



## **Misión de asistencia técnica llevada a cabo entre el 10 y el 14 de junio de 2019 al DANE, en el marco del Comité de Expertos evaluador del Censo Nacional de Población y Vivienda – CNPV 2018 realizado en Colombia**

### **Informe de Misión Integrado CELADE-CEPAL, UNFPA LACRO y UNFPA Colombia <sup>1</sup>.**

27 de junio de 2019

#### **1. Antecedentes y organización de la misión**

El censo colombiano de 2018 fue un proceso especial por varias razones, entre ellas la implementación de varias modalidades de diligenciamiento incluyendo la autodeclaración por internet (e-censo). Lo anterior supuso un conjunto de desafíos novedosos que se pusieron a consideración del Comité de Expertos (CE), convocado por el DANE en noviembre de 2018 para analizar la cobertura, calidad y consistencia de los resultados del Censo 2018.

El DANE solicitó una misión de CELADE-UNFPA para apoyar el proceso de seguimiento y revisión del censo y sus resultados a cargo del CE, en particular en materia de evaluación de calidad y cobertura de los resultados censales (en el Anexo 1 se adjuntan los términos de referencia de la misión). Para ello, elaboró una serie de tabulados solicitados previamente por el CE, obtenidos mediante procesamiento de diferentes versiones de la base de datos del censo, ninguna de ellas la definitiva. La última versión generada (denominada “base 4”) es bastante cercana a ser la definitiva.

La premura con la que se definió la misión limitó el trabajo previo con los datos enviados, así como el intercambio más sustantivo con el Comité de Expertos y los especialistas del DANE, todo lo cual se tuvo que efectuar en Bogotá durante la misión. Por lo mismo, se procedió a revisar la agenda original para acomodarla a este escenario y maximizar las posibilidades de cumplir los objetivos de la misión (Anexo 2).

Cabe subrayar que, durante todo el trabajo realizado por la misión, el equipo del DANE colaboró activamente, respondió preguntas, reaccionó ante solicitudes, acordó acoger sugerencias y observaciones varias y mostró la mejor disposición para apoyar a la misión, lo que se agradece. Con independencia de algunas observaciones que la misión tiene sobre el funcionamiento del DANE <sup>2</sup> y de los equipos vinculados al censo, solo cabe reconocer la voluntad mostrada por el DANE durante la estadía de la misión y felicitar al equipo del DANE con que trabajó la misión por su compromiso, entusiasmo y capacidades técnicas.

#### **2. Desarrollo de la misión**

---

<sup>1</sup> Este informe fue elaborado por Fabiana Del Popolo, Jorge Rodríguez y Helena Cruz, del CELADE-División de Población de la Comisión Económica para América Latina y el Caribe (CEPAL), Pablo Salazar y Daniel Macadar, de la Oficina Regional para América Latina y el Caribe del Fondo de Población de las Naciones Unidas (UNFPA LACRO), y Paulo Lara y Carlos Ramírez, de la Oficina Colombia del UNFPA. Las opiniones expresadas en este documento, que no ha sido sometido a revisión editorial, son de exclusiva responsabilidad de los autores y pueden no coincidir con las de la respectiva Organización.

<sup>2</sup> Ver sugerencia sobre conformación de equipos mixtos en la sección de recomendaciones.



La misión se desarrolló en la semana del 10 a 14 de junio de 2019 según una agenda enviada previamente, la cual se ajustó para optimizar el funcionamiento de la misión y el logro de sus objetivos, y para adecuarla a las condiciones de disponibilidad de datos censales.

La agenda incluyó reuniones ampliadas (equipo de la misión e integrantes del Comité de expertos nacionales y de UNFPA-Colombia) con el equipo del DANE (Juan Daniel Oviedo, Director del DANE, Liliana Acevedo, Directora Técnica de Censos y Demografía, Humberto Cote, Asesor de la Dirección General y técnicos de diversas secciones) para abordar diferentes tópicos censales y reaccionar frente a los resultados enviados, tanto para llamar la atención sobre algunos potenciales problemas, plantear dudas sobre algunos números e identificar aparentes inconsistencias, como para solicitar nuevos tabulados más elaborados o procesados con otras versiones de la base de datos.

La agenda también incluyó sesiones de trabajo con los analistas del DANE, mediante la conformación de subgrupos temáticos, poniendo énfasis en el análisis de la cobertura y de variables seleccionadas. Cada subgrupo solicitó al DANE tabulados específicos, tanto de la base 2 (aquella sin imputaciones ni reglas de validación aplicadas) como de la base 4 (tomando la última versión de la base de datos), de modo de posibilitar la evaluación de la calidad de las variables y de los efectos de las reglas de validación e imputación aplicadas. Asimismo, uno de los subgrupos se abocó a revisar resultados y metodologías utilizadas para la estimación de los componentes demográficos (fecundidad, mortalidad y migración) y la conciliación demográfica. El DANE colaboró permanentemente con el equipo, atendiendo requerimientos, procesando información y realizando presentaciones temáticas cuando se necesitaban.

Cabe señalar que también hubo espacios para que la misión sesionara de manera autónoma e instancias de reunión exclusiva con los expertos nacionales para la discusión de diferentes puntos específicos. Finalmente, la misión expuso los principales hallazgos y conclusiones preliminares de su trabajo, el último día de la estadía.

### **3. Hallazgos**

#### **3.1 Considerandos**

-La misión se adelantó una semana respecto de la solicitud original, por lo cual la tarea de producción de los tabulados solicitados para ser analizados tuvo que acelerarse y no se alcanzó a concluir antes de la llegada de la misión. Por lo tanto, parte importante de la misión se destinó a producir los tabulados necesarios para el logro de los objetivos. Si bien fue priorizando el procesamiento de las variables clave de la dinámica demográfica, también pudieron ser abordadas algunas variables importantes de viviendas y de personas<sup>3</sup>.

-Un punto que restó posibilidades de profundizar en el análisis y agilizar algunos procesamientos fue no contar con la Base 4<sup>4</sup> en REDATAM. Esto en parte se debió a que aún ésta se estaba revisando (por ejemplo, durante el fin de semana previo a la misión se encontraron problemas en algunos resguardos indígenas y se afinaron reglas de consistencia e imputación). Se reconoce que el DANE dedicó su máximo

---

<sup>3</sup> Más adelante se detallan estas variables en las secciones específicas.

<sup>4</sup> La Base 4 es la base completa (con todas las personas) donde ya se han aplicado las reglas de validación e imputación.



esfuerzo para tenerla lista, sin embargo, la Base 4 en REDATAM estuvo disponible recién el jueves 13, no quedando tiempo para su procesamiento.

-Además, *in situ* fue evidente que era necesario aprovechar la oportunidad para discutir junto con los expertos nacionales del CE y el DANE, aspectos más amplios y generales del censo que los previstos en los objetivos de la misión, tales como las tres modalidades de recolección de datos y su integración, el e-censo, el período de recolección de datos, la definición del momento censal, entre otros.

### 3.2 Bases de datos

La base de datos actual del censo ha sido sometida a un proceso complejo, minucioso y exigente de elaboración con el propósito de evitar errores e inconsistencias (aunque estas últimos son virtualmente imposibles de eliminar completamente en cualquier censo). Se destaca el hecho de que la misión haya podido revisar este proceso y haya tenido acceso directo a la base de datos, dando muestras de un nivel de transparencia que cabe reconocer y valorar. Los procesos de revisión permitieron identificar y corregir algunas incongruencias y decisiones discutibles en materia de imputación, identificados a partir de la comparación de las diferentes versiones (“sin” y “con” validaciones/imputaciones) de la base de datos censal.

La combinación de tres procedimientos de recolección de datos (papel, DMC<sup>5</sup>, e-censo) generó varias complicaciones que afectaron al censo y complejizó la generación de una base de datos consistente y de forma eficiente. Entre otras cosas, hubo casos duplicados que debieron ser identificados a través de los diferentes medios de captura, para luego ser filtrados en la construcción de la base de datos final. Este es un proceso muy complejo y que si bien fue enfrentado con rigor -mitigando significativamente los errores- no está exento de problemas.

