

Recomendaciones para la publicación de estadísticas oficiales a partir de encuestas de hogares frente a la coyuntura de la enfermedad por coronavirus (COVID-19)

Abril de 2020

Resumen

En el contexto de la pandemia global del COVID-19, la CEPAL está elaborando una serie de publicaciones cortas con recomendaciones de políticas relevantes para este período. En la presente nota se plantean una serie de recomendaciones para la recopilación de información relacionada con el mercado laboral mediante encuestas de hogares, dedicando especial atención a los diseños muestrales requeridos para este fin.

A. Introducción

A partir de la emergencia sanitaria por la pandemia global del COVID-19, que ha propiciado que la mayoría de los países de la región hayan declarado el estado de emergencia, las oficinas nacionales de estadística (ONE) han debido interrumpir abruptamente la recolección de información primaria en muchas de sus operaciones estadísticas, entre ellas las encuestas de hogares. En esta nota se identifican los posibles escenarios de recopilación y análisis de datos en las encuestas de hogares y se provee un conjunto de recomendaciones con respecto a la extracción, publicación y difusión de estadísticas oficiales provenientes de dichas encuestas en el contexto de la coyuntura internacional causada por el COVID-19.

En el marco de la declaración del estado de emergencia en los países, y con el fin de detener el contagio entre las personas y favorecer a los más vulnerables frente a esta pandemia, los gobiernos han tomado difíciles decisiones que han impactado en la sociedad y el normal funcionamiento de instituciones gubernamentales, entes educativos, industria, transporte y establecimientos de comercio, entre otros. En algunos casos, la misma sociedad ha empezado a colaborar con cuarentenas voluntarias; no obstante, ante el rápido avance de la pandemia, los gobiernos han debido decretar toques de queda, restricciones de la movilidad y cuarentenas obligatorias, entre otras medidas sanitarias, y se han implementado o desarrollado metodologías virtuales tanto en el ámbito educativo como en el laboral.

Teniendo en cuenta esta situación, la División de Estadísticas de la CEPAL dirigió un breve cuestionario a las ONE y bancos centrales de América Latina y el Caribe para conocer los efectos que estaba generando la emergencia sanitaria en el funcionamiento de las instituciones estadísticas. Las respuestas obtenidas de 20 países indican que, al 20 de marzo, todos los países pasaban por algún tipo de emergencia de salud debido al COVID-19. Las operaciones estadísticas más afectadas por la situación eran las encuestas y, en segundo lugar, los registros administrativos y los censos. Con respecto a las encuestas de hogares, en la mayoría de los casos se afirmó que la recopilación de información se pospondría, y en la mayoría de los casos se había previsto su reprogramación durante el año en curso. Entre las principales dificultades para recopilar información se encontraban las restricciones de movilidad y el cierre de establecimientos.

Resumen

- A. Introducción
- B. Recomendaciones para marzo de 2020
- C. Recomendaciones para los meses subsiguientes
- D. Conclusiones

Desde el 24 de marzo se han venido realizando una serie de videoconferencias para evaluar los impactos de la pandemia del COVID-19 sobre las operaciones estadísticas¹. Estas videoconferencias han sido convocadas por la CEPAL en conjunto con la Organización Internacional del Trabajo (OIT) y el Instituto Nacional de Estadísticas (INE) de Chile, en su calidad de coordinador de las charlas virtuales de la Red de Transmisión del Conocimiento (RTC) de la Conferencia Estadística de las Américas (CEA) de la CEPAL, presidida por el Departamento Administrativo Nacional de Estadística (DANE) de Colombia. Entre el 24 y el 26 de marzo se celebraron tres videoconferencias: dos con los países de América Latina, sobre encuestas de hogares y precios, respectivamente, y una con los países del Caribe. El 1 de abril se llevó a cabo otra videoconferencia enfocada en las encuestas de empleo.

