

ANÁLISIS DE LA SITUACIÓN DE SALUD EN COLOMBIA 2002-2007



TOMO VIII ANÁLISIS DE LA SITUACIÓN DE SALUD DE COLOMBIA 2002-2007: GESTIÓN DE DATOS.



Libertad y Orden

Ministerio de la Protección Social
República de Colombia

MINISTERIO DE LA PROTECCIÓN SOCIAL

UNIVERSIDAD DE ANTIOQUIA

FACULTAD NACIONAL DE SALUD PÚBLICA

ANÁLISIS DE LA SITUACIÓN

DE SALUD EN COLOMBIA, 2002-2007

Bogotá, Diciembre de 2010

Título de la obra: ANÁLISIS DE LA SITUACIÓN DE SALUD EN COLOMBIA, 2002-2007

ISBN obra completa: 978-958-8717-04-3

- Tomo I: Características sociodemográficas de la población colombiana. ISBN: 978-958-8717-05-0
- Tomo II: Relaciones familiares y salud sexual y reproductiva. ISBN: 978-958-8717-06-7
- Tomo III: Morbilidad y mortalidad de la población colombiana. ISBN: 978-958-8717-07-4
- Tomo IV: Situación de salud de la infancia. ISBN: 978-958-8717-08-1
- Tomo V: Prestación de servicios de salud. ISBN: 978-958-8717-09-8
- Tomo VI: Análisis de desigualdades e inequidades en salud. ISBN: 978-958-8717-10-4
- Tomo VII: Situación de salud de Colombia 2002-2007: Resumen. ISBN: 978-958-8717-11-1
- Tomo VIII: Análisis de la situación de salud de Colombia 2002-2007: Gestión de Datos.
ISBN: 978-958-8717-12-8

© Ministerio de la Protección Social y Universidad de Antioquia, Facultad Nacional de Salud Pública.
© Los Autores.

Año de la publicación: 2010

Lugar y fecha de la edición: Imprenta Nacional de Colombia. Bogotá, diciembre de 2011

Queda prohibida la reproducción parcial o total de este documento, por cualquier medio escrito o visual, sin previa autorización del Ministerio de la Protección Social, quien autorizará la reproducción total o parcial siempre y cuando se mantenga la integridad del texto, se cite la fuente y no tenga fines comerciales.

Interventoría: Dirección General de Salud Pública – Ministerio de la Protección Social

ANÁLISIS DE LA SITUACIÓN DE SALUD EN COLOMBIA, 2002-2007



Libertad y Orden

Ministerio de la Protección Social
República de Colombia

MAURICIO SANTA MARÍA SALAMANCA

Ministro de la Protección Social

BEATRIZ LONDOÑO SOTO

Viceministra de Salud y Bienestar

GERARDO LUBÍN BURGOS BERNAL

Secretario General

LENIS ENRIQUE URQUIJO VELÁSQUEZ

Director General de Salud Pública

VÍCTOR HUGO ÁLVAREZ CASTAÑO

Coordinador de Vigilancia en Salud Pública

ANÁLISIS DE LA SITUACIÓN DE SALUD EN COLOMBIA, 2002-2007



**UNIVERSIDAD
DE ANTIOQUIA**
1803

Facultad Nacional de Salud Pública
"Héctor Abad Gómez"

ALBERTO URIBE CORREA

Rector

ÁLVARO CARDONA SALDARRIAGA

Decano Facultad Nacional de Salud Pública U. de A.

LUIS ELISEO VELÁSQUEZ LONDOÑO

Vicedecano

CARLOS ROJAS ARBELÁEZ

Director Centro de Investigaciones

MARÍA PATRICIA ARBELÁEZ MONTOYA, PhD.

Directora General

CARLOS ALBERTO TANGARIFE VILLA

Coordinación de Bases de Datos

GLORIA MOLINA MARÍN

JUAN LUIS LONDOÑO FERNÁNDEZ

MARÍA PATRICIA ARBELÁEZ MONTOYA

FÉLIX LEÓN MARTÍNEZ MARTÍN

DIANA MARCELA MARÍN PINEDA

LUCELLY LÓPEZ LÓPEZ

YAMILE ANDREA BAENA HENAO

Edición Académica

LETICIA BERNAL VILLEGAS

Revisión de estilo

ALINA GIRALDO YEPES

Diagramación

ANÁLISIS DE LA SITUACIÓN DE SALUD EN COLOMBIA, 2002-2007

Participantes en el desarrollo del Proyecto ASIS Colombia 2002-2007

MINISTERIO DE LA PROTECCIÓN SOCIAL

Lenis Enrique Urquijo, Director General de Salud Pública, Comité Técnico ASIS
Víctor Hugo Álvarez Castaño, Coordinador de Vigilancia en Salud Pública - Coordinador de ASIS

