

Coronasurveys: encuestas indirectas en línea para monitorizar la evolución del COVID-19

Coronasurveys: indirect online surveys to monitor the evolution of COVID-19

Antonio Fernández-Anta^a, Jose Aguilar^{a,c,d}, Juan M. Ramírez^a, Rosa Elvira Lillo^b, Sergio Díaz-Aranda^a

^a IMDEA Networks Institute, España

^b Departamento de Estadística, Universidad Carlos III de Madrid, España

^c Centro de Estudios en Microelectrónica y Sistemas Distribuidos, Universidad de Los Andes, Venezuela

^d Grupo I+D+I en Tecnologías de la Información y las Comunicaciones, Universidad EAFIT, Colombia

Resumen

Introducción: Seguir la evolución del COVID-19 ha sido esencial para la toma de decisiones sanitarias. Pero ello requiere de cifras fiables de los infectados, muertos y hospitalizados, muchas veces complicadas de tener por múltiples razones. El uso de técnicas de estimación rápidas y fiables, como las encuestas indirectas, es una opción. **Objetivos:** Presentar el proyecto *CoronaSurveys*, el cual monitorea el COVID-19 combinando encuestas indirectas con el método de ampliación de red (*Network Scale-up Method*, NSUM). **Metodología:** El sistema usa encuestas anónimas indirectas en línea para consultar sobre el COVID-19, y el método NSUM para estimar los casos. Las encuestas están en múltiples idiomas con preguntas para rastrear los casos activos, nuevos y de muertes, entre otros. **Resultados:** *CoronaSurveys* está operando desde marzo de 2020, y sigue recopilando datos, con más de cien mil respuestas en la actualidad (millones de muestras). El sistema ha hecho buenas estimaciones para España y el Reino Unido, entre otros países. **Conclusión:** Este sistema es adecuado en países con infraestructura sanitaria limitada o en situaciones de desconocimiento de la pandemia, como al inicio del COVID-19, porque su coste de implementación es pequeño, requiere dispositivos simples para usarlo, y de pocos participantes para obtener buenas estimaciones.

Palabras clave: COVID-19; encuestas indirectas; métodos NSUM; inferencia estadística.

Abstract

Introduction: Monitoring the evolution of COVID-19 has been essential for health decision-making. But this requires reliable numbers of the infected, dead and hospitalized, often difficult to have for multiple reasons. The use of fast and reliable estimation techniques, such as indirect surveys, is an option. **Objectives:** To present the *CoronaSurveys* project, which monitors COVID-19 by combining indirect surveys with the *Network Scale-up Method* (NSUM). **Methodology:** The system uses indirect online anonymous surveys to inquire about COVID-19, and the NSUM method to estimate cases. The system offers multi-language surveys with questions to track active cases, new cases, and fatalities, among others. **Results:** *CoronaSurveys* has been operating since March 2020, and continues to collect data, with over one hundred responses to date (over one million samples). The system has made good estimates for Spain and the United Kingdom, among other countries. **Conclusion:** This system is suitable in countries with limited health infrastructure or in situations of lack of knowledge about the pandemic, such as at the beginning of COVID-19, because its implementation cost is small and it requires simple devices to use it and few participants to obtain good estimates.

Keywords: COVID-19; indirect surveys; NSUM methods; Statistical inference.

Introducción

Monitorizar las cifras que caracterizan la evolución del COVID-19 durante la pandemia fue esencial para la toma de decisiones estratégicas de muchos países, tales como cuándo confinar a la población, cerrar colegios, etc. Normalmente, en las primeras etapas de una pandemia hay poca capacidad para determinar los casos reales, porque no se tienen los recursos y la logística necesarios para realizar pruebas, por lo que los datos oficiales de las autoridades sanitarias pueden estar muy lejos del número real de casos (Maxmen, 2020; Ruppert et al., 2018). Esto es aún más real en los países de ingresos bajos y medianos. No tener acceso a datos fiables claramente afecta a la calidad de las decisiones tomadas por las autoridades, poniendo a la población en mayor riesgo. Incluso en etapas posteriores de la pandemia, puede seguir ocurriendo que los datos oficiales estén muy lejos de los reales por diferentes motivos, como descuido de las autoridades o pasividad de la población, entre otras razones (Nishiura et al., 2009).

