencia, de acuerdo a su acepción en epidemiología, refiere a la la proporción de individuos en la población que hayan contraído la enfermedad. El Índice fue acompañado por la construcción de un Mapa de Prevalencia Potencial del Chagas cubriendo en detalle todas las localidades del país, desarrollado para tanto para comunicar los resultados como para informar la toma de decisiones de políticas públicas.

6. Fuente: Curso virtual "Chagas, del conocimiento a la acción" de la Fundación Bunge y Born.

Sobre la enfermedad de Chagas⁶

La enfermedad de Chagas o Tripanosomiasis americana es producida por el *Trypanosoma cruzi*, parásito unicelular, que se transmite principalmente por un insecto hematófago, llamado popularmente "vinchuca" o chinche (en otros países recibe otras denominaciones). Esta endemia se extiende desde el sur de EE.UU. hasta aproximadamente el paralelo 42 de Argentina y Chile.

En Argentina, la especie más importante es el *Triatoma infestans*, presente en las viviendas (construida con paredes de adobe, techos de paja, una o dos habitaciones y una galería al frente) y peridomicilios (comprenden el patio, las construcciones no contiguas al techo del domicilio como los depósitos, cocinas y corrales, y otros posibles refugios) de las zonas endémicas. Este conjunto es considerado parte del área de actividad humana.

Si bien la enfermedad de Chagas es una enfermedad asociada principalmente a las malas condiciones de vida (pobreza, escasa educación, vivienda inadecuada e insuficientes acciones del sistema de salud), los recientes movimientos migratorios provenientes de áreas rurales de nuestro país y países limítrofes ha producido un importante cambio epidemiológico de esta endemia. Hoy se entiende que la enfermedad de Chagas está urbanizada: en las ciudades libres de vectores viven un número importante de infectados, muchos de ellos provenientes de área endémica, otros que adquirieron la enfermedad por transmisión vertical.

El ser humano adquiere esta infección principalmente por formas: vectorial y no vectorial.

a) Vía vectorial: Cuando el insecto pica a un ser humano o animal infectado, el protozooario ingresa a su tubo digestivo donde se reproduce. Cuando vuelve a picar a otro huésped deposita en su piel las heces con los parásitos. Esta vía de transmisión sigue siendo la más frecuente en algunas regiones de América.

b) No vectoriales: pueden ser por vía transplacentaria (transmisión de la madre infectada a su hijo durante el embarazo, tanto en la fase crónica como aguda), por transfusiones de sangre o trasplante de órganos, si el donante está infectado, por vía oral, por ingesta de alimentos contaminados con las deyecciones de vinchucas que contienen parásitos (no demostrado aún en nuestro país), o bien por accidente de laboratorio.

En Argentina se estima que viven aproximadamente 1.500.000 de niños y adultos infectados.

Generación del Mapa de Alta Precisión de zonas de riesgo de Chagas en Argentina

Como entregable clave del proyecto, el mapa de riesgo fue pensado para señalar con precisión zonas “calientes” en el territorio argentino, es decir, aquellas que presentan a la vez alta afinidad con la zona endémica del Chagas y alta vulnerabilidad sanitaria. Fue conceptualizado y desarrollado como insumo para diferentes usuarios: investigadores, especialistas en salud pública, decisores de políticas públicas en salud y funcionarios públicos, entre otros. En particular para los decisores, este mapa facilitaría el desarrollo de estrategias costo-efectivas para mejorar el acceso al diagnóstico y tratamiento de la Enfermedad de Chagas.

La unidad de análisis es el radio censal, la unidad estadística más pequeña de la que se dispone información sociodemográfica pública, por lo que la información presentada posee un nivel de desagregación elevado. En la Argentina, cada unidad político-administrativa se desagrega en fracciones y cada una de ellas se desagrega a su vez en radios censales. El tamaño de los radios en áreas urbanas se determina según la cantidad de viviendas: en promedio, los radios contienen unas 300 residencias. En los bordes de localidades el radio urbano puede bajar a 200 viviendas, aproximadamente, y en localidades aisladas a 100 viviendas. En zonas rurales, los radios se determinan por la conjunción de distintos factores: características del terreno, accesibilidad y distancia entre las viviendas. En zonas urbanas los radios censales resultan muy precisos para la proyección territorial de indicadores: en zonas densamente pobladas, cada uno coincide con una cuadra. Sólo en zonas rurales (en las que la densidad poblacional es menor) los radios censales pueden adquirir dimensiones considerables, llegando a varios km². En Argentina, según datos del Censo 2010, existen algo más de 52.000 radios censales distribuidos en todo el país.

La estimación del Índice de Prevalencia Potencial del Chagas se obtuvo al cruzar tres sub-indicadores previamente elaborados, todos ellos calculados a nivel de radio censal. Un índice de “cohabitación”, que mide en la zona endémica el riesgo de convivencia con el vector de infección, la vinchuca. También se estimó un índice de afinidad con la zona endémica de la enfermedad, cuantificando el volumen de llamadas entre sitios de toda la Argentina con el Gran Chaco, para inferir patrones migratorios con su relocalización de población infectada. Y por último un indicador de vulnerabilidad sanitaria, para calificar áreas no sólo por su grado de afinidad con la zona endémica, sino también por la fragilidad de su población local frente a problemas de salud.

Figura 1
Límites de radios censales (en color rosado) sobreimpuestos sobre la grilla de calles de Rosario, Santa Fe.



Índice de Receptividad al Vector de Transmisión

La estimación de este indicador representa un paso más respecto a estudios anteriores, que toman a la zona endémica como un área homogénea⁷. Es habitual considerar a la totalidad de la superficie de la ecorregión del Gran Chaco como el área endémica de la enfermedad de Chagas, ya que es donde ocurren las condiciones ambientales necesarias para la existencia de su vector de transmisión, el *Triatoma infestans*, conocido también como vinchuca.

Sin embargo, es razonable pensar que dentro del territorio abarcado las probabilidades de transmisión de la enfermedad no se encuentran distribuidas de manera uniforme. En efecto, el Gran Chaco no es una zona de transmisión vectorial

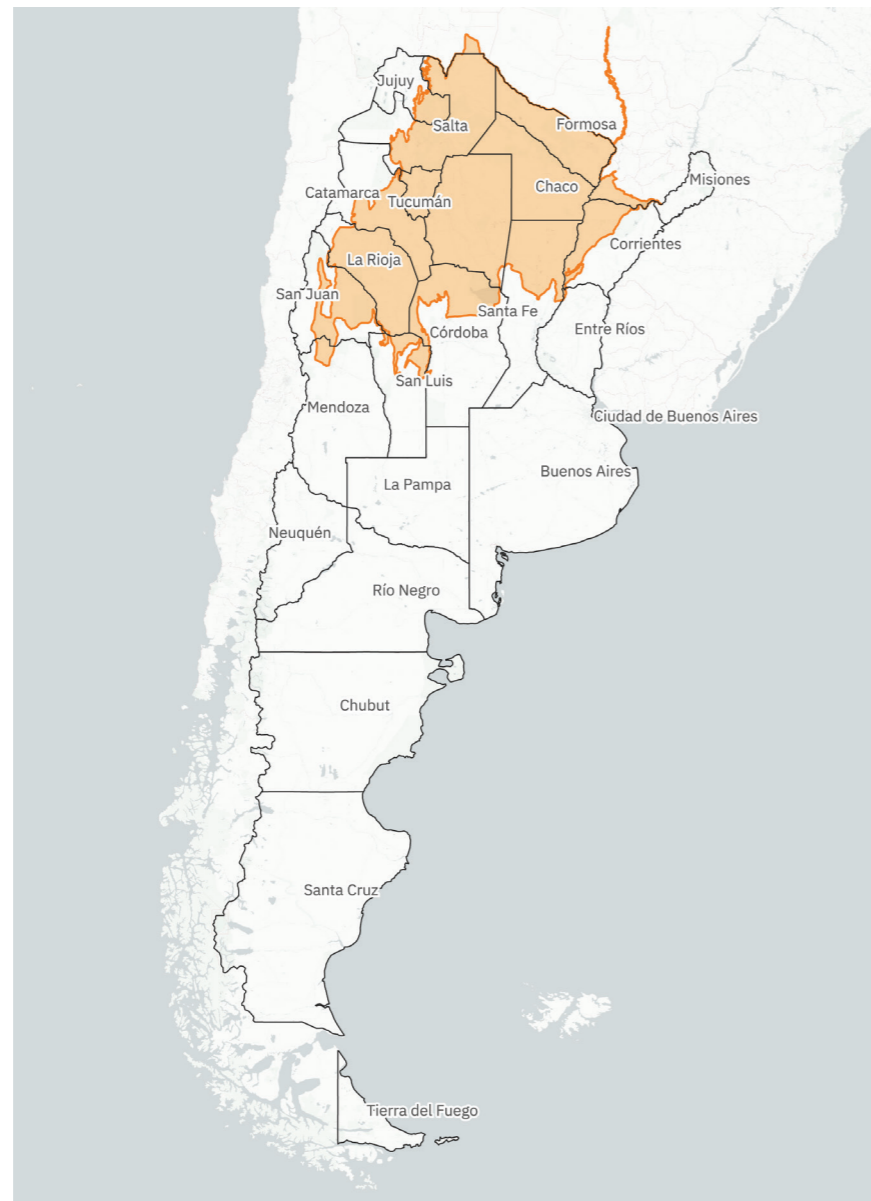
7. Monasterio, Salles, Lang, Weinberg, Minnoni, Travizano, Sarraute. *Analyzing the spread of Chagas disease with mobile phone data* (2016). Monasterio, Salles, Lang, Weinberg, Minnoni, Travizano, Sarraute. (2016). *Analyzing the Spread of Chagas Disease with Mobile Phone Data*. 607-612. 10.1109/ASONAM.2016.7752298. Sarraute, Lang, Monasterio, Weinberg. (2015) *Descubriendo Chagas con big data*.