María Alexandra Durán, Profesional Especializada Comité Técnico ASIS
Patricia Cifuentes, Profesional Especializada Comité Técnico ASIS

FACULTAD NACIONAL DE SALUD PÚBLICA, UNIVERSIDAD DE ANTIOQUIA

María Patricia Arbeláez Montoya, Directora del Proyecto
Germán González Echeverri, Decano 2006-2009
Gloria Molina Marín, Jefe Centro de Investigación 2007-2009
Carlos Eduardo Castro Hoyos, Jefe Centro de Extensión 2006-2009
Carlos Alberto Tangarife Villa, Coordinador Gestión de Datos
Félix Martínez, Investigador Social FEDESALUD

SISTEMAS ESPECIALIZADOS DE INFORMACIÓN SEI S.A.

Yimer Yezid Botiva Gutiérrez, Coordinador Integración ENS 2007- ENDS 2005
Hanner Sánchez García, Consultor en Estadística

RECONOCIMIENTOS

Ministerio de la Protección Social

Diego Palacio Betancourt, Ministro de la Protección Social 2003-2010
Gilberto Álvarez Uribe, Director General de Salud Pública 2006-2010

Departamento Administrativo de Ciencia, Tecnología e Innovación - COLCIENCIAS

Juan Francisco Miranda, Director General Colciencias 2006-2010
Clara Inés Vargas, Subdirectora de Investigaciones en Salud, Comité Técnico ENS 2007

CENDEX – Pontificia Universidad Javeriana

Jesús Rodríguez García, Director del Proyecto ENS 2007
Fernando Ruiz Gómez, Director de CENDEX, Coordinador de Análisis del Proyecto ENS 2007
Rolando Enrique Peñaloza Quintero, Coordinador Técnico del Proyecto ENS 2007

Observatorio de Salud Pública de Santander

Rafael Ortiz, Director OSPS, Coordinador Recolección RIPS Proyecto ENS 2007

Sistemas Especializados de Información SEI S.A.

José Rafael Arenas Mantilla, Gerente de Operaciones y Logística de encuestas ENS 2007
Luis Carlos Gómez Serrano, Consultor en Muestreo y Estimaciones del Proyecto ENS 2007
Marcela Valencia Suescum, Asesora Diseño y Ajuste de protocolo y formularios ENS 2007

Universidad Nacional de Colombia

Rosa Helena Flórez Mosquera, Directora Salud Pública, Facultad de Medicina UNAL 2003-2005
Carlos Arturo Sarmiento Limas, Coordinador UNAL Protocolo Proyecto ASIS 2003-2005