En ese sentido, se requiere de otras estrategias y/o técnicas, más allá de las pruebas de laboratorio, que puedan estimar el número de casos y su evolución. Por ejemplo, técnicas que recopilen masivamente y fácilmente datos para hacer estimaciones razonablemente precisas, tales como las encuestas directas. Varias encuestas directas

para obtener datos de salud sobre el COVID-19 se han implementado, tales como las reportadas en los trabajos de Facebook y Oliver et al., 2020. Ahora bien, las encuestas directas necesitan una gran cantidad de participantes para lograr estimaciones fiables, además de que recopilan información de salud personal confidencial, lo que puede disuadir de responder a personas preocupadas por su privacidad.

Una alternativa a las encuestas directas son las indirectas, donde las preguntas que responde un participante no son sobre sí mismo, sino sobre sus contactos. Este enfoque tiene al menos tres ventajas importantes con respecto a las encuestas directas. Primero, la encuesta se puede diseñar de modo que no se recopile información personal del participante para preservar la privacidad de los participantes (es decir, es completamente anónima). Segundo, tiene un efecto multiplicador, ya que reduce el número de respuestas necesarias para lograr una cobertura de población específica, es decir, alcanzar una cobertura amplia en un período de tiempo corto. Tercero, permite

obtener estimaciones que converjan más rápido al verdadero valor del parámetro a estimar usando las técnicas que se conocen en la literatura como el método de ampliación de red (Network Scale-up Method, NSUM) (Bernard et al., 2010; Laga et al., 2021), que permiten realizar estimaciones con las respuestas recopiladas. Así, en el caso del COVID-19, las respuestas individuales actúan como conocimiento instantáneo de la situación actual de la pandemia desde un punto de vista personal. Cuando estas respuestas se analizan colectivamente, a través del tiempo y ubicaciones geográficas, se puede inferir una visión combinada de la pandemia. De esta manera, la pérdida de precisión debido a que los encuestados no siempre tienen información exacta sobre el estado de salud de los demás, puede compensarse con el aumento significativo de la cobertura (es decir, el número que cada encuestado reporta) inspirado en el “paradigma de la multitud” o la “sabiduría de los grupos”, por lo que ha sido utilizado con éxito en salud pública (Nishiura et al., 2009).

Particularmente, en el proyecto CoronaSurveys (<https://coronasurveys.org/>) se ha creado un sistema para estimar la cantidad de casos de COVID-19 en base a encuestas indirectas abiertas. El sistema usa el método NSUM a partir de encuestas anónimas indirectas en línea, para realizar las estimaciones (The CoronaSurveys research team, 2020 <https://coronasurveys.org/>). El sistema CoronaSurveys ha estado operando desde marzo de 2020, comenzando en sólo tres países inicialmente (España, Portugal y Chipre) y desplegando en unas semanas encuestas para todos los países del mundo. Este sistema cuenta con encuestas en 60 idiomas y permite reportar datos sobre la incidencia del COVID-19 en todos los países. El sistema sigue recopilando datos y ha recopilado más de cien mil de respuestas hasta el momento, reportando el estado de millones de personas. Todos los datos recopilados están disponibles para su uso abierto. Hasta donde sabemos, este es el sistema basado en NSUM de mayor escala jamás implementado y el único que ha recopilado datos continuamente durante un período de más de tres años mediante encuestas abiertas.

En este documento se comentan los aspectos más relevantes del proyecto CoronaSurveys. Para ello, se presenta antes que nada la metodología que está detrás del proyecto, posteriormente se analizan las estimaciones resultantes obtenidas, para culminar con una discusión sobre los resultados.

En el proyecto CoronaSurveys (<https://coronasurveys.org/>) se ha creado un sistema para estimar la cantidad de casos de COVID-19 en base a encuestas indirectas abiertas

Metodología

El proyecto Coronasurveys ha desplegado encuestas indirectas en línea que se pueden contestar a través de un navegador web o una aplicación móvil. La encuesta consulta a los encuestados sobre la evolución del COVID 19, en particular, la encuesta original tenía sólo dos preguntas básicas: 1) ¿Cuántas personas hay en su entorno cuyo estado de salud conoce? y 2) ¿Cuántas de ellas fueron diagnosticadas o tuvieron síntomas del COVID-19?