Las restricciones a la movilidad impuestas en los países han llevado a suspender la recolección presencial de datos para las encuestas de hogares hasta nuevo aviso. Las ONE están recurriendo a la realización de encuestas por vía telefónica o por Internet. Este cambio en el modo de recolección es necesario para poder seguir generando cifras sobre el empleo y la pobreza. Estos indicadores son particularmente importantes en el contexto de la pandemia, dado el gran impacto que tienen las restricciones a la movilidad y las cuarentenas en la ocupación de las personas.

B. Recomendaciones para marzo de 2020

Para analizar el efecto de las restricciones a la movilidad en las encuestas de hogares que se llevan a cabo de manera continua es pertinente separar el mes de marzo de 2020 de los meses subsiguientes. En muchos países fue posible recabar información de manera normal aproximadamente hasta la segunda semana de marzo, momento en que se interrumpieron las actividades habituales de recolección. En cambio, para los meses subsiguientes no será posible obtener datos recabados de la manera habitual, mientras no termine este período excepcional.

Dependiendo de la cobertura parcial de las encuestas de marzo, las ONE enfrentan tres posibles escenarios:

- Niveles de cobertura deficientes, ante lo cual se recomendaría no utilizar la información recolectada para la producción de ninguna estadística oficial.
- Niveles medios de cobertura, ante lo cual se recomendaría modificar los factores de expansión en la encuesta para acotar los niveles de sesgo y publicar estadísticas oficiales únicamente a nivel nacional.
- Niveles aceptables de cobertura: ante lo cual se recomendaría continuar con el proceso normal de publicación.

Teniendo en cuenta el diseño transversal o rotativo de las encuestas de hogares en la región, y que estas se han planeado con el fin de lograr representatividad a nivel subregional (urbana y rural, por área geográfica, ciudades principales u otros aspectos), el tamaño de muestra de una encuesta en un mes específico en condiciones normales es mucho mayor que el tamaño de muestra necesario para alcanzar representatividad nacional. Las ONE deben calcular cuáles son estos tamaños de muestra mínimos.

Un ejemplo podría ser una encuesta planeada para lograr representatividad en los dominios urbano y rural cuya muestra correspondiente a marzo fuera de 8.000 hogares repartidos en todo el territorio nacional. A partir del diseño de la encuesta, es posible extraer el tamaño de muestra necesario para alcanzar únicamente representatividad nacional: en este caso podrían ser 1.800 hogares.

El primer escenario (niveles de cobertura deficientes) implicaría que la cobertura en la recolección de información de las dos primeras semanas de marzo es tan baja —menos de 1.800 hogares— que la información disponible no puede emplearse para brindar ningún tipo de estadística, ni siquiera a nivel nacional, y mucho menos regional. En este contexto, la mejor alternativa es no publicar ni difundir ningún tipo de información y comunicar a la

¹ Véase [en línea] <https://rtc-cea.cepal.org/es/conectados-rtc/planes-de-contingencia-adoptados-por-las-oficinas-nacionales-de-estadistica-ante-la>.

sociedad las limitaciones de la situación. Evidentemente, las series históricas creadas a partir de estos estudios tendrán un punto de discontinuidad correspondiente al tercer mes del año.

En el segundo escenario (niveles medios de cobertura), la información recolectada —más de 1.800 hogares en la muestra nacional— sí permitiría publicar cierta información, sujeta a algunas restricciones y a los ajustes descritos a continuación.

En primer lugar, no se recomienda recurrir a los métodos de imputación para completar la información faltante (que puede llegar al 50% de la muestra original). Aun si se dispone de información suficiente para realizar imputaciones, estas adoptan el supuesto de que el fenómeno bajo estudio se comporta de la misma manera durante todo el mes de marzo, y no reflejarían el comportamiento real del fenómeno bajo la realidad del COVID-19. Por otra parte, generar modelos predictivos válidos para completar toda la información inicialmente planeada implicaría una labor considerable en términos de la medición conjunta del error asociado a la imputación y el error asociado al muestreo con que se seleccionaron las unidades de muestreo. Además, el paradigma de las imputaciones y el ajuste de los modelos predictivos para proveer estadísticas oficiales en un país se aparta del paradigma de la muestra expandida y el principio de representatividad que predomina en las ONE.

