



**Departamento Administrativo Nacional de
Estadística
DANE**

**Censo de población y vivienda de Colombia
Año 2005**

**Análisis de la estructura y composición de
las principales variables demográficas y
socioeconómicas del Censo 2005**

Informe final

Grupo Mixto nro. 2.2

**Integrantes:
Edgar Baldión Waldron
Lina Marcela Salamanca
Clara Inés González**

**Bogotá, D.C.
30 de septiembre de 2008**

CONTENIDO

Introducción	3
1. Marco general.....	4
1.1 Objetivos.....	4
1.2 Metodología.....	4
1.3 Procesos de depuración.....	5
1.4 Proceso de análisis.....	5
1.4.1 Determinación de la línea de base	5
1.4.2 Utilidad del análisis básico.....	6
1.4.3 Definición de los rangos de aceptación o rechazo (tolerancias):	6
1.4.4 Integridad global	6
1.4.5 Coherencia con información de otras fuentes	7
2. Resultados.....	7
2.1 Variables demográficas	7
2.1.1 Distribución espacial.....	7
2.1.2 Edad	14
2.1.3 Sexo.....	20
2.1.4 Parentesco.....	27
2.1.5 Estado conyugal	28
2.1.6 Pertenencia étnica	29
2.1.7 Fecundidad	30
2.1.8 Mortalidad	37
2.1.9 Migración	39
2.2 Variables socioeconómicas	43
2.2.1 Educación.....	43
2.2.2 Actividad en la semana previa al Censo.....	55
2.2.3 Rama de actividad económica.....	57
2.3 Variables de la vivienda y el hogar	58
2.3.1 Variables de la vivienda.....	58
2.3.2 Variables del hogar.....	60
Resumen	61
Conclusiones y recomendaciones	66
Anexos.....	69
Anexo A. Metodología para determinar los niveles de subregistro de los nacimientos de las mujeres censadas en 2006	69
Anexo B. Respuesta institucional al problema detectado en las variables educativas y en las relacionadas con la actividad de las personas en la semana anterior al Censo 2005..	77

INTRODUCCIÓN

Una de las recomendaciones de la Comisión Internacional de Evaluación fue la creación por parte del DANE de varios grupos mixtos (con representación interna y externa) con el objetivo de estudiar a fondo temas derivados del Censo para los cuales subsisten dudas y con el objetivo adicional de contribuir a la documentación final del Censo.

La conformación de los grupos mixtos de trabajo toma como punto de referencia las 7 áreas temáticas recomendadas por la Misión y la selección preliminar de 4 áreas grandes de trabajo, que recogen los aspectos centrales del censo: (1) análisis de los procesos previos al procesamiento de la información; (2) procesamiento de la información e impacto sobre los resultados; (3) análisis de la cobertura censal; y (4) información recolectada sobre las unidades económicas independientes y la actividad agropecuaria asociada a la vivienda.

Como complemento e insumo para precisar el alcance de los grupos de trabajo, se tomó el inventario de los eventuales puntos críticos en cada uno de los componentes de los procesos centrales del censo.

Con base en los fundamentos presentados se configuraron 5 grupos de trabajo:

1. Análisis de los procesos previos al procesamiento de la información
2. El procesamiento de la información e impacto sobre los resultados, que dada la especificidad de los temas que encierra, dio lugar a la conformación de dos grupos de trabajo:
 - a. Tecnología de la información
 - b. Análisis de la integridad y calidad de la información
3. Análisis de la cobertura censal
4. Información recolectada sobre las unidades económicas independientes y la actividad agropecuaria asociada a la vivienda.

En este documento, se abordan temas correspondientes a la segunda parte de los alcances especificados para el grupo 2: análisis de la integridad y calidad de la información. En síntesis, se analizan los aspectos relacionados con la depuración de la información, una vez que ha sido transcrita de los DMC e integrada a bases de datos objeto de procesos de limpieza de algunos datos incoherentes o sobre los cuales no se obtuvo respuesta. Estos procesos de depuración están orientados a mejorar la calidad de la información obtenida durante el recuento; no obstante, de no ser utilizados correctamente, pueden causar un efecto negativo en la calidad de los resultados definitivos del censo y que por tanto, ameritaría rehacer algunos de los ejercicios implementados.

En este contexto, se analizan y valoran los cambios experimentados por los datos básicos, como producto del tratamiento de la información censal; se mira la correspondencia de los resultados censales con los obtenidos mediante otras fuentes, y se sugieren algunas medidas orientadas a evitar que los errores detectados se vuelvan a producir en futuros censos.

El grupo de trabajo agradece el apoyo brindado por el personal de la Dirección de Censos y Demografía, pero en particular, el soporte técnico y colaboración de los doctores Édgar Sardi, Guillermo Hernández y Raúl Pérez.

1. MARCO GENERAL

Cuatro principios orientan el análisis:

1. En todos los censos existen errores de cobertura y de contenido: los primeros cuando hogares y personas no quedan incorporados en el censo; y los segundos, cuando el encuestador o el declarante cometen errores al anotar o proporcionar los datos sobre características individuales o de los hogares.
2. La corrección de los datos incoherentes (consistencia interna) y la adición de datos que no existen en el cuestionarios, pero que admiten que se formulen supuestos razonables al respecto (imputación), son práctica común en los censos.
3. Los procesos de depuración (consistencia interna, imputación, otros) no deben alterar la distribución espacial de la población y tener efectos muy poco significativos o inexistentes sobre la estructura y composición de la población.
4. La integralidad de la información tiene que preservarse durante todas las fases del procesamiento de la información.

1.1 Objetivos

- Verificar que la *integridad de la información* permanece una vez que experimenta procesos de depuración durante el procesamiento.
- Valorar el impacto que los procesos de depuración tienen sobre los resultados del Censo 2005.
- Analizar la correspondencia entre las estructuras del Censo 2005 con las que se obtienen a través de otras fuentes.
- Plantear recomendaciones para la utilización de las bases de datos.
- Establecer lineamientos útiles para futuros censos.

1.2 Metodología

En síntesis, la metodología radica en analizar los cambios que experimenta la información entre la base inicial (cruda) y la base resultado de todos los procesos de depuración (final)¹. De encontrarse variaciones que ameriten profundización, se procede a analizar las bases intermedias correspondientes a cada uno de los procesos de depuración (control de estructura 1 y 2, imputación), con el fin de precisar la fuente de error.

¹ Incluye únicamente los procesos normales de depuración, control de estructuras e imputación; no contempla los ajustes realizados mediante la conciliación censal.

Con base en los resultados de otras fuentes de información (censos anteriores, encuestas), analizar la coherencia de los resultados del Censo 2005 para cada una de las variables seleccionadas.

1.3 Procesos de depuración

Los procesos de depuración que experimenta la información censal corresponden a:

Control de estructura 1: busca lograr la integridad de los datos para permitir la reenumeración de las edificaciones y unidades censales y así garantizar la integridad del registro y codificación de los mismos. También, permite suprimir los registros de datos sobrantes y las duplicaciones y controlar las variables de área de las unidades agropecuarias, la unicidad de los registros de LEA, la codificación de las unidades económicas según CIU a 4 dígitos y la codificación de los grupos étnicos.

Control de estructura 2: es complemento del control de estructura 1, que busca la integridad de las diferentes unidades censales y la corrección de algunas variables respuesta que hacen parte de la estructura, como el caso de las Unidades Censales –UC– de vivienda o de hogar, o de las variables de parentesco, sexo, edad y estado civil, y centrarse esencialmente en lo que corresponde a la edad válida, la unicidad tanto del jefe como de cónyuge y la edad válida de este último.

Imputación: este proceso permite depurar los datos censales, para generar los resultados básicos del Censo, al validar la consistencia de las variables mediante procesos lógicos de imputación. Corresponde al proceso más importante de la depuración de la información, se utiliza para resolver problemas de no respuesta “missing”, inválidas o inconsistentes, identificadas en el proceso de edición (identificación de registros potencialmente erróneos).

1.4 Proceso de análisis

1.4.1 Determinación de la línea de base

Corresponde a un conjunto reducido de variables de las personas, las viviendas y el hogar, investigadas en el Censo 2005 y a algunos indicadores seleccionados; a todos ellos se les realiza el análisis descrito en la metodología. Su aplicación toma como fundamento la base de datos “crudos”, producto de la recolección de la información.

La selección de las variables que conforman la línea de base responden a: (1) las variables fundamentales del Censo 2005; (2) aquellas de interés producto de la revisión documental, o de particular atracción por parte de los investigadores; (3) las que surgen por recomendaciones de profesionales internos y externos al DANE; y (4) otras de interés por la magnitud del proceso de depuración que experimenta.

Línea de base

Variables demográficas:

- Volumen y distribución espacial

- Edad
- Sexo
- Parentesco
- Estado conyugal
- Etnia
- Fecundidad
- Mortalidad
- Migración interna y al exterior.

Variables socioeconómicas:

- Educativas: analfabetismo, asistencia escolar, nivel educativo y último año aprobado
- Actividad y rama de actividad económica.

Variables de la vivienda y el hogar:

- Vivienda: tipo y tenencia de la vivienda y servicios públicos básicos
- Hogar: tamaño del hogar.

1.4.2 Utilidad del análisis básico

- Analizar las características de la información cruda: valores válidos, valores perdidos
- Determinar las variables objeto de profundización
- Focalizar geográficamente la incidencia de las variaciones: clase, departamentos y municipios, si se considera necesario
- Orientar la definición de indicadores.

1.4.3 Definición de los rangos de aceptación o rechazo (tolerancias):

Corresponde al establecimiento de los niveles de tolerancia aceptados para las variaciones en cada una de las variables y sus categorías, que conforman la línea de base. Se aceptó como rango válido un punto porcentual de diferencia.

1.4.4 Integridad global

La comparación de los resultados de la línea de base, calculados con la base cruda y los que arroja la base final, permite determinar si la información de cada una de las variables bajo análisis, se encuentra en los niveles de tolerancia esperados; de ser así, la valoración de integridad es positiva; si los resultados del análisis se encuentran por fuera de los niveles de tolerancia, se acude a las bases intermedias (estructura 1, estructura 2, imputación), con el fin de establecer la fuente que genera la variación.

Las variables de la línea de base pueden experimentar cambios significativos debido a que: (1) los procesos de depuración diseñados no son los adecuados; y (2) el software elaborado no corresponde a las normas de depuración planteadas. Cualquiera que sea el caso, debe analizarse y valorarse el proceso realizado.

La base construida con los resultados del proceso CSPRO de depuración, permite obtener las estadísticas del proceso realizado en estructura 2, edición e imputación. De aquí pueden salir otros indicadores, considerando el nivel de imputación de cada variable.

1.4.5 Coherencia con información de otras fuentes

Con la finalidad de realizar algunos alcances sobre la calidad de la información, se analiza la coherencia entre las estructuras básicas de variables o indicadores seleccionados con las que suministran otras fuentes; para el efecto, se lleva a cabo:

- Definición de variables e indicadores con base en el análisis documental y las recomendaciones de expertos.
- De encontrarse diferencias, se plantean las probables fuentes que explican la variación.

2. RESULTADOS

2.1 Variables demográficas

2.1.1 Distribución espacial

Distribución de la población por clase²: las diferencias por clase no deberían existir; sin embargo, aunque los cambios en los volúmenes totales son relativamente muy pequeños, estas diferencias se presentan; su impacto sobre los resultados es desestimable, pero a nivel municipal puede tener significado (cuadro 1).

Cuadro 1
Distribución de la población, por base, según clase
2005

Clase	Base cruda		Base final		Diferencia
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	
Total	41 174 853	100,0	41 174 853	100,0	0
Urbano	31 279 657	76,0	31 282 708	76,0	-3 051
Centro poblado	2 302 581	5,6	2 297 386	5,6	5 195
Rural disperso	7 592 615	18,4	7 594 759	18,4	-2 144

Fuente: DANE –Censo General 2005

² Clase corresponde a:

Cabecera municipal: es el área geográfica que está definida por un perímetro urbano, cuyos límites se establecen por acuerdos del Concejo Municipal. Es donde se localiza la sede de la Alcaldía municipal.

Centro poblado: se define como una concentración de mínimo 20 viviendas contiguas, vecinas o adosadas entre sí. Dicha concentración presenta características urbanas tales como la delimitación de vías vehiculares y peatonales.

Rural disperso: es el área que se caracteriza por la disposición dispersa de viviendas y explotaciones agropecuarias. No cuenta con un trazado o nomenclatura de calles, carreteras, avenidas, y demás. (DANE. Censo General 2005. Nivel nacional. Bogotá. 2008. Pág. 446).

73 municipios, el 7% del total en el país, presentan en la base final población total diferente a la observada en la base cruda, lo cual se produjo porque algunas áreas geográficas pertenecientes a una localidad, fueron inicialmente incluidas como pertenecientes a otra. Así, las diferencias se compensan, 37 de ellas son positivas, mientras que

Cuadro 2
Municipios con diferencias entre la base cruda y la final
2005

Municipio	Diferencia	Municipio	Diferencia
Total	18 293		-18 293
Medellín	3 799	Itagüí	-3 799
Sucre	2 862	Sabanas San Ángel	-1 769
Lorica	1 919	Cimitarra	-1 281
Pivijay	1 769	San Pelayo	-1 237
Guamal	1 006	Vélez	-1 088
La Cruz	909	San Sebastián	-1 006
Puracé	765	Moñitos	-944
Tarso	763	La Plata	-765
Calima	604	Venecia	-763
El Tambo	447	El Peñón	-712
Bolívar	423	Yotoco	-618
Quípama	410	La Unión	-553
Cotorra	395	Timbío	-447
Villa Nueva	243	Belén	-419
El Guacamayo	226	Otanche	-410
Caramanta	225	Aguada	-323
Caldas	210	Valparaiso	-301
Silos	181	Jordán	-243
Jesús María	181	Simijaca	-206
Vista Hermosa	179	Zarzal	-191
Caicedonia	164	San Juan de Arama	-179
Totoró	160	Piedecuesta	-178
Villeta	111	Silvia	-160
La Pintada	97	Purísima	-159
Colón	63	La Belleza	-149
El Dovio	48	Guavatá	-134
Sevilla	27	Bituima	-111
Chima	26	La Unión	-48
La Paz	14	Florián	-31
Vijes	14	Fredonia	-21
Santa Bárbara	13	Guaca	-19
Landázuri	12	Málaga	-9
San José de Miranda.	9	Paime	-6
San Cayetano	6	Peque	-5
Ituango	5	San Carlos	-5
Montería	5	Carmen de Carupa	-4
Chitagá	3	Vijes	

Fuente: DANE –Censo General 2005

36 presentan el signo contrario (cuadro 2). Por ejemplo, Medellín en la base cruda tenía 3 799 personas más de las que aparecen en la base final; mientras que Itagüí mostraba el caso contrario. Sin embargo, la correspondencia no siempre es uno a uno, ni Sucre ni Loricá tienen su contraparte exacta, lo que seguramente muestra es que los excedentes en la base cruda correspondían a más de un municipio.

El ejercicio de restitución de las AG³ a su municipio de origen asegura que la población corregida sea la que le corresponde; para estos municipios, el impacto de la corrección es positivo. Este proceso se realizó de acuerdo con la evaluación de la base geográfica desarrollada por la Dirección de Geoestadística –DIG–, el cual está completamente documentado.

Los centros poblados: la DIG entregó al Censo 2005 la base *homologada* que contiene todos los centros poblados incluidos en la cartografía; de igual forma, la DIG cuenta con la Divipola del año 2007 que incluye no sólo los registros actualizados de la base homologada, sino también aquellos que las autoridades municipales reportaron como existentes en sus localidades.

El cruce de las anteriores bases con la respectiva información incorporada en el Censo General 2005, da como resultado que 298 centros poblados están en la Divipola y en la base homologada, pero no aparecen en el Censo 2005 (cuadro 3). De esta cifra, deben descontarse 102, que corresponden a municipios en donde, por distintas razones, no se llevó a cabo el censo en la zona rural (centro poblado y rural disperso) o su cobertura fue muy baja.

Los 196 centros poblados restantes pueden no aparecer en el Censo 2005 por distintas razones: (1) pueden haberse clasificado como clase 3, rural disperso ante la no existencia de cartografía para dicho centro poblado; (2) aunque la actualización de la Divipola reduce significativamente esta posibilidad, puede haberse dado que el crecimiento de los centros urbanos haya absorbido e integrado a su casco algunos de ellos (la actualización de 2004 no fue sobre todos los 1 119 entes territoriales); (3) se convirtieron en municipios o pasaron a formar parte de municipios recién conformados y (4) pueden no haber sido detectados en el operativo de campo.

La mayor parte de los departamentos experimentan algún grado de no detección o no identificación de centros poblados. El mayor número de ellos se observan en el Valle del Cauca, Antioquia, La Guajira y Arauca.

Aún si se considerara el caso extremo en que ninguno de los 196 centros poblados fue detectado durante el operativo, el impacto sobre los resultados a nivel nacional o departamental es prácticamente inexistente (el tamaño promedio de los centro poblados a nivel nacional es de 160 personas). No obstante, a nivel municipal, la cobertura en algunos de ellos podría verse afectada.

³ *Área Geográfica (AG):* corresponde a una manzana en las cabeceras municipales y en los centros poblados; y a una sección rural, en el área rural dispersa.

Cuadro 3
Centro poblados inexistentes en el Censo 2005,
pero sí en la Divipola y en la base homologada

Departamento	Centros poblados
Total	196
Antioquia	17
Bogotá, D.C., D.C.	2
Bolívar	10
Boyacá	5
Caldas	7
Caquetá	7
Cauca	7
Cesar	4
Córdoba	6
Cundinamarca	8
Chocó	4
Huila	9
La Guajira	17
Magdalena	8
Nariño	8
Norte de Santander	3
Quindío	2
Risaralda	13
Sucre	4
Tolima	8
Valle del Cauca	27
Arauca	15
Casanare	1
Putumayo	1
Guainía	3

Fuente: DANE –Censo General 2005

Las áreas geográficas (AG): la perfecta identificación y localización de las áreas geográficas en terreno es fundamental para la organización, ejecución del operativo de recolección de la información y el aseguramiento de la cobertura.

Para el Censo General 2005, la Dirección de Geoestadística suministró una base con 401 130 registros (AG); de ellos, el 77% en la zona urbana, el 15% en centros poblados y el restante 8% en el rural disperso. El cruce de esta base de AG con las existentes en el Censo, informa que el 17% de ellas no pueden ser identificadas en la base censal, como áreas en las que existen viviendas y personas. A nivel de clase, el diferencial de las AG inexistentes es del 12% en el urbano; el 24% en los centros poblados y el 47% en el rural disperso. Estos resultados no implican que las áreas no hayan sido cubiertas, en la medida en que en las AG se contemplan, también, parques, zonas verdes y áreas no desarrolladas sin viviendas ni personas (cuadro 4).

Cuadro 4
AG en la base DIG existentes en el Censo 2005, según clase

Clase	Total	Existen		No existen	
		AG	Porcentaje	AG	Porcentaje
Total	401 130	333 853	83,2	67 277	16,8
Urbano	308 603	271 519	88,0	37 084	12,0
Centro poblado	58 471	44 427	76,0	14 044	24,0
Rural disperso	34 056	17 907	52,6	16 149	47,4

Fuente: DANE –Censo General 2005

La simbología que debe acompañar a cada AG (1: parque; 2: zona verdes; 3: parque vivienda; 4: zona verde y vivienda; 5: parque y zona verde; 6: parque, zona verde y vivienda; 7: propiedad horizontal; 8: área sin desarrollo) permitiría precisar en cuál existen viviendas; no obstante, esta simbología existe únicamente para 10 554 de ellas, todas de la zona urbana, pertenecientes a sectores de 113 municipios, lo cual limita su utilidad, primero, para la explotación adecuada de su potencial en el Censo 2005; y segundo, para que sirva de referente válido en la estimación de la no cobertura censal (cuadro 5).

Se debe establecer un control adicional, 221 municipios fueron *compensados*⁴, 65 de ellos en la zona urbana, y por tanto, las AG pertenecientes a esta zona no deben tomarse en consideración en el análisis. No obstante, ninguno de los 65 aparece dentro de los 113 con información de la simbología.

Cuadro 5
AG con simbología, según la existencia de viviendas

Concepto	Frecuencia	Porcentaje
Total	10 554	100,0
AG sin viviendas	5 562	52,7
AG con viviendas	4 992	47,3

Fuente: DANE –Censo General 2005

Los estimativos realizados con las 10 554 AG que cuentan con la simbología, arroja los siguientes resultados:

- En las 10 554 áreas con simbología el porcentaje de AG sin vivienda en la zona urbana es de un 53%, lo cual requiere que la DIG genere un proceso de actualización en miras al próximo Censo 2005.
- Al tomar únicamente las AG reportadas con viviendas (4 992), un 13% no existe en el Censo 2005; sin embargo, el total de AG en esta condición es de únicamente 644 (cuadro 6).
- Las 644 AG inexistentes en el Censo 2005 corresponden tanto a municipios grandes, por ejemplo, Medellín, Cali, Barranquilla, como a ciudades intermedias y pequeñas (cuadro 7).

⁴ La compensación es un proceso de ajuste por no cobertura geográfica en aquellos municipios donde el censo no se pudo realizar y se recogió información secundaria a través de las entidades locales.

Cuadro 6
AG con viviendas, según existencia en el Censo 2005

Concepto	Frecuencia	Porcentaje
Total	4 992	100,0
No existen	644	12,9
Existen	4 348	87,1

Fuente: DANE –Censo General 2005

Cuadro 7
AG inexistentes en el Censo 2005, según municipio

Municipios	Frecuencia	Porcentaje
Total	644	100,0
Medellín	39	6,1
Apartadó	1	0,2
Barranquilla	36	5,6
Baranoa	3	0,5
Malambo	3	0,5
Sabanalarga	6	0,9
Soledad	12	1,9
Cartagena	3	0,5
Arjona	1	0,2
El Carmen de Bolívar	2	0,3
Magangué	7	1,1
Tunja	9	1,4
Chiquinquirá	1	0,2
Sogamoso	1	0,2
Manizales	28	4,3
Chinchiná	2	0,3
Villamaría	4	0,6
Florencia	11	1,7
Puerto Tejada	5	0,8
Santander de Quilichao	5	0,8
Valledupar	44	6,8
Aguachica	4	0,6
Agustín Codazzi	1	0,2
Montería	11	1,7
Cereté	1	0,2
Lorica	1	0,2
Planeta Rica	1	0,2
Girardot	84	13
La Mesa	1	0,2
Madrid	1	0,2
Mosquera	1	0,2

Cuadro 7 (conclusión)
AG inexistentes en el Censo 2005, según municipio

Municipios	Frecuencia	Porcentaje
Tena	1	0,2
Zipacquirá	5	0,8
Quibdó	2	0,3
Neiva	1	0,2
Pitalito	1	0,2
Riohacha	5	0,8
Maicao	9	1,4
Santa Marta	5	0,8
Ciénaga	5	0,8
Villavicencio	9	1,4
Acacías	1	0,2
Pasto	21	3,3
Ipiales	1	0,2
Cúcuta	2	0,3
Los Patios	14	2,2
Ocaña	2	0,3
Villa del Rosario	4	0,6
Armenia	8	1,2
Calarcá	1	0,2
Montenegro	2	0,3
Bucaramanga	19	3
Barrancabermeja	9	1,4
Floridablanca	15	2,3
Girón	4	0,6
Piedecuesta	1	0,2
Sincelejo	9	1,4
Corozal	1	0,2
Ibagué	18	2,8
Cali	88	13,7
Buenaventura	2	0,3
Buga	4	0,6
Cartago	5	0,8
Jamundí	3	0,5
Palmira	1	0,2
Pradera	1	0,2
Tuluá	18	2,8
Yumbo	7	1,1
Arauca	6	0,9
Yopal	4	0,6
San Andrés	1	0,2

Fuente: DANE –Censo General 2005

No se considera procedente que con menos de 1 000 AG no detectadas en el Censo 2005 y sin que se haya realizado una mayor profundización en el problema, hacer inferencias sobre la cobertura censal.

Fuente probable de error: en la distribución de la población por clase, el error probable podría atribuirse a la cartografía, ya fuera por desactualización o porque no permitía la delimitación exacta del municipio, pero también a problemas en la organización y realización del operativo.

En el segundo de los casos, a pesar de las razones enunciadas (conversión en municipio, desaparición, contado como parte del rural disperso), no puede descartarse que eventualmente exista un problema de cobertura asociado a la no detección de los mencionados centros.

En la no identificación de AG en el Censo 2005, además de la cartografía y la organización y ejecución del operativo, surge otra fuente probable de error, el que la mayor parte de las AG no cuenta con simbología.

Recomendaciones:

- *De aplicación inmediata:* en coordinación con la DIG, debe adelantarse un trabajo de escritorio para validar los resultados obtenidos; y de acuerdo con los resultados, determinar si se considera procedente llevar a cabo investigación en terreno, mediante una muestra de municipios, para determinar la existencia de centros poblados o AG no existentes como tal en el Censo 2005.
- *Para futuros censos:* actualizar la base de georreferenciación, de forma tal que incorpore las AG con creación cercana al Censo 2005; y que todas las AG cuenten con simbología. A partir de esa base y de otros insumos, generar los mecanismos para el aseguramiento de la cobertura, paralelo a la recolección de la información.