La primera pregunta determinaba el alcance del participante (RI) y la segunda el número de casos conocidos (CI). Los datos se han venido agregando y mezclando diariamente, para hacer un pre-procesamiento para eliminar los valores atípicos, basado en los siguientes criterios: Si RI está fuera de 1,5 veces el rango inter-cuartil sobre el cuartil superior; Si CI/RI es $> 1/3$ (para excluir a los participantes con contacto de casos excepcionalmente alto).

El cuestionario de CoronaSurveys se amplió con preguntas adicionales para mejorar su granularidad y estimar más parámetros de la pandemia (como fallecimientos), sobre las personas con síntomas. Algunas de estas preguntas son: 3) ¿Cuántas siguen enfermas?; 4) ¿Cuántas comenzaron con síntomas en los últimos 7 días?; 5) ¿Cuántas murieron?

Con estas preguntas adicionales, es posible rastrear los casos activos (Pregunta 3), los casos nuevos (Pregunta 4), y el número acumulado de muertes (Pregunta 5).

Desde el principio del proyecto, el mayor desafío ha sido conseguir participantes que rellenen la encuesta. Inicialmente se diseminó usando redes sociales en línea (Twitter, Facebook, WhatsApp, etc.), lo cual fue efectivo en los primeros meses de la pandemia. Más tarde, se usaron campañas de anuncios, principalmente en Facebook, en España y Brasil, con relativamente baja efectividad. Actualmente, desde julio de 2022 se ha utilizado la plataforma Prolific (<https://www.prolific.co/>), que cuenta con potenciales encuestados en diferentes países, a los que se les paga por completar encuestas. Los encuestados con la plataforma Prolific en varios países (inicialmente 11 países y ahora solamente España, Reino Unido y EEUU) rellenan una encuesta indirecta sobre lo que saben en su contexto sobre la evolución del COVID-19.

La parte superior de la Figura 1 muestra los países de los que se ha recolectado más información en el 2022, (con el mayor número de registros), con las cantidades recolectadas, debido principalmente a que, desde el proyecto, se han optimizado los fondos para obtener datos en países estratégicos. En concreto, España es el país con mayor participación, y casi duplica al siguiente país (Reino Unido). Por otro lado, la parte inferior de la Figura 1 muestra la participación por país en los últimos seis meses del 2022. Varios aspectos a destacar son los siguientes: i) prácticamente toda la participación de algunos de los países se concentra en esos 6 meses (por ejemplo, Reino Unido, Sudáfrica y Estados Unidos), ii) hay países en los que ha cesado la recogida (por ejemplo, Ucrania y Chipre) y, iii) España ya no es el país del que se recopila la mayor parte de la información, siendo este el Reino Unido.