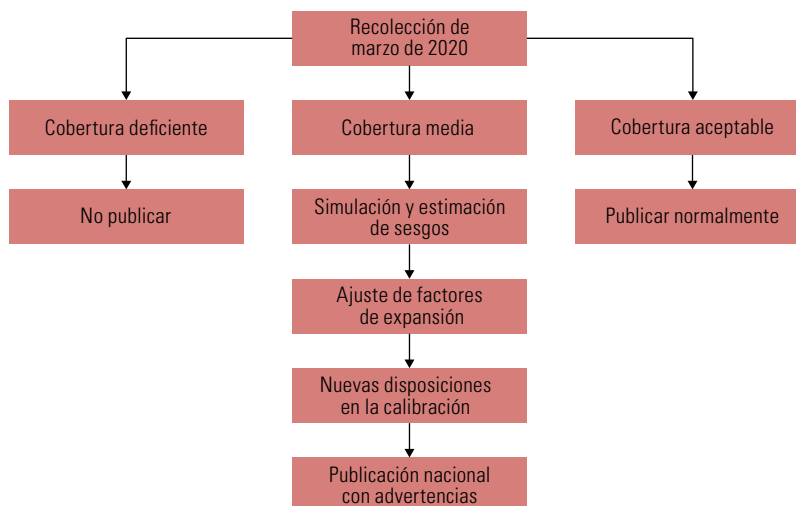
En este segundo escenario se recomienda utilizar los datos recolectados como insumo principal para obtener estimaciones nacionales, siguiendo la misma lógica que se utiliza en las ONE para la publicación de estadísticas oficiales. Es posible realizar una modificación sustancial de los factores de expansión para corregir el sesgo generado por la ausencia de respuesta debida al estado de emergencia declarado en los países.

La simulación para la estimación del sesgo deberá tomar la estructura final de las unidades primarias de muestreo (UPM) que fueron efectivas en marzo de 2020 (tanto en el ámbito geográfico como socioeconómico, en aquellos casos en que el marco de UPM se encuentre estratificado) y tratar de imitarla en la muestra de los meses anteriores (por ejemplo, en marzo de 2019). En este caso se contará con una base de datos recortada que se usará para obtener una estimación de marzo de 2019 como contrafactual a marzo de 2020. Con esta base de datos recortada se deberán ejecutar los procedimientos de estimación usuales para obtener estimaciones contrafactuales de marzo 2019. La diferencia entre las estimaciones publicadas y las estimaciones contrafactuales dará una idea de la magnitud del sesgo en cada uno de los constructos de interés de las encuestas de hogares. Teniendo en cuenta estas magnitudes de sesgo, los factores de expansión deberán ser ajustados de manera acorde en marzo de 2020 para corregir esta diferencia. El procedimiento podría ser iterativo y requerir nuevas disposiciones en la calibración de los factores de expansión, cuyas restricciones deberán adaptarse a los totales nacionales.

Se recomienda que la publicación y difusión de la información se realice únicamente a nivel nacional y que no se presente desagregada por las categorías del diseño de muestreo, dado que el fenómeno asociado a la ausencia de respuesta no puede clasificarse como aleatorio y, por ende, el sesgo existente crecerá a medida que se desagreguen las estimaciones. Es recomendable que las ONE comuniquen de forma oportuna que se abstendrán de publicar estadísticas desagregadas. También deberían explicitar que las cifras nacionales no corresponden a una recolección regular de información y pueden haber perdido la comparabilidad con los anteriores registros de la serie estadística. Por último, se recomienda especificar el período de recolección de la información del mes de marzo, que en la mayoría de los casos corresponderá a la primera quincena del mes, así como las tasas de cobertura nacional, por dominios subregionales y por sexo. El diagrama 1 clarifica las recomendaciones sugeridas para los tres escenarios referidos en esta sección.