8. R. Chuit, E. Subias, P. AC., I. Paulone, C. Wisnivesky-Colli, and E. Segura. *Usefulness of serology for the evaluation of trypanosome cruzi transmission in endemic areas of chagas's disease* (2003).

homogénea debido a diferencias geográficas, a la historia de acciones de control de la enfermedad, y a condiciones sociodemográficas y edilicias de sus pobladores. Si bien se carece de registros completos y confiables, la transmisión (su incidencia anual) se ve facilitada en parte por una pobre infraestructura de las viviendas y el peridomicilio, receptivas al anidamiento de la vinchuca⁸.

Por este motivo, se buscó identificar la receptividad del hábitat humano al vector de la enfermedad. Si una vivienda presenta ciertas características (como materiales precarios en techo y paredes), presentará una mayor probabilidad de alojar al vector, y por ende, una mayor probabilidad de transmisión de la enfermedad. A tal efecto se construyó una métrica para cuantificar en cada radio censal la cantidad de viviendas que presentan condiciones habitacionales y de materiales que son propicias para que la vinchuca se asiente y forme colonia.

Figura 2
Ecorregión del Gran Chaco, área endémica de la Enfermedad de Chagas.



9. INDEC. *Hábitat y vivienda por medio de datos censales. Calidad de los materiales de la vivienda (CALMAT) (2003)*

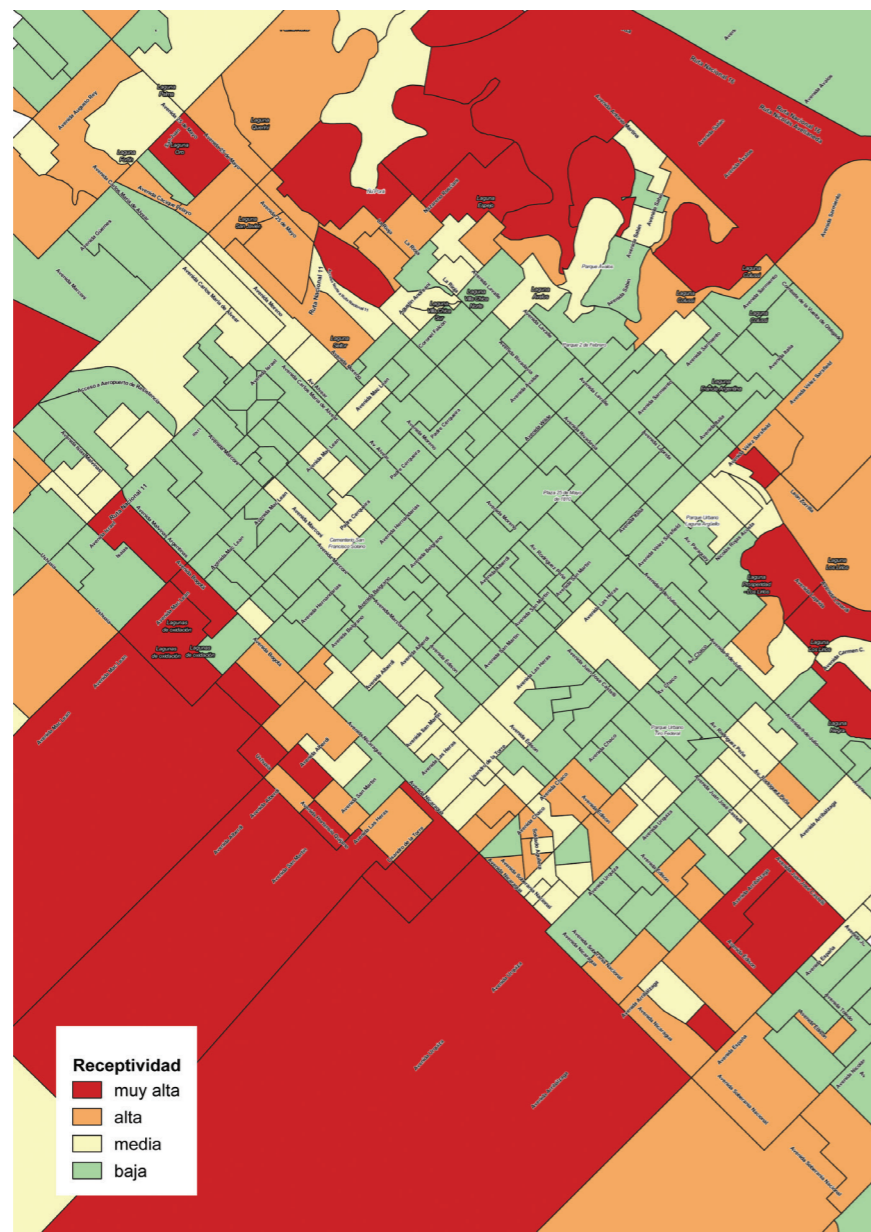
Condiciones de la vivienda

Dado que es posible flexibilizar el supuesto acerca de las chances de transmisión uniformes en toda la ecorregión del Gran Chaco, y considerando el componente social de la enfermedad, las condiciones de habitabilidad de las viviendas y su relación directa con las condiciones propicias para el alojamiento del vector, se propuso como objetivo desagregar la región en múltiples sub-áreas. Para ello, se utilizó información proveniente del Censo Nacional de Población, Hogares y Viviendas del año 2010 analizando calidad de materiales para construir un índice que cuantifica, para cada vivienda, sus condiciones de viabilidad para el alojamiento de la vinchuca.

Las viviendas con la tipología de rancho, cuyos materiales de construcción incluyen paredes de adobe sin revoque y techos de caña, barro o paja favorecen la domiciliación del vector. Como se reportó en el informe técnico del INDEC⁹ sobre calidad de materiales de la vivienda: "La caña, tabla o paja con barro, paja sola son elementos que suelen ser buenos aislantes térmicos, pero no son aislantes hidrófugos. Se usan en lugares secos y se construyen con mucha pendiente a fin de acelerar el escurrimiento del agua. Asimismo, tienen peligro de incendio, no son higiénicos, e inclusive favorecen a la propagación del Mal de Chagas [sic] a través de la vinchuca".