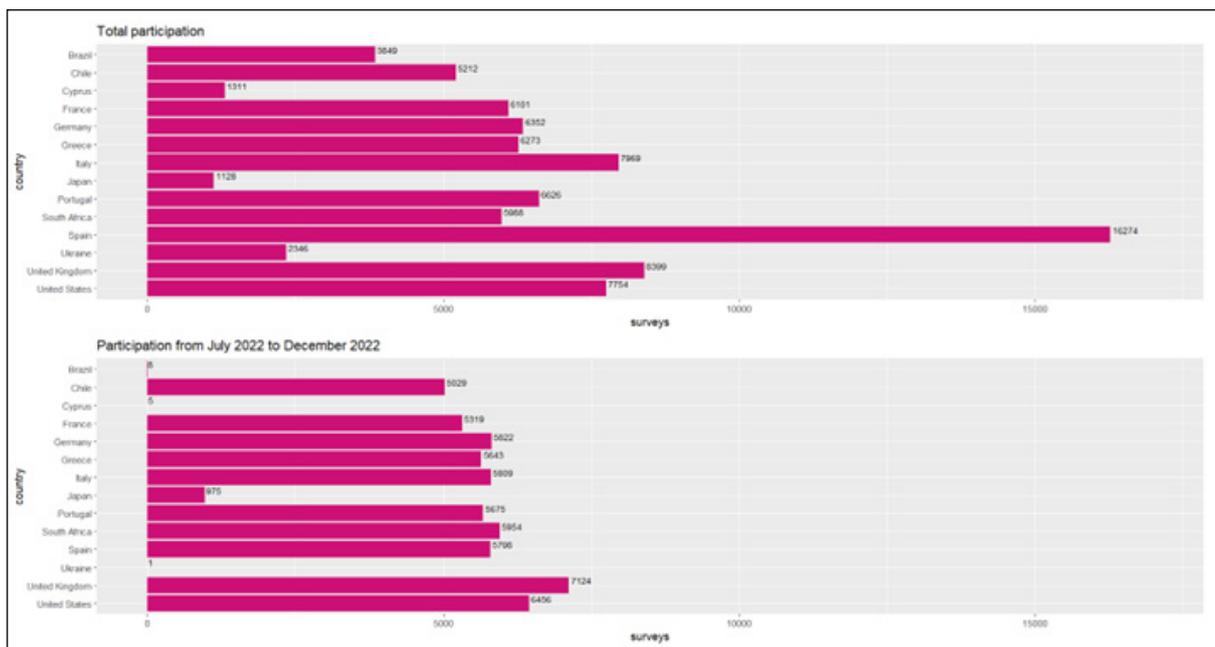


Figura 1. Participación por país.

Finalmente, con el conjunto de datos agregados por región, se puede realizar una estimación sólida de la prevalencia del COVID-19. Así, para una región i , obtenemos n_i respuestas, y cada respuesta j tiene una variable de alcance (RI_j), que es el número de personas (red) cuyo participante j conoce (primera pregunta de la encuesta); y una variable de conteo (CI_j) que es el número de personas conocidas por el encuestado con síntomas compatibles con COVID-19 (segunda pregunta de la encuesta). Entonces, una forma de estimar la proporción de personas infectadas con síntomas en la región i (I_i) es de la siguiente manera:

$$I_i = \frac{\sum_1^{n_i} CI_i}{\sum_1^{n_i} RI_i}$$

Ha de dejarse claro que esta no es la única manera de hacer las estimaciones. El conjunto de métodos de estimación de este tipo son conocidos como métodos NSUM, un tema de mucha relevancia en la comunidad científica actualmente (Bernard et al., 2010; Laga et al., 2021). Básicamente, estos métodos surgen para poder tener información sobre fenómenos que son considerados socialmente tabús como por ejemplo, VIH, muertes en terremotos,...., donde los encuestados se sienten más cómodos si no responden sobre sí mismos.

Resultados

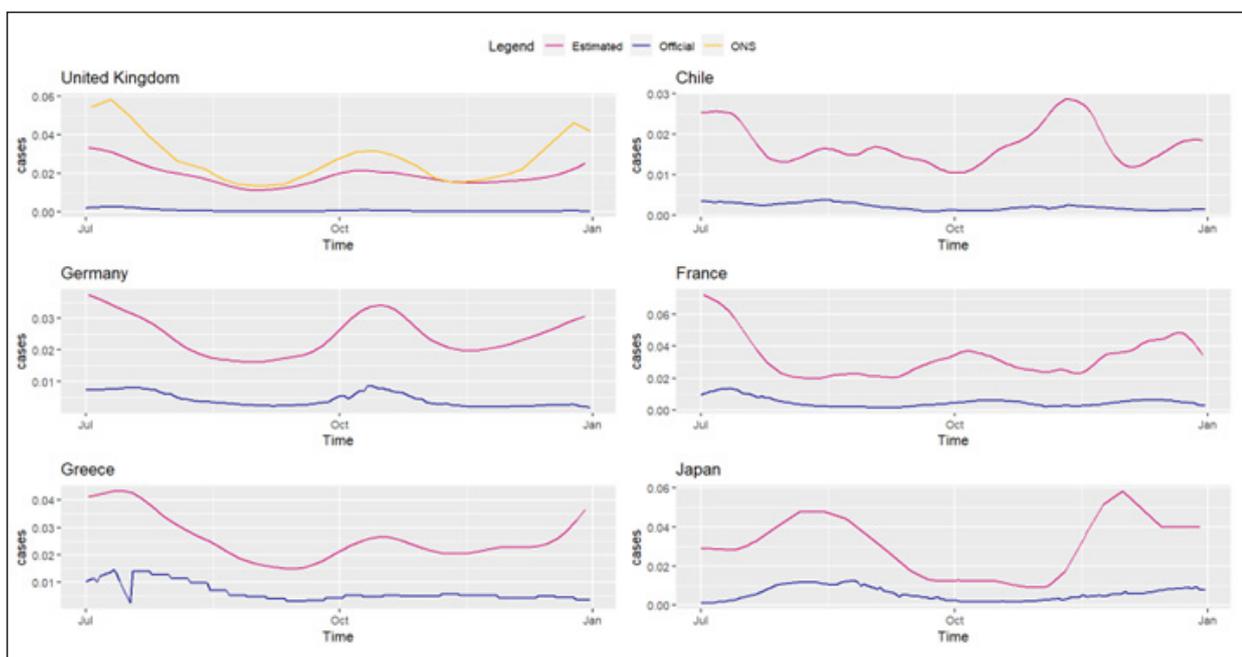
La Tabla 1 muestra un ejemplo del total de datos recopilados por país, donde la primera columna representa al país. Específicamente, se muestran tres variables: alcance, muertos y hospitalizaciones. Para algunos valores, las desviaciones típicas son muy grandes (por ejemplo, los valores de hospitalización para casi todos los países). Una desviación típica grande significa que los datos están muy dispersos con respecto a la media. La primera columna de la Tabla 1 describe el número de contactos (encuesta indirecta) declarados por el encuestado. Como se puede observar, se obtiene una media aproximada de 25 contactos por cada encuestado, que es un número muy por debajo del que Dunbar estableció como límite cognitivo de relaciones sociales estables que un ser humano puede mantener, el cual, según Dunbar, sería 150 (Dunbar, 1992). Dunbar (1992) propuso que en el mundo de las relaciones sociales existe una jerarquía de círculos definidos a partir de la intensidad y la frecuencia de las relaciones, tal que si el entramado social está bien estructurado, en cada uno de los círculos sucesivos se incluye más del doble de gente que en el anterior. En términos generales, esos círculos serían progresivamente 5, 15, 50, 150, 500 y 1500 individuos. Así, el círculo más pequeño estaría compuesto por cinco personas (nuestros seres queridos), después por 15 amigos íntimos, 50 amistades, 150 contactos importantes. 500 conocidos, y 1500 personas que quizás conozcas. De esta manera, 25 está entre el tamaño de amigos íntimos y el de amistades. Con esta media de contactos en CoronaSurveys y más de 100.000 respuestas recopiladas, el proyecto ha tenido hasta ahora un alcance de varios millones de personas en total.