ANÁLISIS DE LA SITUACIÓN DE SALUD EN COLOMBIA, 2002-2007

Investigadores y miembros de los equipos técnicos participantes

Profesores Universidad de Antioquia

Alexandra Restrepo H.	Andrés Agudelo S.	Julián Vargas J.
Blanca Miriam Chaves G.	Andrés Camilo Jiménez H.	Laura Lucía Cano G.
Blanca Susana Ramírez P.	Aura Ruiz G.	Leidy Cristina Jaramillo M.
Carlos Mario Quiroz P.	Beatriz E. Arroyave P.	Lina María Martínez F.
Carlos Rojas A.	Bibiana Andrea Castro M.	Lorena Andrea Gómez G.
Eliana Martínez H.	Cindy Maryei Crespo S.	Lorena Sánchez D.
Elkin Martínez L.	Carlos Mario Arcila E.	Lucas Restrepo M.
Erika María Montoya V.	Catalina Lopera C.	Lucelly López L.
Fabio León Rodríguez O.	Catalina María Arango A.	Luis Fernando Duque R.
Fernando Peñaranda C.	Clara Cecilia Zuluaga M.	Manuel Antonio Céspedes S.
Germán González E.	Cristina Mejía M.	Marda Ucaris Zuluaga A.
Gloria Escobar P.	Dalila Patricia García R.	Margarita Inés Quiroz A.
Gloria Molina M.	Diana Carolina Ocampo R.	Margarita María Pérez O.
Héctor Byron Agudelo G.	Diana Catalina Mesa L.	Margarita Rosa Díaz B.
Humberto Ramirez G.	Diana Marcela Marín P.	María Isabel Jaramillo A.
Juan Fernando Saldarriaga F.	Diego León Bastidas L.	María Osley Garzón D.
Leonardo Uribe A.	Diego Alejandro Jiménez D.	Martha Claudia González P.
Lisardo Osorio Q.	Diego Mauricio Montoya R.	Martha Lucía Olarte L.
Luis Eliseo Velásquez L.	Dillanelid Cañas B.	Miguel Ángel Mafla P.
Luz Nelly Zapata V.	Doralba Aristizábal J.	Mónica L. Vargas B.
María Isabel Lalinde A.	Doris Cardona A.	Nilton E. Montoya G.
María Luisa Montoya R.	Elsa María Vásquez T.	Nydia S. Caicedo M.
María Patricia Arbeláez M.	Esteban Páez Z.	Omar Ramírez R.
Miriam Bastidas A.	Félix Martínez M.	Paula Andrea Díaz V.
Mónica Lucía Soto V.	Fredy Suárez P.	Paula Andrea Uribe R.
Nora Adriana Montealegre H.	Gilma Stella Vargas P.	Piedad Estrada A.
Rubén Darío Gómez A.	Hernando Guerrero M.	Piedad Roldán J.
Samuel Andrés Arias V.	Ilduara Peña G.	Rosmary Aristizábal G.
Sergio Cristancho Marulanda	Iván Darío Flórez G.	Rubén Darío Manrique H.
Yolanda Lucía López A.	Jairo Ramírez G.	Sandra Milena Agudelo L.
Yomaira Higuita H.	Janeth Pérez G.	Verónica Romero A.
	Javier Mauricio Sierra A.	Walter Alonso Zuluaga R.
	Jenifer C. Lopera C.	Yadira Estrada L.
	Jhojandelli Congote E.	Yamile Andrea Baena H.
	Jhovanny Andrés Cañas P.	Yaneth Pérez G.
	Jorge A. Pulgarín T.	Yomaira Viñas S.
	Jorge Emilio Salazar F.	Yuliana Castaño L.

Otros participantes

Adriana Milena Espinosa L.
Adriana Rocío Roa P.
Alejandra Agudelo M.
Ana Daniela González E.

PRESENTACIÓN

Con el propósito de satisfacer los requerimientos de información sobre la situación de salud de los colombianos, el Ministerio de la Protección Social, en asociación con Colciencias, se dio a la tarea de diseñar y ejecutar un estudio de cobertura nacional que brindara información sobre la morbilidad sentida e institucional, los factores de riesgo y la discapacidad que caracterizan la oferta y demanda de servicios de salud.

Para el efecto se encargó a un equipo de expertos de la Universidad Nacional el diseño del Modelo de Análisis de la Situación de Salud, del que se obtuvo el protocolo de la Encuesta Nacional de Salud que la Unión Temporal SEI S.A. y el Centro de Proyectos para el Desarrollo CENDEX de la Pontificia Universidad Javeriana aplicó durante el 2007 a lo largo de la geografía nacional, de lo cual se ofreció una publicación con los resultados más relevantes.

Con el fin de lograr el propósito de realizar un análisis en profundidad de la situación de salud del país correspondiente al período 2002-2007, el Ministerio de la Protección Social estableció el contrato interadministrativo de cooperación con la Facultad Nacional de Salud Pública de la Universidad de Antioquia, número 519-2008. Dicho análisis se realizó con base en la información de la ENS-2007, la ENDS-2005, las estadísticas vitales, los RIPS, el censo de población de 2005, la encuesta de Calidad de Vida, el Estudio Nacional de Salud Mental Colombia 2003 y la Encuesta Nacional de Situación Nutricional Colombia ENSIN 2005, entre otros. Para ello se realizó un control de calidad a los datos y se organizaron bases de datos de tal modo que permitieran un análisis integral de los diferentes aspectos de la salud.