Diagrama 1
Recomendaciones para tres escenarios de recolección de información en marzo de 2020



Fuente: Elaboración propia.

C. Recomendaciones para los meses subsiguientes

Mientras continúen las restricciones a la movilidad, es casi imposible asegurar la recopilación presencial de datos para las encuestas de hogares. La modificación de la forma de recabar información para estos instrumentos, ahora mediante entrevistas telefónicas o por Internet, supone desafíos en diversos ámbitos.

Uno de ellos es la inaccesibilidad de los hogares de la muestra. Las encuestas de hogares de la mayoría de los países se caracterizan por basarse en visitas repetidas en esquemas rotativos de muestreo. En cada período (mes o trimestre), se sustituye una parte de las viviendas que ya han sido visitadas en ocasiones anteriores por otras nuevas.

Dado que los números de teléfono de los hogares que forman parte de la muestra de la encuesta se recogen en la primera visita, no se dispondría de estos datos en el caso de los hogares que debían incluirse en la muestra de abril. Es decir, no habría forma de contactar a estos hogares seleccionados de forma probabilística. Por lo tanto, es necesario tomar decisiones sobre la muestra de hogares que se utilizará para continuar la recolección de información mientras se mantengan las limitaciones que afectan a los operativos de campo.

Un segundo desafío tiene que ver con la extensión y el contenido del cuestionario. Si la entrevista se realiza de manera telefónica, el tiempo disponible para llevarla a cabo es mucho menor que en una visita presencial. Resulta necesario reducir la extensión del cuestionario de forma de que centre en el subconjunto de preguntas de mayor relevancia en este contexto, que además requiere incluir nuevas preguntas que ayuden a determinar si las personas que no están trabajando mantienen su vínculo laboral con el empleador, así como a clasificar adecuadamente la condición laboral².

Este tipo de modificaciones plantean un escenario en que las autoridades nacionales deben asumir la pérdida de comparabilidad de la serie de indicadores de empleo e ingresos. Esta pérdida de comparabilidad puede remediarse de manera parcial más adelante, aplicando instrumentos que recojan la información retrospectivamente. No obstante, ello no implica que se deba interrumpir la recolección de información sobre el empleo y las demás estadísticas sociales. Incluso limitadas por las restricciones planteadas, estas cifras son cruciales para informar la política pública y la toma de decisiones en medio de un contexto de alta incertidumbre.

² Véase una caracterización general sobre el impacto del COVID-19 en la recolección de estadísticas del mercado laboral en Organización Internacional del Trabajo (OIT), "El impacto de COVID-19 en la recopilación de estadísticas del mercado laboral" [en línea] <https://ilostat.ilo.org/es/topics/covid-19/covid-19-impact-on-labour-market-statistics/>.

En estos meses excepcionales se debe priorizar por encima de todo que la publicación de estadísticas oficiales se base en la selección probabilística de una muestra de hogares o personas. Ante la imposibilidad de contactar con los hogares previstos para la muestra de abril, se recomienda que las ONE adopten un cambio temporal en el diseño de las encuestas de hogares y opten por el seguimiento de un panel en los meses subsiguientes; la rotación de paneles y la aplicación de los diseños de muestra usuales deberá congelarse hasta que se recupere la normalidad y se pueda obtener información de forma presencial.

Este panel estaría compuesto por las recopilaciones efectivas inmediatamente anteriores; por ejemplo, enero y febrero de 2020 y los últimos meses de 2019. Evidentemente, el mes de marzo de 2020 quedaría descartado al no corresponder a una recopilación usual o completa.

Si la oficina nacional de estadísticas no está preparada para llevar a cabo encuestas telefónicas de forma masiva para cubrir la muestra completa del mes elegido (por ejemplo, enero de 2020), es posible diseñar una submuestra nacional mediante un muestreo en dos fases. Cualquier modificación a este seguimiento en panel solo se considera válida si la muestra telefónica corresponde a una muestra probabilística cuya probabilidad de selección se conoce de antemano.