Usando los conjuntos de datos recopilados en el proyecto, se han realizado varias pruebas para estimar la incidencia de COVID-19 en trabajos anteriores. Por ejemplo, el trabajo de Baquero et al., 2021 muestra la evolución del número de casos activos, casos nuevos diarios, y casos contagiosos, estimados con algunos de los métodos del proyecto Coronasurveys, utilizando datos de Portugal. Un segundo trabajo evaluó la calidad del conjunto de datos, analizando la precisión de varios métodos para estimar el número acumulado de casos de COVID-19 (García-Agundez et al., 2021). Se compararon las estimaciones de la encuesta con los resultados de un estudio serológico reportado en Pollán et al., 2020 para España, realizado entre el 27 de abril y el 11 de mayo de 2020. Particularmente, el coeficiente de determinación (R^2) de las estimaciones en todas las regiones (comunidades autónomas) de España en el periodo de tiempo del estudio serológico arrojó 0,89, resultado bastante bueno. Una conclusión interesante de este trabajo es que un valor de 50 respuestas por región permitió realizar una estimación razonable de casos, muy alineada con los datos oficiales, los cuales tienen un coste mucho mayor para su obtención. Incluir más respuestas podría aumentar la precisión, pero los números se mantienen razonablemente estables.

Tabla 1. Algunas variables del conjunto de datos.

País	Media de Alcance	Desviación típica de alcance	Media de Infectados	Desviación típica de Infectados	Media de Hospitaliz.	Desviación típica de Hospitaliz.
Chile	25.32	4.11	0.31	1.01	2.29	9.67
Francia	14.48	2.59	0.39	1.05	3.84	15.21
Alemania	24.47	3.83	0.52	1.08	15.34	27.47
Grecia	29.33	4.74	0.63	1.44	0.95	1.47
Italia	31.28	4.90	0.61	1.28	1.03	3.85
Japón	14.58	2.82	0.47	1.10	0.28	1.12
Portugal	32.80	5.05	0.24	0.98	22.13	24.12
Suráfrica	33.36	5.27	0.51	1.34	82.01	46.33
España	29.97	4.85	0.35	1.22	0.78	1.29
Inglaterra	29.95	4.84	0.56	1.46	8.46	15.31
Estados Unidos	27.82	4.96	0.38	1.13	7.61	20.28