Para la elaboración del Índice de Contacto con el Vector de Transmisión se contó con variables para cada vivienda de la región relacionadas al material predominante en los pisos (cerámica, madera, cemento, ladrillo, tierra, etc), en la cubierta exterior del techo (membrana, losa, teja, chapa metálica, chapa plástica, caña, paja, barro, etc) y también sobre la presencia de revestimiento interno o cielorraso en los techos. Dado que se trata de variables discretas y categóricas, para combinarlas en una medida unificada se utilizó el método de análisis estadístico conocido como Análisis de Correspondencias Múltiples. La técnica permite combinar múltiples "dimensiones" en los datos -como en este caso, la calidad de pisos, de techos y de revestimientos- en un conjunto menor de variables, incluso una sola, perdiendo la menor capacidad posible de información. Una vez caracterizados todos los radios censales de la zona endémica con su Índice de Receptividad, fue posible identificar las áreas donde las condiciones favorecen en baja, media o alta medida la presencia de la vinchuca, e indirectamente su grado de riesgo de infección.

Figura 3
Índice de Receptividad al Vector de Transmisión (vinchuca) de las viviendas en los alrededores de Sáenz Peña, Chaco.



10. G. Busso. *Migración interna y desarrollo territorial en Argentina a inicios del siglo xxi. Brechas e impactos sociodemográficos de la migración interna interprovincial* (2007)

Índice de Afinidad con la Zona Endémica de Chagas

El Índice de Afinidad cuantifica el grado de vinculación entre áreas endémicas de la enfermedad de Chagas y el resto del país. Este índice se basa en los registros de llamadas de telefonía móvil como fuente de datos. Se extrajo para cuantificar la afinidad entre áreas endémicas y no endémicas, es decir, en qué medida las diferentes localidades del país se vinculan con la ecorregión del Gran Chaco.

Dado que fuera del área endémica no existen condiciones propicias para la reproducción del principal vector de la enfermedad (la especie *Triatoma infestans* de vinchuca), el principal vehículo de propagación de la enfermedad fuera de su origen natural se vincula a las corrientes migratorias.

Este hecho es aún más relevante teniendo en cuenta que buena parte de las provincias que forman el Gran Chaco han sido productoras históricas de fuerza de trabajo para la economía argentina, cuanto menos entre principios y fines del siglo XX¹⁰. Por ello, la posibilidad de detectar zonas con alta prevalencia potencial de enfermedad de Chagas requiere una correcta detección y mapeo de los flujos migratorios entre el Gran Chaco y el resto del país.

Un primer obstáculo para identificar flujos migratorios siempre ha residido en la falta de información demográfica con un nivel de desagregación adecuado. La fuente clásica (y única del Sistema Estadístico Nacional con cobertura nacional rural y urbana) para el estudio de la migración interna son los Censos Nacionales de Población. Sin embargo, en general, se limitan a cuantificar los flujos migratorios internos a nivel provincial. Si bien se indaga por los lugares de nacimiento, residencia habitual y residencia previa (cinco años atrás), los datos se publican a nivel provincial: sólo se dispone de datos referidos a cuánta población recibió o emitió cada provincia, sin detalle sobre los destinos ni sobre los orígenes del flujo migratorio.

Existen otros métodos que pueden ser utilizados para la estimación indirecta de flujos migratorios con un nivel de desagregación menor al provincial. Los registros de Estadísticas Vitales se basan en la llamada "ecuación compensadora" de la población, que partiendo de la cantidad de habitantes en un período dado, suma nacimientos e inmigraciones y resta defunciones y emigraciones para obtener el cambio total. El saldo migratorio puede entonces obtenerse como la diferencia entre el crecimiento "vegetativo" de la población (restando las defunciones a los nacimientos) y el crecimiento total observado. Si la diferencia es negativa, el saldo migratorio es negativo, ya que han emigrado más personas de las que inmigraron; y viceversa. Este método tiene como ventaja su simplicidad: sólo requiere conocer la población al inicio y al final del período y la cantidad de defunciones y nacimientos del período, todos datos disponibles en la Argentina a nivel Departamento, en los registros de Estadísticas Vitales¹¹. Sin embargo, si bien eso permite una mayor desagregación geográfica en comparación con los datos del Censo Nacional, se mantiene la incertidumbre sobre el origen y destino de las migraciones. Eso reduce la utilidad de éstas métricas para estudios epidemiológicos.

Estas limitaciones llevan a la necesidad de explorar otras fuentes de información para cuantificar los flujos migratorios, y es cuando resultan de utilidad las bases de registros de llamadas. Se parte de la siguiente premisa: un alto volumen de comunicaciones móviles entre dos áreas (debidamente filtrados y procesados, como se verá a continuación) constituye un indicador aproximado válido de la

11. Dirección de Estadísticas e Información en Salud, Ministerio de Salud de Argentina. *Estadísticas Vitales*. Disponible en <http://www.deis.msal.gov.ar/index.php/estadisticas-vitales/>

existencia de flujos migratorios entre ellas. Un alto grado de afinidad entre un área endémica y un área externa a ella podría ser considerado entonces como un indicador indirecto de la presencia de población migrante de área endémica que vive al momento del análisis en un área no endémica.

De este modo, cuantificar con alta precisión espacial el grado de afinidad de cada región permitirá identificar aquellas zonas donde reside una población con mayores chances de haber contraído Chagas, por la dinámica de transmisión en Argentina. Tanto si la infección se produjo en un período anterior a la migración, cuando la población habitaba en área endémica, o por vía transplacentaria, en cuanto aquellas zonas con población que proviene de zona endémica tendrán también mayor probabilidad de tener mujeres en edad fértil infectadas, que pueden transmitir la enfermedad a su prole¹².

Call Detail Records (CDRs)

El conjunto disponible de datos de telefonía móvil utilizado para el análisis consta de un registro con la ubicación de las antenas que brindan la cobertura de servicio, y una tabla de llamados (en inglés conocidos como *Call Detail Records* o CDRs), donde cada llamado está asociado a una antena particular, que es la que se utilizó para efectuar la conexión.

El registro de antenas geolocalizadas tiene aproximadamente 3.500 antenas con ubicación válida¹³. La base de registros de llamadas cuenta con aproximadamente 70 millones de observaciones por día, es decir unos 2100 millones por mes y más de 10 mil millones durante el período disponible, de 5 meses entre noviembre de 2011 y marzo de 2012.

Para contabilizar interacciones entre usuarios, éstos fueron identificados unívocamente utilizando un código anonimizado con la técnica conocida como "hashing". Este paso, que se realiza en forma previa a cualquier análisis de los datos, consiste en transformar los identificadores de usuario utilizando una función resumen que toma la secuencia de caracteres del identificador original y devuelve una nueva secuencia. La operación es irreversible: no es posible determinar la información de entrada -es decir, el código que identifica a un usuario- en base su código anonimizado resultante¹⁴. Esto permite resguardar la anonimidad de los usuarios registrados en la base, manteniendo a la vez la capacidad de discernir entre usuarios distintos.

Los CDRs están organizados en registros (filas), y cada uno aporta los siguientes datos (columnas): identificador anonimizado de usuario emisor, identificador anonimizado de usuario receptor, día del llamado, hora, duración en segundos y código de la antena usada.

Metodología

El proceso de análisis de los CDRs para la extracción del Índice de Afinidad con la Zona Endémica de Chagas se realizó tomando a la antena como unidad de análisis. Dado que cada antena representa un área de cobertura dada, el índice es un atributo de *lugares*, y no de *personas*. En otras palabras, el índice permite clasificar y comparar entre sitios, pero de ningún modo identifica individuos.

El procesamiento de los datos requerido para estimar el índice puede resumirse en cuatro etapas: detección de antena hogar de cada usuario, estimación del área de cobertura de cada antena, cálculo de receptividad al vector de la zona de cada antena, y por último cálculo de afinidad con zona endémica de chagas de cada antena.

1) Detección de "antena hogar" de cada usuario

El primer paso consistió en identificar la antena en la que se registraron más llamados de cada usuario durante el período estudiado. Dado que el objetivo fue determinar el área de residencia de los usuarios, se consideraron sólo los llamados realizados los días de semana por la noche: de lunes a jueves, entre las 8 de la noche y las 6 de la mañana del día siguiente. Se optó por ese rango asumiendo que coincide con la franja horaria durante la cual la mayor parte de las personas permanece en su residencia. De existir más de una antena con una cantidad máxima de llamados, se seleccionó en forma aleatoria una de ellas.

La cantidad de usuarios para los cuales se definió una antena hogar ronda los 15 millones.