2.1.2 Edad

La recolección de la información mediante DMC y la incorporación de la fecha de nacimiento como control de la edad, mejora significativamente la calidad de la información de la variable básica del Censo 2005: la edad.

Los tres índices analizados muestran que la información capturada en los DMC se encuentra en rangos de buena calidad. Cuando la captura es en papel, y posteriormente grabada en DMC; en dos de los índices (Naciones Unidas y Myers) la información aún se encuentra en los límites de aceptabilidad, más no en el tercero (Whipple), cuando el valor (110,85) corresponde a “datos aproximados” (tabla 1).

Tabla 1
Indicadores de la calidad de la información relacionada con la edad

Indicador de calidad	Definición del indicador	Parámetro de comparación	Resultados de calidad censal			
			Total nacional	Captura en DMC	Cuestionario en papel y grabada en DMC	Papel y escáner
Naciones Unidas	Muestra la mala declaración de la edad y la preferencia de dígitos	<20 Satisfactorio	10,30	10,40	11,70	20,00
		20 y 40 De calidad intermedia				
		>40 Indican información deficiente				
Myers resumen	Evalúa la atracción o rechazo que se dio en la declaración de la edad en los dígitos finales	0,0 a 5,0 Bajo	1,93	1,77	4,45	6,28
		5,1 a 15,0 Mediano				
		15,1 a 30,0 Alto				
		30,1 y más Muy alto				
Whipple	Muestra la mala declaración de la edad, respecto a la preferencia de los dígitos 0 y 5	100 a 105 Muy precisos	103,49	103,03	110,85	116,67
		105 a 110 Relativamente precisos				
		110 a 125 Datos aproximados				
		125 a 175 Datos malos				
		175 y más Datos muy malos				

Fuente: DANE –Censo General 2005

Cuando la recolección se realiza mediante papel y luego se graba mediante escáner, la información, según los tres indicadores es de calidad media. Con la última estrategia únicamente se recopiló y grabó el 2,3% de la información; mientras que con sólo DMC se capturó cerca del 95% (tabla 2).

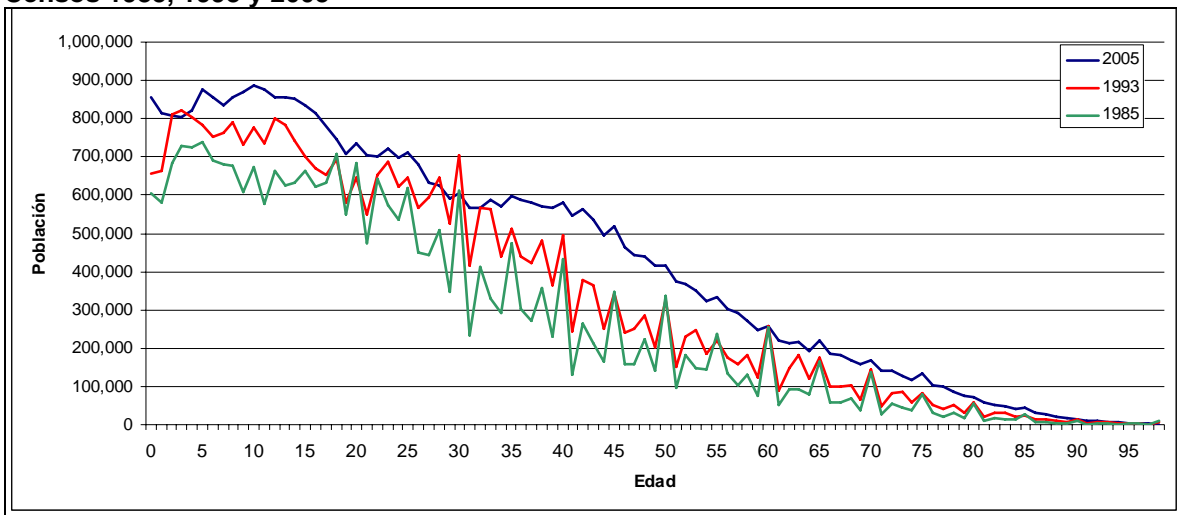
Tabla 2
Personas en la base, según alternativa de recolección y grabación

Medio de captura	Personas	Porcentaje
Captura en DMC	39 248 930	94,5
Papel trascripción DMC	1 253 533	3,0
Papel - escáner	965 921	2,5

Fuente: DANE –Censo General 2005

En el gráfico 1 puede observarse que la distribución por edades del Censo 2005 es mejor en comparación con la del Censo 1993, al no ser tan marcada las preferencias de dígitos.

Gráfico 1
Distribución de la población, por edades simples
Censos 1985, 1993 y 2005



Fuente: DANE

A pesar de las notorias mejoras en la calidad de la información sobre la edad, existen aspectos de interés que llaman la atención:

- A pesar de que se controló por fecha de nacimiento, lo cual reduce significativamente la preferencia de dígitos, aún se observan pequeños abultamientos en las edades terminadas en 0 y 5.
- El primer grupo de edad presenta un comportamiento atípico, no explicado por mortalidad infantil o en eventuales incrementos de la fecundidad en el año antes del Censo 2005: alto en el grupo de edad 0, baja en los tres siguientes (1, 2, 3) y se incrementa para el de 4 años.

La comparación base cruda y base final: el primer ejercicio corresponde a validar los cambios experimentados por la estructura de edades entre la base cruda y la base final una vez se ha realizado el proceso de depuración de la información. De esta comparación se desprende:

- Los valores perdidos (*missing*) de la edad en la base cruda son únicamente 31 530 (perdidos: 21 636; código 99: 9 888; código 999: 6) que corresponden al 0,08% del total (cuadro 8).
- No existe diferencia significativa por edad entre los valores de la base cruda y la final.
- La población menor de 1 año se reduce en 9 671 personas entre la base cruda y la final.
- De existir sobrestimación en la edad 0, esta no proviene de los procesos de depuración de la información.

Cuadro 8
Distribución de la población, por base, según grupos de edad

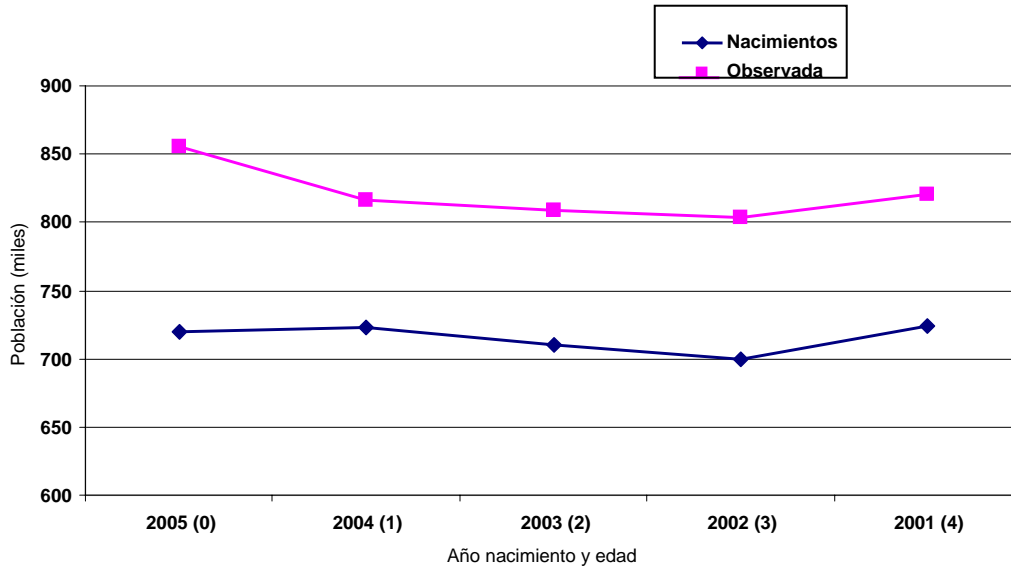
Edad (años)	Base cruda		Base final	
	Personas	Porcentaje	Personas	Porcentaje
Total	41 143 329	100,0	41 174 853	100,0
0	865 602	2,1	855 931	2,1
1 - 4	3 244 372	7,9	3 250 052	7,9
5 - 9	4 291 303	10,4	4 290 594	10,4
10 - 14	4 325 904	10,5	4 325 890	10,5
15 - 19	3 879 808	9,4	3 885 257	9,4
20 - 24	3 554 928	8,6	3 560 869	8,6
25 - 29	3 236 136	7,9	3 241 664	7,9
30 - 34	2 890 379	7,0	2 895 677	7,0
35 - 39	2 898 382	7,0	2 903 378	7,1
40 - 44	2 716 091	6,6	2 721 239	6,6
45 - 49	2 278 243	5,5	2 283 328	5,5
50 - 54	1 824 693	4,4	1 829 635	4,4
55 - 59	1 441 171	3,5	1 446 203	3,5
60 - 64	1 094 873	2,7	1 100 559	2,7
65 - 69	910 696	2,2	916 133	2,2
70 - 74	692 096	1,7	696 740	1,7
75 - 79	493 288	1,2	497 309	1,2
80 - 84	269 844	0,7	272 275	0,7
85 y más	235 520	0,6	202 120	0,5

Fuente: DANE –Censo General 2005

Coherencia de la información: los nacimientos ocurridos en el año censal y en los 4 años anteriores, a pesar del subregistro que experimentan, son un buen referente para mirar la coherencia de los resultados censales en los primeros años de vida. Bajo el supuesto que el porcentaje de subregistro permanece en niveles similares en el período, se observa que el patrón entre nacimiento y población censal, es similar a partir del primer año de vida (la diferencia con respecto a los nacimientos, oscila entre el 13% y el 15%); no así entre los nacidos vivos en 2005 y la población menor de 1 año, en donde la brecha se amplía y la diferencia relativa alcanza el 19% (gráfico 2).

Aunque la evidencia externa indica sobrestimación de la población menor de un año, es preciso aunar mayor evidencia para poder precisar si existen desviaciones significativas frente al patrón de comportamiento esperado, relacionado anteriormente.

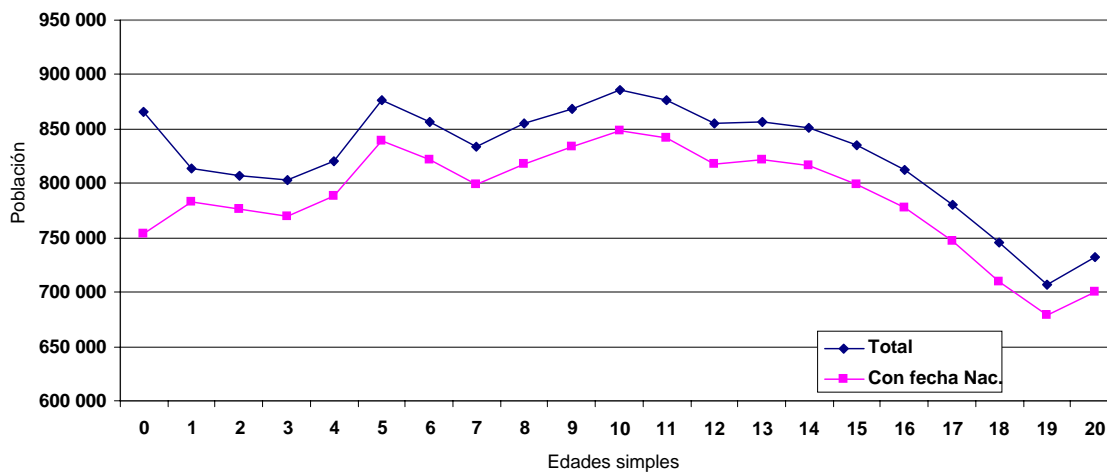
Gráfico 2
Nacimientos en los últimos años (2001-2005) y población de 0 a 4 años
Base final



Fuente: DANE –Censo General 2005

Fuente probable de error: de no existir error en la asignación de la edad, se esperaría que el patrón por edad de la población que no reportó la fecha de nacimiento, fuera similar al de la población que sí lo hizo.

Gráfico 3
Base cruda. Distribución de la población de 20 años y menos según la presencia de valores perdidos 2005



Fuente: DANE –Censo General 2005

En la base cruda existen cerca de 1,9 millones de personas para las que no se cuenta con la fecha de nacimiento. Seleccionando la población de 20 años y menos, se observa que la estructura por edades simples de la población total (incluye a quienes no declararon fecha de nacimiento), es similar, a partir de la edad 1, con la de la población que no considera los valores perdidos para la fecha de nacimiento; pero no es así para la edad 0, en donde el patrón es contrario al esperado.

La evidencia externa y la derivada del Censo 2005, muestra la existencia de sobrestimación en la edad cero (0), las fuentes probables o hipotéticas de error y difícilmente demostrables, se encontrarían en:

- Problemas de capacitación que impedían al encuestador conocer cabalmente el tratamiento de la población para la que se desconocía la fecha de nacimiento.
- Asignación consciente por parte del encuestador de la edad 0, para no realizar preguntas sobre otras variables y acortar así el tiempo de la encuesta.
- Problemas de supervisión para detectar oportunamente la calidad de la información que se estaba recogiendo.
- Problemas de configuración, especialmente en el componente regional, que define separación de decimales, fecha, etc., lo cual es importante para la validación de la edad declarada.

Balance e impacto sobre algunos indicadores:

- La utilización de la tecnología mejora significativamente la distribución por edad de la población y por tanto, puede afirmarse que la información es, en términos generales, superior a la que arrojan los censos anteriores.
- Los procesos de depuración de la información están bien realizados y contribuyen a mejorar la calidad de la información.
- Cuando no se suministra la fecha de nacimiento, eventuales fallas en los DMC, la capacitación y supervisión conducen a que a algunas personas se les asigne edad 0 (cero), sin que esa sea su verdadera edad. El error no se corrige por procesos de imputación, en la medida en que la formulación de este tipo de normas no es posible, dado que implicaría realizar intervenciones técnicamente no aceptables.
- El primer impacto importante de una sobrestimación de la población menor de un año es sobre las proyecciones de población. De no corregirse o suavizarse la estructura por edad, van a generarse errores acumulativos en el tiempo.
- Para algunos programas de gobierno, por ejemplo vacunación de infantes, puede llevar a una sobrestimación de los requerimientos reales.
- Algunos de los indicadores que se generan con base en la población menor de 5 años pueden verse afectados, como:
 - Relación hijos-mujer
 - Hijos sobrevivientes/hijos nacidos vivos

- Estimación de la mortalidad infantil por métodos indirectos que toman como referencia la población de 0 años.

Recomendaciones:

Para la utilización de la información:

- La sobrestimación de la población de 0 años es de particular importancia en el momento de realizar proyecciones de población; se recomienda suavizar o corregir las estructuras mediante multiplicadores u otras técnicas diseñadas para el efecto. Es preciso señalar que la suavización fue realizada en el proceso de conciliación para garantizar la estructura de la población base de las proyecciones.
- En la construcción y análisis de indicadores que utilizan la población de 0 años, deben tomarse en consideración las limitantes que trae la sobrestimación detectada.

Para futuros censos:

- La tecnología y software de captura que vaya a utilizarse debe estar lo suficientemente probada, validada y configurada antes de iniciar el operativo censal.
- Prestar especial atención a los procesos de capacitación y supervisión.

2.1.3 Sexo

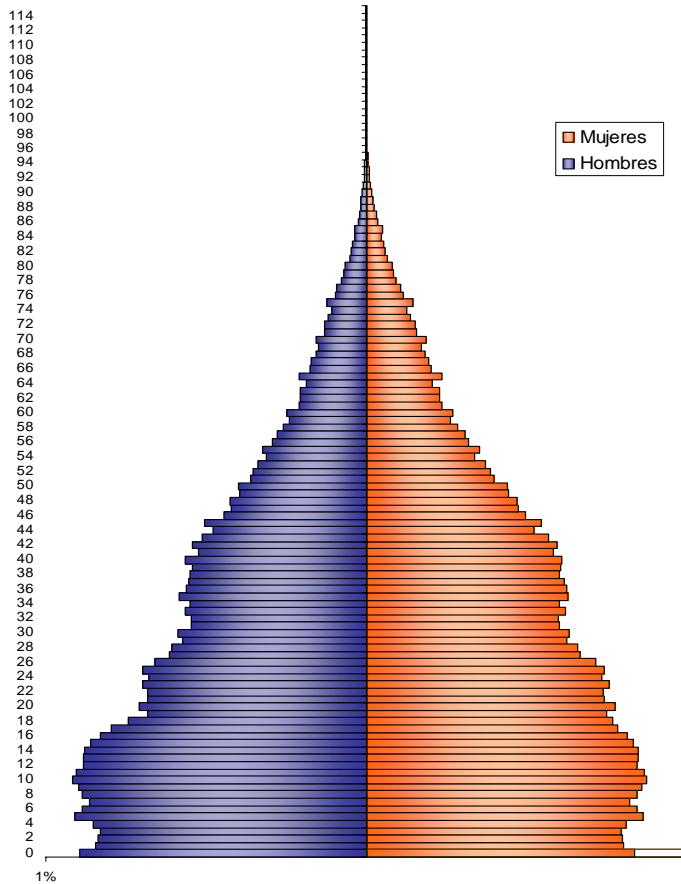
El principal punto de interés relacionado con la calidad de los datos sobre el sexo de las personas, está relacionado con la cobertura diferencial de ellos; al respecto, la pirámide de población por edades simples (gráfico 4) refleja algunos aspectos de interés:

- Muestra la regularidad señalada en el análisis de la edad
- Refleja la sobrestimación de la población menor de 1 año
- Entre los 20 y los 40 años, la regularidad del descenso en la importancia relativa de la población, se pierde y aparecen muescas o hendiduras en los dos sexos. La violencia, que afecta en especial a los hombres en estas edades y la emigración al exterior, con mayor intensidad en las mujeres, pueden ser algunas de las causas de este comportamiento de la pirámide.

La comparación base cruda y base final: la probabilidad de que se presente errores de la variable sexo en la incorporación en los DMC o en el papel, es mucho menor que en las restantes variables. En la base cruda existen únicamente 9 888 (0,02%) valores perdidos.

Una vez efectuada la depuración de la información, la estructura de la población por sexo no cambia entre una base y otra (cuadro 9). Se puede concluir que el proceso de depuración destinado a corregir problemas de la variable sexo, por no coherencia con otras variables o no reportado, fue adecuado.

Gráfico 4
Pirámide de población, por edades simples y sexo
Censo 2005



Fuente: DANE –Censo General 2005

Cuadro 9
Distribución de la población, por sexo y base, según grupos de edad
2005

Grupos de edad (años)	Hombres (%)		Mujeres (%)	
	Final	Cruda	Final	Cruda
Total	100,0	100,0	100,0	100,0
0 - 4	10,5	10,5	9,5	9,5
5 - 9	10,9	10,9	9,9	10,0
10 - 14	11,0	11,0	10,0	10,1
15 - 19	9,6	9,6	9,3	9,3
20 - 24	8,5	8,5	8,8	8,8
25 - 29	7,7	7,7	8,0	8,0
30 - 34	6,9	6,9	7,2	7,2
35 - 39	6,9	6,9	7,2	7,2
40 - 44	6,4	6,4	6,8	6,7
45 - 49	5,4	5,4	5,7	5,7
50 - 54	4,3	4,3	4,5	4,5

Cuadro 9 (conclusión)

Distribución de la población, por sexo y base, según grupos de edad 2005

Grupos de edad (años)	Hombres (%)		Mujeres (%)	
	Final	Cruda	Final	Cruda
55 - 59	3,4	3,4	3,6	3,6
60 - 64	2,6	2,6	2,7	2,7
65 - 69	2,1	2,1	2,3	2,3
70 - 74	1,6	1,6	1,8	1,8
75 - 79	1,1	1,1	1,3	1,3
80 - 84	0,6	0,7	0,7	0,8
85 - 89	0,4	0,4	0,6	0,5

Fuente: DANE –Censo General 2005

Coherencia de la información con fuentes externas: existen diferentes fuentes que pueden arrojar indicadores sobre la calidad de la información recolectada en el Censo 2005; el primero de ellos corresponde a la relación de masculinidad a la edad 0, la cual es una aproximación que muestra por cada 100 niñas, cuántos niños nacen. De acuerdo con el Censo 2005, la relación es de 106,6, valor que es mayor, comparativamente con lo observado en el Censo 1993 y en países para los cuales se pudo obtener información, en donde, la relación es inferior a 105 hombres por cada 100 mujeres (tabla 3). El diferencial observado estaría posiblemente evidenciando, en comparación con el género masculino, un subregistro de mujeres en el Censo 2005.

Tabla 3

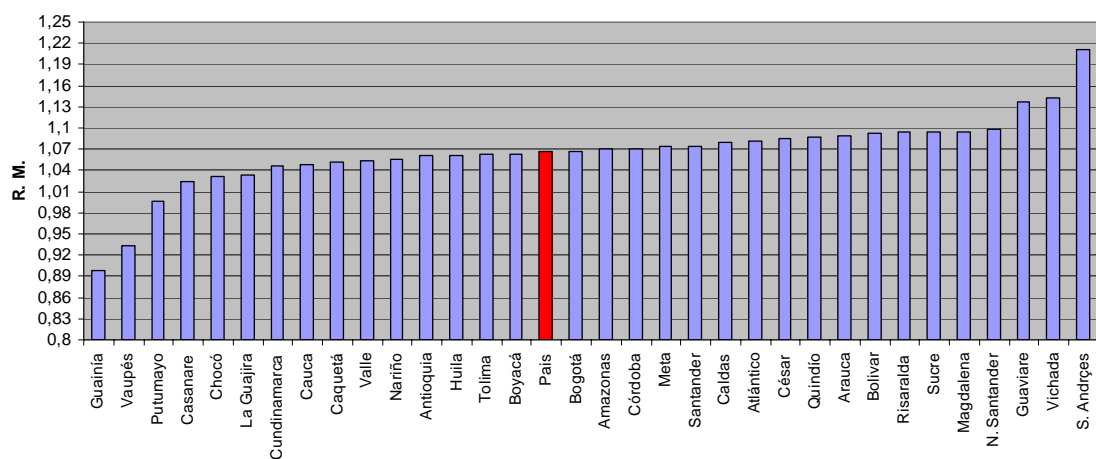
Relación de masculinidad a la edad 0, según países seleccionados y censos

País	Año del censo	Relación de masculinidad edad 0
Colombia	2005	106,6
Colombia	1993	104,8
Argentina	2001	103,8
España	2001	104,9
Costa Rica	2000	104,5
Chile	2002	104,4
Uruguay	2004	103,9
Guatemala	2002	101,9
Chile	1992	104,7
Paraguay	1992	103,2

Fuente: DANE –Censo General 2005

El subregistro es diferencial por departamento, 18 de ellos presenta valores superiores al promedio nacional y en 3, la relación supera los 113 hombres por cada 100 mujeres. En el otro extremo, en Guainía, Vaupés y Putumayo, la relación es inferior a 100.

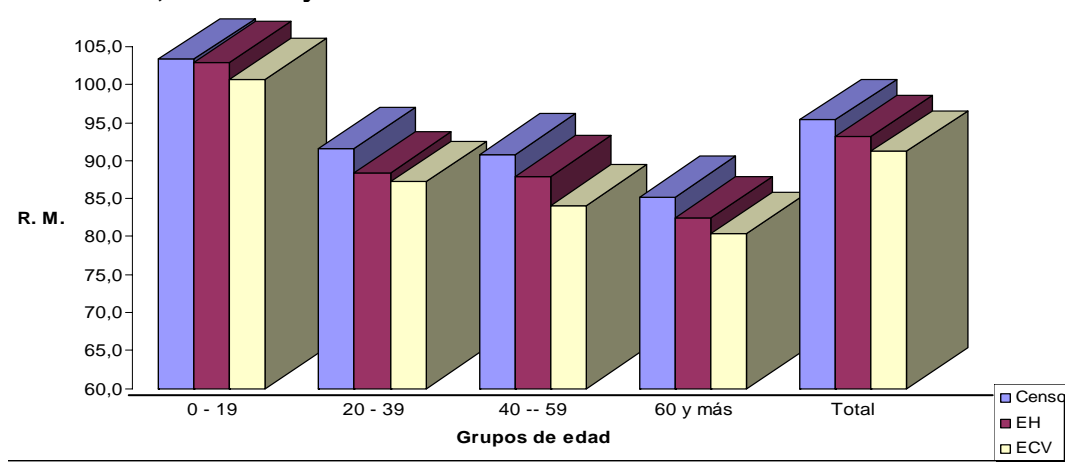
Gráfico 5
Relación de masculinidad a la edad 0, por departamento
2005



Fuente: DANE –Censo General 2005

Un segundo indicador externo corresponde a la comparación entre las relaciones de masculinidad observadas en el Censo, con las que se derivan de la Encuesta Continua de Hogares 2005 y de la Encuesta de Calidad de Vida 2003⁵. En la primera de ellas, se tomó información de un poco más de 151 000 hogares; en la segunda, de únicamente un poco más de 10 000 (gráfico 6).

Gráfico 6
Relación de masculinidad para grupos de edad, según fuente de información
Censo 2005, ECH 2005 y ECV 2003



Fuente: DANE –Censo General 2005

⁵ La comparación de la información del Censo 2005 con la que se obtiene en encuestas por muestreo, se lleva a cabo para observar similitud o no en los patrones de comportamiento de las variables, ya que no pueden estimarse diferenciales precisos, debido a los errores de muestreo asociados a las encuestas y las distintas estructuras implícitas en cada una de las fuentes.