El seguimiento de un panel es una manera sencilla y efectiva de contar con una muestra probabilística para estimar las estadísticas necesarias. Esta estrategia está libre de los sesgos de selección, como cualquier muestra representativa, pero a su vez permitiría evaluar el efecto del sesgo no muestral debido al cambio del modo de recolección. De esta forma, es posible realizar una estimación insesgada de flujos brutos (seguimiento mensual de la situación respecto a la ocupación de los miembros del hogar), así como evaluar posteriormente el efecto de la pandemia en las estadísticas derivadas de la encuesta.

Si bien la recolección de información por vía telefónica está expuesta al sesgo de selección, debido a que no todos los hogares contactados en enero tienen teléfono o van a contestar la llamada de la ONE, este puede ser corregido mediante modelos estadísticos, técnicas de calibración y el ajuste de los factores de expansión.

En el diagrama 2 se exponen dos ejemplos de cómo un país puede hacer seguimiento a un panel proveniente de recolecciones anteriores. En el escenario de la izquierda se presenta un país que a partir de abril de 2020 tiene la capacidad para realizar un seguimiento telefónico (o por Internet) de los mismos hogares que se seleccionaron en enero de 2020. El segundo escenario ilustra un caso en que se realiza el seguimiento a una submuestra de hogares (o personas) basada en la muestra probabilística de febrero de 2020. En ambos casos, el seguimiento del panel está supeditado únicamente a la duración de las restricciones a la movilidad.

Diagrama 2
Escenarios de seguimiento de una muestra panel basada en recolecciones de información anteriores



Fuente: Elaboración propia.



Al finalizar este período de confinamiento, es posible volver a incorporar las particularidades del diseño de muestreo usual. Esta reincorporación puede ser paulatina, si se desea hacer un análisis profundo de los sesgos generados por el cambio de modo de recolección. Por ejemplo, es posible obtener un porcentaje de la información presencialmente y otro porcentaje de la información de manera telefónica (o por Internet). Esta información permitirá definir los niveles de comparabilidad de las series de tiempo y determinará el tipo de empalme que estas puedan tener con las series a mediano plazo.

D. Conclusiones

En esta nota se propone una estrategia para hacer frente a las restricciones a la movilidad en la ejecución de las encuestas de hogares que se llevan a cabo de manera continua. Se realiza una distinción entre el mes de marzo de 2020, en el cual se logró recolectar una parte de la información de manera normal, y los meses subsiguientes, que requieren implementar un modo distinto de recolección.

En el caso del mes de marzo, se propone utilizar la parte recolectada de la muestra para publicar solamente estadísticas a nivel nacional y sin desagregaciones. En cuanto al mes de abril, y mientras dure el período excepcional, se recomienda a las ONE cambiar el modo de recolección, interrumpir la aplicación de los diseños de muestreo usuales, realizar el seguimiento de un panel en los meses de cuarentena (o de estado de emergencia) e introducir nuevamente el diseño original a la vuelta a la normalidad.

Esta estrategia permite generar estadísticas no sesgadas sobre el mercado laboral basadas en la selección probabilística de una muestra de hogares o personas. Estas estadísticas proporcionan información crucial para la política pública y la toma de decisiones sobre el mercado de trabajo.

Este documento es parte de un conjunto de informes elaborados por la Comisión Económica para América Latina y el Caribe (CEPAL) sobre la evolución y los efectos de la pandemia del COVID-19 en América Latina y el Caribe. Fue preparado por la División de Estadísticas, dirigida por Rolando Ocampo, bajo la coordinación general de Alicia Bárcena, Secretaria Ejecutiva de la CEPAL.

Copyright © Naciones Unidas, 2020



Comisión Económica para América Latina y el Caribe (CEPAL)
Economic Commission for Latin America and the Caribbean (ECLAC)
www.cepal.org