Esta información, así como las bases de datos que quedan a disposición de la comunidad científica para realizar análisis adicionales, permitirá conocer el perfil de salud-enfermedad de la población colombiana, identificar los problemas de salud más relevantes del país, así como los factores asociados. Este análisis también facilitará el reconocimiento de las desigualdades que se presentan por razones de sexo, edad, raza, distribución geográfica, acceso, uso y calidad de los servicios de salud, lo cual permitirá orientar más efectivamente la política de salud en los diferentes niveles de gobierno, así como las acciones de las entidades públicas y privadas del sector, con el fin de reducir las inequidades existentes y garantizar los derechos en salud de los colombianos.

MAURICIO SANTA MARÍA SALAMANCA

Ministro de la Protección Social

INTRODUCCIÓN

El Análisis de la Situación de Salud ASIS¹ fue implementado por el Ministerio de la Protección Social como prioridad en el área de la salud pública en los Planes Estratégicos de Salud, formulados desde el año 2002, con el propósito de establecer y mantener un conocimiento actualizado e integral de la situación de salud de los colombianos.

Este enfoque metodológico propuesto por la OPS/OMS en 1999, como el mejor abordaje para comprender la situación de salud de una población, ha sido incorporado en Colombia con iniciativas que van desde la actualización y publicación periódica, desde el año 2000, de la serie de Datos Básicos de Salud, hasta el planteamiento de un sistema de encuestas periódicas dentro del nuevo Sistema Integral de Información de la Protección Social SISPRO, lo cual permite desarrollar un modelo de análisis basado en el conocimiento, caracterización y evaluación de la salud de la población colombiana y de los subgrupos de población que la constituyen, en función de tres dimensiones fundamentales: los daños a la salud, la funcionalidad y la discapacidad de la población, y el comportamiento y la distribución de los factores determinantes del perfil de salud en el país.

La adopción de la propuesta de ASIS como metodología de actualización permanente, sistemática y completa de la información en salud a nivel nacional ha contado con estudios integrales sobre las condiciones de salud de la población colombiana. Entre estos estudios se encuentran el Estudio de Recursos Humanos para la Salud y Educación Médica en Colombia e Investigación Nacional de Morbilidad, realizados por el Ministerio de Salud y ASCOFAME en 1965-1966, y el Estudio Nacional de Salud realizado por el Instituto Nacional de Salud, el Ministerio de Salud y

¹ El Análisis de la Situación de Salud –ASIS– es un proceso analítico-sintético que abarca diversos tipos de fuentes de información y estrategias de análisis, que permite caracterizar, medir y explicar el perfil de salud-enfermedad de una población, incluyendo los daños y problemas de salud, así como sus determinantes, sean éstos competencia del sector salud o de otros sectores. Los ASIS se basan en el estudio de la interacción de las condiciones de vida y el nivel existente de los procesos de salud de un país u otra unidad geográfico-política, facilitando la identificación de necesidades y prioridades en salud, así como la identificación de intervenciones y programas apropiados y la evaluación de su impacto en salud (Definición tomada del Boletín Epidemiológico de la Organización Panamericana de la Salud, volumen 20 número 3, septiembre de 1999).

ASCOFAME en 1977-80, seguidos en la década de los 90 por otras investigaciones como los Estudios Nacionales de Factores de Riesgo de Enfermedades Crónicas (ENFREC I en 1993 y ENFREC II en 1998), el III Estudio Nacional de Salud Bucal (ENSAB III en 1998), el Estudio Nacional de Salud Mental en 2003 y la serie de Encuestas Nacionales de Demografía y Salud (ENDS 1990, 1995, 2000 y 2005), entre otros.

Como parte de este proceso, el Ministerio de la Protección Social propuso el Análisis de la Situación de Salud en Colombia para el período quinquenal 2002-2007, para lo cual se planteó la construcción de un modelo y plan de análisis, cuyo diseño estuvo a cargo de la Universidad Nacional de Colombia.

El modelo del ASIS, definido para dar soporte conceptual al análisis de salud planteado en la primera ilustración,² supone la aproximación sucesiva a la realidad a través de cuatro niveles de análisis de carácter descriptivo.

El primer nivel busca identificar y caracterizar los problemas de salud y los niveles de discapacidad de acuerdo con su magnitud, severidad y estructura demográfica, para priorizar los problemas de salud de cada subgrupo poblacional describiendo su comportamiento epidemiológico. A partir de la información contenida en este nivel se identificaron los problemas de mayor interés en salud pública que debían ser medidos, y se obtuvieron los elementos necesarios que luego sirvieron para definir los contenidos del protocolo y los formularios que fueron diseñados por la Universidad Nacional para la ENS-2007.