El Censo ha sido siempre el referente para evaluar la calidad de otras fuentes de información; no obstante, no debe descartarse los indicadores generados con base en esas otras fuentes, en particular, cuando se desea analizar la calidad de la información censal.

A pesar de las limitaciones anotadas sobre la validez de las comparaciones con los resultados de las encuestas por muestreo (efecto estructura), se observa que en todos los grupos de edad seleccionados, el valor de las relaciones de masculinidad es menor al obtenido en el Censo 2005 (cuadro 10). En una primera instancia podría hablarse de un subregistro de mujeres mayor que el de los hombres en el Censo 2005; no obstante, es necesario profundizar en el análisis de otros indicadores que sustenten o nieguen este supuesto.

Cuadro 10
Relaciones de masculinidad, por fuente, según grupos de edad
Censo 2005, ECH 2005 y ECV 2003

Grupos de edad (años)	Censo 2005			Encuesta de Hogares 2005			Calidad de Vida 2003		
	Hombres	Mujeres	RM	Hombres	Mujeres	RM	Hombres	Mujeres	RM
Total	20 101 180	21 073 673	95,4	21 508 375	23 067 356	93,2	20 856 959	22 860 619	91,2
0 - 19	8 441 527	8 166 197	103,4	9 024 776	8 766 202	102,9	9 012 994	8 944 137	100,8
20 - 39	6 023 475	6 578 113	91,6	6 249 289	7 063 563	88,5	6 205 505	7 113 155	87,2
40 - 59	3 940 108	4 340 297	90,8	4 219 552	4 796 320	88	3 804 023	4 522 274	84,1
60 y más	1 696 070	1 989 066	85,3	2 014 758	2 441 272	82,5	1 834 438	2 281 054	80,4

Fuente: DANE –Censo General 2005

Los jefes de hogar que se declaran en unión y que su pareja no se encontraba en el momento del Censo 2005 tenían varias opciones: (1) declararla como residente habitual; (2) declarar su estado conyugal real y no declarar la pareja; y (3) cambiar el estado conyugal para no tener que declarar el cónyuge. Las opciones 2 y 3 se convierten en pérdida de personas en el Censo 2005 y, por tanto, en no cobertura. El Censo 2005 permite tener una aproximación únicamente al segundo de los efectos.

En el Censo 2005, 6,6 millones de jefes de hogar se declararon en unión; de ellos el 5,4% (359 458) informó que el(la) cónyuge no estaba presente en el momento del censo. Este eventual subregistro corresponde casi en su totalidad a mujeres; ellas representan únicamente el 5,5% (364 137) del total de jefes de hogar en unión con cónyuge presente cuando se censó (cuadro 11).

Cuadro 11
Jefes en unión sin pareja presente en el momento del Censo 2005

Presencia de la pareja	Frecuencia	Porcentaje
Total	6 649 901	100,0
Hombres con pareja presente	5 926 306	89,1
Mujeres con pareja presente	364 137	5,5
Jefes sin pareja presente	359 458	5,4

Fuente: DANE –Censo General 2005

Cuadro 12
Jefes en unión, por presencia del cónyuge, según departamento
2005

Departamento	Total	Con pareja		Sin Pareja	
		Recuento	Porcentaje*	Recuento	Porcentaje*
Total	6 649 901	6 290 443	94,6	359 458	5,4
Antioquia	883 338	834 396	94,5	48 942	5,5
Atlántico	331 645	312 067	94,1	19 578	5,9
Bogotá, D.C.	1 147 862	1 101 921	96,0	45 941	4,0
Bolívar	283 036	265 173	93,7	17 863	6,3
Boyacá	201 862	191 856	95,0	10 006	5,0
Caldas	151 159	142 738	94,4	8 421	5,6
Caquetá	49 238	45 895	93,2	3 343	6,8
Cauca	182 468	171 136	93,8	11 332	6,2
Cesar	136 549	127 812	93,6	8 737	6,4
Córdoba	227 174	217 048	95,5	10 126	4,5
Cundinamarca	383 355	368 325	96,1	15 030	3,9
Chocó	46 137	39 837	86,3	6 300	13,7
Huila	167 686	160 352	95,6	7 334	4,4
La Guajira	80 881	72 251	89,3	8 630	10,7
Magdalena	177 195	167 265	94,4	9 930	5,6
Meta	117 143	109 718	93,7	7 425	6,3
Nariño	226 633	216 078	95,3	10 555	4,7
Norte de Santander	191 098	181 656	95,1	9 442	4,9
Quindío	85 502	80 017	93,6	5 485	6,4
Risaralda	141 159	132 098	93,6	9 061	6,4
Santander	321 210	304 405	94,8	16 805	5,2
Sucre	119 327	112 304	94,1	7 023	5,9
Tolima	219 775	207 112	94,2	12 663	5,8
Valle del Cauca	643 218	603 189	93,8	40 029	6,2
Arauca	22 325	20 910	93,7	1 415	6,3
Casanare	47 154	44 531	94,4	2 623	5,6
Putumayo	30 711	27 706	90,2	3 005	9,8
Archipiélago de San Andrés, Providencia y Santa Catalina	9 886	9 258	93,6	628	6,4
Amazonas	6 396	5 928	92,7	468	7,3
Guainía	2 748	2 556	93,0	192	7,0
Guaviare	7 459	6 932	92,9	527	7,1
Vaupés	2 428	2 327	95,8	101	4,2
Vichada	6 144	5 646	91,9	498	8,1

Fuente: DANE –Censo General 2005

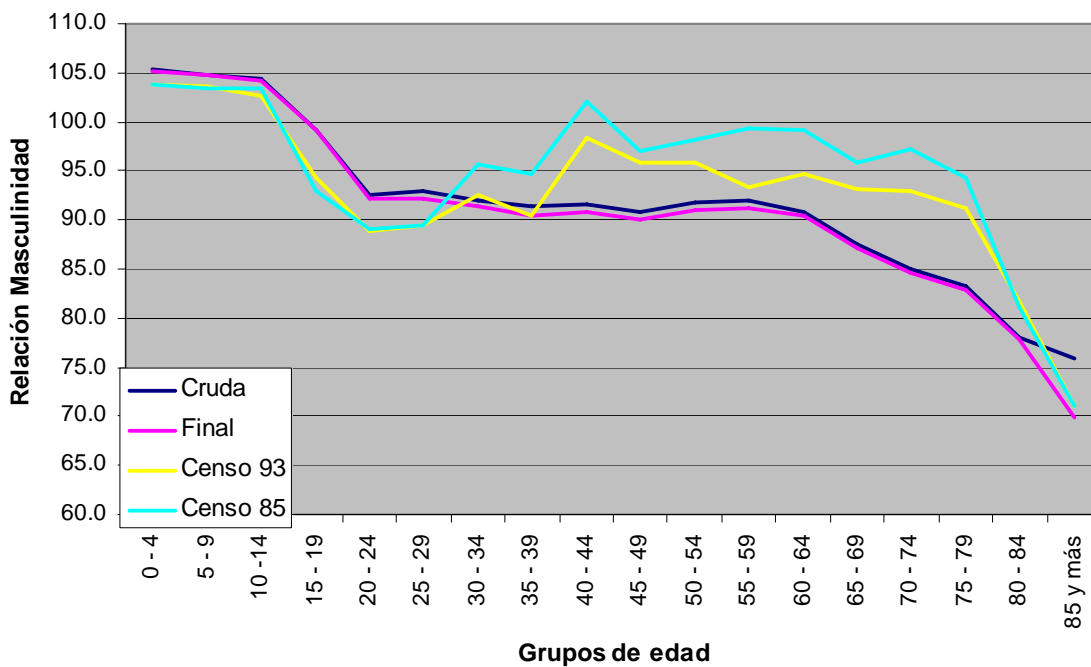
* Porcentaje de fila

Por departamento el diferencial cambia: en Bogotá, D.C., Cundinamarca, Nariño, Huila, Norte de Santander y Vaupés, la omisión estaría por debajo del 5%; mientras que en la mayor parte de los nuevos departamentos (antiguos territorios nacionales), el porcentaje supera el 7% y en el Chocó y La Guajira supera el 10% (cuadro 12).

La comparación con censos anteriores muestra que la relación de masculinidad crece en los primeros 6 grupos de edad, esto refleja el efecto acumulado de las cada vez mejores condiciones de salud, tanto en tecnología, como en la afiliación de la población a los sistemas de salud; factores que tienden a cerrar la brecha entre la mortalidad masculina y femenina, principalmente, en los primeros años de vida.

En los grupos de edad 30-34 y 35-39 años, las relaciones de masculinidad son similares, a partir de los 40 años el patrón cambia, las relaciones de masculinidad son significativamente más altas en los censos anteriores que en el de 2005 (gráfico 7); la explicación de este comportamiento no se ha precisado y se considera que amerita la realización de un estudio específico para su análisis.

Gráfico 7
Relación de masculinidad, por grupos quinquenales de edad, según base Censos 1985, 1993 y 2005



Fuente: DANE

Fuente probable de error: en relación con el primero de los problemas detectados, la fuente de error es la omisión consciente o involuntaria del informante, que afecta en particular el registro de las niñas menores de un año, lo cual indicaría una cobertura diferencial por sexo, siendo mayor el subregistro de mujeres.

De la comparación con otras fuentes no surge la suficiente evidencia sobre la existencia de limitaciones de la información del Censo 2005 relacionada con el sexo.

Balance e impacto sobre algunos indicadores: los procesos de depuración de la información relacionada con el sexo están bien realizados. El impacto importante de un subregistro diferencial entre hombres y mujeres, en particular en la población menor de un año, es

sobre las proyecciones de población, en la medida en que, de no corregirse las desigualdades, van a mantenerse en el tiempo.

El problema afecta a la mayor parte de los departamentos del país; en unos (la mayoría), porque la relación de masculinidad a la edad 0 es muy alta, y en otros, porque están por debajo de cualquier consideración sobre los valores esperados de la relación.

Recomendaciones: al igual que con la edad, altas relaciones de masculinidad a la edad 0, a nivel del país, y significativos diferenciales a nivel departamental, afectan las líneas de base para la realización de las proyecciones de población. La recomendación, ya considerada en el proceso de conciliación, es la de corregir el desbalance que se presenta en la distribución por sexo.

Adelantar un estudio específico en el que se profundice sobre la explicación de las relaciones de masculinidad, particularmente las que se observan después de los 30 años.

2.1.4 Parentesco

Comparación base cruda y base final: en la base cruda aparecen 140 mil valores no válidos para la variable parentesco (valores perdidos: 93 239; código 0: 46 845). Al pasar de la base original a la final, la estructura se conserva, indicativo de un adecuado proceso de depuración (cuadro 13).

Cuadro 13
Relación de parentesco, por base
Censo 2005

Parentesco	Base cruda		Base final	
	Personas	Porcentaje	Personas	Porcentaje
Total	41 034 769	100,0	41 174 853	100,0
Jefe	10 532 918	25,7	10 570 899	25,7
Cónyuge	6 644 689	16,2	6 534 002	15,9
Hijo/hijastro	17 207 967	41,9	17 216 983	41,8
Yerno/nuera	462 927	1,1	455 505	1,1
Nieto	2 632 456	6,4	2 639 352	6,4
Padre/suegro	591 164	1,4	578 533	1,4
Hermano/hermanastro	717 368	1,7	716 606	1,7
Otro pariente	1 189 710	2,9	1 302 198	3,2
Empleado	165 264	0,4	164 765	0,4
No pariente	890 306	2,2	996 010	2,4

Fuente: DANE –Censo General 2005

Coherencia de la información con fuentes externas: al compararse los resultados del Censo con otras fuentes, los de la Encuesta de Hogares y los de la de Calidad de Vida son prácticamente iguales (cuadro 14).

Cuadro 14
Relación de parentesco, por fuente de información y sexo

Parentesco	Censo 2005		Encuesta de Hogares 2005		Calidad de Vida 2003	
	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
Total	48,8	51,2	48,3	51,7	47,7	52,3
Jefe	70,1	29,9	70,5	29,5	69,1	30,9
Cónyuge	6,7	93,3	5,7	94,3	5,0	95,0
Hijo/hijastro	52,0	48,0	52,7	47,3	51,2	48,8
Otros parientes	48,4	51,6	46,4	53,6	47,2	52,8
No parientes	47,2	52,8	41,3	58,7	43,5	56,5

Fuente: DANE

Balance e impacto sobre los resultados: los procesos de depuración efectuados y su coherencia con fuentes externas indican que la información sobre el parentesco no presenta limitantes para su utilización confiable.

2.1.5 Estado conyugal

Comparación base cruda y base final: para la variable estado conyugal, en la base cruda se presentan 729 621 valores perdidos y 26 489 con código 0. Una vez se realiza la depuración permanecen 695 211 registros como “no informa” (cuadro 15).

Cuadro 15
Distribución de la población de 10 y más años, por base, según por estado conyugal Censo 2005

Estado conyugal	Base cruda		Base final	
	Personas	Porcentaje	Personas	Porcentaje
Total	31 985 942	100,0	32 083 165	100,0
Unión libre (2 o más)	6 514 515	20,4	6 777 319	21,1
Unión libre (menos de 2)	608 197	1,9	629 419	2,0
Separado	1 642 900	5,1	1 601 764	5,0
Viudo	1 361 164	4,3	1 344 019	4,2
Soltero	14 471 862	45,2	14 316 559	44,6
Casado	7 387 304	23,1	7 414 085	23,1

Fuente: DANE –Censo General 2005

Coherencia de la información con fuentes externas: la información sobre el estado conyugal es coherente con la que arrojan otras fuentes de información, lo cual es un buen indicativo de la consistencia y calidad de esta variable (cuadro 16).

Balance e impacto sobre los resultados: los procesos de depuración efectuados y su coherencia con fuentes externas indican que la información sobre el estado conyugal no presenta limitantes para su utilización confiable.

Cuadro 16
Distribución de la población de 10 y más años, por fuente, según estado conyugal

Estado Conyugal	Censo 2005		Encuesta de Hogares 2005		Calidad de Vida 2003	
	Población	Porcentaje	Población	Porcentaje	Población	Porcentaje
Total	31 985 942	100,0	35 614 633	100,0	34 703 329	100,0
Unión libre	7 122 712	22,3	7 601 464	21,3	7 486 428	21,6
Casado	7 387 304	23,1	8 423 664	23,7	8 219 662	23,7
Separado	1 642 900	5,1	1 747 380	4,9	1 752 758	5,1
Viudo	1 361 164	4,3	2 699 446	7,6	2 749 666	7,9
Soltero	14 471 862	45,2	15 142 679	42,5	14 494 815	41,8

Fuente: DANE

2.1.6 Pertenencia étnica

Comparación base cruda y base final: los valores relativos están dentro de los márgenes de error establecido (inferiores a 1 punto porcentual); no obstante, la población absoluta por pertenencia étnica cambia sustancialmente, así: los Rom⁶, que en la base cruda superaban las 45 mil personas, en la base final son sólo 4 700; de la misma forma, los palenqueros⁷ en la primera base sumaban más de 27 000, en la segunda, se reducen a 7 459 (cuadro 17).

Cuadro 17
Pertenencia étnica, por base 2005

Etnia	Base cruda		Base final	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Total	41 174 853	100,0	41 174 853	100,0
Indígena	1 388 970	3,4	1 388 985	3,4
Rom	45 361	0,1	4 700	0,0
Raizal	30 278	0,1	30 359	0,1
Palenquero	27 378	0,1	7 459	0,0
Negro	4 231 405	10,3	4 251 019	10,3
Ninguna	34 590 686	84,0	34 631 579	84,1
No informa	860 775	2,1	860 752	2,1

Fuente: DANE –Censo General 2005

Para el caso del autorreconocimiento étnico, opción Rom, en las primeras revisiones de los datos preliminares se identificaron deficiencias en la capacitación, dado que se estaba reportando un volumen de población muy superior al obtenido en los estudios realizados sobre este grupo poblacional, más concretamente en los efectuados por ProRom.

Para corregir este inconveniente se produjeron instrucciones precisas en “capacitación en píldoras” enviadas a los coordinadores operativos para que las socializaran entre el

⁶ Denominación con la que también se conoce al pueblo gitano.

⁷ Población nacida en el centro poblado de San Basilio de Palenque, municipio de Mahates, departamento de Bolívar.

personal de recolección: supervisores y encuestadores. Para corregir la información recopilada, se aplicaron normas de imputación referidas al lugar de residencia, en las que se identificaron los municipios en donde se tenía certeza que allí residían Rom; para el caso de Bogotá, D.C., se seleccionaron las localidades y los barrios. En segundo lugar, se tomó el nivel educativo, dado que, según los documentos disponibles, por tradición y por cultura, este grupo poblacional no alcanza estudios de postgrado y los mayores de 50 años, la secundaria.

La evaluación realizada por el DANE de la información durante el proceso de recolección, permitió ajustar este proceso, lo cual evitó que el total identificado no fuese mayor, esto muestra un factor importante en la realización en varios días de la recolección.

Para el caso de la opción “Palenquero de San Basilio”, se identificó que en los formularios de la fase 0, 1 y 2, esta opción no se había incluido. Una vez se realizó el proceso de consolidación de la base cruda final, se homologó dicha variable para que la población residente en el municipio de Mahates (Bolívar) en clase 2 (San Basilio de Palenque) se considerara como palenquera.

Fuente probable de error: en el caso de la población Rom las fuentes de error fueron de capacitación en los conceptos básicos y de supervisión. Al parecer algunos encuestadores y encuestados no conocían el significado de la palabra Rom; el problema no fue detectado por la supervisión del operativo de recolección, agentes externos a él llamaron la atención sobre lo que estaba sucediendo.

En relación con la tercera fuente de error se sugiere que a la palabra Rom debió adicionársele “o gitano”; sin embargo, es preciso señalar que durante la prueba experimental de Soacha se probaron las diferentes alternativas y se llegó a la conclusión que la mejor fue la que finalmente quedó en el formulario.

Balance e impacto sobre los resultados: el análisis de la información con las organizaciones que representan a la población Rom y estudios realizados, avalan los correctivos realizados.

2.1.7 Fecundidad

Comparación base cruda y base final: tres variables se utilizan para mirar la fecundidad: mujeres en edad reproductiva y si han tenido hijos nacidos vivos o no, el número de hijos nacido vivos (HNV) y el año y mes del último HNV (cuadro 18).

En la base cruda y en la final 11,1 millones de mujeres están en edad reproductiva; de ellas cerca de 370 no informan sobre si han tenido hijos vivos o no. Al igual que con otras preguntas, debería existir un control que obligara a que todas las mujeres en edad reproductiva reportaran la tenencia de hijos nacidos vivos.

Al pasar de la base cruda a la final, los valores absolutos varían muy poco y las estructuras permanecen idénticas. Igual sucede con el año y mes del último hijo nacido vivo, lo cual indica que los procesos de depuración de la información sobre fecundidad fueron adecuados (cuadros 19 y 20).

Cuadro 18
Mujeres en edad fértil y con información sobre HNV, por base, según grupos de edad
Censo 2005

Grupos de edad (años)	Edad fértil				Con información sobre HNV			
	Base cruda		Base final		Base cruda		Base final	
	Recuento	Porcentaje	Recuento	Porcentaje	Recuento	Porcentaje	Recuento	Porcentaje
Total	11 103 599	100,00	11 156 720	100,00	10 732 937	100,00	10 785 629	100,00
15-19	1 947 924	17,54	1 951 871	17,50	1 874 340	17,46	1 882 597	17,50
20-14	1 846 253	16,63	1 853 700	16,62	1 784 699	16,63	1 792 846	16,60
25-29	1 677 759	15,11	1 686 368	15,12	1 624 826	15,14	1 633 125	15,10
30-34	1 505 261	13,56	1 513 679	13,57	1 455 849	13,56	1 463 496	13,60
35-39	1 515 083	13,64	1 524 366	13,66	1 464 949	13,65	1 472 445	13,70
40-44	1 417 493	12,77	1 425 452	12,78	1 375 645	12,82	1 381 815	12,80
45-49	1 193 826	10,75	1 201 284	10,77	1 152 629	10,74	1 159 305	10,70

Fuente: DANE –Censo General 2005

Cuadro 19
Mujeres en edad fértil, por características de fecundidad y base, según grupos de edad
Censo 2005

Grupos de edad (años)	Con hijos nacidos vivos				Sin hijos nacidos vivos			
	Cruda		Final		Cruda		Final	
	Recuento	Porcentaje	Recuento	Porcentaje	Recuento	Porcentaje	Recuento	Porcentaje
Total	7 029 075	100,0	7 063 115	100,0	3 703 862	100,0	3 722 514	100,0
15-19	277 969	4,0	279 588	4,0	1 596 371	43,1	1 603 009	43,1
20-14	875 193	12,5	879 512	12,5	909 506	24,6	913 334	24,5
25-29	1 144 248	16,3	1 149 943	16,3	480 578	13,0	483 182	13,0
30-34	1 207 763	17,2	1 213 782	17,2	248 086	6,7	249 714	6,7
35-39	1 279 603	18,2	1 285 716	18,2	185 346	5,0	186 729	5,0
40-44	1 219 605	17,4	1 224 638	17,3	156 040	4,2	157 177	4,2
45-49	1 024 694	14,6	1 029 936	14,6	127 935	3,5	129 369	3,5

Fuente: DANE –Censo General 2005

Cuadro 20
Último HNV por base, según año y mes de nacimiento
Censo 2005

Fecha último HNV	Cruda		Final	
	Recuento	Porcentaje	Recuento	Porcentaje
Año de nacimiento último HNV				
2000 o antes	4 189 328	59,6	4 204 479	59,5
2001	461 768	6,6	463 621	6,6
2002	492 277	7,0	494 436	7,0
2003	559 569	8,0	562 229	8,0
2004	640 271	9,1	643 320	9,1
2005	656 100	9,3	659 840	9,3
Total	6 999 313	99,6	7 027 925	99,5
Sistema	29 762	0,4	35 190	0,5
Total	7 029 075	100,0	7 063 115	100,0

Cuadro 20 (conclusión)
Último HNV por base, según año y mes de nacimiento
Censo 2005

Fecha último HNV	Cruda		Final	
	Recuento	Porcentaje	Recuento	Porcentaje
Mes de nacimiento último HNV				
Enero	565 034	8,0	565 764	8,0
Febrero	509 186	7,2	510 061	7,2
Marzo	566 680	8,1	567 904	8,0
Abril	555 251	7,9	556 575	7,9
Mayo	584 206	8,3	585 277	8,3
Junio	572 036	8,1	573 112	8,1
Julio	597 597	8,5	598 700	8,5
Agosto	614 730	8,7	616 004	8,7
Septiembre	639 626	9,1	640 431	9,1
Octubre	620 968	8,8	622 682	8,8
Noviembre	587 369	8,4	588 581	8,3
Diciembre	592 104	8,4	593 171	8,4
Total	7 011 042	99,7	7 018 262	99,4
Sistema	24 288	0,3	44 853	0,6
Total	7 029 075	100,0	7 063 115	100,0

Fuente: DANE –Censo General 2005

Coherencia interna de la información y con fuentes externas: el procesamiento de las bases de microdatos cruda y final permite obtener dos indicadores de fecundidad que resumen la coherencia de la información: uno es la paridez media acumulada por edades simples y el otro, la tasa global de fecundidad. Es preciso señalar que el objetivo del análisis no es el de establecer estimativos ajustados del indicador, por lo que los controles llevados a cabo cuidan únicamente que el numerador tenga correspondencia con el denominador.

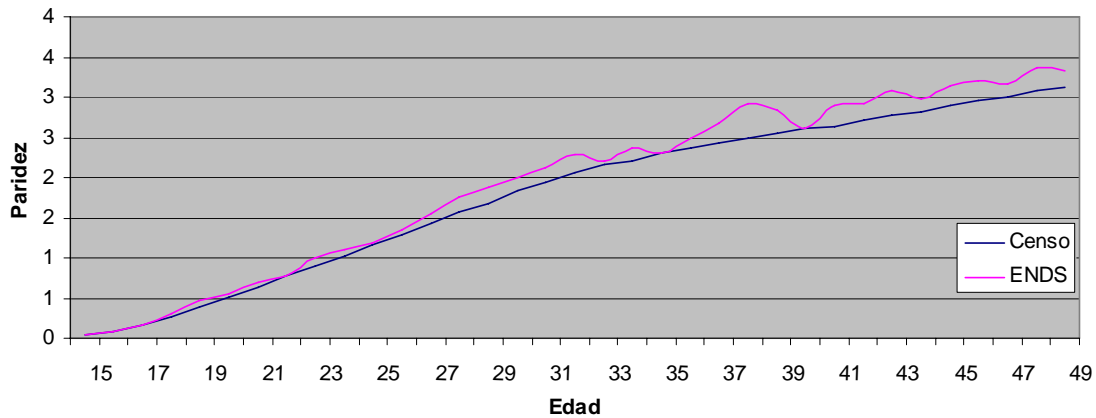
La paridez refleja el número medio de hijos tenidos por las mujeres en las distintas etapas de su vida reproductiva. La comparación con los resultados de la última Encuesta Nacional de Demografía y Salud –ENDS– 2005, indican que la correspondencia entre una y otra fuente es bastante alta hasta los 28 años y después de esta edad se presentan variaciones, y siempre es más alta la paridez que se deriva de la ENDS (gráfico 8).