Por otro lado, la Figura 2 muestra las estimaciones de casos usando el conjunto de datos de CoronaSurveys del último semestre del 2022 para alguno de los países, comparado con los números oficiales (Ojo et al., 2020). En esa base de datos, Japón tiene pocos datos recopilados, y Reino Unido muchos datos recopilados. En general, se puede ver que las estimaciones siguen bastante bien el comportamiento de los datos oficiales en algunos países (como Alemania). En el Reino Unido, también están muy bien alineados con las estimaciones de la Oficina de Estadísticas Nacional (ONS por sus siglas en inglés Office for National Statistics, <https://www.ons.gov.uk/>). El país con mayor diferencia en cuanto a las estimaciones realizadas es Chile.

**Figura 2.** Estimación de casos de COVID.

Finalmente, recientemente se han comparado las estimaciones con los datos oficiales en el Reino Unido y Australia (ver Tabla 2). Las estimaciones de vacunación están muy cerca de los valores oficiales. Las tasas de vacunación en Australia y Reino Unido se han estimado en 76,50 % (73,70 % - 79,29 %) y 78,86 % (intervalo de confianza (IC) del 95 %: 77,00 % - 80,72 %), mientras que los valores oficiales (OWID) son 84,95 % y 79,71 %, respectivamente. En el caso de la mortalidad y hospitalizaciones en el último mes, los valores oficiales se encuentran dentro del IC de las estimaciones para el caso de Australia. En concreto, la tasa de mortalidad es del 0,34% (0,00% - 0,22%) y la oficial es del 0,006%, y la tasa de hospitalización es del 1,02% (0,36% - 1,68%) y la oficial es del 1,327%. Además, en el caso del Reino Unido, los valores oficiales de ONS están dentro del IC de las estimaciones del número de casos, casos en los últimos 7 días y casos en las últimas 24 horas. Para el resto de variables, las diferencias nunca son grandes en los casos en los que existen diferencias entre los valores oficiales y las estimaciones (posiblemente por subregistro en los datos oficiales).

Tabla 2. Métricas de incidencia de COVID-19 en % (95% de IC) obtenidas desde encuestas indirectas e informes oficiales para Australia y el Reino Unido. (1) Personas de 12 años o más que hayan recibido al menos una/dos/tres dosis al 31 de agosto de 2022. (2) Sólo datos de Inglaterra, 5 semanas.

	Australia		Reino Unido		
	Encuestas Indirectas	OWID	Encuestas Indirectas	OWID	ONS
Casos (en el mes pasado)	12.43 (10.26 - 14.60)	1.585	8.67 (7.39 - 9.96)	0.305	9.663
Tasa Vacunación	76.50 (73.70 - 79.29)	84.95	78.86 (77.00 - 80.72)	79.71	93.6/88.2/70.2 (1)
Mortalidad (en el mes pasado)	0.34 (0.00 - 0.72)	0.006	0.43 (0.13 - 0.73)	0.004	0.005 (2)
Hospitalizaciones (en el mes pasado)	1.02 (0.36 - 1.68)	1.327	0.81 (0.40 - 1.22)	0.158	0.044 (2)
Casos (últimas 24 horas)	2.03 (1.10 - 2.96)	0.069	1.30 (0.78 - 1.82)	0.023	1.458
Casos (últimos 7 días)	2.71 (1.64 - 3.78)	0.211	1.30 (0.78 - 1.82)	0.073	1.116

Otro tema interesante que se ha abordado desde CoronaSurveys es la posibilidad de acceder a datos fiables en momento de repuntes de la pandemia como fue el caso de China donde los datos oficiales no ofrecían garantías de ser fiables (Ramírez et al., 2023).

En la web del proyecto <https://coronasurveys.org/> se puede tener acceso a toda la información, documentación y resultados que se han generado desde su existencia en el 2020. Son particularmente interesantes los gráficos que continuamente se han ido generando con las respuestas recibidas, y los trabajos que se han publicado relacionados tanto con la evolución de la pandemia como con los métodos empleados o la eficacia de las vacunas. Por último, resaltar el enfoque internacional del proyecto, y el equipo de investigadores que han hecho y están haciendo posible la innovación en la forma de recoger y analizar datos en momentos críticos como ha sido la pandemia del COVID-19.