2) Estimación del área de cobertura de cada antena

Los datos disponibles para el estudio, así como los índices elaborados en base a ellos, tienen como unidad de análisis el radio censal, representado por polígonos. Ahora bien, la posición de las antenas está representada por puntos fijos en el espacio. Por ello, el primer paso para la generación de los índices a nivel de antena consistió en estimar la superficie de cobertura de cada una de ellas en toda la Argentina. El método utilizado fue el la llamada "partición de Voronoi", que subdivide un plano de acuerdo a la ubicación de un conjunto de sitios, donde a cada sitio le corresponde una y sólo una región o celda. Uno de los atributos de la partición de Voronoi es cada celda abarca los puntos del plano que se encuentran más cercanos a su sitio que a cualquier otro¹⁵. La metodología es utilizada para el análisis de comunicaciones móviles cuando no se dispone de datos oficiales acerca de las zonas de servicio; al calcular la Celda de Voronoi correspondiente a cada antena se obtiene una aproximación a su área de cobertura¹⁶. Debido a su relativa soledad, las antenas en áreas rurales y escasamente pobladas obtienen áreas de cobertura de enorme extensión, llegando a miles de km². En zonas densamente pobladas, donde existen múltiples antenas en estrecha proximidad para atender la demanda de un gran número de abonados, las áreas de cobertura se reducen a una pocas cuadras.

12. R. Chuit, R. E. Gurtler, L. Mac Dougall, E. L. Segura, y B. Singer. *Chagas disease risk assessment by an environmental approach in northern Argentina* (2001).

13. En los datos originales se encontraron antenas con datos inconsistentes, es decir, que no tenían información de localización (latitud/longitud) o bien que con coordenadas fuera del territorio argentino. Estos registros fueron descartados.

14. M. Peyravian and N. Zunic. *Methods for protecting password transmission*. Computers & Security (2000).

15. F. Aurenhammer. *Voronoi diagrams—a survey of a fundamental geometric data structure*. ACM Computing Surveys (1991)

16. J. Steenbruggen, E. Tranos, y P. Nijkamp. *Data from mobile phone operators: A tool for smarter cities?* Telecommunications Policy (2015)

Figura 4
Celdas de Voronoi calculadas en torno a las antenas en el territorio argentino, indicando su área de cobertura.



3) Cálculo de receptividad al vector para cada antena

En este paso se determinó, para cada antena, el grado de riesgo de contagio de chagas de los radio censal abarcados por su área de cobertura, utilizando el Índice de Receptividad al Vector de Transmisión calculado con anterioridad. Como se describió en su apartado, la receptividad es una medida indirecta del riesgo de contraer chagas para la población local. Depende sólo de atributos geográficos y demográficos, y no de aquellos relacionados con la comunicación telefónica.

A cada antena ubicada dentro de la ecorregión del Gran Chaco le correspondió un grado de receptividad, un número entero de 1 a 4 inclusive, en gradiente de

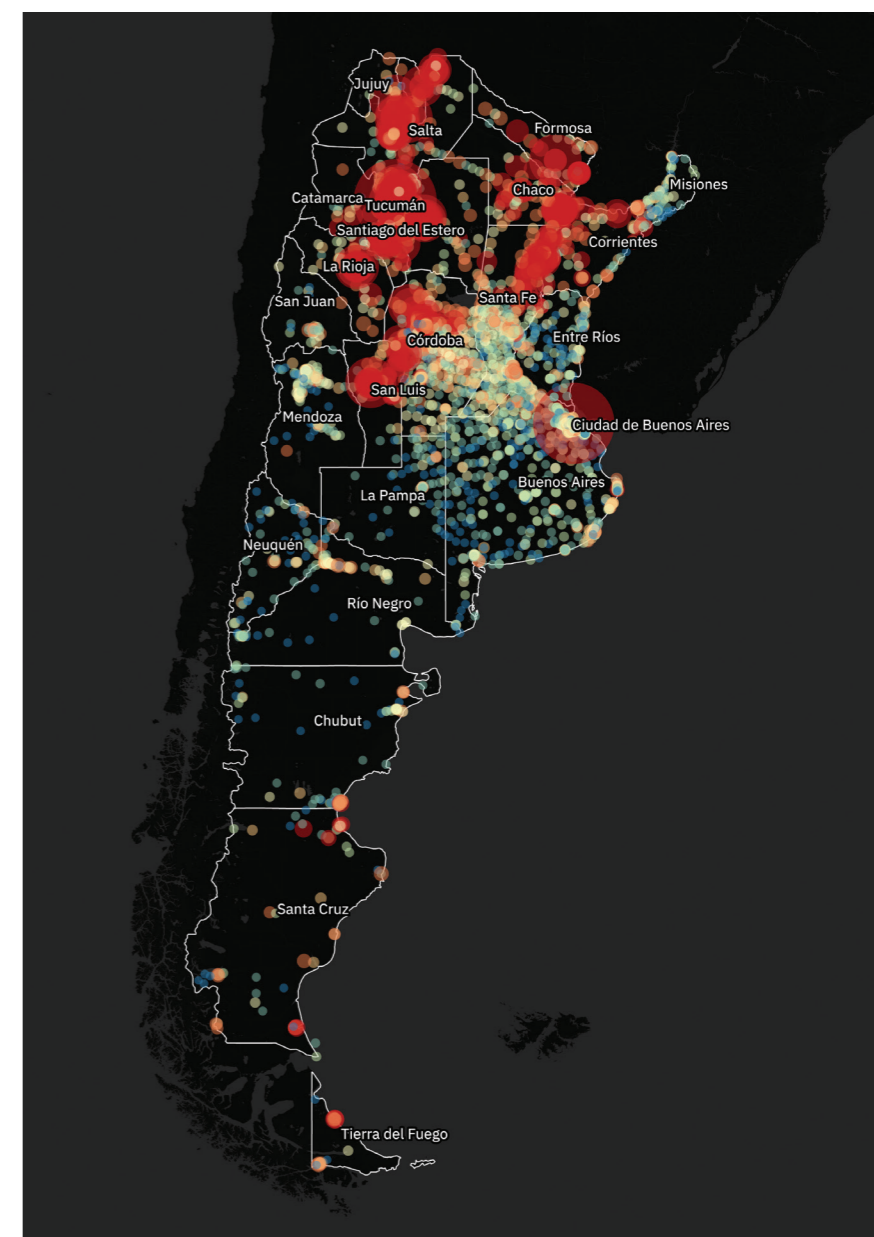
menor a mayor grado de receptividad a la vinchuca de la zona. El valor depende del cuartil del valor de receptividad del radio censal de la antena: del 25% más bajo, que recibió un 1, al 25% más alto, que llegó a un 4. Dado que se descarta la presencia de vinchucas fuera del área endémica, a las antenas ubicadas allí les correspondió un valor de 0.

4) Cálculo de afinidad con zona endémica de Chagas de cada antena

Habiendo identificado las antenas hogar de los usuarios de telefonía móvil, se procedió a generar y analizar el "grafo social", la red que conforman al conectarse con llamados. Para cada llamada, se tomó nota el valor de receptividad al vector de la antena hogar de ambos usuarios conectados por la comunicación. Para cada

Figura 5
Puntos con la ubicación de antenas de telefonía móvil en el territorio argentino

El área de los puntos aumenta según la cantidad de residentes con afinidad al área endémica de Chagas. El color pasa de frío a cálido cuanto más alto es el grado de exposición a la vinchuca de las regiones con las que hay vinculación. En violeta, los límites del área endémica.



usuario, se definió un valor de "afinidad" con el área endémica de Chagas igual al valor máximo de receptividad encontrado en los destinos de todas sus comunicaciones. Una vez se obtuvieron las métricas a nivel usuario, estas se agregaron a nivel antena. Así, a cada antena le correspondió un Índice de Afinidad con la Zona Endémica de Chagas representado por dos valores: la cantidad de usuarios que desde allí se comunicaron con el área endémica, y el nivel máximo de receptividad al vector de las zonas con las que se comunicaron los usuarios.

Índice de Vulnerabilidad Sanitaria

Para cualquier población, las chances de padecer Chagas o cualquier otra enfermedad transmisible no se encuentran asociadas tan sólo a las probabilidades de contacto con el vector, como la residencia o procedencia de zonas endémicas. Existen también determinantes de carácter más general que podrían ser englobados bajo el concepto de "Vulnerabilidad Sanitaria". Dichos determinantes se encuentran asociados al desigual acceso a los servicios y cobertura de salud que experimentan diferentes segmentos de la población.