La Tasa Global de Fecundidad –TGF⁸– es uno de los indicadores que brinda el mayor conocimiento sobre los niveles y variación en el tiempo. La TGF observada en el Censo, sin ningún ajuste, de 0,2 hijos fue inferior a la que entrega la ENDS, aunque la información de esta última corresponde al promedio de los tres años anteriores a la encuesta. La diferencia puede considerarse como no significativa, pero eventualmente altera el patrón de la fecundidad visto a través de las tasas específicas (cuadro 21).

La fecundidad, según el Censo 2005, es inferior en los grupos de edad entre 15 y 29 años, pero similar a partir de los 30 años (gráfico 9).

⁸ Número promedio de hijos que una mujer habría tenido al final de sus años reproductivos, si la misma se hubiera ajustado a la tasa de fecundidad por edad específica, durante cada año de su vida reproductiva (Population Reference Bureau, 2003).

Gráfico 8
Paridez acumulada, por edades simples, según fuente de información
2005



Fuente: Censo General 2005 y ENDS 2005

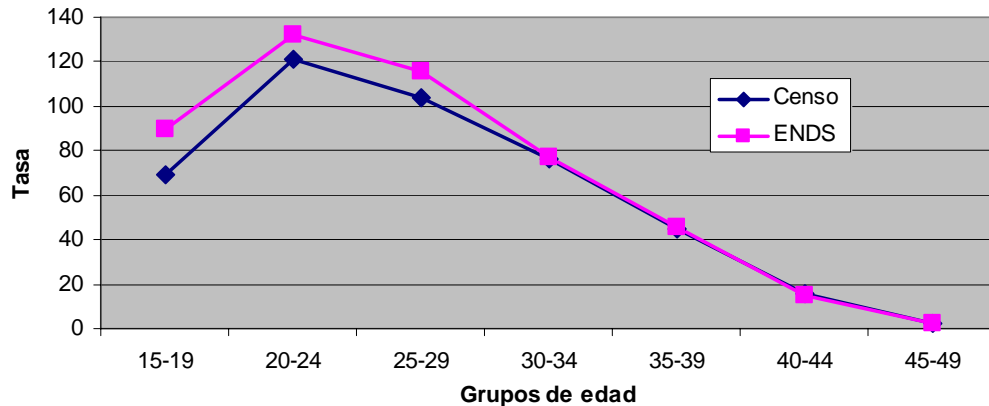
Cuadro 21
Tasas específicas y tasa global de fecundidad, por fuente,
según grupos de edad
2005

Grupos de edad (años)	Mujeres	HNV	Tasas específicas	
			Censo 2005	ENDS 2005*
15-19	1 879 015	130 291	69,0	90,0
20-24	1 786 823	216 162	121,0	132,0
25-29	1 627 021	168 544	104,0	116,0
30-34	1 457 948	111 698	77,0	77,0
35-39	1 467 115	65 566	45,0	46,0
40-44	1 377 367	21 940	16,0	15,0
45-49	1 155 150	2 764	2,0	2,0
Tasa global			2,2	2,4

Fuente: Profamilia, Encuestas de Demografía y Salud, y DANE

*Comprende los tres años anteriores a la encuesta

Gráfico 9
Tasas específicas de fecundidad, por grupos de edad, según fuente
2005



Fuente: Censo General 2005 y ENDS 2005

A pesar de la coherencia entre los resultados que se derivan de las dos fuentes, es preciso reconocer que la información sobre fecundidad puede tener limitaciones, tales como:

- Ni el DMC, ni los formularios permitían captar los nacimientos ocurridos en 2006. Presumiblemente, algunos fueron incorporados en años anteriores, pero otros definitivamente quedaron sin ser consignados.
- De las mujeres con información válida sobre HNV, una cuarta parte fue censada en el año 2006 (cuadro 22).

Cuadro 22
Mujeres en edad reproductiva,
según año del Censo 2005

Años	Frecuencia	Porcentaje
Total	10 785 629	100,0
2005	8 083 950	75,0
2006	2 701 679	25,0

Fuente: DANE –Censo General 2005

- La información de las TGF departamentales no permite hacer inferencia sobre el eventual impacto de departamento. Las mujeres censadas en 2006 varían desde ninguna, como en Bogotá, D.C., hasta el 100%, como en Caquetá o Sucre; no obstante, en ninguno se presentan evidencias sobre el impacto que pueda tener sobre la TGF, el no haber consignado los hijos nacidos en el año 2006 (cuadro 23):

Cuadro 23
Porcentaje de mujeres en edad fértil y TGF, según departamento
2005-2006

Departamento	Mujeres	Año Censo 2005		TGF
		2005 (%)	2006 (%)	
Total	10 785 629	75,0	25,0	2,2
Antioquia	1 484 749	73,9	26,1	2,0
Atlántico	580 981	98,1	1,9	2,2
Bogotá, D.C.	1 918 965	100,0	0,0	1,8
Bolívar	459 052	75,3	24,7	2,5
Boyacá	291 373	66,9	33,1	2,4
Caldas	236 532	79,3	20,7	1,9
Caquetá	74 759	0,0	100,0	2,6
Cauca	285 141	46,1	53,9	2,3
Cesar	220 635	80,4	19,6	2,9
Córdoba	362 904	0,0	100,0	2,7
Cundinamarca	556 111	79,4	20,6	2,2
Chocó	85 926	0,0	100,0	3,0
Huila	244 277	59,6	40,4	2,6
La Guajira	136 317	100,0	0,0	3,1
Magdalena	276 059	76,7	23,3	3,0
Meta	181 692	70,5	29,5	2,4
Nariño	359 766	38,8	61,2	2,2

Cuadro 23 (conclusión)
Porcentaje de mujeres en edad fértil y TGF, según departamento
2005-2006

Departamento	Mujeres	Año Censo 2005		TGF
		2005 (%)	2006 (%)	
Norte de Santander	315 800	81,7	18,3	2,3
Quindío	138 616	100,0	0,0	1,7
Risaralda	232 386	93,5	6,5	1,8
Santander	513 490	79,8	20,2	2,0
Sucre	186 607	0,0	100,0	2,6
Tolima	319 557	72,9	27,1	2,4
Valle del Cauca	1 110 305	88,6	11,4	1,8
Arauca	35 572	0,0	100,0	2,4
Casanare	70 081	0,0	100,0	2,6
Putumayo	55 220	0,0	100,0	2,2
Archipiélago de San Andrés, Providencia y Santa Catalina	16 016	100,0	0,0	2,2
Amazonas	10 470	0,0	100,0	3,3
Guainía	4 237	0,0	100,0	3,1
Guaviare	11 399	0,0	100,0	2,5
Vaupés	3 102	0,0	100,0	3,6
Vichada	7 532	0,0	100,0	3,6

Fuente: DANE –Censo General 2005

- Aunque únicamente el 25% de las mujeres en edad fértil fue censado en 2006, el porcentaje de municipios que las alberga es del 70% del total (775) (cuadro 24). No obstante, tampoco puede inferirse que el valor de las TGF que se observa en cada uno de ellos pueda estar determinado por la ausencia de respuesta correcta de los nacidos en 2006.

Cuadro 24
Municipios, por año de censo, según rango de la TGF
2005-2006

Rango TGF	Censados 2005		Censados 2006	
	Total	Porcentaje	Total	Porcentaje
Menor a 1,5	3	0,9	21	2,7
1,5 - <2,0	62	18,3	43	5,5
2,0 - <2,2	52	15,3	59	7,6
2,2 - <2,5	77	22,7	146	18,8
2,5 - <3,0	89	26,3	218	28,1
3,0 - <4,0	54	15,9	238	30,7
4,0 y más	2	0,6	50	6,5
Total	339	100,0	775	100,0

Fuente: DANE –Censo General 2005

- Aunque esta limitante complementa y explica la anterior, puede señalarse que existe subregistro diferencial en las preguntas de fecundidad, en detrimento del numerador;

así, por ejemplo, hijos menores de un año declarados en las preguntas relacionadas con la edad, por una u otra razón, no aparecen reportados ni en la relacionada con la tenencia de hijos en su vida reproductiva, ni tampoco en la fecha del último hijo nacido vivo.

La población menor de un año en el Censo 2005 corresponde a los nacimientos ocurridos en el año anterior al Censo 2005, descontadas las defunciones infantiles correspondientes a estos nacimientos.

En este marco, la población en edad 0 debe ser menor a los nacimientos ocurridos en el período. En el Censo 2005, esto no ocurre, el número de últimos hijos nacidos vivos en el año anterior a la fecha censal (incluidos los de la población entre 10 y 14 años) es de 720 044, mientras que la población en edad 0, aún teniendo en consideración la sobrestimación que presentó en el análisis de la variable edad, supera los 800 000; de hecho, el número de niños menores de 5 años supera, en cada edad simple, esta misma cantidad de personas (cuadros 25 y 26).

Cuadro 25
Mujeres entre 10 y 49 años con y sin hijos nacidos vivos en el último año

Concepto	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Total	12 848 743	100,0	100,0	
Sin HNV	12 128 699	94,4	94,4	94,4
Con HNV	720 044	5,6	5,6	100,0

Fuente: DANE –Censo General 2005

Cuadro 26
Diferencia relativa entre la población menor de 5 años y los últimos HNV

	0 años	1 año	2 años	3 años	4 años
Población	855 931	815 629	809 219	803 665	821 539
Diferencia %	15,9	11,7	11,0	10,4	12,4

Fuente: DANE –Censo General 2005

Si un porcentaje significativo de este subregistro es producto de la no inclusión de nacidos vivos en 2006, las implicaciones sobre las medidas de fecundidad, en especial a nivel municipal, pueden ser significativas.

Fuente probable de error: el error se genera cuando las directivas del Censo 2005 no esperaban que el período de recolección fuera más allá del año 2005; sin embargo, una vez que hubo la necesidad de extender el período de recolección a 2006, no se tomaron las previsiones para que los DMC y los formularios en papel tomaran en consideración los eventos de fecundidad que ocurren en este año.

Impacto sobre los resultados: para evaluar el eventual impacto que puede haber traído consigo la no consideración de los HNV tenidos por las mujeres en el año 2006, se realizó un ejercicio que se fundamenta en que los hijos sobrevivientes menores de un año, reportados por las mujeres en edad reproductiva, son una buena aproximación a los hijos nacidos que los originan (anexo A).

Los principales resultados del ejercicio indican que:

- El subregistro de los HNV en el último año existe, afecta en mayor medida a los hijos de las mujeres censadas en 2006.
- Las metodologías diseñadas y utilizadas en censos y encuestas por muestreo para ajustar los indicadores de fecundidad, contribuirían a reducir los diferenciales observados.
- El problema mencionado anteriormente, no es de la magnitud indicada. La información puede utilizarse efectuando los ajustes recomendados para ésta.

Recomendaciones

Para la utilización de la información: se sugiere profundizar en el eventual problema con el fin de estimar con mayor precisión: primero, el subregistro de las variables que intervienen en el cálculo de las medidas de fecundidad; segundo, evaluar cómo afecta los estimativos a los diferentes niveles de desagregación geográfica; y tercero, emitir recomendaciones para la utilización de los datos.

Para futuros censos:

- El diseño, alistamiento, ejecución del Censo 2005 deben responder a tiempos previamente establecidos y que deben cumplirse estrictamente.
- Las eventualidades surgidas durante el operativo de ninguna manera pueden ir en contra de la integridad y calidad de la información, todas ellas deben ser contempladas antes de la iniciación del operativo y tener las alternativas de respuesta apropiadas.

2.1.8 Mortalidad

Comparación base cruda y base final: el cuestionario de hogares suministra el número de personas fallecidas en los últimos doce meses (cuadro 27). La comparación de la estructura de las defunciones de la base cruda y la final no muestra alteraciones con el proceso de depuración de la información.

Coherencia interna de la información y con fuentes externa: la comparación con las defunciones consignadas en el sistema de estadísticas vitales, muestra que la información censal presenta un elevado subregistro en la edad cero y después de los sesenta años de edad. Las defunciones en los restantes grupos de edad presentan valor similar; es decir, tendrían subregistro parecido al de las estadísticas vitales (gráfico 10).

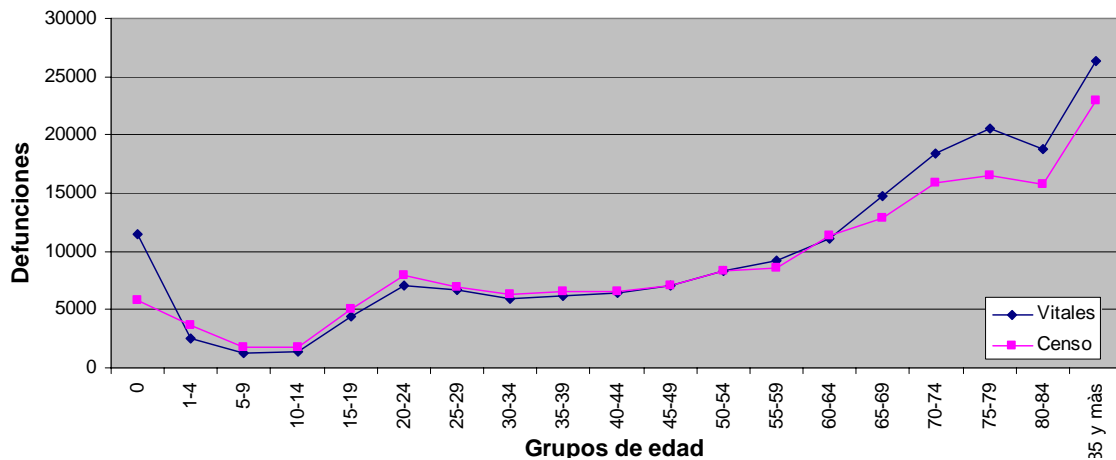
Balance e impacto sobre los resultados: los resultados sobre la población fallecida en los últimos doce meses no permiten realizar mediciones directas de la mortalidad. El significativo subregistro de las defunciones de menores de un año, limitan también el estudio de las relaciones existentes entre la incidencia de la mortalidad y las características socioeconómicas de los hogares.

Cuadro 27
Fallecidos en los últimos doce meses, por base, según
grupos de edad
Censo 2005

Grupos de edad (años)	Cruda		Final	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Total	171 365	100,0	171 287	100,0
Sistema	70	0,0	0	,0
Total	171 435	100,0	171 287	100,0
0	5 753	3,4	5 749	3,4
1-4	3 704	2,2	3 704	2,2
5-9	1 768	1,0	1 767	1,0
10-14	1 717	1,0	1 717	1,0
15-19	4 997	2,9	4 994	2,9
20-24	7 948	4,6	7 942	4,6
25-29	6 943	4,0	6 940	4,1
30-34	6 289	3,7	6 285	3,7
35-39	6 527	3,8	6 524	3,8
40-44	6 506	3,8	6 503	3,8
45-49	7 049	4,1	7 044	4,1
50-54	8 312	4,8	8 308	4,9
55-59	8 563	5,0	8 560	5,0
60-64	11 322	6,6	11 315	6,6
65-69	12 830	7,5	12 822	7,5
70-74	15 853	9,2	15 848	9,3
75-79	16 576	9,7	16 572	9,7
80-84	15 718	9,2	15 713	9,2
85 y más	22 990	13,4	22 980	13,4

Fuente: DANE –Censo General 2005

Gráfico 10
Defunciones, por grupos de edad, según fuente de información
2005



Fuente: DANE –Censo General 2005

Recomendaciones: para la inclusión de esta pregunta en futuros censos se recomienda evaluar la pertinencia de su incorporación en la boleta o dispositivo de captura de la información. Si la respuesta es positiva, debe preguntarse, además, sobre la forma de realizarla y su localización.

2.1.9 Migración

La migración tiene dos componentes: uno relacionado con la migración interna (incluye inmigrantes que vivían en otro país), la cual se capta a través del módulo de personas; y el segundo, correspondiente a la emigración de colombianos al exterior investigada con el módulo de hogares.

Migración interna

Comparación base cruda y base final: la información sobre migración interna comprende, a su vez, dos bloques: el primero relacionado con la residencia 5 años antes del Censo 2005, y el segundo, con la condición de migración durante los 5 años anteriores a la fecha censal.

Residencia 5 años antes del Censo 2005: la comparación de la base cruda con la base final muestra que los procesos de depuración no alteran la estructura por lugar de residencia de las personas 5 años antes de realizarse el Censo (cuadro 28).

Cuadro 28
Distribución de la población, según lugar de residencia hace 5 años
Censo 2005

Lugar de residencia	Base cruda		Base final	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
No había nacido	4 077 732	9,9	4 122 902	10,0
En este municipio	33 619 091	81,6	33 577 118	81,5
En otro municipio	2 671 569	6,5	2 666 142	6,5
En otro país	80 561	0,2	78 931	0,2
Total	40 511 151	98,4	40 445 093	98,2
No sabe, no informa	725 900	1,8	729 760	1,8
Total	41 174 853	100,0	41 174 853	100,0

Fuente: DANE –Censo General 2005

De igual forma, la distribución de los migrantes intermunicipales de los últimos cinco años no cambia al pasar de la base cruda a la final. En síntesis, los procesos de depuración de la información fueron adecuados.

Condición de migración durante los 5 años anteriores a la fecha censal: para analizar la coherencia de la información del segundo bloque, es necesario precisar a qué se hace referencia; de ese modo se pueden identificar 6 grupos de población claramente definidos:

1. Migrantes de los últimos 5 años, de un municipio a otro o de otro país al municipio donde fue censado, y que en el periodo cambiaron de residencia, al menos una vez, dentro del municipio (cuadro 29).

Cuadro 29
Migrantes intermunicipales últimos 5 años, por base, según departamento
Censo 2005

Departamento	Base cruda		Base final	
	Recuento	Porcentaje	Recuento	Porcentaje
Total	2 671 569	100,0	2 666 142	100,0
Antioquia	444 036	16,6	442 909	16,6
Atlántico	96 621	3,6	97 030	3,6
Bogotá, D.C.	335 171	12,5	335 105	12,6
Bolívar	57 773	2,2	57 700	2,2
Boyacá	73 307	2,7	73 409	2,8
Caldas	70 656	2,6	71 035	2,7
Caquetá	31 068	1,2	30 929	1,2
Cauca	54 076	2,0	54 160	2,0
Cesar	46 708	1,7	46 565	1,7
Córdoba	50 298	1,9	50 171	1,9
Cundinamarca	256 432	9,6	257 055	9,6
Chocó	28 890	1,1	28 107	1,1
Huila	66 264	2,5	66 062	2,5
La Guajira	23 482	0,9	21 601	0,8
Magdalena	47 783	1,8	47 688	1,8
Meta	92 920	3,5	92 583	3,5
Nariño	52 944	2,0	52 424	2,0
Norte de Santander	71 132	2,7	70 961	2,7
Quindío	61 830	2,3	61 577	2,3
Risaralda	101 001	3,8	100 822	3,8
Santander	184 290	6,9	184 374	6,9
Sucre	24 777	0,9	24 757	0,9
Tolima	79 569	3,0	79 292	3,0
Valle del Cauca	245 998	9,2	246 009	9,2
Arauca	7 702	0,3	7 629	0,3
Casanare	27 741	1,0	27 736	1,0
Putumayo	25 845	1,0	25 268	0,9
Archipiélago de San Andrés, Providencia y Santa Catalina	1 522	0,1	1 521	0,1
Amazonas	1 380	0,1	1 378	0,1
Guainía	1 101	0,0	1 094	0,0
Guaviare	4 767	0,2	4 728	0,2
Vaupés	1 253	0,0	1 235	0,0
Vichada	3 232	0,1	3 228	0,1

Fuente: DANE –Censo General 2005

2. Migrantes de los últimos 5 años de un municipios a otro o de otro país al municipio donde fue censado, y que en el periodo cambiaron de municipio de residencia, al menos una vez.

3. No migrantes de los últimos 5 años, pero que en el período cambiaron de residencia dentro del municipio
4. No migrantes de los últimos 5 años, que en el período cambiaron de municipio de residencia. Corresponde a los migrantes nacidos en el período.
5. Migrantes de los últimos 5 años que manifiestan no haber cambiado de residencia en el período.
6. No migrantes que manifiestan no haber cambiado de lugar de residencia en el período.

La estructura por condición de migración no cambia significativamente entre la base cruda y la final; sin embargo, en valores absolutos se presentan variaciones en las categorías que pueden incidir en las estimaciones de migración: (1) los migrantes de los últimos 5 años que manifiestan no haber cambiado de residencia en el período (281 346), pasan a engrosar la categoría 9 (no sabe, no responde); (2) los no migrantes que manifiestan no haber cambiado de lugar de residencia en el período, se reducen en 290 682 personas, entre la base cruda y la final; y (3) se incrementan en similar valor los migrantes que cambiaron de municipio (cuadro 30).

Cuadro 30
Condición de migración en los últimos 5 años, por base
Censo 2005

Condición de migración	Base cruda		Base final	
	Recuento	Porcentaje	Recuento	Porcentaje
No sabe, no responde	936 977	2,3	1 206 286	2,9
Migrante - cambió -mismo municipio	786 554	1,9	786 536	1,9
Migrante - cambió -otro municipio	1 671 238	4,1	1 945 667	4,7
No migrante - cambió -mismo municipio	6 153 161	14,9	6 176 005	15,0
No migrante - cambió -otro municipio	404 469	1,0	409 933	1,0
Migrante - no cambió	281 346	0,7	0	0,0
No migrante - no cambió	30 941 108	75,1	30 650 426	74,4
Total	41 174 853	100,0	41 174 853	100,0

Fuente: DANE –Censo General 2005

En síntesis, se corrige la información de la pregunta 30: ¿En dónde vivía hace cinco años? Cuando la persona informa que vivía en este municipio (código 2), pero el departamento o municipio de residencia hace 5 años era diferente al departamento o municipio donde fue censado, se considera como migrante y se le asigna el código 3 y pasan a formar parte de la categoría 2 (migrantes de los últimos 5 años que en el período cambiaron, al menos una vez, de municipio de residencia).

La composición de la migración por la principal causa del movimiento permanece al pasar de la base cruda a la final. Los procesos de depuración fueron adecuados (cuadro 31).

Coherencia con otras fuentes: la única fuente externa de comparación corresponde a las estadísticas sobre desplazados suministrada por Acción Social de la Presidencia de la República, que informa que los desplazados para 2005 son 221 000; o la de ACNUR que para el mismo año, informa de 310 000 en esta condición. Frente a estas cifras, la información censal aparece subregistrada, pero esto es inevitable, el desplazado, por

miedo, prefiere no declarar su condición en censos y encuestas o, atribuyen la migración a razones familiares, ya que en un gran número de oportunidades, al efectuarse el desplazamiento, acuden a ellos para tener vivienda temporal (tabla 4).

Cuadro 31
Principal causa de la migración, según base
Censo 2005

Causa migración	Base cruda		Base final	
	Recuento	Porcentaje	Recuento	Porcentaje
Total	9 389 940	100,0	9 691 488	100,0
Trabajo	1 420 547	15,1	1 510 013	15,6
Desastre natural	107 629	1,1	203 661	2,1
Violencia	310 672	3,3	406 202	4,2
Educación	314 338	3,3	415 055	4,3
Salud	140 828	1,5	244 970	2,5
Razones familiares	4 465 184	47,6	4 563 875	47,1
Otra razón	2 284 724	24,3	2 287 730	23,6
No responde	346 018	3,7	59 982	0,6

Fuente: DANE –Censo General 2005

Tabla 4
Desplazados, según fuente
2001 - 2005

Fuente	2001	2002	2003	2004	2005
Acción Social	170 393	323 461	381 468	194 400	221 681
Acnur					310 387

Fuente: Acción Social, Registro Único de Población desplazada (3/2/2007)

Emigración al exterior

Otra variable de interés en los hogares es el total de emigrantes al exterior; la depuración de los datos no presenta variaciones entre la base cruda y la base final.

En relación con los emigrantes al exterior de toda la vida captados por el Censo 2005 son muy bajos, frente a cualquier estimación, lo cual limita la utilización de esta variable para cuantificar o generar indicadores orientados a la explicación de este proceso (cuadro 32).

Recomendaciones: para el manejo de la información sobre la migración al exterior debe tenerse en consideración las restricciones que presenta. Los valores absolutos seguramente no reflejan la realidad de los flujos permanentes al exterior, pero la composición por países de destino puede ser de alguna utilidad.