El equipo DANE confeccionó un glosario que explica de manera detallada los contenidos y grados de depuración de las cuatro bases de datos existentes. En dicha descripción se indica que la base 2 (que junta todos los casos provenientes de los diferentes medios de recolección) no contiene las características de las viviendas recolectadas en el e-censo, lo que imposibilita la comparación (de los valores de las variables “antes” y “después” de los procesos de validación/imputación) con la base 4 donde si están todas las personas con todas sus características. La información sobre viviendas (procedentes del e-censo) recién se incorpora en la base 3, pero la base 3 también incorpora algunos procesos de validación, crítica e imputación, que imposibilitan tomarla como punto de comparación (base completa “sin correcciones”).

Por esta razón no fue posible hacer una comparación de las características de vivienda de la base completa y sin duplicados antes y después de dicho proceso. Respecto a las variables de personas, sí es posible hacer este análisis. Si bien los totales de la base 2 y la base 4 no eran exactamente los mismos, las diferencias eran poco significativas y por tanto no invalidaron lo observado respecto al comportamiento de las variables.

### 3.3 Cobertura

---

<sup>5</sup> Dispositivo Móvil de Captura



Como todo censo, el de Colombia 2018 tuvo subenumeración, atribuible a diferentes factores. Medir el subregistro es siempre un asunto complejo y ninguna modalidad de estimación de la omisión censal es perfecta. El DANE está utilizando diferentes métodos para estimar la cobertura del Censo, ya que utilizar varias estimaciones derivadas de diferentes métodos o fuentes, permite tener parámetros de comparación para llegar a una estimación más robusta.

Aunque no se realizó una encuesta post censal propiamente dicha, la utilización de la Encuesta de Calidad de Vida 2018 (ECV) fue de utilidad para estimar la cobertura del censo a través de un emparejamiento directo entre la población encuestada y la población censada. Se destaca que la pregunta de documento de identidad en ambas fuentes de datos posibilitó un emparejamiento más preciso y la identificación de una gran mayoría de personas en ambas fuentes. Sin embargo, en el primer ejercicio de *Dual system* presentado al equipo de la misión se obtuvo un importante número de casos que no pudieron ser pareados. Esto se debía a problemas de identificación tanto en la encuesta como en el censo. Una primera opción frente a esta situación consistía en incorporar las personas no pareadas de acuerdo a una probabilidad de ser omitidas similar a las personas con identificación. Siguiendo las recomendaciones de la misión, el equipo DANE analizó las características de edad, estatus migratorio y educación de las personas que no tenían identificación en la ECV 2018 y en el Censo 2018 comparado con las que sí se identificaron.

Los resultados mostraron que en la ECV 2018 el grupo de personas *sin identificación* resultó ser muy diferente respecto al grupo *con identificación*. Específicamente, se observó que el grupo *sin identificación* presenta una estructura etaria más joven y con mayor probabilidad de ser migrante. Personas con esas características tienen mayor probabilidad de omisión censal. Por tanto, si se incorpora al cálculo el grupo de personas sin identificación, con igual probabilidad de omisión, es altamente probable que el porcentaje de subenumeración obtenido por ese método esté subestimado. El último día de la misión, el equipo del DANE presentó un segundo análisis que utiliza el nombre de manera fonética como segundo criterio de identificación para las personas no pareadas. Esta opción mejoró la cantidad de personas pareadas, pero produjo algunos ruidos (por repetición de nombres) que deberán ser subsanados. De todas maneras, esta opción debería ser considerada una vez agotada la posibilidad de mejorar la información de identificación del Censo y de la ECV con las bases espejos de la Registraduría.

Sobre la utilización del método *Bootstrap* para mejorar el estimador del total de población resultante del *Dual system*, se observó que el método cambia de manera significativa los resultados del método directo (en el último ejercicio presentado la cifra pasa de 9,6% a 8,6%, una disminución de 10% del resultado del método directo de emparejamiento). En general, se utiliza esa técnica para estimar intervalos de confianza y no para alterar la media que resulta del método. Además, ambas muestras son grandes y a primera vista no se justifica la utilización del método para corregir el resultado del método directo.

Producto de varios ejercicios de *Dual system* (que fueron mejorando el % de pareo) y de otros ejercicios de estimación de la cobertura censal se obtuvieron resultados que varían en un rango del 8.6% (*Dual system* entregado el 14-06-2019) al 11% (considerando la revisión 2019 de las estimaciones y proyecciones que realiza el CELADE).



Es importante señalar que la cifra nacional de subregistro es un primer valor de referencia, que luego debe ser calculado a escalas más desagregadas para efectos de las estimaciones y proyecciones subnacionales de población.

En el Anexo 3 se presenta un detalle adicional sobre este análisis.

### **3.4 Consistencia y Calidad**

A pesar de no ser posible comparar una base completa “sucía” con una base completa “limpia” para algunas variables, las comparaciones que fueron hechas muestran que los procesos de validación, crítica e imputación, en general, se ajustan a los estándares y no involucran ajustes significativos ni tampoco generan sesgos evidentes. El comportamiento de las variables revisadas muestra consistencia con otras fuentes relevantes, se ajusta a patrones teóricos esperados, se adecúa a patrones de evolución demográfica y socioeconómica del período 2005-2018 y (al menos en el módulo de vivienda) a conocidos diferenciales territoriales (urbano-rural). La misión pudo revisar las reglas de validación e imputación de variables clave. En general, estas reglas mejoran la calidad de la información. Con todo, algunos algoritmos de imputación parecen excesivos y otros perfectibles, lo que se le indicó al DANE de manera precisa, con propuestas de cambio por parte de la misión y del grupo de expertos, en el caso de variables críticas como sexo, edad, parentesco y fecundidad.

El uso de documentos y números de identificación para verificar información del censo es otra novedad del censo de Colombia 2018. Estos identificadores fueron ampliamente usados para tales efectos, beneficiando la calidad de la información y ofreciendo una fuente de datos muy valiosos para los ejercicios de estimación del subregistro. Con todo, la disponibilidad de esta información no implica su uso ilimitado, pues algunas variables claves, como la declaración de la edad o incluso del sexo, normalmente no debieran ser corregidas con apoyo de información externa, ya sea por un asunto de derechos, o porque su utilización implicaría alterar el flujo de preguntas del formulario (saltos condicionales por edad o por sexo). Durante el relevamiento este flujo estuvo indicado por las respuestas de edad y sexo declaradas, si se cambian ex-post, habrá una cantidad de información recogida que desaparecerá y una cantidad de ignorados que aparecerá, debido a que se habrán excluido personas que respondieron o incluido personas que antes no habían respondido por no estar incluidas.

Las variables revisadas (de vivienda y seleccionadas de personas: edad, sexo, fecundidad, mortalidad y migración) mostraron bajos niveles de no respuesta. Sin embargo, estos niveles variaron significativamente según método de captura: los DMC registraron menores índices de no respuesta.

Los análisis fueron realizados a partir de la información suministrada por las bases Bases 0, 2 y 4.

En la base 0, para la gran mayoría de estas variables, los porcentajes de blancos resultaron superiores en el e-censo y en el papel, en algunos casos con diferencias muy significativas respecto al DMC. Fecha de nacimiento y sexo son las dos variables que mostraron mayor completitud en el e-censo. La integración de las bases según fuente de captura mejoró estos porcentajes, por el hecho de que, en los casos de existir duplicados, se optaba por conservar la información lo más completa posible, que solía ser la de los DMC. Por otra parte, las pautas de validación e imputación no cambiaron de manera significativa el porcentaje



de blancos -cuando se compara la Base 2 con la Base 4-, con lo cual, en principio, no se abrían afectado las estructuras.