El segundo nivel parte de los modelos explicativos vigentes y del desarrollo del conocimiento a la fecha sobre los problemas de salud priorizados, para caracterizar los factores asociados a la incidencia de dichos problemas, las intervenciones sociales implementadas para afrontar el problema bajo análisis y el conocimiento y percepción que la población tiene sobre los problemas, los mecanismos de solución y las barreras existentes en su interacción con los servicios del sector de la salud.

El tercer nivel descriptivo de este estudio se refiere al análisis de desigualdades, que tiene relación con la existencia de diferencias en la manifestación de un problema de salud entre poblaciones de áreas geográficas diversas en función de un factor asociado a la presentación del mismo, sin que en la determinación de estas diferencias medie un criterio de justicia, como ocurre con categorías de importancia del individuo afectado (sexo, edad y factores raciales y genéticos), de la ubicación en un medio geográfico (regiones o departamentos), de concentración poblacional (urbana o rural), de factores de riesgo específicos del medio (la vivienda, el lugar de trabajo y la comunidad), de inclusión en salud (acceso y uso) y sociales (inclusión social).

Finalmente, el cuarto nivel de aproximación descriptiva que da cuenta del análisis de inequidades en materia de salud, se refiere a las disparidades sanitarias que se consideran injustas, injustificadas, evitables e innecesarias, y que sistemáticamente agobian a las poblaciones que se han vuelto vulnerables a causa de estructuras sociales fundamentales y de instituciones políticas, económicas y jurídicas que prevalecen, tales como las diferencias relacionadas con el género, el ingreso, el nivel educativo, la posición ocupacional, el estrato social y el nivel de desarrollo de las regiones.

² Tomado del documento *Encuesta Nacional de Salud 2007. Protocolo de la Encuesta. Bogotá, junio de 2006*, elaborado por el equipo de investigadores de la Universidad Nacional de Colombia.