Cuadro 32
Emigrantes permanentes al exterior, por base
Censo 2005

Emigrantes por hogar	Base cruda			Base final		
	Hogares	Personas	Porcentaje	Hogares	Personas	Porcentaje
Total	297 831	464 336	100,0	296 060	462 679	100,0
1	205 377	205 377	44,2	203 863	203 863	44,1
2	54 326	108 652	23,4	54 086	108 172	23,4
3	20 452	61 356	13,2	20 321	60 963	13,2
4	9 448	37 792	8,1	9 448	37 792	8,2
5	4 068	20 340	4,4	4 071	20 355	4,4
6	1 800	10 800	2,3	1 827	10 962	2,4
7	962	6 734	1,5	972	6 804	1,5
8	568	4 544	1,0	586	4 688	1,0
9	274	2 466	0,5	281	2 529	0,5
10	420	4 200	0,9	433	4 330	0,9
11	26	286	0,1	40	440	0,1
12	27	324	0,1	46	552	0,1
13	17	221	0,0	20	260	0,1
14	18	252	0,1	21	294	0,1
15	38	570	0,1	45	675	0,1
16	1	16	0,0	0	0	0,0
20	1	20	0,0	0	0	0,0
26	1	26	0,0	0	0	0,0
30	2	60	0,0	0	0	0,0
60	5	300	0,1	0	0	0,0

Fuente: DANE –Censo General 2005

2.2 Variables socioeconómicas

2.2.1 Educación

Se han seleccionado 4 variables educativas que, en su conjunto, reflejan la consistencia y calidad de la información censal: (1) analfabetismo (sabe leer y escribir); (2) asistencia escolar; (3) nivel educativo; y (4) último grado aprobado.

Analfabetismo

Comparación base cruda y base final: para el total del país, las diferencias entre los valores relativos, tanto para alfabetos y analfabetas, son cercanas a 1 punto porcentual, pero con un incremento en valores relativos de la población de 3 años o más que manifiesta no saber leer ni escribir (cuadro 33).

Cuadro 33
Condición de analfabetismo, por base
Censo 2005

Lee y escribe	Base cruda		Base final	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Total	38 656 788	100,00	38 694 074	100,00
0	38 376	0,10	0,00	0,00
Sí	32 317 948	83,60	32 797 134	84,76
No	5 573 131	14,40	5 880 677	15,20
Perdidos/No informa	727 333	1,90	16 263	0,04

Fuente: DANE –Censo General 2005

Con el fin de profundizar en el análisis se toma la tasa de analfabetismo para la población de 15 años o más como indicador de referencia; cuando se comparan los resultados de la base cruda con la base final, las diferencias presentan similar patrón y magnitud que el observado para la población de 3 años y más; para el país la tasa pasa del 8,9% al 9,6%, producto del comportamiento de la zona urbana, en donde el indicador crece del 5,8% al 6,7%; mientras que los valores de la zona rural son similares.

A nivel departamental y por zona, las diferencias son más notorias: en 15 departamentos (17 zonas urbanas y 15 rurales) el analfabetismo crece, con un agravante, que ocurre tanto en departamentos con problemas de recolección (Vaupés, Vichada), como en departamentos en donde estos no se presentaron, así como en Bogotá, D.C. (cuadro 34).

Cuadro 34
Tasas de analfabetismo por zona y base, según departamento
2005

Departamento	Urbano		Rural		Total	
	Cruda	Final	Cruda	Final	Cruda	Final
Total	5,8	6,7	20	19,9	8,9	9,6
Antioquia	5,2	7,2	17,4	16,8	7,7	9,2
Atlántico	6,2	5,7	20,8	19,4	6,9	6,3
Bogotá, D.C.	2,7	5,2	9	10,1	2,7	5,2
Bolívar	9,9	10,6	26,8	25,9	13,5	13,9
Boyacá	5,3	5,6	17,5	16,8	11,1	10,9
Caldas	5,3	4,9	13,5	13,3	7,5	7,1
Caquetá	9,7	12,9	15,1	19	11,4	14,8
Cauca	7,2	6,2	18,1	17,8	13,4	12,9
Cesar	11,1	10,5	29,3	29,3	16	15,5
Córdoba	11,6	10,6	27,7	26,6	19,2	18,2
Cundinamarca	4,4	5,3	11,9	12,1	7	7,7
Chocó	12,4	14,3	31,2	37,8	20,5	24,2
Huila	7,1	7,4	14,6	14,6	10	10,2
La Guajira	8,7	15,8	56,6	61,4	29,9	35,6
Magdalena	10,7	10,4	27,1	27,1	15,5	15,3
Meta	6,1	6,3	13,1	14,6	7,4	7,8
Nariño	7,5	8	19,8	19,4	13,9	13,9
Norte de Santander	7,8	7,3	23,3	23,5	10,9	10,6

Cuadro 34 (conclusión)
Tasas de analfabetismo por zona y base, según departamento
2005

Departamento	Urbano		Rural		Total	
	Cruda	Final	Cruda	Final	Cruda	Final
Quindío	6,3	5,5	11,2	10,5	7	6,1
Risaralda	5,3	4,4	14,6	13,9	7,3	6,4
Santander	5,2	4,7	16,9	15,9	8,2	7,5
Sucre	14,7	13,4	30,7	29,7	20,3	19,1
Tolima	7,3	7,6	18,9	18,3	10,9	10,9
Valle del Cauca	4,9	4,5	11,8	11,9	5,8	5,4
Arauca	8,5	11,8	18,6	21,5	9,7	12,9
Casanare	7	6,1	15	15,8	9,3	9
Putumayo	6,8	8,7	12	14,5	9,2	11,4
Archipiélago de San Andrés, Providencia y Santa Catalina	2,4	1,4	2,3	1,5	2,4	1,5
Amazonas	4,9	4,7	16,6	19,7	9,8	11
Guainía	14,2	12,7	24,1	26,3	18,1	18,2
Guaviare	7,9	20,8	15,5	19,3	9,5	20,5
Vaupés	5,5	24,5	24,5	29,7	10,3	25,8
Vichada	7,4	17,9	24	29,2	16,7	24,2

Fuente: DANE –Censo General 2005

Coherencia de la información con otras fuentes: cuando se comparan con los resultados finales de los dos últimos censos, se puede concluir en periodo intercensal, que no hubo ningún avance en la lucha contra el analfabetismo: el 9,5% en 1993 vs. el 9,6% en 2005. Peor aún, no existe ninguna explicación válida que permita justificar que Bogotá, D.C., casi duplique su tasa de analfabetismo, o que Antioquia la incrementó en 1,6 puntos porcentuales (cuadro 35).

Cuadro 35
Tasas de analfabetismo por departamento, según fuente
Censo 1993 y Censo 2005

Departamento	Urbano		Rural		Total	
	Censo 1993	Final 2005	Censo 1993	Final 2005	Censo 1993	Final 2005
Total	5,6	6,7	20,0	19,9	9,5	9,6
Antioquia	4,5	7,2	16,7	16,8	7,6	9,2
Atlántico	5,3	5,7	19,0	19,4	6,1	6,3
Bogotá, D.C.	2,4	5,2	9,0	10,1	2,5	5,2
Bolívar	8,8	10,6	26,6	25,9	14,0	13,9
Boyacá	5,7	5,6	17,3	16,8	12,4	10,9
Caldas	5,1	4,9	11,8	13,3	7,4	7,1
Caquetá	9,3	12,9	15,6	19,0	12,6	14,8
Cauca	6,6	6,2	20,2	17,8	14,9	12,9
Cesar	11,4	10,5	26,9	29,3	16,9	15,5
Córdoba	12,7	10,6	31,5	26,6	22,1	18,2

Cuadro 35 (conclusión)
Tasas de analfabetismo por departamento, según fuente
Censo 1993 y Censo 2005

Departamento	Urbano		Rural		Total	
	Censo 1993	Final 2005	Censo 1993	Final 2005	Censo 1993	Final 2005
Cundinamarca	4,9	5,3	13,7	12,1	8,9	7,7
Chocó	12,5	14,3	34,6	37,8	25,5	24,2
Huila	8,0	7,4	15,4	14,6	10,8	10,2
La Guajira	10,6	15,8	51,8	61,4	24,4	35,6
Magdalena	9,9	10,4	26,6	27,1	15,6	15,3
Meta	6,1	6,3	14,6	14,6	9,1	7,8
Nariño	7,7	8,0	21,1	19,4	15,2	13,9
Norte de Santander	8,1	7,3	26,3	23,5	13,0	10,6
Quindío	6,5	5,5	11,7	10,5	7,3	6,1
Risaralda	4,8	4,4	12,7	13,9	6,7	6,4
Santander	5,6	4,7	20,6	15,9	10,0	7,5
Sucre	18,1	13,4	35,6	29,7	23,6	19,1
Tolima	7,2	7,6	18,2	18,3	11,3	10,9
Valle del Cauca	4,3	4,5	12,3	11,9	5,4	5,4
Arauca	8,6	11,8	18,4	21,5	12,0	12,9
Casanare	7,1	6,1	16,1	15,8	11,1	9,0
Putumayo	6,9	8,7	14,4	14,5	11,7	11,4
Archipiélago de San Andrés, Providencia y Santa Catalina	1,5	1,4	1,7	1,5	1,6	1,5
Amazonas	6,0	4,7	17,4	19,7	11,2	11,0
Guainía	10,3	12,7	20,9	26,3	17,5	18,2
Guaviare	7,5	20,8	12,8	19,3	10,8	20,5
Vaupés	5,3	24,5	17,0	29,7	14,0	25,8
Vichada	10,1	17,9	18,8	29,2	16,5	24,2

Fuente: DANE –Censo General 2005

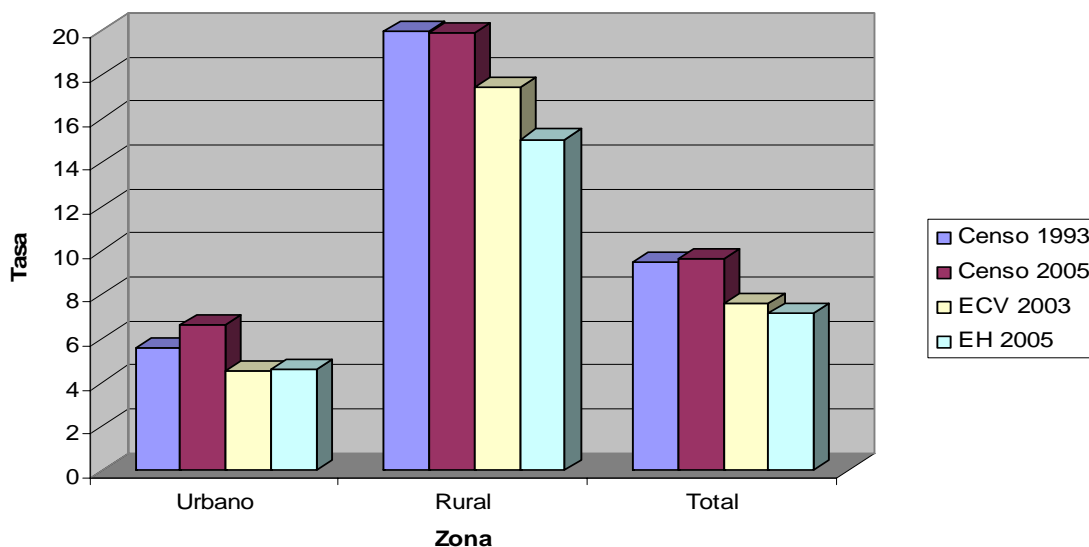
En forma similar, las tasa de analfabetismo para la población de 15 años o más que arroja el Censo, difiere significativamente de las que suministran las encuestas. De acuerdo con estas últimas, en el período 2003-2005, la tasa de analfabetismo vendría en descenso, en particular por el comportamiento observado en la zona rural. Los tasas producto del Censo 2005, para el total del país, superan por más de 2 puntos porcentuales a los de la Encuesta de Hogares 2005; esta diferencia es similar en la zona urbana, pero cercana a los 5 puntos en la zona rural (gráfico 11 y cuadro 36).

Cuadro 36
Tasas de analfabetismo, por fuente, según zona

Zona	Censo 1993		Calidad de Vida 2003		EH 2005		Censo 2005	
	Recuento	Media	Recuento	Media	Recuento	Media	Recuento	Media
Total	21 490 969	9,5	29 905 467	7,6	30 929 789	7,2	28 452 386	9,6
Urbano	15 714 183	5,6	22 672 889	4,5	23 277 377	4,6	22 109 222	6,7
Rural	5 776 716	20,0	7 232 577	17,4	7 652 412	15,1	6 343 164	19,9

Fuente: DANE –Censo General 2005- ECV-EH

Gráfico 11
Tasas de analfabetismo, por zonas, según fuentes



Fuente: DANE –Censo General 2005- ECV-EH

La comparación de las tasas que arroja el Censo 2005 con las suministradas por el Censo 1993 y las de otras fuentes, evidencia que el probable error se genera al pasar de la base cruda a la base ampliada. Los resultados a nivel municipal permiten presumir que el 27% de ellos incrementa su nivel de analfabetismo, algunos con incrementos que de ninguna forma son explicables y que superan los 10 puntos porcentuales (cuadros 37 y tabla 5).

Cuadro 37
Cantidad de municipios que experimentan cambios en la tasa de analfabetismo, al pasar de la base cruda a la ampliada Censo 2005

Incremento o reducción	Recuento	Porcentaje
Más de 10	27	2,4
5 a 10	55	4,9
2 a 5	79	7,1
Más de 0 a 2	142	12,7
Total positivos	303	27,2
0 a -1	325	29,1
-1 a -2	318	28,5
-2 y Más	169	15,2
Total negativos	812	72,8
Total municipios	1 115	100,0

Fuente: DANE –Censo General 2005

Tabla 5
Municipios con incremento en la tasas
de analfabetismo superior a 10 puntos
porcentuales
Censo 2005

Municipios	Tasa de analfabetismo
Argelia	67,9
El Retorno	50,8
Mapiripán	29,8
López de Micay	26,3
Manaure	26,2
Calamar	24,3
Puerto Concordia	23,0
Arauquita	19,3
Vigía del Fuerte	18,3
Medio Atrato	17,4
Mitú	17,1
Bagadó	15,8
Carmen de Atrato	15,5
La Primavera	15,3
Chivatá	14,1
Montecristo	13,7
Carmen del Darién	11,8
La Vega	11,5
Recetor	11,4
Susa	11,1
Nimaima	11,0
Puerto Leguízamo	11,0
La Uribe	10,4
Algeciras	10,4
Cantón de San Pablo	10,2
Venadillo	10,2
Medio San Juan	10,1

Fuente: DANE –Censo General 2005

El problema adquiere mayor relevancia cuando observamos que entre los municipios que incrementan la tasa de analfabetismo están ciudades como Bogotá, Medellín y Cartagena (tabla 6).

Con el propósito de enfocar el problema, se seleccionan las comunas y localidades de Medellín y Bogotá, D.C.. En la primera de las ciudades los problemas son más notorios en las comunas 10, 11, 12 y 15, en los que las tasas de analfabetismo pasan de cerca del 2% en la base cruda del Censo 2005, a valores cercanos o que superan el 10% en la base final (gráfico 12). En la segunda, en la localidad de Barrios Unidos (12) la tasa pasa de menos del 2% en la base cruda a más del 19% en la base final; el problema también es serio en las localidades 14 y 17, en donde los incrementos superan los 8 puntos porcentuales.

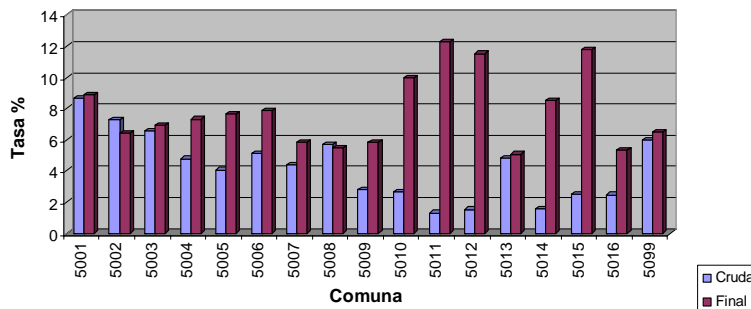
Tabla 6
Capitales que incrementan la tasas de analfabetismo
al pasar de la base cruda a la base final
Censo 2005

Capitales	Base cruda		Base final	
	Recuento	Tasa	Recuento	Tasa
Medellín	1 658 175	4,23	1 655 878	7,58
Bogotá, D.C.	4 936 833	2,68	4 946 310	5,24
Cartagena	628 252	5,83	628 477	7,12
Tunja	106 007	3,78	106 143	4,33
Florencia	85 944	8,31	86 137	12,56
Quibdó	69 302	11,50	69 591	12,05
Riohacha	105 996	14,50	105 763	22,63
Santa Marta	279 934	5,90	279 867	6,32
Pasto	272 806	5,44	272 960	5,94
Ibagué	349 577	5,06	349 733	5,69
Arauca	42 291	8,16	42 435	11,06
San Jose del				
Guaviare	23 532	9,01	23 679	14,95
Mitú	10 343	9,04	10 371	26,17
Puerto				
Carreño	7 371	11,84	7 432	13,59

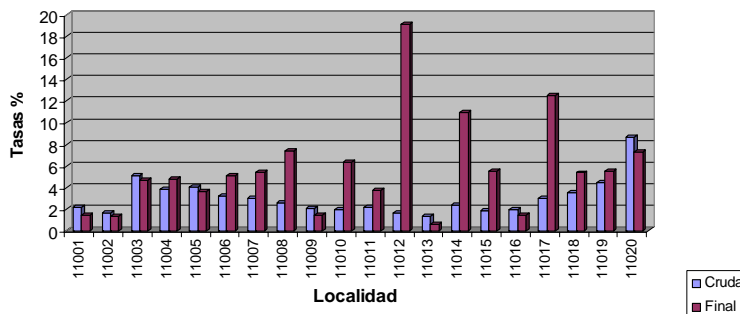
Fuente: DANE –Censo General 2005

Gráfico 12
Tasas de analfabetismo en Medellín y Bogotá, según base
Censo 2005

Medellín. Tasas de analfabetismo por comuna, según base



Bogotá. Tasas de analfabetismo por localidad, según base



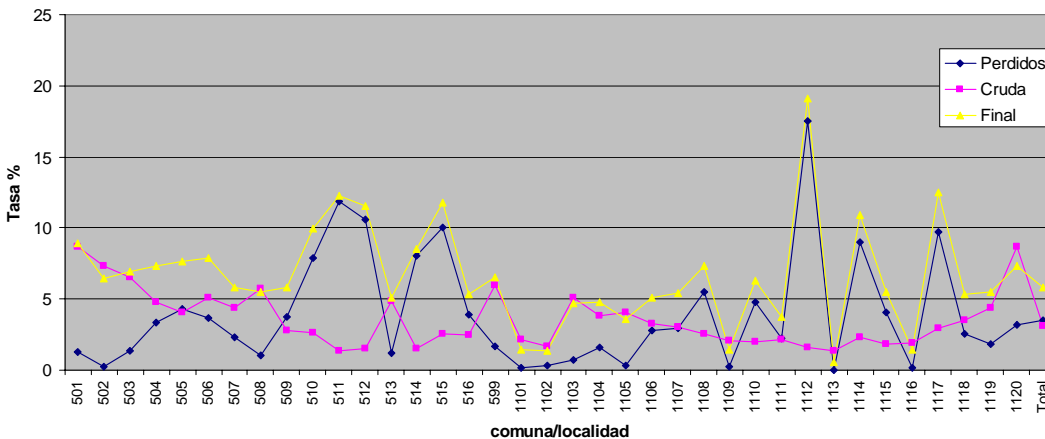
Fuente: DANE –Censo General 2005

Fuente probable de error: las distorsiones se generan cuando existe valor en blanco o nulo (“missing”) en la variable alfabetismo (sabe leer y escribir, pregunta 41) y simultáneamente el nivel aprobado es “missing” (código 13, pregunta 44); los registros con estas características son convertidos a analfabetas en el proceso de imputación.

Por definición operacional, este tipo de situación únicamente podía darse en la información capturada en papel; el dispositivo únicamente permitía que en la pregunta 41 los valores fueran 1 o 2; de otra forma no dejaba continuar con la encuesta. De darse esta situación se esperaría cierta regularidad en el comportamiento de los valores “missing” en relación con los valores válidos y con la aplicación o no de papel para la recolección de la información; sin embargo, esto no ocurre:

- El porcentaje de valores “missing” difiere significativamente en las 37 localidades de Bogotá, D.C.: 15 presentan valores inferiores al 2%; 13, entre el 2% y menos del 5%; 5, entre el 5% y menos del 10%; y 4, con valores superiores al 10% (gráfico 13 y cuadro 38).
- El porcentaje de valores perdidos está directamente asociado al valor que adoptan las localidades en la base final.

Gráfico 13
Tasas de analfabetismo en Bogotá, por localidad, según base
y porcentaje de valores missing
Censo 2005



Fuente: DANE –Censo General 2005

Los valores “missing”, con excepción de 741, fueron grabados en papel y luego transcritos a los DMC, lo cual permite formular 4 hipótesis de necesaria demostración:

1. No existen elementos que expliquen la enorme variación en los porcentajes de valores “missing” entre localidad y localidad.
2. La existencia de valores “missing” en analfabetismo y nivel son los que determinan el valor de la tasa final de analfabetismo.

3. Presuntamente, un gran número de encuestadores no diligenció la información educativa, y por tanto, al transcribirla a la DMC se convierten en valores *missing*.
4. La información estuvo correctamente diligenciada, pero la transcripción a los DMC falló

Las normas de imputación aplicadas fueron correctas, más no así su transformación en el algoritmo desarrollado para implantarlas.

Pruebas preliminares realizadas indican que el problema se origina con los formularios transcritos a DMC, sin ningún proceso de validación; y que la norma de imputación es correcta, pero la lógica del software desarrollado no lo es.

La conclusión es que la variable alfabetismo tiene problemas, con el agravante que está interrelacionada con otras variables educativas que pueden ver afectada su calidad.

Cuadro 38
Bogotá, D.C. Personas con valores *missing* en analfabetismo y nivel educativo, por tipo de grabación, según localidad
Censo 2005

Localidad	Tipo de grabación				Total
	DMC	DIG-DMC	Escáner	Reproceso	
Total	734	155 756	6	1	156 497
Usaquén	79	258	6	0	343
Chapinero	1	303	0	0	304
Santafé	71	232	0	0	303
San Cristóbal	11	4 444	0	0	4 455
Usme	25	586	0	0	611
Tunjuelito	29	3 626	0	0	3 655
Bosa	41	10 103	0	0	10 144
Kennedy	123	36 936	0	0	37 059
Fontibón	78	305	0	0	383
Engativá	44	28 530	0	0	28 574
Suba	112	14 702	0	1	14 815
Barrios Unidos	0	30 471	0	0	30 471
Teusaquillo	1	41	0	0	42
Los Mártires	0	6 452	0	0	6 452
Antonio Nariño	21	3 424	0	0	3 445
Puente Aranda	19	265	0	0	284
Candelaria	3	1 724	0	0	1 727
Rafael Uribe Uribe	42	6 652	0	0	6 694
Ciudad Bolívar	7	6 702	0	0	6 709
Sumapaz	27	0	0	0	27

Fuente: DANE –Censo General 2005

Recomendaciones:

De aplicación inmediata:

- Profundizar en el análisis del problema: su magnitud y fuentes de error.
- Analizar las distintas alternativas y valorar técnicamente el costo de oportunidad para implementarlas.

- Adoptar medidas correctivas de rápida aplicación y que no generen serios traumatismos al Censo 2005.

Una vez conocido el problema, las directivas de la Institución promovieron la realización de los estudios y acciones requeridas para corregir el problema (anexo B, respuesta Institucional).

Para futuros censos:

- No aplicar formularios en papel, de ser necesario, los procesos de capacitación, supervisión y control de calidad deben ser muy exigentes.
- No transcribir los formularios en papel a DMC, se generan dos fuentes de error: el que traen consigo los formularios y el que se genera en la transcripción.

Corrección de la información y resultado final: la corrección de los problemas detectados implicó:

- La revisión del documento de normas, que no generan el resultado final observado.
- Se revisó el código CSPRO que se implementó, y se encontró un cambio de lógica de las normas para el manejo de blancos.
- Se actualiza la base depurada con los datos de la base cruda correspondientes a las variables.
- Se realiza el ajuste a la función que crea la variable auxiliar con el fin de garantizar el funcionamiento adecuado de la imputación en caliente del nivel y año, en aquellos casos que sean inconsistentes.
- Realización de pruebas y conformación de una nueva base después de depurada la información.

El ejercicio de corrección realizado fue el apropiado y la información resultante sobre el alfabetismo de la población es de calidad y puede utilizarse sin ningún tipo de correctivos (anexo B).

Asistencia escolar

Comparación base cruda, base final: a diferencia de lo encontrado para el analfabetismo, la coherencia entre la base cruda y la base final es buena. Para el total de la asistencia escolar, así como para los grupos de referencia de los niveles educativos (primaria, secundaria y superior), las variaciones entre la información original y la final no supera 1 punto porcentual. A nivel departamental y para cada uno de los grupos de edad seleccionados, los resultados entre una y otra base son igualmente coherentes (cuadro 39).