Cuantificar ésta dimensión es útil para complementar al Índice de Afinidad con Zonas Endémicas de Chagas. En particular, conocer el nivel de vulnerabilidad sanitaria de cada zona permite priorizar lugares donde la población no sólo presenta afinidad con el área endémica, sino que además tiene bajo acceso a servicios sanitarios.

El proceso de construcción del Índice de Vulnerabilidad Sanitaria¹⁷ puede resumirse como la combinación de dos factores: Acceso a prestaciones y servicios de salud por parte del estado, medida en base a la cercanía a efectores de salud pública, y Nivel Socioeconómico de la Población, que es calculado usando un método innovador que incorpora una serie de variables relevantes -nivel educativo, indicadores de Necesidades Básicas Insatisfechas, etc.-

Una vez calculados para todos los radios censales del país, los indicadores de acceso a la salud pública y de nivel socioeconómico fueron combinados utilizando una técnica de reducción de dimensionalidad, similar a las descrita para extraer el Índice de Receptividad al Vector. Así se obtuvo una variable continua que varía de 0 (mínima vulnerabilidad entre los radios censales del país) a 1 (máxima vulnerabilidad).

Proyectado sobre el mapa, el Índice de Vulnerabilidad Sanitaria se distribuye de modo tal que se observa la concentración de radios censales de baja vulnerabilidad en la zona central del país, en torno a los núcleos urbanos de las regiones pampeana y litoraleña. En el resto del país, principalmente las zonas menos densamente pobladas, tienen mayor frecuencia los radios con valores críticos. Sin embargo, existen nichos de alta vulnerabilidad sanitaria incluso en el interior de las regiones con mayor nivel de desarrollo, como la CABA.

Figura 6
Tiempo a pie hasta hospital público más cercano, calculado en promedio por radio censal



17. Ver Documento de Trabajo Fundación Bunge y Born 2019-01

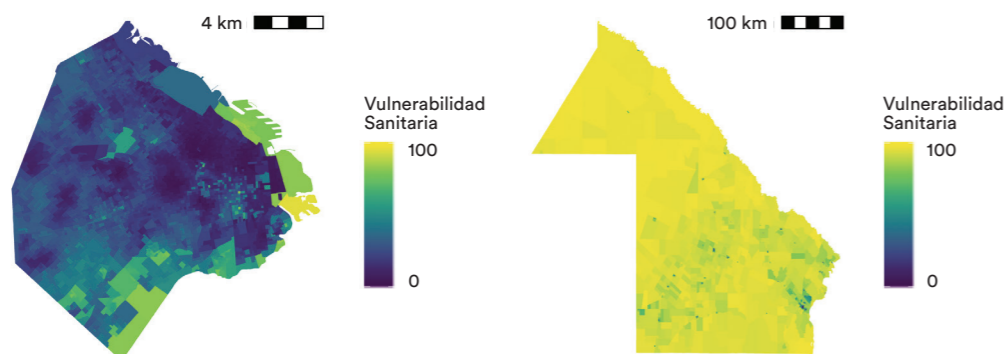


Figura 7
Índice de Vulnerabilidad Sanitaria. CABA (izq.), Chaco (der.)

Índice de Prevalencia Potencial de Chagas (IPPECh)

El IPPECh representa la respuesta al desafío planteado para el proyecto de investigación: "¿Dónde se encuentran los enfermos de Chagas en la Argentina actual?". Fue calculado como el producto de las variables de Afinidad local con zona endémica de Chagas, y de Vulnerabilidad Sanitaria. Multiplicando los términos, mayor será la prevalencia potencial de Chagas donde sean, a la vez, mayores la afinidad con la zona endémica y la vulnerabilidad sanitaria.

El índice predice la prevalencia de la enfermedad en forma relativa: sus valores indican cuánto mayor es la prevalencia potencial de Chagas en un radio censal respecto al valor de la media.

Resultados

El análisis de los resultados que se presenta en esta sección contempla los componentes del índice por separado, con el objetivo de detectar áreas de interés para un futuro trabajo de campo, que permita validar el modelo de prevalencia potencial y la estimación de los parámetros de agregación del IPPECh. Dichas áreas fueron seleccionadas a partir de la identificación de zonas del país donde la afinidad es alta que se ven mediadas por la Vulnerabilidad Sanitaria.

Para tal fin, se consideraron sólo áreas ubicadas fuera de la eco-región del Gran Chaco -aquellas donde un alto nivel de comunicación con el Gran Chaco no sea atribuible al volumen esperable de llamados entre personas físicamente cercanas.

En este punto, debe tenerse en cuenta que la unidad original del mapa (radios censales) no resulta viable para la organización del trabajo de campo: seleccionar radios censales relevantes (con alta afinidad) y sus correspondientes contrastes conduciría a una gran dispersión territorial de las áreas a visitar, razón por la cual se agregó la información a nivel localidad¹⁸.

Así, luego de filtrar aquellas localidades con menos de 350 habitantes, se construyeron dos métricas por localidad:

1. El promedio de afinidad a lo largo de todos los radios que componen la localidad (ponderado por la población en cada radio),
2. El promedio de afinidad de los radios de la localidad que estuvieran entre los valores más altos de la distribución nacional de afinidad¹⁹.

Estas métricas dan lugar a tres tipos de situaciones posibles:

- I. *Localidades* que presentan alta afinidad pero "dispersa", exhibiendo una concentración similar en todos los radios que la componen; las llamaremos *de alta afinidad media*,
- II. *Localidades* que presentan valores de afinidad altos y altamente concentrados: es decir, pocos radios de la localidad presentan valores altos de afinidad; nominadas como localidades *de radios extremos*,
- III. *Localidades que cumplen ambos criterios*.

Además, dentro de cada localidad se descartaron aquellos radios censales donde la densidad de población no alcanza los 350 habitantes por km², en tanto es necesario alcanzar una concentración de población local mínima para garantizar la viabilidad operativa de una eventual intervención en terreno. Para favorecer contrastes y permitir variabilidad, fueron seleccionadas (selección estratificada) al interior de cada provincia localidades de los tres tipos mencionados.

18. Para el INDEC una localidad "...es una porción de la superficie de la tierra caracterizada por la forma, cantidad, tamaño y proximidad entre sí de ciertos objetos físicos artificiales fijos (edificios) y por ciertas modificaciones artificiales del suelo (calles), necesarias para conectar aquellos entre sí." Es decir, "la localidad se concibe como concentración espacial de edificios conectados entre sí por una red de calles." Este es el criterio implícito en todos los censos argentinos, y explicitado a partir del censo de 1991." Operativamente, cada radio censal pertenece a una y solo una localidad, por lo cual es posible agregar la información de cada radio a nivel localidad.

19. Este umbral fue definido luego de comprobar que la distribución nacional de la afinidad es bimodal, y más específicamente, una mixtura de distribuciones. De esta forma, el umbral es el valor que mejor logra separar ambas distribuciones. En este caso, se usó un método clásico basado en el boxplot: el umbral fue calculado como $Q_{75} + 1,5 \times IQR$, es decir, el cuarto cuartil más 1,5 veces el rango intercuartilico. Todas las localidades cuyo valor de afinidad estuviera por encima de dicho umbral fueron consideradas en el cálculo de esta segunda métrica.

1. **Localidades de alta afinidad media:** fueron seleccionadas las tres con mayor afinidad media ponderada por la población de cada radio en cada provincia,
2. **Localidades de radios extremos y localidades con ambos tipos:** se siguió el mismo criterio, la diferencia es que en este caso, el promedio se calcula sobre los radios que sobreviven al filtro anterior; esto puede hacer que en algunas provincias no presenten localidades de estos tipos.