### 3.4.1 Viviendas

Se obtuvo la información sobre vivienda con valores absolutos para las bases 2A (que no incluye imputación para vivienda, pero sí elimina duplicados) y la base 4 (que sí la incluye). La variable de “paredes” presenta una cifra marginal de ignorados desde la base sin imputar (85 mil viviendas), lo que sugiere una calidad satisfactoria de la recolección en el terreno. Por otra parte, la estructura de las respuestas y la comparación intercensal resulta razonable. Respecto de la imputación, baja a la mitad la cantidad de ignorados, pero básicamente por asignación a la categoría MODAL, que no es precaria (y la no respuesta puede ser sesgada hacia categorías más bien precarias). Como la imputación se basa en la moda territorial (que no parece un criterio particularmente sólido) y son muy pocos los casos sin información, entonces se sugiere mejor NO imputar, sobre todo si la imputación también cambia valores captados en terreno (caso concreto vaciado), aunque no es seguro que la imputación de ignorados sea la responsable de este cambio.

PISO: Presenta una cifra marginal de ignorados desde la base sin imputar (88 mil viviendas), lo que sugiere una calidad satisfactoria de la recolección en el terreno. Por otra parte, la estructura de las respuestas resulta razonable, sobresaliendo netamente una categoría (amplia moda nacional): Baldosa, vinilo, tableta, ladrillo, laminado. Asimismo, la comparación intercensal también arroja resultados razonables. Respecto de la imputación, baja a la mitad la cantidad de ignorados, pero básicamente por asignación a la categoría MODAL (y la no respuesta puede ser sesgada hacia categorías más bien precarias). Como la imputación se basa en la moda territorial y son muy pocos los casos sin información, entonces también la sugerencia sería mejor NO imputar, sobre todo si la imputación también cambia valores captados en terreno (como alfombra), aunque no es seguro que la imputación de ignorados sea la responsable de este cambio.

ENERGÍA ELÉCTRICA: Presenta una cifra marginal de ignorados desde la base sin imputar (88 mil viviendas), lo que sugiere una calidad satisfactoria de la recolección en el terreno. Por otra parte, la estructura de las respuestas resulta razonable, sobresaliendo netamente una categoría (amplia moda nacional): Tiene acceso a electricidad. Lo mismo ocurre con la comparación intercensal: caída de sin acceso de 6.4 al 3.6. En este caso la imputación con la moda territorial parece menos riesgosa y más razonable en términos teóricos y en sus resultados.

ESTRATO: Presenta una cantidad no menor de ignorados en la base 2 (615 mil), así como muchos más casos de valores inválidos (sobre todo por una categoría 9 que no es fácil de interpretar), lo que tal vez no es tan extraño con una variable tan peculiar y propia de Colombia como ésta. En la imputación, se corrigen los valores inválidos (incluyendo el valor 9) y se reduce la cantidad de ignorados en un 12%. De cualquier manera, la cifra de ignorados llega a 540 mil, la variable con mayor cantidad de ignorados del módulo de vivienda; que, aun así, es solo del orden del 4% de las viviendas. No es fácil tener un parámetro adicional de calidad de la información de esta variable, aunque si llama la atención que el valor 0, que aparentemente significa que no hay medidor de electricidad que es la fuente de esta información de estrato, es bastante inferior a las casi 500 mil viviendas que no tienen acceso de electricidad.



**ACCESO A SERVICIOS:** Acceso a acueducto, Acceso a Alcantarillado, Acceso a gas natural y Acceso a recolección de basura. Con estas cuatro consultas ocurre algo parecido a la materialidad: baja no respuesta y resultados consistentes con lo esperado, consistentes con otras fuentes y consistente con la distinción rural-urbana (menor acceso a servicios en las zonas rurales). La imputación también tiende a reducir los ignorados en torno a un 50%, pero a diferencia de lo observado en el caso del acceso a energía eléctrica, la imputación en la conexión a acueducto y alcantarillado tiende a aumentar los casos de conexión y a reducir (muy ligeramente) los sin conexión, lo que resulta contra-intuitivo y no procedente, porque no resulta evidente cuál regla de imputación generó este efecto. Por lo mismo, se sugiere revisar o hasta revertir esta imputación. En los otros dos casos, la imputación no tiene este efecto, pero tampoco parece justificarse.

**NÚMERO DE VECES QUE PASA LA BASURA:** Pocos casos ignorados y bastantes no aplica, que provienen principalmente de quienes respondieron previamente que no tenían acceso a recolección de basura. La categoría 0 tiene muy pocos casos y no es compatible con la respuesta SI en la pregunta filtro, a menos que aquí se hayan incluido los casos con una frecuencia inferior a la semanal, pero son muy pocos para ello. El último tabulado recibido no permite mayor conclusión sobre esta variable.

**WC:** Poco ignorados y distribución consistente con gradiente urbano rural y otras fuentes como la ECV 2018. Curiosamente en este caso, la imputación (si acaso esa es la única diferencia entre las dos bases del último tabulado) aumenta la cantidad de ignorados (de 93 a 126 mil), por lo cual no queda claro el sentido de la regla de imputación.

**INTERNET:** Tabulado con tabla 4 no incluyó esta variable. Pocos ignorados (100 mil) en tabulados con tabla 2A. Penetración del internet inferior al 50%, básicamente por el muy bajo acceso en zona rural. Cabría chequear este dato (aunque no para cambiarlo sino para entenderlo), porque tal vez la pregunta a escala de vivienda puede haber originado respuesta a escala de vivienda invisibilizando acceso a internet vía celulares personales.

### **3.4.2 Características demográficas de las personas**

En el caso de la edad y el sexo se realizó un examen algo más detallado por la relevancia central que tienen estas dos variables. Respecto a la completitud, en la base 0 la variable sexo presentaba muy bajos valores sin información (0,10% de blancos en DMC; 0,07% en e-censo y 1,43% en papel). En todos los departamentos se observan bajos niveles de no respuesta, las cifras varían entre 0,06% a 1,69%. Respecto a la edad declarada, la base 0 también muestra porcentajes muy bajos de blancos, tanto en DMC como en el e-censo, sin embargo, en el papel, el porcentaje de ignorados es cercano al 10% y en un departamento llega al 22%. La fecha de nacimiento, por su parte, registró mayores problemas: sumando las categorías “blancos” + “no sabe”, se tiene un 11,2% para DMC, 8,4% para el e-censo y 16,7% para el papel. La integración de las bases provenientes de las 3 fuentes de recolección mejora significativamente estos indicadores (en el caso de los duplicados entre DMC y e-censo se tomó la información más completa, y, además, los blancos se subsanaron con la información derivada del registro de identidad).

Cabe notar que, no obstante, tanto en la base 2 como en la base 4, en un 4,5% de los casos no se sabe la fecha de nacimiento. Colombia no es el único país de la región en tener mayores dificultades respecto a la captación de la fecha de nacimiento respecto a la edad declarada. Sin embargo, es recomendable tener



ambas variables puesto que de todas maneras aporta en el control de calidad de la edad y, en el caso colombiano, también para corroborar la identificación de la persona con el registro de identidad.

Se compararon los resultados de la población por edades simple y sexo, obtenidas a partir de 4 bases de datos: Base 2; Base 2 con imputación de sexo a partir del registro de identificación; Base 4 con reglas de consistencia aplicadas a la edad; y Base 4 con estas reglas más el reemplazo de las edades declaradas por las estimadas en casos especiales.

Las conclusiones son: a) Se concuerda con la decisión de imputar el sexo a partir del registro de identidad cuando la variable estaba en blanco; y, pese a que no se recomienda cambiar el sexo declarado, las diferencias observadas entre las bases 2 con y sin imputación del sexo es insignificante y no altera la estructura de esta variable ni en el total ni por edades simples; b) la aplicación de las reglas de consistencia en base 4 sin imputar la edad declarada por la edad estimada, tiene ciertos impactos esperados, principalmente en los menores de 15 años, pero estos no llegan al 1% (0,86% cambio neto para mujeres de 1 año a 0,03% en mujeres de 13 años); c) comparando la base 4 con imputación versus sin imputación de la edad declarada a partir de la edad calculada, los cambios son irrelevantes. Solo se observan cambios en los menores de 4 años que tampoco son significativos (0,1% cambio neto mujeres de 4 años; a 0,03% mujeres de 3 años). En síntesis, puede decirse que la información censal respecto al sexo y edad de las personas fue editada adecuadamente.