Cuadro 39
Asistencia escolar, por base y grupo de edad, según departamento
Censo 2005

Departamento	Grupos de edad (años)					
	Base cruda			Base final		
	7 a 11	12 - 17	18 - 24	7 a 11	12 - 17	18 - 24
Total	91,0	76,9	27,1	92,0	77,8	27,4
Antioquia	90,9	75,6	27,3	91,8	76,4	27,6
Atlántico	89,7	81,0	29,5	90,6	81,9	29,8
Bogotá, D.C.	94,0	85,5	36,1	94,9	86,4	36,5
Bolívar	90,2	79,3	27,7	91,1	80,1	28,0
Boyacá	93,8	75,0	26,8	94,7	75,8	27,1
Caldas	91,8	75,3	28,8	92,7	76,1	29,1
Caquetá	89,4	69,4	21,5	90,3	70,1	20,3
Cauca	89,8	65,8	21,0	90,9	66,9	21,4
Cesar	86,6	74,1	22,8	87,4	74,8	23,1
Córdoba	90,9	78,6	23,2	91,9	79,4	23,5
Cundinamarca	93,8	79,6	23,8	94,8	80,5	24,1
Chocó	77,9	70,9	29,6	80,0	73,0	29,5
Huila	90,3	67,5	19,7	91,2	68,1	19,9
La Guajira	69,2	64,8	25,6	69,5	65,1	26,2
Magdalena	88,4	77,3	22,6	89,3	78,0	22,8
Meta	90,9	75,9	22,2	92,0	76,8	22,4
Nariño	89,5	64,4	19,3	90,7	65,9	19,6
Norte de Santander	90,1	72,4	24,5	91,0	73,2	24,7
Quindío	91,0	78,7	26,7	91,9	79,5	27,0
Risaralda	90,6	76,4	25,1	91,5	77,2	25,4
Santander	92,3	74,8	28,8	93,3	75,6	29,2
Sucre	92,5	81,8	24,1	93,4	82,6	24,3
Tolima	89,0	73,6	22,9	89,9	74,4	23,1
Valle del Cauca	93,8	79,7	26,1	94,8	80,6	26,3
Arauca	92,3	77,3	18,4	93,6	78,3	18,7
Casanare	92,0	74,7	20,4	92,9	75,5	20,4
Putumayo	91,9	67,5	19,2	93,6	70,7	17,1
Archipiélago de San Andrés, Providencia y Santa Catalina	95,1	88,5	22,2	96,0	89,4	22,4
Amazonas	90,7	78,7	21,9	91,8	80,2	22,5
Guainía	83,7	77,0	31,1	85,2	79,0	33,1
Guaviare	90,7	76,4	17,3	91,7	77,0	16,6
Vaupés	91,9	84,2	39,7	92,7	85,0	40,0
Vichada	74,8	57,3	15,0	76,5	59,6	16,0

Fuente: DANE –Censo General 2005

Coherencia con otras fuentes: cuando se compara la asistencia escolar por grupos de edad, que arroja el Censo 2005 y las encuestas tomadas como referencia, no se observa un patrón definido; mientras que para el grupo 7 a 11 años, la tasa de asistencia en la Encuesta Continua de Hogares es del 97%, en la de Calidad de Vida y el Censo, es cercana al 92%; el comportamiento del grupo 12 a 17 años es similar, más no en el de 18 a 24 años, en el cual las mayores tasas son las del Censo (tabla 7).

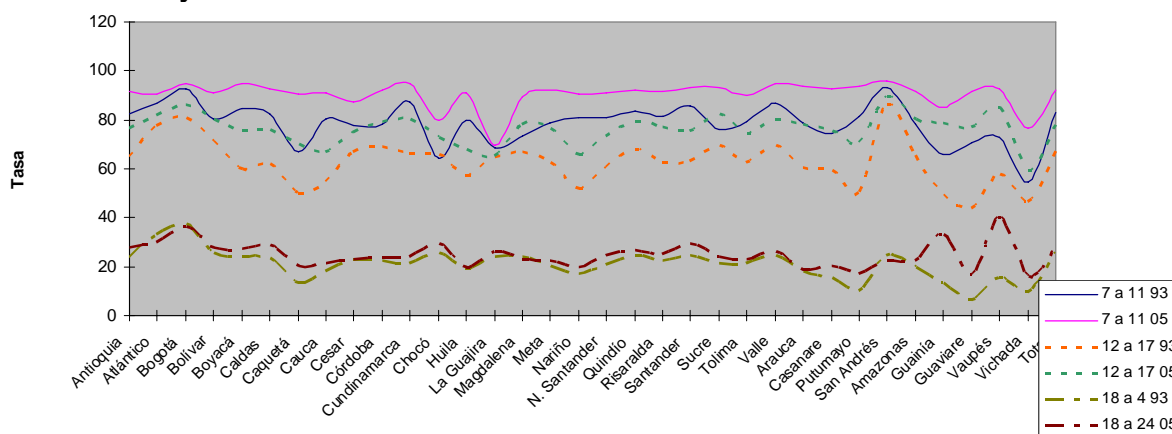
Tabla 7
Asistencia escolar, por fuente, según grupos de edad

Grupos de edad (años)	ECV 2003		EH 2005		Censo 2005	
	Recuento	Media	Recuento	Media	Recuento	Media
7-11	4 866 706,92	92,80	4 884 851,17	96,70	4 243 559,00	92,00
12-17	5 420 058,62	76,20	5 243 379,67	81,20	4 903 234,00	77,80
18-24	5 445 457,35	24,70	5 441 950,33	25,10	4 924 408,00	27,40

Fuente: DANE –Censo General 2005- ECV-EH

La comparación con el Censo 1993, a diferencia de lo que se observa con el analfabetismo, muestra el patrón esperado de acuerdo con los logros que ha experimentado el país en los últimos doce años. En cada uno de los grupos de edad y en todos los departamentos, se incrementa la asistencia escolar (gráfico 14).

Gráfico 14
Tasas de asistencia escolar, por grupos de edad, según departamento
Censos 1993 y 2005



Fuente: DANE – Censo 1993- Censo General 2005

Fuente probable de error: no se evidencian errores en el tratamiento de esta variable.

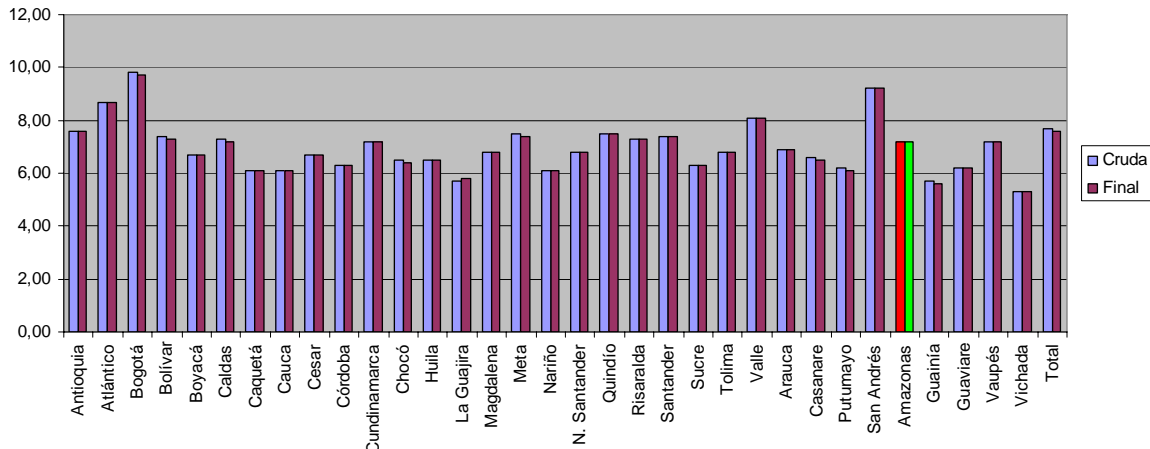
Recomendaciones: no se formulan

Nivel educativo y grado aprobado

El proceso para resolver problemas de respuestas *missing*, inválidas o inconsistentes en la información educativa, primero, imputa el nivel educativo y el grado alcanzado; y segundo, con base en esta información, depura los datos sobre analfabetismo. En la medida en que la traducción de las normas al software presenta limitaciones, los indicadores que toman como referencia las primeras variables, van a verse distorsionados. El principal de ellos son los años promedio de educación que resumen el comportamiento de estas variables.

La primera comparación no evidencia la incidencia del inadecuado tratamiento de las variables educativas. En todos los departamentos los años promedio de educación son similares en la base final que en la base cruda (gráfico 15).

Gráfico 15
Población de 15 años y más. Años promedio de educación, por departamento, según base
Censo 2005



Fuente: DANE –Censo General 2005

A pesar de la similitud en la información, es necesario profundizar en el problema y corregir el software de imputación para después analizar la coherencia interna y con fuentes externas de las variables educativas.

Corrección de la información y resultado final: se esperaba que la revisión y corrección de los programas para imputar la información educativa, trajeran consigo variaciones en la escolaridad de la población, pero esto no ocurrió; una vez se homologan numeradores y denominadores (los valores en blanco en el nivel educativo de la base cruda, se convierten en 0 en la base imputada), la correspondencia en los niveles de escolaridad es buena; los cambios entre la base cruda y las generadas posteriormente (base con errores y base final) no son significativos (anexo B).

El ejercicio de corrección y comparación no es inútil; permite tener la certeza que la información final es de buena calidad y utilizable sin restricciones.

2.2.2 Actividad en la semana previa al Censo

Comparación base cruda y base final: de la comparación entre estas bases surgen varios hechos que ameritan profundización:

- Al parecer a todos los valores perdidos (o a la mayor parte) observados en la base cruda se les asigna código cero en la base final.
- No existe razón aparente por la cual la población estudiante, en particular la menor de 18 años, pierde representatividad; mientras que la población ocupada, la gana (cuadro 40).

A pesar que se reconoce que la información directa (sin ajustes) de los censos no es la más apropiada para medir los indicadores de la fuerza de trabajo, la depuración de los datos cambia radicalmente el valor de ellos (tabla 8).

Cuadro 40
Actividad en la semana anterior al Censo, por grupos de edad, según base
Censo 2005

Actividad semana anterior	5-11 años		12-17 años		18 y más años	
	Recuento	Porcentaje	Recuento	Porcentaje	Recuento	Porcentaje
Base cruda						
Total	5 903 808	100,0	4 885 089	100,0	25 203 039	100,0
0	14 831	0,3	7 856	0,2	23 463	0,1
Trabajó	23 090	0,4	354 684	7,3	12 650 794	50,2
No trabajó, tenía trabajo	12 298	0,2	19 594	0,4	305 433	1,2
Buscó trabajo, había trabajado	1 028	0,0	26 260	0,5	872 407	3,5
Buscó trabajo por primera vez	2 664	0,0	25 540	0,5	179 304	0,7
Estudió, no trabajó, no buscó	4 492 070	76,1	3 281 217	67,2	1 211 015	4,8
Oficios del hogar	147 226	2,5	419 214	8,6	6 528 687	25,9
Incapacitado	22 689	0,4	21 768	0,4	446 572	1,8
Jubilado, rentista	2 863	0,0	2 150	0,0	739 610	2,9
Otra situación	1 185 049	20,1	726 806	14,9	2 245 754	8,9
Base final						
Total	6 052 950	100,0	4 994 283	100,0	26 021 637	100,0
0	164 389	2,7	115 404	2,3	814 036	3,1
Trabajó	95 995	1,6	401 838	8,0	12 691 353	48,8
No trabajó, tenía trabajo	12 323	0,2	19 636	0,4	305 722	1,2
Buscó trabajo, había trabajado	1 025	0,0	26 224	0,5	872 266	3,4
Buscó trabajo por primera vez	2 595	0,0	25 014	0,5	176 883	0,7
Estudió, no trabajó, no buscó	4 421 652	73,0	3 239 324	64,9	1 200 833	4,6
Oficios del hogar	147 148	2,4	418 221	8,4	6 525 065	25,1
Incapacitado	22 689	0,4	21 764	0,4	446 853	1,7
Jubilado, rentista	0	0,0	0	0,0	739 571	2,8
Otra situación	1 185 134	19,6	726 858	14,6	2 249 055	8,6

Fuente: DANE –Censo General 2005

Tabla 8
Tasas globales de participación y de desempleo, por grupo de edad, según fuente
Censo 2005

Base	5 - 11 años		12 - 17 años		18 y más años	
	TGP	Desem-pleo	TGP	Desem-pleo	TGP	Desem-pleo
Cruda	0,8	9,4	10,3	12,2	61,1	7,5
Final	2,4	3,2	11,4	10,8	61,2	7,5

Fuente: DANE –Censo General 2005

Fuente probable de error: indagaciones preliminares indican que a la población que manifestaba estar en cualquiera de las categorías ocupacionales (pregunta 49) y no estuvieran ocupados, se les asignaba código 1 (trabajando). Esto, de acuerdo con el flujo de la información, no cubría a la población inactiva (los remitía a la pregunta 51).

Recomendaciones: las recomendaciones y acciones emprendidas son similares a las planteadas para la información educativa:

- Profundizar en el análisis del problema: su magnitud y fuentes de error.
- Adoptar medidas correctivas de rápida aplicación y que no generen serios traumatismos al Censo 2005.

Corrección de la información y resultado final: el proceso para la corrección de la información sobre la actividad en la semana previa al Censo, es similar al realizado con los datos educativos:

- Recuperación de los registros originales contenidos en la base cruda.
- Aplicación correcta de las normas de imputación.
- Conformación de la nueva base con información de los registros objeto de imputación.

La corrección realizada permite asegurar que la información sobre la actividad realizada por las personas en la semana anterior al Censo, no presenta problema y que puede utilizarse en forma confiable. Así mismo, los correctivos aplicados corrigen simultáneamente la variable categoría ocupacional, al reducir especialmente a la población menor de edad estudiante y antes clasificada como trabajadores familiares sin remuneración.

2.2.3 Rama de actividad económica

Coherencia de la información: los resultados sobre la rama de actividad económica son muy pobres. El DANE no debería presentarla en ninguno de sus resultados, ya que preguntas con niveles del 70% de no respuesta es de muy poco valor para los investigadores del mercado de trabajo (cuadro 41).

Cuadro 41
Población ocupada, por zona de residencia, según actividad económica
Censo 2005

Actividad económica	Total	Cabecera	Porcentaje	Resto	Porcentaje
Total general	13 214 721	10 575 861	100,0	2 638 860	100,0
No informa	8 560 679	7 347 092	69,5	1 213 587	46,0
Agricultura, ganadería, caza y silvicultura	1 349 710	273 539	2,6	1 076 171	40,8
Pesca	29 368	11 235	0,1	18 133	0,7
Explotación de minas y canteras	42 801	19 780	0,2	23 022	0,9
Industrias manufactureras	429 344	382 566	3,6	46 778	1,8
Suministros de electricidad, gas y agua	13 297	12 177	0,1	1 120	0,0
Construcción	268 595	215 408	2,0	53 187	2,0
Comercio al por mayor y al por menor	915 604	849 791	8,0	65 813	2,5
Hoteles y restaurantes	123 671	112 111	1,1	11 560	0,4
Transporte, almacenamiento y comunicaciones	317 246	290 082	2,7	27 163	1,0

Cuadro 41 (conclusión)
Población ocupada, por zona de residencia, según actividad económica
Censo 2005

Actividad económica	Total	Cabecera	Porcentaje	Resto	Porcentaje
Intermediación financiera	53 276	52 274	0,5	1 003	0,0
Actividades inmobiliarias, empresariales y de alquiler	282 163	263 640	2,5	18 523	0,7
Administración pública y defensa	124 327	115 346	1,1	8 981	0,3
Educación	249 348	222 694	2,1	26 654	1,0
Servicios sociales y de salud	165 685	154 078	1,5	11 607	0,4
Otras actividades de servicios comunitarios	145 202	135 010	1,3	10 191	0,4
Hogares privados con servicio doméstico	143 824	118 481	1,1	25 343	1,0
Organizaciones y órganos extraterritoriales	581	558	0,0	23	0,0

Fuente: DANE –Censo General 2005

2.3 Variables de la vivienda y el hogar

2.3.1 Variables de la vivienda

Comparación base cruda y base final: de las características de la vivienda se seleccionan tres variables: tipo de vivienda, disponibilidad de servicios públicos y condición de ocupación. Las dos primeras se utilizan para visualizar los cambios en la integridad de la información al pasar de la base cruda a la final; mientras que la tercera permite efectuar algunos alcances sobre la cobertura censal (cuadro 42).

Cuadro 42
Distribución de las viviendas, por base, según tipo
Censo 2005

Tipo de vivienda	Base cruda		Base final	
	Recuento	Porcentaje	Recuento	Porcentaje
Total	10 390 785	100,0	10 390 207	100,0
Perdidos sistema	7 285	0,1		
Total	10 383 500	99,9		
0	20 292	0,2		
Casa	7 106 464	68,4	7 135 521	68,7
Casa indígena	100 216	1,0	102 745	1,0
Apartamento	2 547 705	24,5	2 577 356	24,8
Tipo cuarto	563 765	5,4	535 948	5,2
Otro tipo	45 058	0,4	38 637	0,4

Fuente: DANE –Censo General 2005

La distribución de las viviendas según tipo ni la tenencia de servicios públicos experimentan cambios en su estructura, al pasar de la base cruda a la base final (cuadro 43).

Coherencia con otras fuentes: la comparación con el Censo 1993 muestra que el período extendido de recolección y el esquema de visitas múltiples para obtener el máximo número de entrevistas completas, permiten reducir significativamente la pérdida de información; así, el 2,4% de viviendas ocupadas con personas ausentes observado en el Censo 1993, desaparece en el Censo 2005.

Cuadro 43
Tenencia de servicios públicos básicos, por base, según tipo
Censo 2005

Servicio	Base cruda		Base final	
	Recuento	Porcentaje	Recuento	Porcentaje
Energía eléctrica				
Total	10 390 785	100,0	10 390 207	100,0
No informa	547 142	5,3	647 279	6,2
Total	9 843 643	94,7		
0	45 466	0,4		
Tiene	9 211 290	88,6	9 119 787	87,8
No tiene	586 881	5,6	623 141	6,0
3	5	0,0		
4	1	0,0		
Acueducto				
Total	10 390 785	100,0	10 390 207	100,0
No informa	548 719	5,3	647 279	6,2
Total	9 842 066	94,7		
0	46 316	0,4		
Tiene	8 189 660	78,8	8 125 908	78,2
No tiene	1 606 090	15,5	1 617 020	15,6
Alcantarillado				
Total	10 390 785	100,0	10 390 207	100,0
No informa	549 040	5,3	647 279	6,2
Total	9 841 745	94,7		
0	54 343	0,5		
Tiene	7 192 489	69,2	7 117 781	68,5
No tiene	2 594 913	25,0	2 625 147	25,3

Fuente: DANE –Censo General 2005

Si bien, para las zonas urbanas la evidencia parece ser incontrovertible, para la zona rural, el panorama no es tan claro: (1) el censo rural por barrido se realizó en 1 019 municipios del país, y permitió la recolección de la información censal de toda el área rural, mediante la visita a la totalidad de las veredas existentes en cada municipio. Para cada uno de estos, se efectuó de manera previa a la recolección, una ubicación aproximada de las viviendas en la cartografía rural, es decir, se visitaba casi que exclusivamente las viviendas que se sabía estaban ocupadas; (2) independiente del período de recolección, la población rural no se inmoviliza, o lo hace en menor medida que la urbana, para atender la realización del Censo; y (3) en el rural disperso no se efectuaron revisitas (cuadro 44).

Los tres factores mencionados permiten suponer que las viviendas ocupadas con personas ausentes, encontradas en terreno, fueron clasificadas como desocupadas. Así, mientras que el porcentaje de viviendas ocupadas es el mismo en los dos censos, el correspondiente a las desocupadas, en el período, se incrementa en 2,7 puntos porcentuales. De ser esto así, la evidencia estaría reflejando un subregistro de viviendas ocupadas rurales y, por tanto, un problema de cobertura.

Cuadro 44
Viviendas particulares, por área, según condición de ocupación
Censos 1993 y 2005

Condición de ocupación	Cabecera		Resto		Total	
	Recuento	Porcentaje	Recuento	Porcentaje	Recuento	Porcentaje
Censo 1993						
Total	4 819 944	100,0	2 104 001	100,0	6 923 945	4 819 944
Ocupada, personas presentes	4 481 080	93,0	1 845 092	87,7	6 326 172	4 481 080
Ocupada, personas ausentes	103 999	2,2	62 126	3,0	166 125	103 999
Desocupada	234 865	4,9	196 783	9,4	431 648	234 865
Censo 2005						
Total	7 826 422	100,0	2 563 785	100,0	10 390 207	7 826 422
Ocupada, personas presentes	7 488 834	95,7	2 254 094	87,9	9 742 928	7 488 834
Ocupada, personas ausentes	603	0,0	273	0,0	876	603
Desocupada	336 985	4,3	309 418	12,1	646 403	336 985

Fuente: DANE

Fuente probable de error: no se evidencian errores en el tratamiento de la información sobre las características de la vivienda.

Recomendaciones: la información sobre las características de la vivienda puede utilizarse en forma confiable.

2.3.2 Variables del hogar

Comparación base cruda y base final: en términos absolutos y relativos, la mayor parte de los valores “perdidos” y algunos correspondientes a los tamaños que pierden representatividad (2 y 3), se agregan a hogares con 6 y más personas. No obstante, su incidencia sobre el tamaño medio del hogar es desestimable (3,8 personas en la base cruda y 3,9 personas en la final) (cuadro 45).

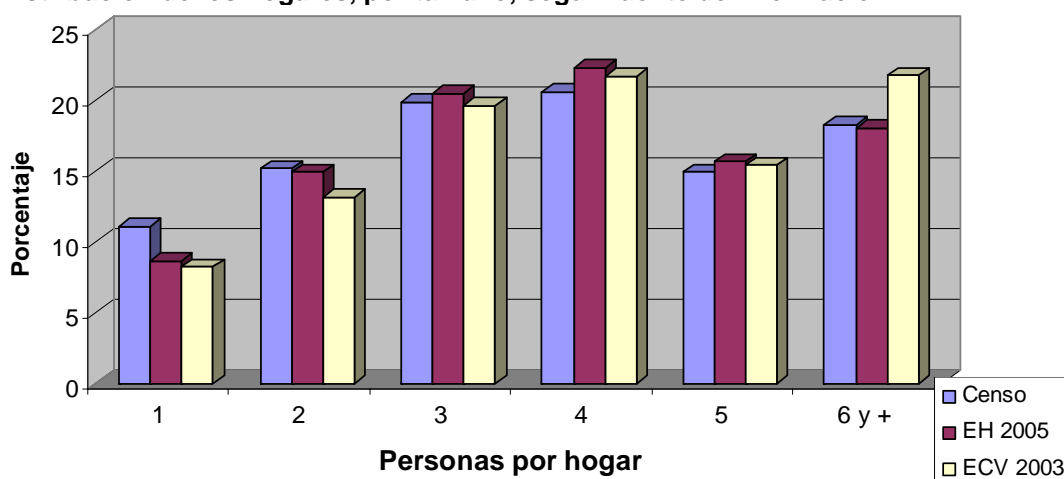
Cuadro 45
Distribución de los hogares, por base, según tamaño

Tamaño (personas por hogar)	Cruda		Final	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Total general	10 735 948	100,0	10 570 899	100,0
Sistema	314 529	2,9	0	0,0
Total	10 421 419	97,1	10 570 899	100,0
1	1 161 792	10,8	1 176 558	11,1
2	1 640 526	15,3	1 604 082	15,2
3	2 155 316	20,1	2 099 506	19,9
4	2 215 093	20,6	2 173 801	20,6
5	1 533 002	14,3	1 580 188	14,9
6 y más	1 715 690	16,0	1 936 764	18,3

Fuente: DANE –Censo General 2005

Coherencia de la información: lo que muestra la comparación entre los resultados censales y los de las encuestas cercanas a la fecha del empadronamiento, es que en el Censo son relativamente mayores los hogares con una o dos personas, lo cual, lo que se podría explicar, dado que las encuestas por muestreo no están diseñadas para captar suficientemente ciertas especificidades, o estar indicando la existencia de subregistro al interior de los hogares y que se produce cuando uno o más de sus miembros no atienden el llamado de inmovilización y no son declarados como residentes habituales del hogar (gráfico 16).

Gráfico 16
Distribución de los hogares, por tamaño, según fuente de información



Fuente: DANE –Censo General 2005- ECV-EH

Fuente probable de error: no se evidencian errores en el tratamiento de la información sobre las características del hogar.

RESUMEN

Balance del análisis de la depuración de la información

El objetivo central del ejercicio realizado era el de verificar que los procesos de corrección de los datos incoherentes (consistencia interna) y la adición de datos que no existen en el cuestionario, pero que admiten que se formulen supuestos razonables al respecto (imputación), no alteraran en forma significativa el volumen, estructura y composición de la información producto del recuento censal.

La metodología radicó en determinar los cambios que experimenta la información entre la base inicial (cruda) y la base resultado de todos los procesos de depuración (final). De igual forma, acorde con los resultados de otras fuentes de información (especialmente censos anteriores y encuestas), se analizó la coherencia de los resultados del Censo 2005 para cada una de las variables seleccionadas para conformar la línea de base.

No todas las variables fueron incluidas en la comparación entre la base cruda y la final, se

seleccionaron aquellas de mayor relevancia en el análisis demográfico y socioeconómico, así: (1) variables demográficas : volumen y distribución por clase, edad, sexo, parentesco, estado conyugal, etnia, migración, fecundidad, mortalidad; (2) variables educativas: analfabetismo, asistencia escolar, años promedio de educación; (3) variables de la actividad y participación económica: actividad la semana previa al censo y rama de actividad económica; (4) variables de la vivienda y el hogar: tipo de vivienda, condición de ocupación y disponibilidad de servicios públicos, tamaño y composición de los hogares y emigrantes del hogar al exterior.