Considerando todo lo anterior, los resultados pueden visualizarse como sigue:

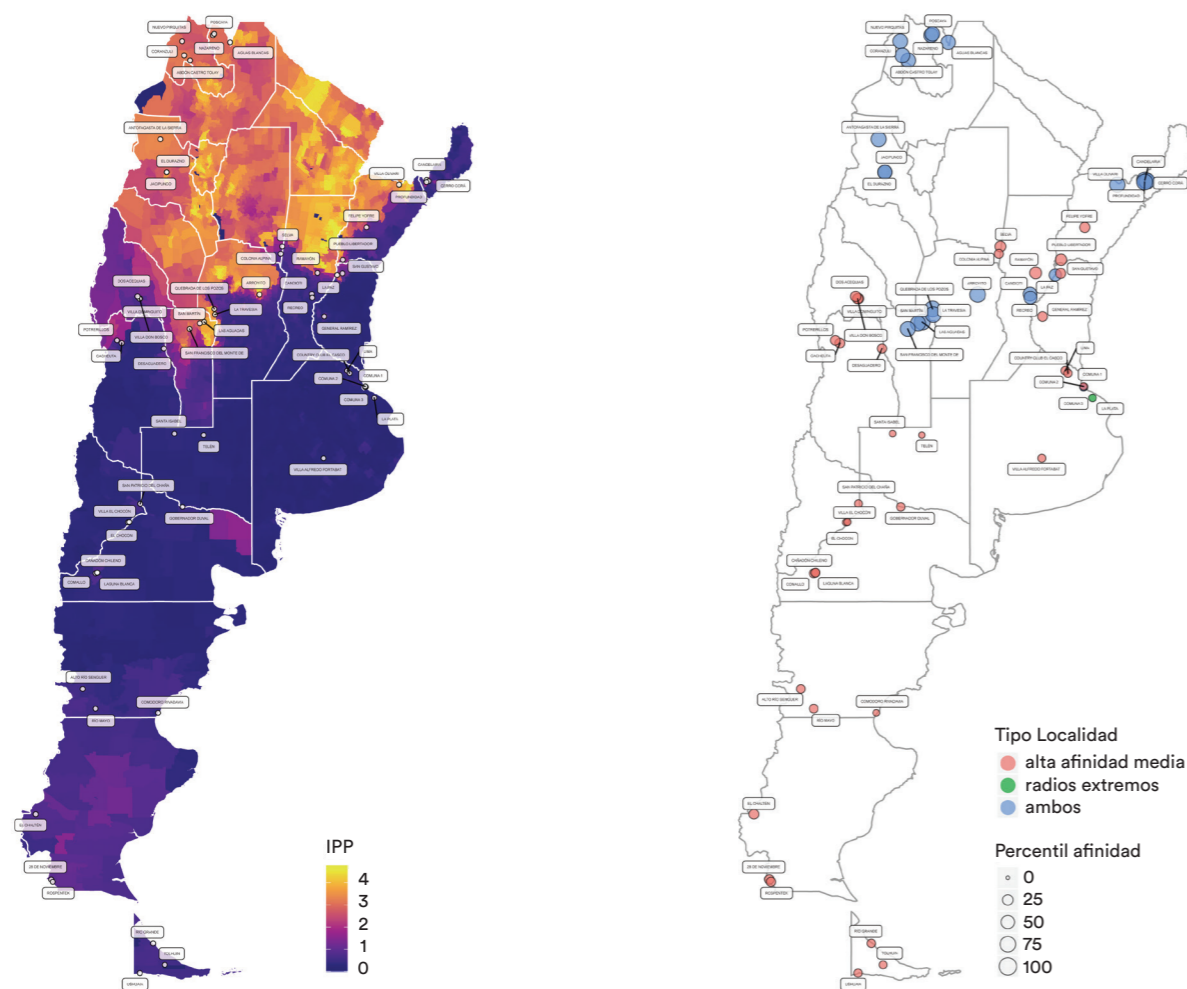


Figura 8
Localidades (homogéneas y concentradas) seleccionadas como áreas potenciales de intervención según niveles del IPPECh y según percentil de afinidad.

La figura 8 muestra las localidades seleccionadas, proyectadas sobre el mapa del IPPECh.

Tal como se estableciera en [11], las áreas con mayores índices de Afinidad se concentran en la Región Metropolitana de Buenos Aires (incluyendo a la ciudad de La Plata) y en las provincias de la Patagonia Argentina -Neuquén, Río Negro, Chubut, Santa Cruz y Tierra del Fuego-. También aparecen como relevantes algunas localidades de áreas fronterizas al Gran Chaco, en provincias como Misiones, Entre Ríos y Mendoza.

Un primer punto a notar es que existe solamente una localidad (La Plata, Buenos Aires) que se caracteriza por presentar radios extremos pero no alta afinidad media. A su vez, puede notarse que entre las zonas fronterizas al Gran Chaco, predominan dentro de las localidades seleccionadas aquellas que combinan radios extremos y localidades de alta afinidad media. Estas localidades se caracterizan, como era esperable, por mayores valores en los percentiles de afinidad.

En cambio, en las zonas centro y Patagonia, la situación parece ser diferente: las localidades seleccionadas se caracterizan por ser de afinidad media alta y por menores valores en los percentiles de afinidad.

El cuadro 1 resume los atributos de las áreas más salientes, representando 20 localidades seleccionadas en función de los criterios antes mencionados. Puede consultarse en el Anexo B una tabla extendida con las tres localidades fuera del área endémica de Chagas con mayor Prevalencia Potencial, por provincia.

Cuadro 1
Localidades fuera del área endémica con mayor Prevalencia Potencial de Enfermedad de Chagas, por provincia²².

Provincia	Departamento	Localidad	Población	Afinidad	Vulnerabilidad Sanitaria	Usuarios con afinidad
San Luis	Lib. San Martín	Las Aguadas	487	81.4%	96.6%	94.9%
Córdoba	San Javier	Quebrada de los Pozos	398	78.9%	76.7%	93.8%
Misiones	Candelaria	Candelaria	13871	78.2%	65.9%	78.5%
Corrientes	Ituzaingó	Villa Olivari	1276	72.2%	88.3%	94.2%
Catamarca	Antofagasta	Antofagasta	730	70.7%	75.9%	91.4%
Jujuy	Cochinoca	Abdón Castro Tolay	376	66.2%	86.6%	88.0%
Salta	Santa Victoria	Nazareno	916	63.9%	93.8%	97.8%
Santa Fe	La Capital	Candioti	888	55.7%	72.9%	74.2%
Entre Ríos	La Paz	La Paz	24307	38.5%	60.7%	42.2%

Provincia	Departamento	Localidad	Población	Afinidad	Vulnerabilidad Sanitaria	Usuarios con afinidad
Santiago del Estero	Rivadavia	Selva	2878	27.5%	75.0%	33.5%
San Juan	San Martín	Villa Dominguito	619	24.4%	88.2%	89.0%
Santa Cruz	Lago Argentino	El Chaltén	1627	20.6%	89.0%	25.7%
Mendoza	Luján de Cuyo	Cacheuta	701	20.1%	93.5%	25.6%
Buenos Aires	Zárate	Lima	10219	14.7%	75.6%	19.1%
Río Negro	Pilnaniyeu	Comallo	1497	14.3%	45.1%	17.6%
Chubut	Río Senguer	Río Mayo	2791	13.7%	83.5%	18.4%
Tierra del Fuego	Río Grande	Río Grande	66475	12.5%	36.8%	16.5%
La Pampa	Curacó	Gobernador Duval	413	12.2%	95.1%	17.2%
Neuquén	Añelo	San Patricio Del Chaña	6474	9.4%	35.0%	14.4%
CABA	Comuna 01	Comuna 1	205886	8.6%	25.9%	9.2%

22. Afinidad y Vulnerabilidad Sanitaria son percentiles: un valor de 80% indica que la localidad tiene un valor más alto que el del 80% de todas las localidades del país. "Usuarios con afinidad" indica el porcentaje de los abonados a telefonía móvil residentes que han realizado o recibido llamados contactando con el área endémica: un valor de 80% indica que 8 de cada 10 de los residentes encontrados en los registros de llamados han contactado con el área endémica.

Si bien comparten altos niveles de afinidad, las zonas no exhiben patrones similares en cuanto a su vulnerabilidad sanitaria. En esa dimensión la variabilidad es alta, abarcando áreas en todo el rango desde la baja vulnerabilidad en lo relativo al resto del país (CABA, Vicente Lopez, La Plata), hasta altos niveles: El Durazno y Jacipunco (Catamarca), Pueblo Libertador (Corrientes), Coranzulis (Jujuy), Poscaya (Salta).