El comportamiento por edad y sexo de la población derivada del censo (analizando resultados de la Base 4), muestra alta coherencia con la Base de Datos Única de Afiliados (BDUA) y con las proyecciones del CELADE que se publicaron la semana próxima a la misión en el *World Population Prospect 2019*. Tanto las estructuras etarias como el índice de masculinidad por edad muestran tendencias muy similares. Las diferencias respecto a las proyecciones se asocian más bien con la omisión de personas en el censo, cuya magnitud final habrá que determinar una vez que se realice la conciliación censal.

Respecto a las defunciones en el hogar, se examinó la estructura por edad y sexo con las estadísticas vitales del 2017, mostrando comportamientos muy similares. Incluso, pese a que no son fuentes directamente comparables, las magnitudes se asemejan. Asimismo, se analizó el porcentaje de hogares sin fallecidos, total nacional y por departamento, comparándolo con el censo 2005. Al respecto, los resultados del censo 2018 se ven más coherentes en términos de las desigualdades esperadas entre departamentos según sus niveles de desarrollo, mientras que el censo 2005 no muestra diferencias entre los mismos.

En el caso del módulo de fecundidad, sí se encontraron algunos resultados que nos ponen en alerta. Como punto de partida se observó la completitud de las preguntas P54. *¿ha tenido algún hijo o hija que haya nacido vivo(a)?*; P55. *De los hijos e hijas, que nacieron vivos(as) . . . ¿cuántos están vivos actualmente?*; y P57. *¿Cuál es el mes y el año de nacimiento del último hijo o hija nacido(a) vivo(a) de . . . ?* según fuente de recolección. El porcentaje de blancos era demasiado elevado, sobre todo en la P55 y P57, con lo cual se supone que los cálculos incluían casos para los que no aplica la pregunta (al no tener las cifras absolutas se dificultó la posibilidad de análisis). En un primer análisis se observaron comportamientos que llamaban la atención respecto a los cambios en el porcentaje de no respuesta de la base 2 y de la base 4, junto con ello, resultados atípicos en el comportamiento de la no respuesta por edad (relativamente baja en las



primeras edades y se incrementaba en las edades más avanzadas) como así también en los porcentajes de mujeres sin hijos comparados con la ENDES 2015 (resultaban bastante más elevados en el censo y se incrementaban con la edad), Una primera hipótesis era que la no respuesta se asignó a 0 hijos. Lo anterior llevó a revisar las pautas de validación en conjunto con el DANE. Al respecto se brindaron una serie de sugerencias que fueron bien recibidas por el equipo para la revisión de la base.

Sin perjuicio de ello, el equipo del DANE volvió a procesar la base 4 (revisando la programación), entregando nuevos resultados, tanto valores absolutos como relativos. Por una parte, los porcentajes de blancos en el total de hijos nacidos vivos por edad de la madre mostraban el comportamiento esperado, siendo más elevados en las edades más jóvenes, aunque resultando bajos (4,2% en el grupo 10-14 a 2,4% en el grupo 45-49).

En cuanto a la comparación del porcentaje de mujeres según número de hijos con la ENDES 2015, en términos generales presentan un comportamiento similar según la edad de las mujeres, pero difieren en magnitud. Por ejemplo, en el caso del porcentaje de mujeres sin hijos, éste es bien elevado en 10-14 y 15-19 (incluso con magnitudes similares en censo y encuesta) y disminuye con la edad de las mujeres. No obstante, a partir del grupo 20-24 más allá del descenso con la edad, las cifras del censo siempre son algo más altas que las de la encuesta. En contraste, si se mira el porcentaje de mujeres con 3 hijos, éste aumenta sistemáticamente con la edad de manera similar en el censo y en la encuesta, pero las cifras son siempre algo más elevadas en la encuesta. Por tanto, a partir de estos resultados es altamente probable que el nivel de fecundidad esté subestimado comparado con la ENDES.

En cuanto a los hijos sobrevivientes, el nuevo procesamiento entregado por el DANE mostraba niveles de no respuesta más bajos aún que el total de hijos nacidos vivos (3,3% en 10-14 y entre 0,06% y 0,20% en el resto de las edades). En términos más sustantivos, de la comparación con la ENDES, el promedio de hijos sobrevivientes por edad de las mujeres resulta siempre más bajo en el censo que en la encuesta.

Con todo, se sugiere revisar las reglas de validación e imputación y volver a evaluar las preguntas de este módulo, en la base 0, y antes y después de aplicar dichas reglas.

### **3.4.3 Migración**

El Módulo de migración corresponde a las preguntas retrospectivas sobre residencia anterior constituidas por las preguntas P39 a P41 sobre municipio/país de nacimiento, municipio/país de residencia 5 años antes del censo y municipio/país de residencia 12 meses antes del censo.

Un primer análisis de la calidad de la información recogida, basado en frecuencias de varias versiones de la base de datos censal, sugiere que, desde la primera base válida disponible, los niveles de no respuesta fueron bajos (inferiores al 2%) en preguntas que en el terreno suele presentar dificultades y estar expuesta a riesgos de no respuesta o respuestas inválidas. Es probable que el uso de dispositivos (DMC) y aplicativos (e-censo) haya contribuido a estos bajos niveles de no respuesta. Los casos ignorados en la última versión de la base (que excluye LEAS) bordean el 1% (magnitud que aumenta marginalmente al considerar los ignorados en la especificación del municipio o país de residencia anterior). Sí se detectaron unos pocos casos inválidos (entre 15 mil y 7 mil dependiendo de la pregunta), que debieran ser revisados y asignados a alguna categoría, probablemente a la categoría de Ignorados.



Un examen más sustantivo de la información sugiere que los datos de migración se ajustan a patrones esperados, sea de perfiles por edad, de orígenes y destinos y/o de intensidad. En este sentido, se calcularon los porcentajes de migrantes internos por edad y la curva de las mismas se ajusta al patrón esperado de migración (Rogers y Castro, 1982; White, 2016) con cúspides en edad inicial y juvenil (sin salto en edad avanzada como en los países desarrollados). La matriz de migración (5 años) por Departamento y Municipios arrojó flujos esperados que sirvieron de chequeo (como el que hay entre Bogotá y Cundinamarca) o el que hay entre el municipio de Cali y varios municipios de su entorno (Candelaria, Jamundí, Palmira). Respecto de los niveles de migración, el total de migrantes es algo mayor que los registrado por el censo de 2005, en parte explicado, aunque solo parcialmente, por el aumento de la migración internacional.

Finalmente, sobre la migración internacional, el censo capta de manera muy elocuente la principal novedad del período: la inmigración desde Venezuela. En número de personas que nacieron en Venezuela es 813,260, el número de personas que vivían en Venezuela hace cinco años es de 813,260 y el número de personas que vivían y llegaron hace un año es de 398,781, lo que es coherente con el alto flujo de entrada regulares de venezolanos observados en 2017 y 2018 en el país en los datos de Migración Colombia. Adicionalmente, en la comparación con el censo de 2005, resulta evidente el aumento muy significativo de la población nacida en Venezuela.

### **3.5 Conciliación demográfica**

El DANE comenzó los ejercicios del proceso de conciliación censal y presentó resultados preliminares. Un resultado sorprendente fue la casi coincidencia de la población conciliada en 2005 con la censada, lo que lleva a pensar que la base de datos del censo de 2005 tiene un nivel de imputación de población que no ha sido documentado. Este ejercicio fue revisado y comentado por la misión, identificando algunos aspectos metodológicos y de fuentes que pueden ser mejorados para la continuación del proceso.

En el ejercicio de conciliación presentado, el DANE ocupaba series de mortalidad y fecundidad previamente estimadas y hacia simulaciones con el nivel de omisión censal para el 2018 basado en las diferentes fuentes utilizadas. Se resaltó al equipo del DANE la importancia de hacer nuevas estimaciones de fecundidad y mortalidad para el periodo de referencia censal y también realizar análisis utilizando diferentes técnicas demográficas con las variables de defunciones, de fecundidad y de migración del Censo.