Discusión

Durante la pandemia, se ha evidenciado la necesidad de tener datos fiables de los casos confirmados y muertes para determinar su evolución. Pero también se ha podido ver que obtenerlos a veces ha sido muy complicado. En particular, cuando no se tienen todos los recursos para establecer los casos reales de infección. Así, se requieren de enfoques de estimación alternativos, como el que se ha usado en el proyecto Coronasurys para monitorizar y estimar los posibles casos de personas infectadas con COVID-19. Esto es aún más relevante en países con

un mal sistema de salud. Particularmente, las encuestas anónimas indirectas abiertas proporcionan estimaciones relativamente cercanas a los valores reales, sin embargo, los métodos de estimación a partir de ellas siguen siendo un tema de investigación para garantizar estimaciones más precisas.

Si bien aquí solo se presenta uno de los métodos de estimación de NSUM sobre la cantidad de personas infectadas con COVID-19, el equipo de investigación ha seguido probando y diseñando nuevos métodos de estimación, aplicándolos a nuevas encuestas en otros ámbitos. El grupo está explorando otros enfoques, algunos basados en la teoría de grafos, otros usando técnicas de aprendizaje automático, así como agregando preguntas de control a las que interesan para validar los datos.

En el grupo de investigación se es consciente de las debilidades y sesgos de las estimaciones usando encuestas indirectas, ya que normalmente los métodos NSUM se basan en tres condiciones: todos en la población tienen la misma probabilidad de conocer a alguien en una subpoblación, para cada persona en la red de un encuestado, el encuestado conoce cada subpoblación a la que pertenece la persona, y los encuestados pueden recordar completamente a todos en su red social en el momento de la encuesta.

Estos supuestos se cumplen raramente en la práctica de manera precisa, produciéndose en general sesgos de 4 tipos:

- Error de transmisión, que es producido por el hecho de que el participante no conoce el estado de todos sus conocidos, es decir, no saben todo acerca de las personas en sus redes sociales, violando la condición 2;
- Error de recuerdo, que ocurre cuando los participantes recuerdan incorrectamente el número de personas en algunas subpoblaciones, violando la condición 3. El encuestado puede estar sobreestimando o subestimando;
- Efecto barrera, es el sesgo que se produce porque la probabilidad de saber algo de una subpoblación depende del individuo. Esto significa que es más probable que un participante conozca más de ciertas subpoblaciones que de otras, violando la condición 1.
- Finalmente, sesgo de respuesta, que es un error que se produce por una declaración deliberadamente errónea de los participantes.

Todo lo anterior puede conducir a una estimación inexacta. Propuestas para manejar estos sesgos son un tema interesante de investigación de interés para nuestro grupo.

Finalmente, creemos que un sistema de encuestas como el proyecto CoronaSurveys es especialmente adecuado en países de bajo ingreso, ya que el costo de preparar e implementar una encuesta es extremadamente pequeño, los participantes pueden usar dispositivos muy simples para completar la encuesta (ya que está basada en la web), y el número de los participantes que deben tener información sobre el tema de estudio (por ejemplo, de una pandemia) es bastante bajo. También ha servido para homogeneizar conceptos a nivel internacional cuando la definición de caso era diferente por país, e incluso comunidad autónoma. Particularmente concluimos que las encuestas indirectas abiertas anónimas, en combinación con el método NSUM, ofrecen una opción económica y flexible para monitorear epidemias.

Concluimos que las encuestas indirectas abiertas anónimas, en combinación con el método NSUM, ofrecen una opción económica y flexible para monitorear epidemias

Contribuciones de los autores

Los autores participaron igualmente en la elaboración del manuscrito y aprobaron la versión final presentada.

Financiación

Este trabajo fue parcialmente financiado por la subvención CoronaSurveys-CM, financiada por IMDEA Networks y Comunidad de Madrid, España, subvenciones COMODIN-CM y PredCov-CM financiadas por la Comunidad de Madrid y la Unión Europea a través del Fondo Europeo de Desarrollo Regional (FEDER), subvenciones TED2021-131264B-I00 (SocialProbing) y PID2019-104901RB-I00 financiadas por el Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación, España (MCIN/AEI/10.13039/501100011033) y la Unión Europea (NextGenerationEU/PRTR), y donaciones individuales al Proyecto CoronaSurveys <https://coronasurveys.org>.