El balance general indica que, al igual que la mayoría de los censos, el de 2005 presenta errores producto de la mala información suministrada por el encuestado o el mal manejo de los conceptos y los medios de recolección de la información, por parte del encuestador. Algunos de estos errores se corrigen en los procesos de depuración de la información (corrección de datos incoherentes y formulación de supuestos razonables sobre el valor de los datos), pero otros permanecen y deben ser tratados mediante metodologías especialmente diseñadas para dar coherencia a la información. En el Censo 2005 se presentan estos errores, pero ninguno de ellos reviste la importancia que conduzca a descalificar la calidad de la información censal.

En el Censo 2005 aparece otro tipo de errores generados por el manejo inadecuado de algunos procesos o por una mala interpretación y traducción de las normas de imputación en algoritmos útiles para depurar la información. Una vez se detectan estos problemas los programas se corrigen y se reprocesa la información. Los procesos realizados respetan la integridad de la información y la lógica de los procesos de imputación, por tanto, los resultados de las variables tratadas son de calidad y dejan de ser una limitante para la utilización de la información del Censo 2005.

A continuación se presenta un resumen de los principales hallazgos del análisis realizado:

Variables demográficas

Distribución espacial de la población: tres desagregaciones geográficas se tomaron para el análisis de la distribución espacial de la población: (1) distribución de la población por clase; (2) los centros poblados; y (3) las áreas geográficas (AG). En la primera de ellas, al comparar los resultados de la base cruda con la base final, se encontraron pequeñas variaciones en 73 municipios. Para corregir las desviaciones, el DANE llevó a cabo un ejercicio de restitución de AG que asegura que la población corregida es la que le corresponde al municipio.

Pocos centros poblados y algunas AG existentes no aparecen consignados en la base censal, varios factores podrían contribuir a la explicación de los faltantes (Centro poblados: conversión en municipio, desaparición, contado como parte del rural disperso; áreas geográficas: la población fue censada en otras AG); sin embargo, no puede descartarse que exista un problema de no cobertura.

El análisis de las tres desagregaciones geográficas evidencia problemas de desactualización cartográfica y eventualmente, problemas de organización del operativo. No obstante, ninguna de ellas genera limitantes para la utilización confiable del Censo.

Edad: la información sobre la edad es un buen ejemplo de cómo la utilización de los DMC y el software (controles), mejoran la calidad de la información censal. Los índices de calidad analizados muestran que la información capturada en los DMC (95% del total) se encuentra en los rangos de la información de buena calidad; cuando la captura es cuestionario en papel (3% del total) y posteriormente grabada en DMC, la información aún se encuentra en los límites de aceptabilidad; y cuando la recolección se realiza mediante papel y luego se graba mediante escáner (2%), es de calidad media.

De igual forma, la distribución por edades simples del Censo 2005 es, en balance con la que suministran los censos anteriores, significativamente mejor; ya no son tan marcadas las distorsiones producto del subregistro en algunas edades (0-4) y las ocasionadas por la preferencia de dígitos.

A pesar de los logros mencionados, la evidencia externa y la derivada del Censo 2005, confirman la existencia de sobrestimación de la población menor de 1 año; el impacto importante de este comportamiento es sobre las proyecciones de población. En los ejercicios de conciliación censal adelantados por el DANE, este efecto ya ha sido considerado y controlado.

Sexo: el proceso de depuración de la variable fue apropiado, pues una vez realizada la depuración de la información, la estructura de la población por sexo, entre la base cruda y la base final, no cambia.

No obstante, la comparación con fuentes externas y el análisis de la coherencia de la información muestra un subregistro mayor de niñas que de niños en la edad 0. Nuevamente, el efecto de este comportamiento es sobre las proyecciones de población y al igual que con la edad, se ha considerado en el ejercicio de conciliación censal

Las relaciones de masculinidad en otros grupos de edad, cuando se comparan con otras fuentes, no permiten explicación directa, por tanto, se considera que amerita la realización de un estudio específico para su análisis.

Parentesco: los procesos de depuración efectuados y su coherencia con fuentes externas indican que la información sobre el parentesco no presenta limitantes para su utilización.

Estado conyugal: esta información presenta consistencia y calidad; y al igual que la de parentesco, no presenta limitantes para su utilización.

Pertenencia étnica: los valores absolutos de la población Rom y los palenqueros, al pasar de la base cruda a la final, experimentaron cambios significativos. A pesar de las dificultades para la imputación de la población Rom, los resultados pueden utilizarse en forma confiable.

Fecundidad: la evidencia presentada muestra la existencia de subregistro de los nacimientos ocurridos en el año anterior a la fecha cuando la mujer en edad reproductiva fue censada, la causa se atribuye a que los hijos de las mujeres censadas en el año 2006 y

tenidos en este año, por restricciones del instrumento de recolección, no podían quedar incorporados en el año de nacimiento o, en su defecto, quedar en años diferentes al de la ocurrencia del evento.

Un primer ejercicio realizado para determinar la incidencia del aparente subregistro, muestra que sí existe y que afecta en mayor medida a los nacidos vivos de madres censadas en 2006. El subregistro no limita la utilización de la información, pero recomienda la utilización de metodologías específicamente diseñadas para corregir este tipo de errores comunes en censos y encuestas por muestreo.

Mortalidad: la fuente de información sobre mortalidad proviene del módulo de hogares. La comparación de la estructura de las defunciones de la base cruda y la final no muestra alteraciones con el proceso de depuración de la información.

La comparación con otras fuentes muestra que la información censal presenta un elevado subregistro en la edad cero y después de los sesenta años de edad. Las defunciones en los restantes grupos de edad presentan similar valor; es decir, tendrían subregistro parecido al de las estadísticas vitales.

Migración interna: la estructura por condición de migración no cambia significativamente entre la base cruda y la final. La información general de migración puede utilizarse en forma confiable.

La comparación con otras fuentes muestra que la migración por causas atribuibles al desplazamiento forzado, encubre la real incidencia del problema. Este problema no es del Censo 2005, está asociado a las dificultades que existen para obtener este tipo de información en censos y encuestas.

Emigración al exterior: la depuración de los datos no presenta variaciones entre la base cruda y la base final, pero los emigrantes al exterior de toda la vida captados por el Censo 2005 son, frente a cualquier estimación, muy bajos, lo cual limita la utilización práctica de esta variable.

Variables socioeconómicas

Alfabetismo: cerca del 3% de la información recolectada en papel se transcribió a los DMC; para permitir el ingreso de la información se levantaron los controles que aseguraban los flujos adecuados y la validación de los datos incorporados. El proceso afectó la calidad de las variables educativas.

A los problemas relacionados con la transcripción de la información recolectada en papel a los DMC, se agrega que, a pesar de que las normas de imputación eran las adecuadas, la lógica del software desarrollado no se correspondía con ellas. Esto generó significativas distorsiones en los niveles de analfabetismo y afectan, aunque en menor medida, los niveles de escolaridad, tanto del país, como de los departamentos y los municipios.

El ejercicio de corrección realizado fue el apropiado y la información resultante sobre el alfabetismo de la población es de calidad y puede utilizarse sin ningún tipo de correctivos.

Asistencia escolar: la coherencia entre la base cruda y la base final y de esta última con otras fuentes de información, es buena. No se evidencian errores en el tratamiento de esta variable.

Nivel educativo y grado aprobado: existía el temor de que el proceso que afectó la variable relacionada con el alfabetismo impactara, igualmente, las variables relacionadas con el nivel educativo y el grado alcanzado en ese nivel, pero no fue así, los cambios observados en el indicador de escolaridad calculados con base en la información original y los estimados en las dos bases siguientes (anterior y final) no son significativos, pero el haber hecho el ejercicio de corrección del alfabetismo, confieren seguridad sobre la calidad de la información educativa de nivel y grado.

Actividad realizada por las personas la semana anterior al censo: se observó que la población estudiante pierde representatividad, en particular la menor de 18 años; mientras que la población ocupada, la gana. Se detectó que en la información recolectada en papel y transferida mediante escáner (rutas), se le asignó, también, la condición de trabajador familiar sin remuneración a un gran número de menores de 18 años que estaba estudiando. El problema se genera cuando a estos últimos, por una inadecuada interpretación de la norma de imputación, se les transforma en trabajadores.

La corrección realizada, similar a la hecha en la variable alfabetismo, permite asegurar que la información sobre la actividad realizada por las personas en la semana anterior al censo, no presenta problema y que puede utilizarse en forma confiable. La corrección en la variable actividad corrige automáticamente la variable categoría ocupacional.

Rama de actividad económica: los resultados sobre la actividad económica son muy pobres, pues se presentan preguntas con niveles del 70% de no respuesta en la zona urbana y del 46%, en la rural, lo que limitan su utilización confiable.

Variables de la vivienda y el hogar

De la vivienda y el hogar se contemplaron cuatro variables: tipo de la vivienda, condición de ocupación, servicios públicos básicos y tamaño del hogar. La estructura de ninguna de ellas experimenta cambios significativos al pasar de la base cruda a la final, lo que elimina cualquier restricción para su uso.

Respuesta institucional a los problemas de depuración detectados

Una vez detectados los problemas en las variables educativas y en la actividad en la semana anterior al censo, las directivas del DANE suministraron el recurso humano y tecnológico para profundizar en la fuente del problema y llevar los correctivos del caso para asegurar que la información resultante tuviera los niveles de integridad y de calidad esperados.

Una vez precisada la fuente del problema en cada una de las variables, se llevó a cabo la corrección de la lógica del software de depuración, se realizaron pruebas con base en municipios críticos y se generó la base nacional con las correcciones requeridas. Los correctivos aplicados cambian radicalmente el panorama de las variables e indicadores tratados y se observa coherencia con los derivados de la base cruda.

En la medida en que los errores detectados no eran problemas del Censo, sino de los procesos de grabación y lógica del software, la información de las variables, objeto de tratamiento e incorporadas en la nueva base, dejan de ser una limitante para la utilización confiable del Censo General 2005.

CONCLUSIONES Y RECOMENDACIONES

- El Censo continúa siendo la mejor y, la mayor parte de las veces, única fuente para conocer las características demográficas y socioeconómicas de la población y herramienta imprescindible para la planeación económica y social a nivel nacional, departamental y municipal.
- Al igual que la mayoría de los censos, el de 2005 presenta errores producto de la mala información suministrada por el encuestado o el mal manejo por parte del encuestador de los conceptos y los medios de recolección de la información. Adicional a ellos, en el Censo aparece otro tipo de error generado por el manejo inadecuado de algunos procesos o por una mala interpretación y traducción de las normas de imputación en algoritmos útiles para depurar la información.
- Existen algunas características de la población de difícil investigación a través de censos y encuestas no especializadas; por ejemplo, la emigración al exterior, el desplazamiento forzado, la rama de actividad económica y otras. Su inclusión debe evaluarse cuidadosamente.
- La calidad de la información recolectada en el Censo 2005, por la utilización de los DMC en la recolección es, en términos generales, superior a la que arrojan censos anteriores. El cambio tecnológico incorporado es irreversible; de ninguna forma puede regresarse a la recolección de la información en papel; por el contrario, para futuros eventos censales debe utilizarse los últimos desarrollos, tanto en tecnología blanda, como dura.
- Una variable que demandó tratamiento especial fue la fecundidad; problemas en el proceso de recolección de la información, determinaron la necesidad de profundizar en su análisis. El ejercicio realizado permite señalar que existe el subregistro de los hijos nacidos vivos tenidos por las mujeres en el último año, pero aplicando metodologías específicas diseñadas para ajustar información por falta de respuesta, la información sobre esta variable puede utilizarse en forma confiable.
- La revisión y corrección de la lógica del software utilizado para la aplicación de las normas de imputación, debida a los problemas detectados en el proceso de depuración de la información, permiten que las variables educativas y la correspondiente a la actividad realizada en la semana anterior al censo, se puedan utilizar en forma confiable.
- Una vez obtenida información precisa sobre las variables educativas y de la actividad de la persona, no existe evidencia de problemas que invaliden el Censo. Las limitaciones menores que se detectaron, afectan en particular al estudio de conciliación censal; en él se les dio el tratamiento requerido.

- Es necesario integrar y complementar los resultados del análisis llevado a cabo con el ejercicio de conciliación censal, en la medida que algunos de los hallazgos pueden condicionar los parámetros y variables fundamentales utilizados en este ejercicio.
- El análisis de integridad y calidad de la información, como el efectuado en el presente estudio, debe institucionalizarse como proceso básico del Censo.
- Los niveles de cobertura siguen siendo una de las principales inquietudes sobre el Censo; sin embargo, y sin restar la importancia que este parámetro tiene en los resultados, puede señalarse que la información censal puede utilizarse en forma confiable.

Algunas lecciones aprendidas

Del análisis de la información y de los problemas detectados, surgen indicadores relacionados con el uso de la tecnología y la calidad de algunos procesos que deben tomarse en consideración para futuros censos.

- Los errores detectados en el Censo General 2005 deben minimizarse en futuros censos; para el efecto, se requiere analizarlos, sistematizarlos y diseñar los controles necesario para que estos no ocurran o su incidencia sea muy baja.
- En relación con las características de la población de difícil investigación, para futuros censos es recomendable, no sólo revisar los resultados de pasadas experiencias, sino también realizar las suficientes pruebas de campo que aseguren la calidad final de la información investigada.
- El avance tecnológico más importante incorporado en el Censo 2005 corresponde a la recolección de la información mediante los DMC. Su utilización mejora significativamente la calidad de los resultados. No obstante, para eventos futuros, surgen algunas recomendaciones:
 - La tecnología y software de captura que vaya a utilizarse debe estar lo suficientemente validada y configurada antes de iniciar el operativo censal.
 - El uso de la tecnología no está exenta de enfrentar problemas emergentes en el proceso de recolección de la información (utilización de formularios en papel). Ante la eventualidad de que esto ocurra, los procesos deben estar plenamente definidos, incluyendo la necesaria supervisión y control de calidad.
- El diseño, alistamiento y ejecución del Censo, deben responder a tiempos previamente establecidos y que deben cumplirse estrictamente. Las eventualidades surgidas durante el operativo, no pueden ir en contra de la integridad y calidad de la información.
- Los procesos de transcripción de la información de papel a DMC y la depuración de la información deben estar sujetos a controles exhaustivos de calidad.

- Prestar especial atención a los procesos de capacitación y supervisión: en el primero de los casos debe darse énfasis al manejo de los conceptos básicos, no sólo para la captura mediante los DMC, sino también para el empadronamiento en papel; en el segundo, para que pueda llevarse a cabo una adecuada supervisión, racionalizar las actividades propias que debe realizar el supervisor.
- No es suficiente contar con cartografía, lo que es valioso para el Censo es su actualización. El DANE debe emprender el proceso de actualización teniendo en cuenta los requerimientos de los censos para la organización del operativo y, conjuntamente con otros instrumentos, apoyar el aseguramiento de la cobertura censal.
- El análisis de integridad de la información, como el efectuado en el presente estudio, debe institucionalizarse como proceso básico del Censo.

ANEXOS

Anexo A. Metodología para determinar los niveles de subregistro de los nacimientos de las mujeres censadas en 2006

1. Fundamento

Los hijos sobrevivientes menores de un año, reportados por las mujeres en edad reproductiva, son una buena aproximación a los hijos nacidos que los originan. La mortalidad infantil a la que estos últimos se ven sometidos, no alteran en forma significativa las tasas de fecundidad estimadas con base en ellos.

2. Limitaciones

- Para no toda la población menor de un año se reportó la fecha de nacimiento: subregistro de la población.
- Existe una sobrestimación de la población menor de un año cuando esta se estima con base en la edad reportada.
- Se producen ligeras distorsiones producto de malas declaraciones de la edad o la fecha de nacimiento por efecto recordación.

3. Metodología

Para tener en cuenta las dos primeras limitaciones, la propuesta metodológica toma como fundamento las mediciones de las tasas de las mujeres censadas en 2005 y 2006; para el efecto se incluyen 3 diferentes mediciones para cada conjunto de población:

1. Tasa Global de Fecundidad –TGF– originada por las mujeres en edad fértil y la población menor de un año, estimada con base en la fecha de nacimiento (límite inferior).
2. TGF de las mujeres en edad fértil y la población menor de un año estimada con base en la edad (límite superior).
3. TGF calculadas con base en la información de año y mes cuando la mujer tuvo su último Hijo Nacido Vivo –HNV.

La comparación entre los estimativos para las mujeres censadas en 2005 tienen como objetivo validar o negar la metodología; mientras que los relacionados con las encuestas en 2006 proporcionan indicativos sobre los niveles de subregistro de los hijos nacidos en 2006.

4. Proceso

Para obtener los diferentes estimativos se llevó a cabo el siguiente proceso:

Para las mujeres censadas en 2005 y 2006 con información sobre hijos sobrevivientes

- Identificación en la base nacional de la población menor de un año, según edad y fecha de nacimiento.
- Asignación a cada hogar del número de hijos menores en las dos dimensiones.
- Conformación de dos grupos de mujeres en edad fértil: las censadas en 2005 y 2006.
- Para cada uno de los grupos, en lo posible, identificar las madres de los menores, para lo cual se aplican los siguientes filtros secuenciales y excluyentes entre sí:
 - Se consideran sin hijos aquellas que informan que no han tenido hijos vivos durante su vida reproductiva (pregunta 56, opción 2) (cuadro A2).

Cuadro A1

Mujeres, según número de menores en el hogar

Menores en el hogar	Mujeres	
	Frecuencia	Porcentaje
Total	2 819 367	100,0
Sin menores	2 484 449	88,1
1	303 455	10,8
2	18 872	0,7
3	2 742	0,1
4	8 478	0,3
5	690	0,0
6	264	0,0
7	108	0,0
8	185	0,0
9	36	0,0
10	38	0,0
11	3	0,0
12	44	0,0
13	1	0,0
14	2	0,0

Fuente: DANE

**Cuadro A2
Corrección por HNV**

Menores en el hogar	Mujeres	
	Frecuencia	Porcentaje
Total	2 819 367	100,0
Sin menores	2 533 865	89,9
1	258 901	9,2
2	15 657	0,6
3	2 216	0,1
4	7 600	0,3
5	574	0,0
6	203	0,0
7	95	0,0
8	150	0,0
9	32	0,0
10	30	0,0
11	3	0,0
12	38	0,0
13	1	0,0
14	2	0,0

Fuente: DANE

- Se incluyen todas las mujeres en hogares en los que el número de ellas en edad fértil es igual al número de menores reportados (cuadro A3).
- Se consideran con hijos las que informaron que tuvieron un HNV en 2005 (cuadro A4).

Cuadro A3
Corrección por mujeres en el hogar

Mujeres/niños en el hogar	Frecuencia	Porcentaje
Total	2 819 367	100,0
0	2 533 865	89,9
1	114 209	4,1
2	10 228	0,4
3	1 448	0,1
4	1 947	0,1
5	328	0,0
6	110	0,0
7	43	0,0
8	108	0,0
9	32	0,0
10	30	0,0
11	3	0,0
12	38	0,0
13	1	0,0
14	2	0,0
1 mujer, 1 menor	144 692	5,1
2 mujeres, 2 menores	5 429	0,2
3 mujeres, 3 menores	768	0,0
4 mujeres, 4 menores	5 653	0,2
5 mujeres, 5 menores	246	0,0
6 mujeres, 6 menores	93	0,0
7 mujeres, 7 menores	52	0,0
8 mujeres, 8 menores	42	0,0

Fuente: DANE

Cuadro A4
Consideración por último HNV

Mujeres/niños en el hogar	Frecuencia	Porcentaje
Total	2 819 367	100,0
0	2 533 865	89,9
1	82 829	2,9
2	7 383	0,3
3	1 081	0,0
4	1 722	0,1
5	282	0,0
6	92	0,0
7	42	0,0
8	92	0,0
9	24	0,0
10	27	0,0
11	3	0,0
12	32	0,0
13	1	0,0
14	2	0,0
1 mujer, 1 menor	144 692	5,1
2 mujeres, 2 menores	5 429	0,2
3 mujeres, 3 menores	768	0,0
4 mujeres, 4 menores	5 653	0,2
5 mujeres, 5 menores	246	0,0
6 mujeres, 6 menores	93	0,0
7 mujeres, 7 menores	52	0,0
8 mujeres, 8 menores	42	0,0
Con último HNV	34 915	1,2

Fuente: DANE

- Cuando el grupo quinquenal de edad de mujeres contiguas en un mismo hogar es el mismo y se consideran como información replicada, a una de ellas se le considera con hijo y a la pareja, sin él. Para efectos de la medición de las tasas el procedimiento es correcto (cuadro A5).
- No todas las personas inicialmente consideradas como información duplicada, triplicada, lo continúan siendo, por lo que el ejercicio para identificar las réplicas reporta algunas personas que dejaron de ser replicas y pueden incluirse en el grupo de mujeres con niños menores (54 186) (cuadro A6).
- Finalmente, quedan personas que tienen una o más réplicas y que no pudieron identificarse (30 906); con ellas se conforma un archivo del cual se extraen en forma aleatoria aquellas personas que hacen falta para completar el total de menores observados inicialmente (3 955) (cuadro A7).

Cuadro A5
Parejas en igual grupo quinquenal de edad

Mujeres/niños en el hogar	Frecuencia	Porcentaje
Total	2 819 367	100,0
0	2 537 361	90,0
1	76 134	2,7
2	6 407	0,2
3	769	0,0
4	1 444	0,1
5	151	0,0
6	54	0,0
7	21	0,0
8	54	0,0
9	19	0,0
10	19	0,0
11	3	0,0
12	14	0,0
13	1	0,0
14	2	0,0
1 mujer, 1 menor	144 692	5,1
2 mujeres, 2 menores	5 429	0,2
3 mujeres, 3 menores	768	0,0
4 mujeres, 4 menores	5 653	0,2
5 mujeres, 5 menores	246	0,0
6 mujeres, 6 menores	93	0,0
7 mujeres, 7 menores	52	0,0
8 mujeres, 8 menores	42	0,0
Con último HNV	34 915	1,2
Mismo grupo de edad	5 024	0,2

Fuente: DANE

Cuadro A6
Replicas que permanecen en el archivo

Concepto	Frecuencia	Porcentaje
Total	85 092	100,0
Sin réplica	54 186	63,7
Una réplica	29 275	34,4
Dos réplicas	1 631	1,9

Fuente: DANE

Cuadro A7
Resumen

Concepto	Mujeres	
	Frecuencia	Porcentaje
Total	2 819 367	100,0
Sin menores	2 537 361	90,0
Con hijos	251 100	8,9
Réplicas	30 906	1,1

Fuente: DANE

- Con base en el archivo depurado se estima la TGF derivada de los niños menores.

Cuadro A8
Tasas específicas de fecundidad y TGF, con base en la edad al nacimiento

Grupos de edad (años)	Mujeres	Tasa específica	Menores
15-19	570 756	0,08	44 460
20-24	470 674	0,14	66 973
25-29	426 706	0,12	50 731
30-34	381 972	0,09	35 059
35-39	369 797	0,07	26 402
40-44	324 501	0,06	18 310
45-49	274 961	0,05	13 120
Sumatoria		0,61	255 055
TGF		3,03	

Fuente: DANE

Para las mujeres censadas en 2005 y 2006 con información sobre hijos año y mes del último hijo nacido vivo

Es preciso generar el promedio de dos estimativos de las TGF:

- *Alto*: año y mes de nacimiento +12>= a mes del censo (sobrestima los nacimientos)
- *Bajo*: año y mes de nacimiento +12> a mes del censo (subestima los nacimientos)

Cuadro A9
TGF con base en Último HNV, por estimativo
Mujeres censadas en 2006

Grupos de edad (años)	Alto	Bajo	Medio
15-19	0,09	0,08	0,09
20-24	0,16	0,14	0,15
25-29	0,13	0,11	0,12
30-34	0,09	0,08	0,09
35-39	0,06	0,05	0,05
40-44	0,02	0,02	0,02
45-49	0,00	0,00	0,00
Sumatoria	0,55	0,50	0,52
TGF	2,74	2,48	2,61

Fuente: DANE

5. Resultados

- Los resultados para la población censada en 2005 muestran que la metodología, aunque puede afinarse, suministra resultados consistentes que permiten evaluar el impacto de la no inclusión de los nacimientos ocurridos en 2006, en el año correspondiente (cuadros A10 y gráficos A1).
- El subregistro de los HNV en el último año existe, afecta en mayor medida a los hijos de las mujeres censadas en 2006.
- El denominador de los menores de un año corresponde a todas las mujeres en edad fértil entre 15 y 49 años; para el último hijo nacido vivo, toma a las mujeres que informaron haber tenido o no algún hijo nacido vivo (excluye código 9 en la pregunta 55: 117 688 mujeres). Este control reduce la importancia de la no inclusión de todos los nacidos en 2006 (cuadro A11).