Se discutió la posibilidad presentada por el DANE de realizar imputaciones en las variables sin respuesta sobre hijos nacidos vivos y sobrevivientes. Dichas imputaciones no son recomendadas por el equipo de la misión. Es preferible mantener la categoría “ignorado”.

Debe subrayarse que las estimaciones de fecundidad y mortalidad son hechas basada en las diferentes piezas de un rompecabezas que incluye todas las fuentes posibles y los resultados de diferentes métodos indirectos para definición de la serie temporal de cada componente demográfico analizado. En ese sentido, se destacó la importancia de estimar el nivel de completitud asumido con las estimaciones finales de mortalidad y fecundidad, a fin de tener resultados coherentes con lo esperado del nivel de completitud y el mejoramiento del sistema de registros del país en el tiempo.



Fueron aclaradas también algunas dudas específicas respecto al tratamiento de datos sin información de edad en el cálculo de la fecundidad y mortalidad. Se recomendaron métodos y material de referencia para complementar el análisis hecho por el equipo DANE y se recomendó no hacer diagrama de Lexis para estimar población 0-4 del Censo 2018 sin ajustar los nacimientos y defunciones por subregistro y considerar los registros tardíos.

Respecto a la estimación del saldo migratorio, la técnica utilizada por el DANE basado en entrada y salida de personas en el país es innovadora y debe ser continuada y refinada. El método se refiere a las entradas y salidas registradas por Migración Colombia. Las fronteras de Colombia, sin embargo, son muy porosas, por lo cual las entradas y salidas regulares muchas veces no son registradas y menos las entradas y salidas irregulares.

El método utilizado para estimar el saldo migratorio, por el momento, elimina personas con secuencias incoherentes (doble entrada o doble salida) en el registro de migración y considera inmigrante las personas que está en el país hacen más de 12 meses, o sea, que hayan entrado y no poseen un registro de salida en ese período. Las cifras de entradas de venezolanos en el país muestran que las entradas de venezolanos pasaron de 378,965 en 2016 a 796,234 en 2017 a 1,359,815 en 2018. O sea, la cifra de entrada de venezolanos de manera regular en Colombia más que triplicó entre 2016 y 2018. Para captar el saldo migratorio del período de 2018 se necesitaría la base de 2019 que muy probablemente no estará lista para la conciliación demográfica a ser publicada a fines de este año. Sin embargo, aunque la base esté lista, esa estimación sería apenas del saldo migratorio regular/legal ocurrido en el país.

Por otro lado, el Censo 2018 proporciona resultados de inmigrantes (venezolanos y otros) en el país, que como se vio en el punto previo, pueden ser útiles en la estimación del saldo migratorio para la conciliación demográfica.

### **3.6 Otros: Cartografía, tecnología y operativo censal**

En general, se observa que algunas decisiones estratégicas afectaron de forma importante la operacionalización del censo y, por lo tanto, su cobertura. Se identificaron varios puntos críticos que no deberían repetirse en censos futuros.

**Ausencia de pre-censo:** la falta de un pre-censo estándar y, por consecuencia, de una actualización cartográfica efectiva es considerado un factor problemático para una operación de terreno de la magnitud del censo de población y vivienda y es uno de los factores que conspiraron contra la cobertura efectiva del censo. Adicionalmente la falta de información sobre el uso de las estructuras y el número de ocupantes de la vivienda, solo identificable si se logró la visita durante el operativo, limita la información que se cuenta para el análisis y la identificación de la distribución del error de cobertura.

**E-censo:** el e-censo presentó varios problemas relacionados con el diseño del aplicativo, el soporte de conectividad, la declaración de la información, la mantención de conceptos básicos propios de los censos y la integración con las otras bases de datos, entre otros. Además, se mencionó que la obligatoriedad de utilizar el e-censo para los funcionarios públicos y militares generó la percepción de que era obligatorio solo para dichas personas y no necesariamente para todo el hogar al cual pertenecían. Esto se ve reflejado en un porcentaje muy alto de personas e-censadas en hogares unipersonales.



La decisión de incluir a los militares sobre la marcha del operativo tiene consecuencias sobre el vínculo con los hogares de pertenencia de las personas y genera posibles inconsistencias sobre la información proveniente del empadronamiento de los Lugares Especiales de Alojamiento (LEAs).

Se constató que tanto el operativo censal con DMC y con papel incluyeron una pregunta que indagaba si el informante o algún miembro del hogar había diligenciado el e-censo. Pero en caso de una respuesta afirmativa se culminaba la entrevista sin preguntar por el número de personas que integraban ese hogar. También hubiera sido deseable indagar cuantas de ellas habían respondido por internet. Estas 2 preguntas hubieran permitido recuperar población omitida dentro de los hogares (que a su vez disminuirían el número de hogares unipersonales observados).

El e-censo terminó tardando más tiempo que el previsto, cubrió una población inferior a la planeada inicialmente, captó la información con una calidad menor que la estimada y menor que la recogida con DMC, provocó desafíos complejos para la georreferenciación de los hogares que lo usaron, registró altas tasas de “no finalización”, obligó a procesos complejos (y no perfectos) para evitar duplicación de casos con el levantamiento vía DMC, no siguió algunos conceptos establecidos para el censo, y generó algunos casos inconsistentes sin solución aparente. Las razones subyacentes a estos problemas son múltiples, pero las principales parecen haber sido la subestimación de los desafíos de esta modalidad de empadronamiento (por ejemplo, la comprensión de los conceptos censales por partes de los informantes), la sobreestimación de las ventajas de este procedimiento (como la calidad de la autorespuesta, la idea de que los formularios estarían completos y la facilidad para su monitoreo) y la superposición con el procedimiento de recolección vía DMC. A lo anterior, cabe añadir algunos problemas prácticos y técnicos derivados de la novedad del procedimiento, tales como la adecuación tecnológica requerida para facilitar el soporte de conectividad de la plataforma durante épocas de intenso uso de la misma.

**Fecha de referencia población censada:** la larga extensión del período censal que comenzó en enero de 2018 con el e-censo y se extendió oficialmente hasta agosto del mismo año con los otros procedimientos de recolección de datos, pero que luego continuó hasta diciembre con la fase de recuperación de cobertura. La falta de una fecha de referencia del censo y la temporalidad diferencial de la gran mayoría de las preguntas, en función de la fecha de levantamiento, tiene efectos sobre algunas variables que suelen modificarse a lo largo del tiempo y que sin duda varían durante un período tan prolongado. Adicionalmente, aumentó la probabilidad de que en la recolección con DMC se encontraran hogares y personas e-censadas que habían cambiado de residencia. La forma de proceder con estos casos presenta dificultades técnicas y conceptuales y genera mayores riesgos de errores de duplicación o de desajustes entre la información de persona y de vivienda de un mismo individuo. Así, el largo período de recolección de datos del censo de 2018 complejiza la definición de un momento censal representativo, afecta algunas variables y conlleva problema de comparación por la misma dinámica de la población. Además, la no inclusión de una fecha censal de referencia en el cuestionario del censo juntamente con el e-censo y la actualización de las novedades en terreno dificultó la fotografía censal del país en un momento específico, posibilitado por la referencia a una fecha censal durante el empadronamiento.

**Software:** de misiones previas, se observa que el costo de desarrollar un software propio para captura de datos puede no haber sido justificado por la lentitud del programa y procesamiento de mapas en terreno



-lo que dificultó la operación de terreno-, y por el hecho de haber software libres ampliamente utilizados y validados internacionalmente por otros institutos de estadística. Para la fase de crítica e imputación de datos, el DANE prefirió utilizar un software de alto costo (SAS) aunque de menor performance que otros softwares específicos, por ejemplo, el CSPRO creado por el Bureau de Censos de EEUU, que es un programa gratuito y de los más conocidos y utilizados en otros países. Por otro lado, en la etapa de tabulación de datos, el DANE también utilizó el SAS, que también muestra mucha menor performance que otros softwares como el REDATAM (también gratuito), diseñado para procesar información Censal de manera mucho más rápida y eficiente. La actual versión de REDATAM permite además procesar las variables de diferentes bases de datos interconectadas por sus diccionarios, lo que en este caso hubiera permitido comparar ágilmente los valores de las variables “antes” y “después” de las reglas de validación e imputación. Con ello sería posible analizar no sólo los eventuales cambios en la estructura de la variable sino también identificar los flujos de cambio, es decir las categorías de origen y destinos de los casos que cambian al aplicar las reglas.