Declaración de disponibilidad de datos

Los conjuntos de datos presentados en este estudio se pueden encontrar en línea en un repositorio abierto: <https://github.com/GCGImdea/coronasurveys/data/aggregate/>.

Conflictos de interés

Los autores declaran que no hay conflicto de interés.

Referencias bibliográficas

- Baquero, C., Casari, P., Fernandez Anta, A., García-García, A., Frey, D., Garcia-Agundez, A., ... Sanchez, I. (2021) The CoronaSurveys System for COVID-19 Incidence Data Collection and Processing, *Frontiers in Computer Science.*, <http://dx.doi.org/10.3389/fcomp.2021.641237>
- Bernard, H. R., Hallett, T., Iovita, A., Johnsen, E. C., Lyerla, R., McCarty, C., et al. (2010). Counting Hard-To-Count Populations: the Network Scale-Up Method for Public Health. *Sex. Transm. infections* 86 (Suppl. 2), ii11–ii15. <http://dx.doi.org/10.1136/sti.2010.044446>
- Dunbar, R. (1992). Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* 22, 469–493, [https://doi.org/10.1016/2560047-2484\(92\)90081-J](https://doi.org/10.1016/2560047-2484(92)90081-J).
- Facebook Data for Good. COVID-19 symptom survey {request for data access. <https://dataforgood.fb.com/docs/covid-19-symptom-survey-request-for-data-access/>. Accessed: 2020-05-20.
- García-Agundez A., Ojo O., Hernandez H., Baquero C., Frey D., Georgiou C., ... Fernandez Anta A. (2021). Estimating the COVID-19 Prevalence in Spain with Indirect Reporting via Open Surveys. *Frontiers in Public Health*, 9. <http://dx.doi.org/10.3389/fpubh.2021.658544>
- Laga I, Bao L., & Niu X. (2021) Thirty Years of The Network Scale-up Method, *Journal of the American Statistical Association*, 116(535), 1548-1559, <http://dx.doi.org/10.1080/01621459.2021.1935267>
- Maxmen, A. (2020). How Much Is Coronavirus Spreading under the Radar?. *Nature* 10. <http://dx.doi.org/10.1038/d41586-020-00760-8>. Available at: <https://www.nature.com/articles/d41586-020-00760-8>
- Nishiura N., Klinkenberg D., Roberts, M., & Heesterbeek, J. (2009) Early epidemiological assessment of the virulence of emerging infectious diseases: a case study of an influenza pandemic. *PLoS One*, 4(8), 2009. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0006852>.
- Ojo O., García-Agundez A., Girault B., Hernandez H., Cabana E. Garcia, A., ... Fernandez A. (2020). CoronaSurveys: Using Surveys with Indirect Reporting to Estimate the Incidence and Evolution of Epidemics, arXiv:2005.12783 v2 Presented in KDD Workshop on Humanitarian Mapping, San Diego, California USA, August 24, 2020
- Oliver N., Barber X., & Roomp, K. (2020). The covid19 impact survey: Assessing the pulse of the COVID-19 pandemic in Spain via 24 questions. *CoRR*, abs/2004.01014.
- Pollán, M., Pérez-Gómez B., Pastor-Barriuso, R., Oteo, J., Hernán, M.A., Pérez- Olmeda, M., et al. (2020) Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet*. 396, 535–44. [http://dx.doi.org/10.1016/S0140-6736\(20\)31483-5](http://dx.doi.org/10.1016/S0140-6736(20)31483-5)

Ruppert, E., Grommé, F., Upsec-Spilda, F., & Cakici, B., (2018). Citizen Data and Trust in Official Statistics. *Economie Statistique/Economics Stat.* (505-506), 171–184. <http://dx.doi.org/10.24187/ecostat.2018.505d.1971>

The CoronaSurveys research team. (2020). CoronaSurveys: Monitoring the incidence of COVID-19 via open surveys. <https://coronasurveys.org/>. Accessed: 2020-05-20.

Ramírez, J., Diaz Aranda, S., Aguilar, J., Ojo, O., Lillo, R., & Fernández Anta, A. (2023). A Snapshot of COVID-19 Incidence, Hospitalizations, and Mortality from Indirect Survey Data in China in January 2023. medRxiv 2023.02.22.23286167 Presented in the 6th International workshop on Epidemiology meets Data Mining and Knowledge discovery. Long Beach, California, USA, August, 2023.