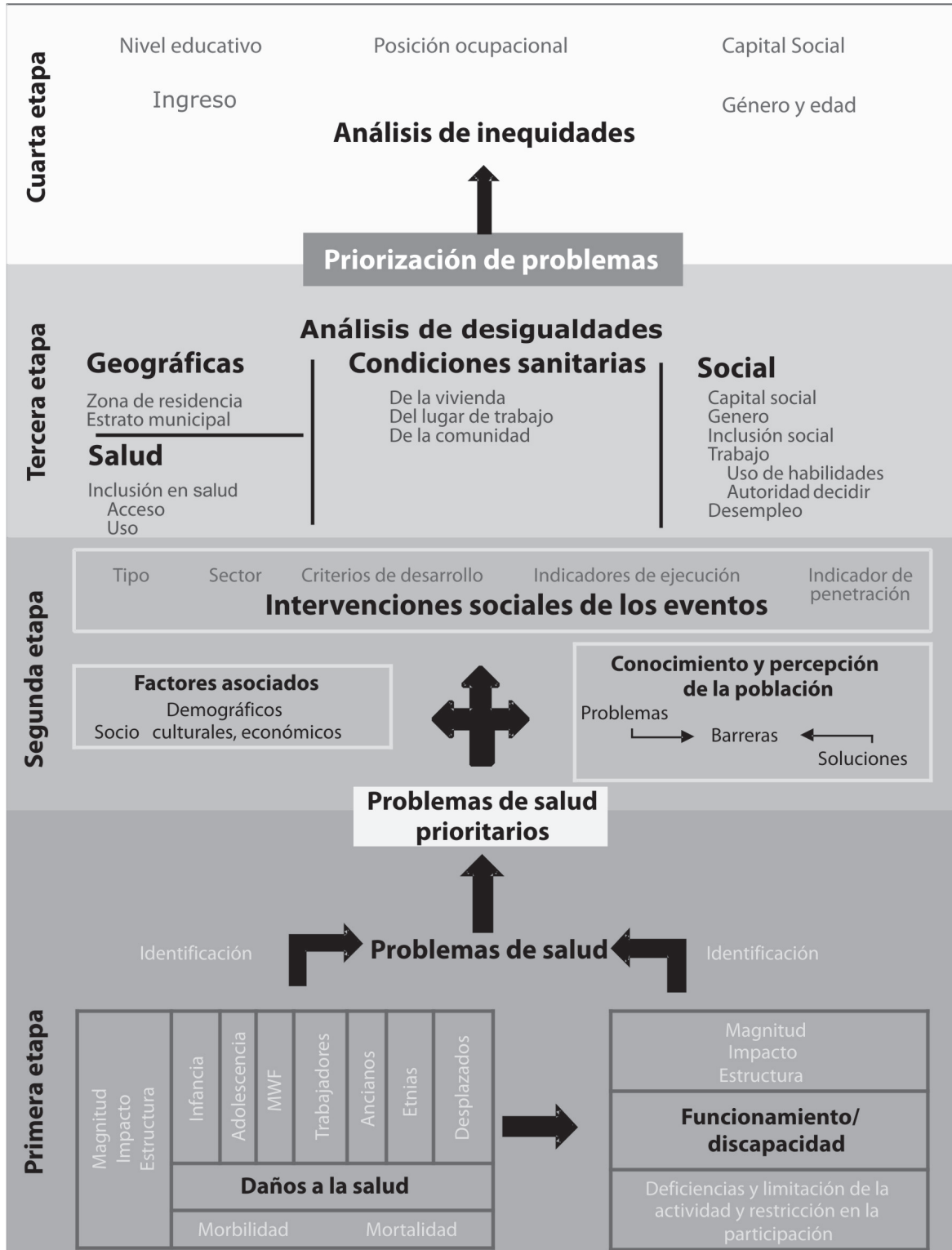


Ilustración 1. Modelo de análisis de la situación de salud

Fuente: Universidad Nacional de Colombia, 2006.

En este marco, la Encuesta Nacional de Salud ENS-2007 fue concebida como una fuente más de información para el ASIS enfocada al levantamiento de datos poblacionales sobre morbilidad sentida e institucional, funcionamiento y discapacidad; caracterización de la oferta y la demanda de servicios de salud, para complementar el conjunto de información necesaria en la caracterización y análisis de la situación de salud en Colombia.

Para el efecto, el Ministerio de la Protección Social, mediante convenio con Colciencias, contrató con la Pontificia Universidad Javeriana y la firma Sistemas Especializados de Información SEI S.A. de Bogotá la realización de la Encuesta Nacional de Salud ENS-2007, con el objetivo principal de recolectar y analizar información de los hogares, usuarios e instituciones del sistema colombiano de salud para la identificación de la situación de salud del país en los ámbitos nacional, regional y departamental.

Con el fin de lograr el propósito de realizar un análisis en profundidad de la situación de salud del país del 2002-2007 el Ministerio estableció el contrato interadministrativo de cooperación entre el Ministerio de la Protección Social y la Universidad de Antioquia-Facultad Nacional de Salud Pública número 519-2008. Dicho análisis se realizó con base en la información de la ENS-2007, la ENDS-2005, las estadísticas vitales, los RIPS, el censo de población de 2005, la Encuesta de Calidad de Vida, el Estudio Nacional de Salud Mental Colombia 2003, la Encuesta Nacional de Situación Nutricional Colombia ENSIN-2005, entre otros. Previamente al análisis, se realizó un control de calidad a la información, se organizaron las bases de datos y se generó una base de datos que permite un análisis integral de los diferentes aspectos de la salud. Esta base de datos estará alojada en la plataforma informática del MPS y, mediante el aplicativo diseñado por la FNSP dentro de los términos del convenio, quedará a disposición de diferentes usuarios para obtener tanto los meta como los micro datos de cada fuente y generar archivos que pueden ser procesados en los paquetes estadísticos convencionales para así incentivar la generación de información y la investigación útil para el país.

El análisis de la ENS-2007 incluyó diferentes unidades de observación y su nivel de desagregación estuvo sujeto a la prevalencia de los problemas de salud priorizados por grupo étnico. Los hogares y sus habitantes formaron parte de muestras probabilísticas y para los análisis se aplicaron los correspondientes factores de expansión así como el cálculo de los errores estándar. Las IPS y los usuarios de los servicios de consulta externa, de urgencias y de hospitalización, estuvieron circunscritos a los entes territoriales que hicieron parte de la ENS-2007 y por lo tanto dichas muestras no fueron representativas de sus respectivos universos. Para este análisis también se utilizaron todas las estadísticas vitales proporcionadas por el DANE y la información de RIPS del año 2007.