**Sistemas de monitoreo y control (SMCs):** los sistemas de monitoreo y control, bien diseñados en teoría, parecen no haber funcionado adecuadamente. Según el informe de Carlos Ardila los sistemas de monitoreo cotidianos de la operación censal en terreno, así como los procedimientos de supervisión del empadronamiento, parecen no haber dado abasto para la magnitud de situaciones complejas que tuvo el censo. Lo anterior se dedujo de un conjunto de señales de problemas que estaba teniendo el operativo y que no fueron solventadas con la rapidez necesaria y por la ausencia de una bitácora de este sistema, más allá de que haya estado activo y haya sido usado por el DANE durante el operativo. El Censo implementó cuatro sistemas para el seguimiento y respaldo de la información transmitida: SMS, Sistema de Monitoreo y Control (SM&C), Geo Visor (GV) y respaldo en nube electrónica y en la DMC. Basado en el análisis hecho por el consultor, el único sistema que parece haber funcionado fue el respaldo (para más detalles véase el informe de Carlos Ardila).

**Rotación de personal:** la rotación de personal crítico vinculado al censo y las limitaciones presupuestarias, también constituyen factores que debilitaron el censo y obligaron a decisiones que terminaron afectando algunos procesos y capacidades, por ejemplo, las tecnológicas (como la falta de paquetes de datos en la DMC).

Algunas decisiones de índole estratégico-coyuntural como la coincidencia del censo con un período electoral, también implicaron adversidades para el censo.

#### 4. Consideraciones finales y recomendaciones

Los hallazgos obtenidos durante la misión, que se centraron en identificar los aspectos mejorables del proceso, permiten decir que la omisión, aunque alta, no invalida el uso de los resultados censales para el análisis poblacional y como insumo de las políticas públicas, de estudios académicos y de las próximas tareas del DANE en materia de estimaciones y proyecciones. La calidad de la información recogida en terreno -considerando las variables analizadas- no parece tener errores sistemáticos o problemas graves.



En general, ésta fue sometida a procesos de validación, crítica e imputación estándares que mejoraron la consistencia de los datos, sin introducir sesgos evidentes ni una manipulación tendiente a aparentar una perfección que no se condice con la realidad.

Se recomienda definir una hoja de ruta para la etapa post censal que incluya los requerimientos para proyecciones nacionales y subnacionales, uso y difusión de la información censal, capacitación a usuarios, entre otras. A partir de esta hoja de ruta, se sugiere identificar eventuales requerimientos de asistencia técnica que deberían asegurarse para su desarrollo. Al respecto cabe indicar que durante la misión DANE y UNFPA acordaron un procedimiento para avanzar en la definición de esta hoja de ruta, con una priorización especial en los requerimientos identificados por el DANE para lo que resta del año 2019 y 2020.

Si bien es recomendable el modelo del Comité Nacional de Expertos que hasta ahora viene acompañando el proceso, se requieren condiciones de sostenibilidad que permitan su eventual continuidad.

Es preciso documentar todos los procesos del censo, con las experiencias y decisiones tomadas, de tal forma de poder visualizar cómo resultaron en la práctica las directrices y recomendaciones establecidas por los manuales.

También se recomienda realizar una evaluación pormenorizada del e-censo, habida cuenta de los problemas detectados y sus implicaciones para la cobertura. En particular, resulta crítico evaluar el tema de los hogares unipersonales y los efectos de la ambigüedad generada por el empadronamiento electrónico para los Lugares Especiales de Alojamiento (LEAs).

Respecto a la evaluación de cobertura se recomienda:

- Seguir mejorando la capacidad de pareo de personas en ECV 2018 y Censo 2018 para aplicar el *Dual system*, especialmente por el mejoramiento de la identificación con las bases espejo. De no lograrse el pareo total, realizar un ejercicio de inclusión en la matriz de los casos que no tienen información a partir de las probabilidades de cada grupo de edades de las personas censadas, de haber sido captado por la encuesta (porcentaje de si-si respecto al marginal del censo). Este ejercicio igualmente puede realizarse para controlar su consistencia con los resultados de los procedimientos de pareo que hayan sido realizados con claves más complejas.
- Consultar un técnico especializado sobre la utilización del Bootstrap para el *Dual system*, en principio la misión no recomienda su utilización para corregir el resultado observado del método.
- Realizar el ejercicio de conciliación censal incorporando los resultados de la omisión estimada por el método directo (*Dual system*) y analizar la consistencia con la evolución de los otros componentes del cambio demográfico.
- Considerar el procedimiento aplicado por Magda Ruiz, a efectos comparativos con la omisión nacional derivada de otros métodos, una vez obtenida la estimación total de vivienda provenientes del análisis cartográfico.
- Considerar las fuentes y metodologías discutidas durante la misión de manera integral para estimar las variables demográficas, antes de concluir el ejercicio de conciliación demográfica. Dialogar con otras instituciones, en especial, el registro civil para obtener registros administrativos



de nacimientos y defunciones y con Migración Colombia para estimación coherente del saldo migratorio e inclusión de estimaciones de migrantes irregulares en el saldo migratorio.

En lo que se refiere a la calidad de las variables, se recomienda:

- Continuar con la revisión de las reglas de validación e imputación, considerando las sugerencias específicas realizadas durante la misión. Para ello, es preciso fortalecer la relación entre analistas y procesadores<sup>6</sup> y estar atentos a los efectos secundarios y acumulados de toda imputación.
- Evitar imputar ignorados sobre la base de criterios de baja certeza.
- Analizar y documentar los porcentajes de cambio de las categorías de las variables antes y después de la aplicación de las reglas de validación, consistencia e imputación.
- Documentar todo el proceso y decisiones en términos de los cambios hechos en la base original e incluir una variable que indique si la información fue imputada (se puede ver el ejemplo de la base de datos del censo de Brasil).
- Finalmente, se insiste en que es necesario seguir las recomendaciones internacionales en materia de censos desde el inicio de la operación censal y con la mayor atención posible.

## 5. Seguimiento

Hacia el final de la misión el DANE brindó información relativa a la pregunta de autoidentificación étnica. El CELADE enviará al DANE y al CE observaciones y comentarios sobre la información recibida. Asimismo, el CELADE enviará al DANE los porcentajes de omisión censal derivados de la revisión 2019 del WPP, para los países de América Latina y otros países seleccionados, con el fin de tener un marco de comparación regional.

El CELADE y el UNFPA continuarán apoyando al CE en la elaboración del Informe final de evaluación (el cual está a cargo de los y las especialistas nacionales), como así también atenderá los requerimientos del DANE, todo en la medida de sus posibilidades.

UNFPA apoyará al DANE en la identificación de una hoja de ruta para el cierre de las actividades de análisis y procesamiento y la posterior fase post-censal. Este plan incluirá el apoyo para la buscar alternativas de movilización de recursos y facilitación de un mecanismo para el financiamiento oportuno de la asistencia técnica y los demás procesos de apoyo requeridos.

---

<sup>6</sup> Este fortalecimiento es necesario tanto para el análisis de calidad como para el futuro análisis específico de interpretación de las variables. La tarea de los analistas hasta ahora se mostró dependiente de procesamientos demasiado lentos y con herramientas que en algunos casos ellos no manejan. En esta dirección se recomienda adoptar herramientas de procesamiento de fácil manejo y rápido procesamiento, como el REDATAM, que permita dar más autonomía a los analistas para el manejo de los datos, de modo de que puedan seguir de cerca las consecuencias de las decisiones adoptadas para la validación, la imputación y el procesamiento de los datos.