Cuadro A10
Tasas específicas y global de fecundidad, según fuente de estimación
Nacidos en 2005 y 2006

Grupos de edad (años)	2005			2006		
	Edad 0	Fecha de nacimiento	Último HNV	Edad 0	Fecha de nacimiento	Último HNV
15-19	0,06	0,06	0,06	0,08	0,08	0,09
20-24	0,10	0,09	0,10	0,14	0,13	0,15
25-29	0,09	0,09	0,09	0,12	0,11	0,12
30-34	0,07	0,07	0,07	0,09	0,08	0,09
35-39	0,05	0,04	0,04	0,07	0,06	0,05
40-44	0,03	0,02	0,01	0,06	0,04	0,02
45-49	0,02	0,02	0,00	0,05	0,03	0,00
Sumatoria	0,42	0,39	0,38	0,61	0,53	0,52
TGF	2,11	1,97	1,89	3,03	2,67	2,61

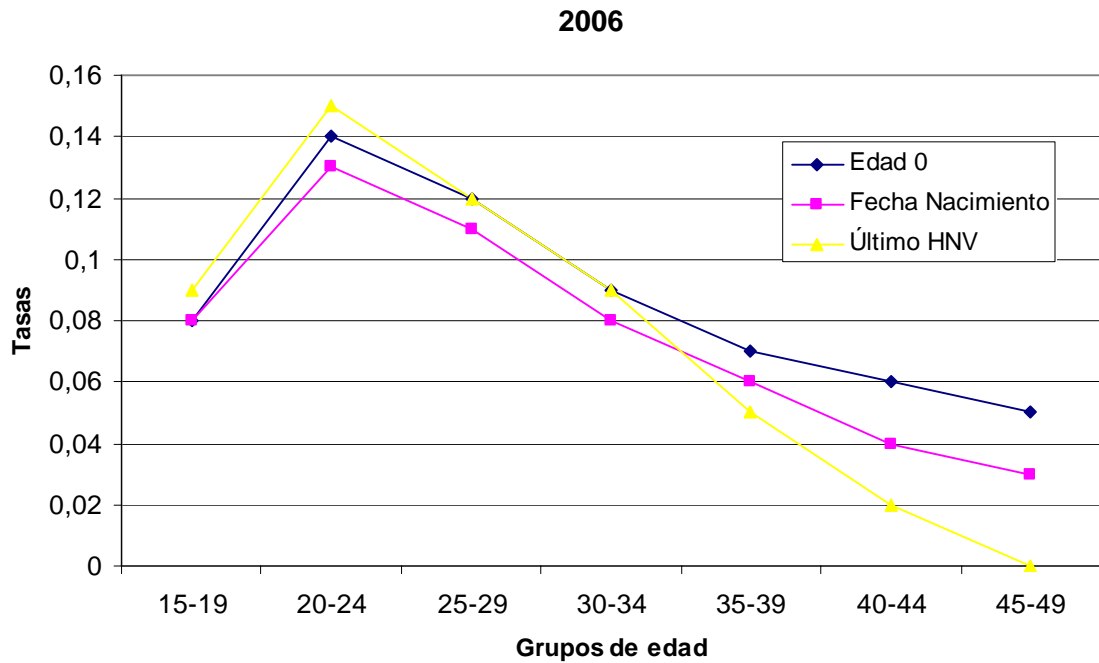
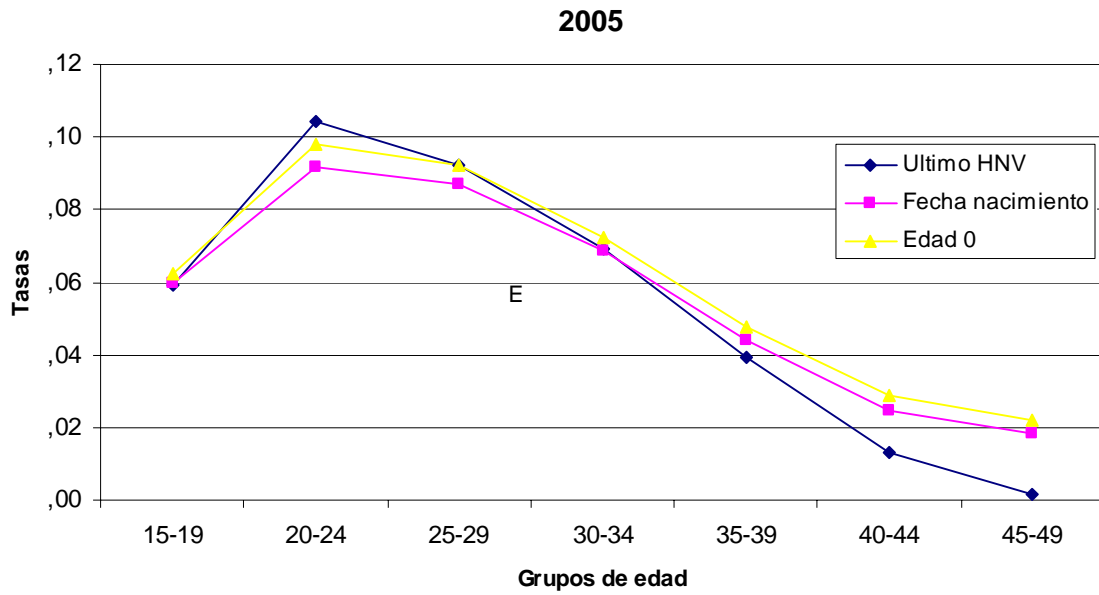
Fuente: DANE. Censo General 2005

Cuadro A11
Condición de paridez
Censo 2005

Concepto	Mujeres	
	Frecuencia	Porcentaje
Con HNV	1 863 632	66,1
Sin HNV	838 047	29,7
Ns/Nr	117 688	4,2
Total	2 819 367	100,0

Fuente: DANE. Censo General 2005

Gráfico A1
Tasas específicas de fecundidad, según fuente de estimación
MEF censadas en 2005 y 2006



Fuente: DANE. Censo General 2005

6. Balance

- Las metodologías diseñadas y utilizadas en censos y encuestas por muestreo, para ajustar los indicadores de fecundidad, contribuirían a reducir los diferenciales observados.
- Aunque debería profundizarse en el análisis, afinando la metodología de estimación y evaluando los impactos a nivel departamental y municipal, puede señalarse que el problema, al que se hizo mención en el documento, no es de la magnitud indicada y que la información puede utilizarse efectuando los ajustes recomendados para esta información.
- Se recomienda profundizar en el análisis tomando en consideración desagregaciones a nivel departamental y municipal.

Anexo B. Respuesta institucional al problema detectado en las variables educativas y en las relacionadas con la actividad de las personas en la semana anterior al censo

Variables educativas

Como resultado de la revisión de las tasas de alfabetismo entre la base cruda y la depurada, se encontró, en algunos departamentos, que la tasa de analfabetismo crece con la información depurada, lo cual no es coherente. Respecto a este caso, se encuentra que dicho incremento ocurre en aquellos departamentos con un porcentaje de blanco superior al 2%. Se realizó la revisión de la información contenida en la base cruda y se encontró lo siguiente:

- Estos casos corresponden a los municipios del grupo tres, en los cuales la recolección se realizó en formulario en papel (no para escáner).
- Su grabación se realizó en DMC sin ningún tipo de validación.

Fuente de error

Durante el operativo y grabación de la información:

- Iniciar operativo de recolección sin haber finalizado el alistamiento. No se contaba con todos los DMC requeridos para la recolección de la información.
- Utilizar formularios en papel para iniciar la recolección de la información.
- No hubo la capacitación ni supervisión suficiente en esta etapa de recolección de la información; por ejemplo, la información sobre la pertenencia étnica.
- Determinar que la información en papel debía transcribirse a los DMC y suprimir los controles para la validación de los datos incorporados. Lo anterior, determinó que la variable nivel quedara en blanco, dado que en el formulario no existe marcación para esta variable y en la grabación no se implementó ningún tipo de supervisión, sumado al hecho de no tener normas de validación de rangos para el DMC.
- Ausencia de control de calidad a la información transcrita a los DMC.

En la depuración de la información:

- Las normas de imputación son las recomendadas; no obstante, la lógica del software desarrollado no corresponde con ellas.
- No hubo el suficiente control de calidad sobre los resultados del proceso de imputación de la información educativa.

Correctivos

En consideración al cambio de la tasa con la información depurada, se realizó el siguiente proceso:

- Se revisó el documento de normas, las cuales no generan el resultado final observado.
- Se revisó el código CSPRO que se implementó, encontrándose un cambio de lógica de las normas para el manejo de blancos, por lo siguiente:
 - Se define una función que estructura una variable que maneja de forma conjunta nivel-año, cuyo rango es de 1 a 42 para nivel_año válido, 43 para nivel ninguno, 98 para último año inválido, y 99 para nivel-año inválido. Este proceso es necesario para facilitar la implementación de las normas definidas; sin embargo, dicha función no hace una evaluación de la información del nivel y/o año y asigna directamente en este caso, blanco a esta variable auxiliar.
 - En una segunda fase, al venir en blanco la variable auxiliar, no entra en la rutina que permite, mediante una imputación en caliente, recomponer la variable nivel_año según la edad y le asigna código 43 (ningún nivel), lo cual determina que en las fases siguientes del proceso se asigne en la pregunta 41 (¿sabe leer y escribir?) código 2 (No) y el último año quedara sin información, lo cual determina una reducción de la escolaridad.

Lo anterior determina que la tasa de analfabetismo se incremente fuertemente en aquellos entes territoriales donde la proporción de blancos sea alta, no por problemas del censo, sino por procesos de grabación y lógica del software de depuración. La falta de tiempo también determina que no se pueda detectar en la fase de análisis, al no poderse mirar este indicador a nivel departamental y municipal.

Proceso y pruebas realizadas

Al identificar el problema, se realizó el siguiente proceso:

1. Se actualiza la base depurada con los datos de la base cruda correspondientes a las variables originales de las preguntas 41 (sabe leer y escribir; libros) y 44 (nivel y años alcanzados). Para esto se realiza un aplicativo en CSPRO utilizando el número de registro que es único y se mantiene en todo el proceso de depuración.
2. Se realiza el ajuste a la función que crea la variable auxiliar con el fin de garantizar el funcionamiento adecuado de la imputación en caliente del nivel y año, en aquellos casos que sean inconsistentes. Esto permite garantizar la imputación adecuada de la pregunta 41 (alfabetismo). Igualmente, se realizan ajustes en la programación para garantizar la lógica con las normas definidas.
3. En la parte de alfabetismo, se corrige el algoritmo que estaba blanqueando de forma genérica la variable *libros leídos*, al considerar que ésta es ampliada. Este ajuste consiste en controlar primero por formulario ampliado, si esta variable trae información, y si la pregunta 41 está en blanco, realizar la correspondiente imputación (si libros es 1 o más y la 41 está en blanco se asigna “alfabeto” (1); en caso contrario,

se tiene en cuenta la información de nivel y años aprobados. En el caso del cuestionario básico, la variable libros es “no aplica”.

- Se realizaron una serie de pruebas con la Localidad 12 –Barrios Unidos– de Bogotá, D.C. con municipios con alta presencia de población indígena y elevadas tasas de analfabetismo y con municipios no indígenas y altas tasas en el indicador.

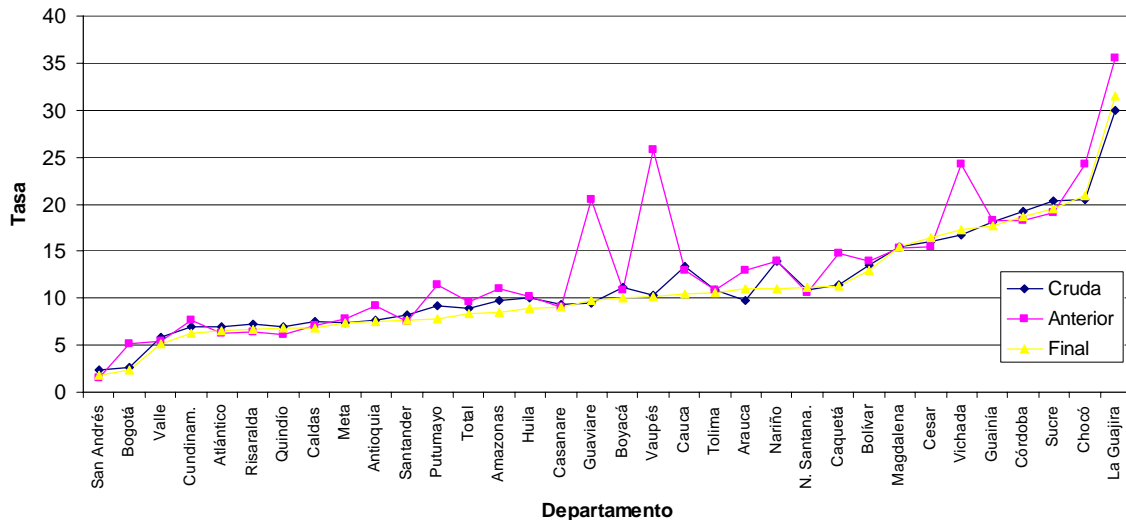
Resultados

Analfabetismo: las correcciones cambian el panorama del analfabetismo obtenido con la base que contenía los errores y es coherente con el obtenido en la base cruda. Para la población de 15 años y más, del nivel nacional, la base cruda muestra un valor del indicador del 8,9%; el tratamiento inadecuado de los valores en el nivel educativo y los años aprobados lo incrementa al 9,6%; una vez se corrige la sintaxis, nuevamente el analfabetismo toma un valor inferior al observado en la base cruda: el 8,3% (gráficos B1 y B2).

Con excepción de los departamentos de los antiguos territorios nacionales, excepto el Archipiélago de San Andrés, Providencia y Santa Catalina y algunos con alta presencia indígena, Cauca y Nariño, los diferenciales en el analfabetismo obtenido con la base cruda y la corregida, desaparecen. De igual forma, en su comparación con la base anterior (con errores), se observan las mejoras que experimentó la mayor parte de ellos.

Gráfico B1

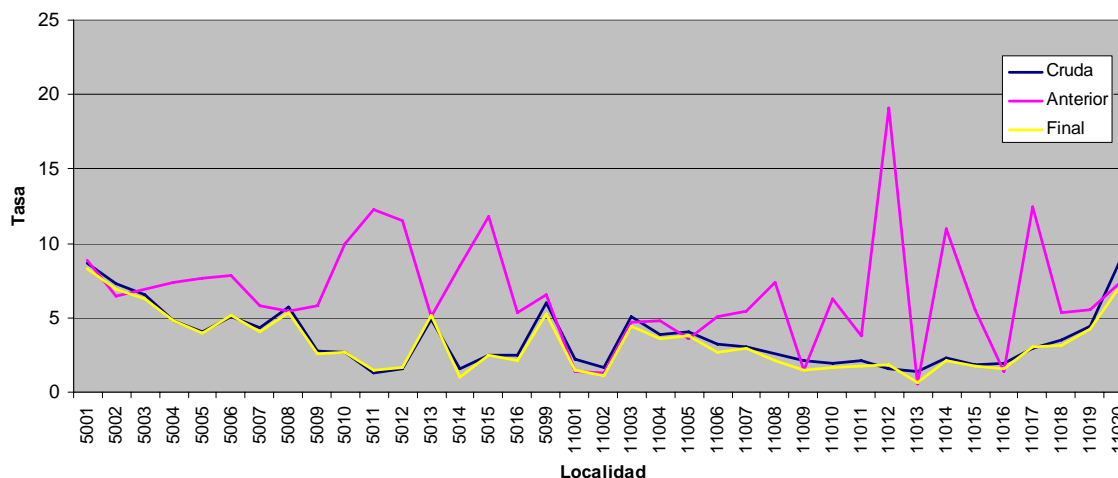
Tasas de analfabetismo para población de 15 años y más, por base, según departamento Censo 2005



Fuente: DANE –Censo General 2005

Por comuna de Medellín y localidad de Bogotá, D.C., los efectos de los correctivos aplicados son más claros, las mejoras en la mayor parte de ellas son significativas y las nuevas tasas de analfabetismo son similares a las observadas en la base cruda (gráfico B2)

Gráfico B2
Tasas de analfabetismo para Bogotá y Medellín, por localidad
Censo 2005



Fuente: DANE –Censo General 2005

En conclusión, puede afirmarse que el ejercicio de corrección realizado fue el apropiado y que la información sobre el alfabetismo de la población es de calidad y puede utilizarse, sin ningún tipo de correctivos, para los efectos que se requiera (cuadro B1).

Cuadro B1
Tasas de analfabetismo para población de 15 años y más,
por base, según departamento
Censo 2005

Departamento	Cruda	Anterior	Final
Total	8,9	9,6	8,3
Antioquia	7,7	9,2	7,3
Atlántico	6,9	6,3	6,5
Bogotá, D.C.	2,7	5,2	2,2
Bolívar	13,5	13,9	12,7
Boyacá	11,1	10,9	10,0
Caldas	7,5	7,1	6,8
Caquetá	11,4	14,8	10,8
Cauca	13,4	12,9	10,2
Cesar	16,0	15,5	16,3
Córdoba	19,2	18,2	18,6
Cundinamarca	7,0	7,7	6,1
Chocó	20,5	24,2	20,0
Huila	10,0	10,2	8,8
La Guajira	29,9	35,6	29,0
Magdalena	15,5	15,3	15,2
Meta	7,4	7,8	7,2
Nariño	13,9	13,9	10,8
Norte de Santander	10,9	10,6	11,0

Cuadro B1 (conclusión)
Tasas de analfabetismo para población de 15 años y más,
por base, según departamento
Censo 2005

Departamento	Cruda	Anterior	Final
Quindío	7,0	6,1	6,8
Risaralda	7,3	6,4	6,7
Santander	8,2	7,5	7,6
Sucre	20,3	19,1	19,4
Tolima	10,9	10,9	10,5
Valle del Cauca	5,8	5,4	5,1
Arauca	9,7	12,9	10,6
Casanare	9,3	9,0	9,0
Putumayo	9,2	11,4	7,6
Archipiélago de San Andrés, Providencia y Santa Catalina	2,4	1,5	1,9
Amazonas	9,8	11,0	8,2
Guainía	18,1	18,2	17,4
Guaviare	9,5	20,5	8,5
Vaupés	10,3	25,8	8,3
Vichada	16,7	24,2	15,6

Fuente: DANE –Censo General 2005

Escolaridad: para hacer comparable los resultados entre la base cruda y restantes versiones, a las personas sin nivel educativo (código 13) y años de educación en blanco en la primera base, se les considera con escolaridad 0. El supuesto planteado en el texto era que podría existir conexión entre los errores observados en el analfabetismo y los niveles de escolaridad; esto no es así, con la consideración hecha, los cambios observados entre la primera y subsiguientes base no son de significancia, pero confieren seguridad sobre la calidad de la información educativa (cuadro B2).

Cuadro B2
Escolaridad para población de 15 años o más, por base,
según departamento

Departamento	Cruda	Anterior	Final
Total	7,7	7,6	7,7
Antioquia	7,6	7,6	7,6
Atlántico	8,7	8,7	8,7
Bogotá, D.C.	9,8	9,7	9,8
Bolívar	7,4	7,3	7,4
Boyacá	6,7	6,7	6,7
Caldas	7,3	7,2	7,3
Caquetá	6,1	6,1	6,1
Cauca	6,1	6,1	6,1
Cesar	6,7	6,7	6,7
Córdoba	6,3	6,3	6,3
Cundinamarca	7,2	7,2	7,2
Chocó	6,5	6,4	6,4

Cuadro B2 (conclusión)
Escolaridad para población de 15 años o más, por base,
según departamento

Departamento	Cruda	Anterior	Final
Huila	6,5	6,5	6,5
La Guajira	5,7	5,8	5,8
Magdalena	6,8	6,8	6,8
Meta	7,5	7,4	7,5
Nariño	6,1	6,1	6,1
Norte de Santander	6,8	6,8	6,8
Quindío	7,5	7,5	7,5
Risaralda	7,3	7,3	7,3
Santander	7,4	7,4	7,4
Sucre	6,3	6,3	6,3
Tolima	6,8	6,8	6,8
Valle del Cauca	8,1	8,1	8,1
Arauca	6,9	6,9	6,9
Casanare	6,6	6,5	6,6
Putumayo	6,2	6,1	6,2
Archipiélago de San Andrés, Providencia y Santa Catalina	9,2	9,2	9,2
Amazonas	7,2	7,2	7,2
Guainía	5,7	5,6	5,6
Guaviare	6,2	6,2	6,2
Vaupés	7,2	7,2	7,2
Vichada	5,3	5,3	5,4

Fuente: DANE –Censo General 2005

Actividad la semana anterior al Censo 2005

El error en la actividad realizada por las personas, en particular de las menores de edad, se generó cuando en algunos de los hogares cuya información se obtuvo mediante formularios en papel, además de la condición de estudiante no trabajador (código 5, pregunta 47) se le consignó, igualmente, la condición de trabajador familiar sin remuneración (código 5, pregunta 49), y sin atender el salto señalado en el formulario (pregunta 7, códigos 4 a 9: pase a pregunta 51).

Las normas de imputación son claras y consistentes; no obstante, la lógica del algoritmo de imputación desarrollada, para interpretarla, contribuyó a incrementar el problema: a los estudiantes que, a la vez, se declararon como trabajadores familiares sin remuneración se les quitó la condición de actividad señalada (código 5, en pregunta 47) y se les convirtió en trabajadores (código 1, pregunta 47). La revisión llevada a cabo implicó:

- La recuperación de los registros originales contenidos en la base cruda.
- Aplicación de las normas de imputación, específicamente para el problema señalado, doble respuestas, así:
 - Si la pregunta 47 = no aplica y pregunta 49 es valida, se asigna código 1 (trabajó) en pregunta 47.

- Si en la pregunta 47 tiene valores entre 4 y 9, no se considera información de preguntas 49 y 50 (se blanquea).
- Conformación de la nueva base con la información de los registros objeto de imputación.

Una vez corregidos los errores, la información obtenida con la base cruda y la final es coherente; los valores absoluto correspondientes a los aspirantes a obtener su primer trabajo y la población en condición de inactividad económica, son prácticamente iguales entre la base original y la corregida.

Las modificaciones incorporadas a la lógica del algoritmo de imputación aseguran que la información sobre la actividad de las personas en la semana anterior al Censo 2005 y por extensión, la relacionada con la categoría ocupacional, que también experimenta modificaciones, pueda utilizarse en forma confiable (cuadro B3).

Cuadro B3
Actividad en la semana anterior al Censo, por grupos de edad, según base
Censo 2005

Actividad semana anterior	5 - 11 años		12 - 17 años		18 y más años	
	Recuento	Porcentaje	Recuento	Porcentaje	Recuento	Porcentaje
Base cruda						
Total	5 903 808	100,0	4 885 089	100,0	25 203 039	100,0
0	14 831	0,3	7 856	0,2	23 463	0,1
Trabajó	23 090	0,4	354 684	7,3	12 650 794	50,2
No trabajó, tenía trabajo	12 298	0,2	19 594	0,4	305 433	1,2
Busco trabajo, había trabajado	1 028	0,0	26 260	0,5	872 407	3,5
Buscó trabajo primera vez	2 664	0,0	25 540	0,5	179 304	0,7
Estudió, no trabajó, no buscó	4 492 070	76,1	3 281 217	67,2	1 211 015	4,8
Oficios del hogar	147 226	2,5	419 214	8,6	6 528 687	25,9
Incapacitado	22 689	0,4	21 768	0,4	446 572	1,8
Jubilado, rentista	2 863	0,0	2 150	0,0	739 610	2,9
Otra situación	1 185 049	20,1	726 806	14,9	2 245 754	8,9
Base anterior						
Total	6 052 950	100,0	4 994 283	100,0	26 021 637	100,0
0	164 389	2,7	115 404	2,3	814 036	3,1
Trabajó	95 995	1,6	401 838	8,0	12 691 353	48,8
No trabajó, tenía trabajo	12 323	0,2	19 636	0,4	305 722	1,2
Busco trabajo, había trabajado	1 025	0,0	26 224	0,5	872 266	3,4
Buscó trabajo primera vez	2 595	0,0	25 014	0,5	176 883	0,7
Estudió, no trabajó, no buscó	4 421 652	73,0	3 239 324	64,9	1 200 833	4,6
Oficios del hogar	147 148	2,4	418 221	8,4	6 525 065	25,1
Incapacitado	22 689	0,4	21 764	0,4	446 853	1,7
Jubilado, rentista	0	0,0	0	0,0	739 571	2,8
Otra situación	1 185 134	19,6	726 858	14,6	2 249 055	8,6

Cuadro B3 (conclusión)

Actividad en la semana anterior al Censo, por grupos de edad, según base Censo 2005

Actividad semana anterior	5 - 11 años		12 - 17 años		18 y más años	
	Recuento	Porcentaje	Recuento	Porcentaje	Recuento	Porcentaje
	Base final					
Total	6 052 950	100,0	4 994 283	100,0	26 021 637	100,0
0	164 389	2,7	115 404	2,3	814 035	3,1
Trabajó	25 440	0,4	359 164	7,2	12 675 904	48,7
No trabajó, tenía trabajo	12 323	0,2	19 636	0,4	305 722	1,2
Busco trabajo, había trabajado	1 025	0,0	26 224	0,5	872 266	3,4
Buscó trabajo primera vez	2 690	0,0	25 546	0,5	179 285	0,7
Estudió, no trabajó, no buscó	4 492 104	74,2	3 281 434	65,7	1 212 735	4,7
Oficios del hogar	147 148	2,4	418 230	8,4	6 525 627	25,1
Incapacitado	22 689	0,4	21 765	0,4	446 901	1,7
Jubilado, rentista					739 620	2,8
Otra situación	1 185 142	19,6	726 880	14,6	2 249 542	8,6

Fuente: DANE