Entre las localidades seleccionadas, en la figura 9 se observan seis con población mayor a 5000 personas, cada una en una provincia distinta y en orden decreciente de afinidad. De forma esperable por su cercanía geográfica a la zona endémica, las localidades en Misiones, Córdoba y Santa Fe exhiben una mayor afinidad respecto a sus contrapartes en el Centro y Sur del país. Sin embargo, cabe notar que la afinidad no disminuye en forma homogénea, es decir en gradiente continuo a medida que las localidades se alejan del área endémica. Al contrario, se detectaron localidades en la Provincia de Buenos Aires y en la Patagonia cuyo grado de afinidad es mucho más alta que el de centros poblados en provincias más cercanas a la zona endémica -como La Pampa-. Esto sugiere la existencia de migración considerable desde regiones endémicas hacia las localidades resaltadas. Un

indicio adicional al respecto es el alto porcentaje de usuarios locales de telefonía celular que llamó, o recibió llamadas desde regiones endémicas: Más de una cuarta parte en 28 de Noviembre, cerca de un 20 por ciento en Lima y un 13 por ciento en Tolhuin.

Figura 9
De izquierda a derecha, y de arriba a abajo: Candelaria (Misiones), Arroyito (Córdoba), Recreo (Santa Fe), 28 de Noviembre (Santa Cruz), Lima (Buenos Aires), Tolhuin (Tierra del Fuego)



Conclusiones

Como primer conclusión, es posible plantear un proceso de revisión, mejora y actualización. En efecto, varios de los indicadores y fuentes de información usadas como insumo son susceptibles de un proceso de mejora relevante.

Por un lado, podrían ampliarse las fuentes de información utilizadas incorporando, por ejemplo, registros obtenidos a lo largo de varios años para identificar dinámicas de cambio relacionadas con patrones migratorios. Así, la incorporación de datos censales que puedan cuantificar dichos patrones migratorios es una tarea pendiente.

Vinculado a este último punto, en el presente trabajo se utilizó el volumen de llamados de residentes de zonas no endémicas con el área endémica como un indicador indirecto de la existencia de movimientos y contactos de población entre ambos tipos de zonas. Esto implica un salto inferencial: se asume que la comunicación es un buen indicador de la existencia de flujos migratorios desde áreas endémicas. A su vez, un segundo supuesto es que parte de la población migrante está o estuvo infectada, al menos, con una probabilidad comparable a la de la población no migrante.

Si bien se trata de un supuesto razonable, también es esperable que los mismos CDRs permitan -a partir de nuevos procesamientos- estimar patrones y volúmenes migratorios efectivos. Es decir, a partir de la sucesión de conexiones de los usuarios a diferentes antenas podrían mapearse movimientos de población reales entre zonas endémicas y no endémicas.

Por otro lado, los datos provenientes de los CDRs presentan limitaciones que es necesario señalar. Quizás la más importante se vincule con el grado de cobertura de las antenas de telefonía. En efecto, esta variable altera la granularidad espacial del análisis. En zonas de baja densidad, la mayor distancia entre antenas hace que el área asignada a cada una sea más extensa, y por lo tanto mayor la dificultad para identificar el radio censal subyacente al que atribuir la actividad detectada. Por ello, que se recomienda como trabajo futuro realizar un análisis espacial de conglomerados con el fin de identificar áreas homogéneas de radios para mitigar dicho problema.

En cuanto a utilización de fuentes adicionales integradas en el Mapa (especialmente, el de Vulnerabilidad Sanitaria), una segunda iteración en la limpieza y consolidación de los datos georreferenciados de efectores de salud emerge como una tarea necesaria. Esta tarea implicaría la incorporación de fuentes que no fueron utilizadas en esta primera aproximación. A su vez, la clasificación según nivel de complejidad de la atención de los efectores de salud merece una revisión. Dichos efectores fueron clasificados en tres niveles en base a su nombre, o tipología

declarada en la fuente original, pero sin precisiones acerca de la capacidad real de cada uno. Es decir, no se determinó para cada instancia si el efector cuenta con los recursos necesarios para diagnosticar o atender enfermos de Chagas.

También la incorporación de nuevas fuentes de datos y dimensiones vinculadas al riesgo y vulnerabilidad ambiental resultan de potencial interés para el trabajo futuro.

El trabajo resultó en la obtención de resultado valioso de carácter general, que tiene que ver con la generación de una metodología de trabajo replicable y aplicable a otros ámbitos. La cuantificación de interacciones entre áreas geográficas dispersas, así como los indicadores de Vulnerabilidad Sanitaria, pueden ser considerados como dimensiones transversales que afectan la evolución y la transmisión de otras patologías además de la ECh. Por ello, una segunda línea de trabajo a futuro se vincula con la posibilidad de considerar mapas como el aquí descrito (y sus posteriores actualizaciones) como un insumo para el estudio de otras enfermedades infecciosas.

Referencias

- F. Aurenhammer. *Voronoi diagrams—a survey of a fundamental geometric data structure*. ACM Computing Surveys (1991)
- J. Candia, M. C. González, P. Wang, T. Schoenharl, G. Madey, y A.-L. Barabási. *Uncovering individual and collective human dynamics from mobile phone records*. *Journal of physics A: mathematical and theoretical*, 41(22):224015, 2008.
- R. Chuit, R. E. Gurtler, L. Mac Dougall, E. L. Segura, y B. Singer. *Chagas disease-risk assessment by an environmental approach in northern Argentina* (2001).
- R. Chuit, E. Subias, P. AC., I. Paulone, C. Wisnivesky-Colli, and E. Segura. *Usefulness of serology for the evaluation of trypanosome cruzi transmission in endemic areas of chagas's disease* (2003)
- Dirección de Estadísticas e Información en Salud, Ministerio de Salud de Argentina. *Estadísticas Vitales*. Disponible en <http://www.deis.msal.gov.ar/index.php/estadisticas-vitales/>
- E. Frias-Martinez, G. Williamson, y V. Frias-Martinez. *An agent-based model of epidemic spread using human mobility and social network information*. En *Privacy, Security, Risk and Trust (PASSAT) y 2011 IEEE Third International Conference on Social Computing (SocialCom)*, páginas 57–64. IEEE, 2011.
- G. Busso. *Migración interna y desarrollo territorial en argentina a inicios del siglo XXI. Brechas e impactos sociodemográficos de la migración interna interprovincial* (2007).
- INDEC. *Hábitat y vivienda por medio de datos censales. Calidad de los materiales de la vivienda (CALMAT)* (2003).
- Monasterio, Salles, Lang, Weinberg, Minnoni, Travizano, Sarraute. (2016). *Analyzing the Spread of Chagas Disease with Mobile Phone Data*. 607-612. 10.1109/ASONAM.2016.7752298. Disponible en: https://www.researchgate.net/publication/310809125_Analyzing_the_Spread_of_Chagas_Disease_with_Mobile_Phone_Data. Último acceso: 03/02/2020
- Monasterio, Salles, Lang, Weinberg, Minnoni, Travizano, Sarraute. (2016). *Uncovering the Spread of an Infectious Disease with Mobile Phone Data*. Disponible en https://www.researchgate.net/publication/319730478_Uncovering_the_Spread_of_an_Infectious_Disease_with_Mobile_Phone_Data. Último acceso: 03/02/2020
- M. Peyravian and N. Zunic. *Methods for protecting password transmission*. *Computers & Security* (2000).
- Sarraute, Lang, Monasterio, Weinberg. (2015) *Descubriendo Chagas con big data*.
- J. Steenbruggen, E. Tranos, y P. Nijkamp. *Data from mobile phone operators: A tool for smarter cities?* *Telecommunications Policy* (2015).

Anexo A Acceso a los datos

Para acceder a información adicional puede recurrirse al sitio en internet dedicado al proyecto, Big Data para la detección de Chagas en zonas no endémicas, accesible en <https://www.fundacionbyb.org/mapa-chagas-argentina>.