## Anexo 1: Términos de Referencia

### EVALUACIÓN DEL CENSO NACIONAL DE POBLACIÓN Y VIVIENDA 2018 TÉRMINOS DE REFERENCIA PARA LA MISIÓN CONJUNTA DE ASISTENCIA TÉCNICA UNFPA-LACRO Y CELADE

Junio 2019

#### 1. OBJETIVO

La misión tiene el objetivo de apoyar al DANE y al Comité nacional de Expertos en el análisis de información disponible dispuesta por el DANE, para su revisión al comité evaluador del Censo Nacional de Población y Vivienda (CNPV) realizado en Colombia en 2018, que contempla, por una parte, el análisis sobre los principales procesos censales y del procesamiento de las diferentes bases de datos; por otra, la evaluación de la cobertura, la calidad y consistencia de los resultados obtenidos.

#### 2. ACTIVIDADES

La misión trabajará participando en un taller de evaluación censal que se llevará a cabo del 17 al 21 de junio de 2019, en la ciudad de Bogotá.

Para el desarrollo del taller se conformarán grupos de trabajo en los siguientes temas:

- Evaluación de calidad y cobertura del censo de viviendas;
- Evaluación de calidad y cobertura del censo de población;
- Evaluación de los aspectos relacionados con el procesamiento e incorporación de innovaciones tecnológicas.

Los expertos de la misión se distribuirán entre los grupos de acuerdo con su interés y perfil técnico, y trabajarán junto con los expertos nacionales y el equipo técnico del DANE en:

- a) Analizar la información proporcionada por el DANE y el Comité de Expertos para la evaluación del tema asignado a cada grupo.
- b) Evaluar los criterios de validación de la cobertura y de la calidad de los resultados censales
- c) Evaluar el procesamiento y las normas de consistencia e imputación de las aplicadas a las bases de datos del CNPV 2018.
- d) Revisar los indicadores demográficos y socioeconómicos utilizados como referencia.
- e) Analizar la información recopilada sobre el tema asignado al grupo y emitir conclusiones y formular recomendaciones.
- f) Emitir un concepto preliminar sobre la calidad de la operación censal en los procesos relacionados con el tema asignado al grupo.
- g) Emitir un concepto preliminar sobre la omisión censal.



### **3. PRODUCTO**

Al finalizar la semana, los expertos internacionales miembros de la misión y el Comité Nacional de expertos, entregarán al DANE los resultados y conclusiones obtenidos en el taller y harán sugerencias para la continuación del proceso de evaluación, el cierre del CNPV 2018 y recomendaciones para futuros censos.

### **4. FECHA Y LUGAR DE REALIZACIÓN DE LA ACTIVIDAD**

- a) El lugar de trabajo serán las instalaciones del DANE.
- b) La misión se realizará del 10 al 14 de junio de 2019.

## Anexo 2: Agenda

	LUNES 10	MARTES 11	MIÉRCOLES 12	JUEVES 13	VIERNES 14
MAÑANA	<p>9:00-9:40 Reunión inicial con DANE y CE</p> <p>9:40-10:45 Reunión CE + misión</p> <p>10:45-12:45 Sesión continua con DANE. Liliana Acevedo sobre el proceso censal y situación de bases de datos</p>	<p>9:00-12:30 Presentaciones del DANE</p> <p>Mariana Ospina: características de las personas que no tuvieron un registro de identidad “emparejable” en la ECV 2018</p> <p>Andryu Mendoza: bases de datos y primeros elementos de normas de validación</p> <p>Alejandro Lelio: resultados iniciales del método directo de captura-recaptura (<i>Dual System</i>)</p>	<p>9:30-12:15 Presentaciones DANE. Oscar Acosta sobre normas generales de imputación. Andryu Mendoza y Juan Camilo Trillos sobre imputaciones de vivienda y hogar</p> <p>12:20-13:00 Presentación de Carlos Ardilla sobre informe de evaluación Tics y recuperación de datos en DMC</p>	<p>9:00-13:00 Reunión Misión, UNFPA y Comité de expertos para acordar conclusiones de lo avanzado, definir actividades y solicitudes pendientes, y discutir sobre el cierre de la misión y su seguimiento</p>	<p>7:40 Reunión de cierre con el Director del DANE</p> <p>8:30 Presentación de Magda Ruiz de procedimiento para estimar omisión censal a escala desagregada</p> <p>10:15-12:45 Elaboración de PPT de cierre por parte de la misión</p> <p>12:55-13:45 Presentación de la PPT de la Misión y discusión final</p>
ALMUERZO				13:00-14:30 REUNIÓN DANE, UNFPA Y OTROS ACTORES POR INDICADOR 3.7.2 ODS	
TARDE	<p>14:15-18:00 Sesión continua con DANE. Lelio Alejandro Arias: metodología de estimación de la omisión censal utilizando método directo de captura-recaptura con la ECV 2018</p> <p>Mariana Ospina, Yenni Marín y Jorge Cabezas: estimación de cobertura mediante ejercicio preliminar de conciliación demográfica</p>	<p>14:30-15:30 Discusión sobre avances presentados por el DANE sobre omisión (<i>Dual System</i> y conciliación demográfica preliminar)</p> <p>15:45-17:30: Trabajo con analistas para revisar resultados y cálculos y obtener nuevos resultados con base 4</p>	<p>14:30-18:00 Presentación DANE. Juan Camilo Trillos sobre imputación de fecundidad, sexo, edad y parentesco</p>	<p>15:00-16:00 Presentación DANE (Alejandro Lelio): nueva aplicación de emparejamiento para estimar la omisión mediante sistema dual</p> <p>16:00-17:30 Video conferencia con Raul Ponce sobre avance de la estimación de problemas de cartografía con datos del levantamiento censal</p>	



### Anexo 3: Detalle del Análisis sobre Cobertura

*Como todo censo, el de Colombia 2018 tuvo subenumeración, atribuible a diferentes factores. Medir el subregistro es siempre un asunto complejo y polémico*

Para el análisis de cobertura realizado desde el DANE fueron implementadas dos estrategias: 1-Encuesta postcensal y 2-comparación con estimaciones de la evolución demográfica. Desde el Comité Nacional de Expertos se está aplicando, además, una estimación de la omisión de personas a partir de la omisión de viviendas (en base a información cartográfica).

#### 1. Encuesta de cobertura

La Encuesta de Post Enumeración (EPE) tiene ciertas características que deben asegurarse: a) lograr la independencia de la información proveniente del censo y de la encuesta, y b) permitir la comparación de la información a nivel individual.

-Dado que se comparan dos fuentes para los mismos datos, en teoría, ambas pueden contener los mismos errores, produciendo resultados perfectamente consistentes, pero sistemáticamente defectuosos. Es importante y no trivial es que los errores en ambas fuentes sean estadísticamente independientes.

-Dado que se comparan datos individuales, es imprescindible que exista suficiente información para realizar un pareo (*matching*) preciso entre los registros de encuesta y el censo.

La independencia entre las dos operaciones busca lograrse mediante el mantenimiento de *Independencia operacional*. Se debe hacer todo lo posible para que los resultados del censo y la EPE sean tan independientes como sea posible. Esto significa tener una unidad específicamente designada para la encuesta que la encuesta sea llevada a cabo por personal diferente, que no tenga ninguna responsabilidad relacionada con el censo. LA EPE debe además usar los mismos mapas que el censo, pero utilizando procesos diferentes para la selección de casos y confección de listados de viviendas a encuestar. Idealmente, el procesamiento de datos de la EPE también debería ser realizado por un equipo diferente, para evitar repetir posibles errores de procesamiento de datos que pudieran ser cometidos en el censo.

Además de procurar la independencia y garantizar el pareo de personas, para la evaluación de contenido, es esencial también que los mismos sean recogidos de la misma forma y con los mismos conceptos y definiciones de base.

En ausencia de una encuesta de post empadronamiento específica, los resultados de otras encuestas de hogares, realizadas en una fecha cercana al censo, también podrían ser utilizados para evaluar sus resultados si cumplen las condiciones mencionadas<sup>7</sup>.

---

<sup>7</sup> Teóricamente, cualquier muestra probabilística de hogares o individuos que cumpla con las condiciones definidas podría ser utilizada como referencia para evaluar cobertura y errores de contenido de un censo.

La Encuesta de Calidad de Vida (ECV) realizada durante los meses de noviembre y diciembre de 2018, cumplen los requisitos de independencia operacional, ya que es dirigida por una unidad de relevamiento continuo del DANE, con personal específico y especialmente capacitado para esta tarea.

La principal observación respecto de la utilización de la ECV para evaluar cobertura y calidad refiere al marco de muestreo utilizado. La ECV utilizó como marco de muestreo el derivado del censo de 2005 y no el derivado de la actualización cartográfica utilizada para el censo. Es decir, que existe una parte de la población censada que por efecto-marco tendrá una probabilidad diferente de ser encontrada en la encuesta, alterando por ello el valor calculado de la omisión<sup>8</sup>.

Respecto al método de pareo (matching) se destaca que la ECV dispuso de métodos de identificación muy confiables para posibilitar el cruce de las personas encuestadas con las personas censadas (el tipo y número de doc. de identidad de las personas, la fonética del nombre y apellido y la edad).

Como primer indicio de cobertura se constató que un 83.1 % de las personas encuestadas (83.9% considerando la población expandida) respondieron que SI habían sido censadas.

**INFORMACIÓN PARA TODOS** **Resultados Muestrales y expandido ECV**

**Durante el año 2018, ¿... respondió al DANE el Censo Nacional de Población y Vivienda?**

	TOTAL (muestra)	TOTAL (expandido)
Si	83.1%	83.9%
No	13.9%	13.5%
No sabe, no responde	3.0%	2.6%
Total	283.012	49.967.281

**¿Cómo fue censado...?**

	TOTAL (muestra)	TOTAL (expandido)
Diligenció el cuestionario por internet (eCenso)	4.4%	7.3%
Lo visitaron en su vivienda	94.5%	91.9%
Lo convocaron a un lugar diferente a su vivienda (Convocatoria)	1.1%	0.9%
Total	233.148	41.953.253

Fuente: DANE Dirección de Metodología y Producción Estadística ECV 2018

Sin embargo, no es el porcentaje de respuestas en la encuesta la que define el nivel de cobertura, sino que es el porcentaje de personas que fue captada por uno y otro relevamiento el que permite realizar una estimación de la población total a través del *Dual system* y calcular la cobertura del operativo censal.

Durante el procedimiento de pareo, a pesar de contar con el tipo y número de documento de identidad, surgieron dificultades para que la población pudiera ser comparada. Se observaron casos donde las personas no tenían documento de identidad o el mismo era inválido<sup>9</sup>. Hubo 20 mil personas encuestadas

<sup>8</sup> Se estudiar el impacto que esto produce en el cálculo final de la omisión.

<sup>9</sup> En el caso de E-censo el documento ingresado por la persona era validado mediante un servicio de *web service* contra la registraduría de personas. Esta validación no fue posible ni con los DMC ni en papel.



y 5.9 millones de personas censadas con problemas de documentación, que no pudieron entrar en el pareo para la implementación del *Dual system* y el cálculo de la población estimada total.

En procura de incorporar las personas con problemas de documentación en la matriz de *Dual System* se sugirió estudiar algunas características de estos subgrupos de población (edad, país de nacimiento, nivel educativo) para poder sustentar su incorporación proporcional al cálculo de la población estimada.

Se recibió esta caracterización para la población encuestada *con* y *sin* problemas de documentación (población que puede “entrar” y población que “no puede entrar” al proceso de pareo).

Asimismo, fue caracterizada (por edades) la población censada que “no pudo entrar” al proceso de pareo. Se solicitó información adicional sobre la distribución por edades quinquenales del total de personas censadas con información para el pareo y del total de personas censadas sin dicha información, a los efectos de calcular la probabilidad diferencial por edades quinquenales de pertenecer al grupo de personas censadas que han sido encuestadas (grupo “si-si”). Dicha información aún no ha sido recibida.

En el último día de la misión, se presentó una nueva versión de la matriz de *Dual system*, donde se pudo parear una mayor cantidad de población (en viviendas particulares) a partir de la consideración de los nombres de las personas.

En esta ocasión se informó por parte de los técnicos del DANE que se habían incorporado a la matriz aquellas personas encuestadas que habían respondido “SI” cuando se les preguntó si habían sido censados.

Los integrantes de la misión solicitaron que se procure incorporar todas las personas encuestadas y no solo aquellas que hayan respondido que SI fueron censadas. Es decir, se solicitó integrar también aquellas que respondieron No y No sé a la pregunta que indaga si fueron censados.

Este ejercicio fue completado sobre el final de la misión generando una población estimada de 48.463.000 personas. Se informó que la inclusión presentó situaciones de pareos múltiple que distorsionan los resultados, ante lo cual se sugirió volver a realizar el ejercicio incorporando la en la clave de pareo (nombre-apellido-edad, y si es necesario alguna otra variable) de modo de eliminar los pareos múltiples que surgen de la consideración solamente del nombre y el apellido.

Todos los ejercicios y simulaciones realizadas generaron una población total estimada entre 47.7 y 48.7 millones de personas, lo que significa un rango de omisión que va de 8.6% <sup>10</sup> (Dual system 14-06-2019) a 11% (Estimación demográfica de CELADE).

---

<sup>10</sup> No se considera en este intervalo el valor 7.1% proveniente del ejercicio de conciliación realizado por el DANE por considerar que aún es muy preliminar.



## 2. Actualización cartográfica / Precenso

Dado que no se realizó un precenso, solo es posible ensayar este análisis a partir de la cartografía disponible y de los antecedentes de censos anteriores.

El análisis se basa en la elaboración de una estimación de personas a partir de la consideración de la cantidad de viviendas y de los promedios de personas por vivienda, a nivel municipal. Esta estimación es comparada con la población censada lo que permite obtener una dimensión de la omisión. La ventaja de esta modalidad de análisis es que se pueden obtener niveles de omisión desagregados a nivel municipal. Cabe señalar que aun cuando la estimación total de omisión que surge de la metodología no fuera aceptada, es posible igualmente descomponer el nivel de omisión total obtenida por los otros métodos, y determinar niveles de omisión a nivel municipal. Estas estimaciones constituyen un valioso aporte de información desagregada que no es posible obtener por los otros métodos de estimación de la omisión.

Los insumos básicos para esta metodología son: la cantidad de viviendas, los niveles de omisión de las mismas obtenidas de revisiones cartográficas, la condición de ocupación de las viviendas y todos los indicios posibles del tamaño medio de los hogares disponibles (encuestas, censo actual y censos anteriores).

Dados los datos obtenidos hasta el momento se ha avanzado en el cálculo de la estimación de la omisión de población (del orden de 5%<sup>11</sup>) debida exclusivamente a la omisión al interior de las viviendas ya censadas y de las mal clasificadas como desocupadas. Para completar el ejercicio se debe incorporar la información de la omisión de viviendas proveniente del análisis de la cartografía por municipio.

Se espera que el análisis de cartografía brinde una estimación de la omisión de viviendas a nivel nacional que al ingresarse al modelo permitirá completar la estimación de la omisión de personas en el censo.

## 3. Comparación con estimaciones de la evolución demográfica (conciliación censal)

Tanto el DANE como la CEPAL han elaborado estimaciones de población al 2018 que permiten la comparación con la población censada y la estimación de una omisión censal. Estos elementos son presentados en el cuerpo principal del informe.

Este ejercicio fue revisado y comentado por la misión, identificando algunos aspectos metodológicos y de fuentes que pueden ser mejorados para la continuación del ajuste de la conciliación.

La misión entiende que aún no se han completado los insumos para cerrar el ejercicio de conciliación censal y que se deben considerar las fuentes y metodologías discutidas durante la misión de manera integral para estimar las variables demográficas.

---

<sup>11</sup> Estimación factible de mejorar a partir de insumos más ajustados.