El sitio lista los congresos y eventos nacionales e internacionales donde el trabajo fue presentado ante el público:

Media Party 2018
ARGENTINA / 22 de agosto de 2018

Reunión Internacional de Inteligencia Artificial y sus Aplicaciones (RIIAA) 2018
MÉXICO / 25 de agosto de 2018

3° Reunión de Proyectos e Investigación de Mundo Sano
ARGENTINA / 12 de septiembre de 2018

3° Digital Summit
ARGENTINA / 5 de octubre de 2018

Cluster AI (UTN)
ARGENTINA / 17 de noviembre de 2018

The Web Conference / DSSGW2019 / Data Science for Social Good
SAN FRANCISCO, USA / 14 de mayo de 2019

Data for Policy - 4th International Conference
LONDRES, UK / 12 de junio de 2019

Encuentro Nacional de Chagas
ARGENTINA / 13 de junio de 2019

XIX Simposio Internacional sobre Enfermedades Desatendidas
ARGENTINA / 22 de agosto de 2019

Además, permite descargar el dataset **Índice de Prevalencia Potencial de Enfermedad de Chagas (por localidad)**. Contiene los componentes del IPPEC de cada radio censal no perteneciente a la Ecorregión del Gran Chaco, agregado por localidad, para las variables de Afinidad para con la zona endémica y Vulnerabilidad Sanitaria.

Anexo B

Tabla extendida. Tres localidades fuera del área endémica de Chagas con mayor Prevalencia Potencial, por provincia.

Provincia	Departamento	Localidad	Tipo	Población	Afinidad	Vulnerabilidad Sanitaria	Usuarios con afinidad
San Luis	Lib. San Martín	Las Aguadas	Ambos	487	81.4%	96.6%	94.9%
Córdoba	San Javier	Quebrada De Los Pozos	Ambos	398	78.9%	76.7%	93.8%
Misiones	Candelaria	Candelaria	Ambos	13871	78.2%	65.9%	78.5%
Misiones	Candelaria	Cerro Corá	Ambos	684	78.2%	51.9%	78.5%
Misiones	Candelaria	Profundidad	Ambos	417	78.2%	82.5%	78.5%
Córdoba	San Javier	La Travesía	Ambos	431	77.1%	90.7%	90.9%
Córdoba	San Justo	Arroyito	Ambos	22147	75.8%	44.4%	96.6%
Corrientes	Ituzaingó	Villa Olivari	Ambos	1276	72.2%	88.3%	94.2%
San Luis	Lib. San Martín	San Martín	Ambos	697	70.7%	35.8%	74.7%
Catamarca	Antofagasta	Antofagasta	Ambos	730	70.7%	75.9%	91.4%
San Luis	Ayacucho	San Francisco Del Monte De	Ambos	3732	69.5%	47.9%	83.9%
Jujuy	Cochinoca	Abdón Castro Tolay	Ambos	376	66.2%	86.6%	88.0%
Jujuy	Susques	Coranzuli	Ambos	520	66.2%	96.4%	88.0%
Jujuy	Rinconada	Nuevo Pirquitas	Ambos	672	66.2%	86.3%	88.0%
Salta	Santa Victoria	Nazareno	Ambos	916	63.9%	93.8%	97.8%
Salta	Santa Victoria	Poscaya	Ambos	470	63.9%	98.7%	97.8%
Salta	Orán	Aguas Blancas	Ambos	2395	61.6%	68.6%	96.8%
Catamarca	Belén	El Durazno	Ambos	785	60.7%	96.6%	98.6%
Catamarca	Belén	Jacipunco	Ambos	785	60.7%	96.6%	98.6%
Santa Fe	La Capital	Candioti	Ambos	888	55.7%	72.9%	74.2%
Santa Fe	La Capital	Recreo	Ambos	13336	46.7%	67.3%	28.2%
Entre Ríos	La Paz	La Paz	Ambos	24307	38.5%	60.7%	42.2%
Santa Fe	San Justo	Ramayón	Alta Afinidad Media	550	36.8%	90.3%	38.5%
Corrientes	Esquina	Pueblo Libertador	Alta Afinidad Media	932	32.7%	97.0%	36.0%
Santiago del Estero	Rivadavia	Selva	Alta Afinidad Media	2878	27.5%	75.0%	33.5%

San Juan	San Martín	Villa Dominguito	Alta Afinidad Media	619	24.4%	88.2%	89.0%
San Juan	San Martín	Villa Don Bosco	Alta Afinidad Media	2728	24.3%	72.9%	89.0%
San Juan	San Martín	Dos Acequias	Alta Afinidad Media	972	23.8%	78.9%	87.8%
Corrientes	Mercedes	Felipe Yofre	Alta Afinidad Media	1435	23.7%	85.7%	31.2%
Entre Ríos	La Paz	San Gustavo	Alta Afinidad Media	1325	22.3%	88.8%	23.0%
Santa Cruz	Lago Argentino	El Chaltén	Alta Afinidad Media	1627	20.6%	89.0%	25.7%
Mendoza	Luján de Cuyo	Cacheuta	Alta Afinidad Media	701	20.1%	93.5%	25.6%
Entre Ríos	Diamante	General Ramírez	Alta Afinidad Media	8868	19.7%	32.5%	20.4%
Mendoza	Luján de Cuyo	Potrerosillos	Alta Afinidad Media	653	19.4%	82.9%	23.7%
Santa Cruz	Güer Aike	28 De Noviembre	Alta Afinidad Media	6145	18.9%	56.4%	25.5%
Santa Cruz	Güer Aike	Rospentek	Alta Afinidad Media	677	17.5%	59.2%	24.3%
Mendoza	La Paz	Desaguadero	Alta Afinidad Media	641	17.3%	96.9%	19.5%
Santiago del Estero	Rivadavia	Colonia Alpina	Alta Afinidad Media	545	15.7%	84.5%	20.3%
Buenos Aires	Zárate	Lima	Alta Afinidad Media	10219	14.7%	75.6%	19.1%
Río Negro	Pilnaniyeu	Comallo	Alta Afinidad Media	1497	14.3%	45.1%	17.6%
Río Negro	Pilnaniyeu	Ca adón Chileno	Alta Afinidad Media	400	14.3%	92.9%	17.6%
Río Negro	Pilnaniyeu	Laguna Blanca	Alta Afinidad Media	400	14.3%	92.9%	17.6%
Chubut	Río Senguer	Río Mayo	Alta Afinidad Media	2791	13.7%	83.5%	18.4%

Chubut	Río Senguer	Alto Río Senguer	Alta Afinidad Media	1570	13.6%	89.2%	18.2%
Tierra del Fuego	Río Grande	Río Grande	Alta Afinidad Media	66475	12.5%	36.8%	16.5%
La Pampa	Curacó	Gobernador Duval	Alta Afinidad Media	413	12.2%	95.1%	17.2%
Tierra del Fuego	Ushuaia	Ushuaia	Alta Afinidad Media	56593	11.7%	33.4%	15.0%
Buenos Aires	Zárate	Country Club El Casco	Alta Afinidad Media	532	11.1%	75.7%	14.2%
Buenos Aires	Olavarría	Villa Alfredo Fortabat	Alta Afinidad Media	3673	10.7%	60.1%	3.7%
Tierra del Fuego	Río Grande	Tolhuin	Alta Afinidad Media	2949	9.9%	85.3%	12.7%
Neuquén	A elo	San Patricio Del Cha a	Alta Afinidad Media	6474	9.4%	35.0%	14.4%
CABA	Comuna 01	Comuna 1	Alta Afinidad Media	205886	8.6%	25.9%	9.2%
CABA	Comuna 03	Comuna 3	Ambos	187537	8.5%	15.8%	9.8%
Buenos Aires	La Plata	La Plata	Radio Extremo	645569	7.7%	33.0%	4.9%
CABA	Comuna 02	Comuna 2	Alta Afinidad Media	157932	7.5%	6.4%	6.2%
Neuquén	Confluencia	El Chocon	Alta Afinidad Media	467	6.7%	75.9%	8.4%
Neuquén	Confluencia	Villa El Chocón	Alta Afinidad Media	707	6.7%	69.7%	8.4%
Chubut	Escalante	Comodoro Rivadavia	Alta Afinidad Media	175196	5.3%	48.9%	8.1%
La Pampa	Chalileo	Santa Isabel	Alta Afinidad Media	2697	4.7%	79.1%	5.8%
La Pampa	Loventué	Telén	Alta Afinidad Media	1240	4.3%	64.6%	5.7%



FUNDACIÓN
BUNGE Y BORN

25 de Mayo 501, 6° Piso (C1002ABK)
Ciudad de Buenos Aires, Argentina
www.fundacionbyb.org