El análisis se presenta en algunos casos desagregado por departamentos y/o por regiones con base en la distribución que se presenta a continuación:

Región	Subregión / Departamentos
1 Atlántica	1 Guajira, Cesar, Magdalena
	2 Barranquilla y el Área Metropolitana
	3 Atlántico, San Andrés, Bolívar Norte
	4 Bolívar Sur, Sucre, Córdoba
2 Oriental	5 Santanderes
	6 Boyacá, Cundinamarca, Meta
3 Central	7 Medellín y el Área Metropolitana
	8 Antioquia sin Medellín
	9 Caldas, Risaralda, Quindío
	10 Tolima, Huila, Caquetá
4 Pacífica	11 Cali y el Área Metropolitana
	12 Valle sin Cali ni el Litoral Pacífico
	13 Cauca y Nariño sin el Litoral Pacífico
	14 Litoral Pacífico
5 Bogotá	15 Bogotá
6 Orinoquia y Amazonía	16 Amazonas, Arauca, Casanare, Guainía, Guaviare, Putumayo, Vaupés, Vichada.

El análisis en profundidad de la situación de salud se presenta en ocho tomos: el primero se ocupa de las características socio-demográficas de la población, incluyendo el análisis de las condiciones de los hogares y su entorno, así como de los aspectos principales del acceso, la cobertura y la calidad de los servicios públicos domiciliarios. El segundo se centra en el análisis de la familia, principalmente en las relaciones de pareja, la violencia intrafamiliar y algunos aspectos de la salud sexual y reproductiva tales como la planificación familiar, las infecciones de transmisión sexual y el cáncer de Cérvico-uterino. El tercer tomo presenta las principales causas de morbilidad y mortalidad de la población general, por grupos de edad y por regiones del país. Igualmente, las características de algunos eventos tales como accidentalidad vial, enfermedades crónicas como Diabetes Mellitus e Hipertensión Arterial y la obesidad. El cuarto tomo se ocupa de la situación de salud de la infancia en aspectos relacionados con el cuidado del menor, la vacunación y enfermedades tales como la Enfermedad Diarreica Aguda –EDA– y la Infección Respiratoria Aguda-IRA. El quinto analiza la prestación de servicios de salud de atención ambulatoria, obstétrica y de urgencias, así como las acciones de salud pública de los entes municipales; el sexto presenta un análisis de las desigualdades e inequidades en salud; el séptimo presenta una síntesis de los aspectos más relevantes de la situación de salud en Colombia 2002-2007. Finalmente, el tomo ocho presenta los documentos técnicos y metodológicos de la gestión de los datos.

ANÁLISIS DE LA SITUACIÓN DE SALUD EN COLOMBIA, 2002-2007

TOMO VIII

GESTIÓN DE DATOS

Investigadores y equipo técnico

CAPÍTULO 1: MARCO CONCEPTUAL

Profesores

Carlos Alberto Tangarife Villa
Héctor Byron Agudelo García
Fabio León Rodríguez Ospina
Nora Adriana Montealegre Hernández
Doris Cardona Arango

Profesionales

Diego León Bastidas Lopera
Dillanelid Cañas Betancur
Diego Mauricio Montoya Restrepo.
Rosmary Aristizabal Guzmán

Estudiantes

Jhojandrelli Congote Espinosa
Cindy Maryei Crespo Sánchez
Yamile Andrea Baena Henao
María Alejandra Agudelo Martínez
Lucas Restrepo Molina
Lorena Sánchez Diosa
Yadira Efigenia Estrada Loaiza
Yuliana Castaño Isaza
Jorge Alberto Pulgarín Tobón

CAPÍTULO 2: DISEÑO METODOLÓ- GICO

Profesores

Carlos Alberto Tangarife Villa
Héctor Byron Agudelo García
Fabio León Rodríguez Ospina
Nora Adriana Montealegre Hernández
Doris Cardona Arango

Profesionales

Diego León Bastidas Lopera
Dillanelid Cañas Betancur
Diego Mauricio Montoya Restrepo.
Rosmary Aristizabal Guzmán

Estudiantes

Jhojandrelli Congote Espinosa
Cindy Maryei Crespo Sánchez
Yamile Andrea Baena Henao
María Alejandra Agudelo Martínez
Lucas Restrepo Molina
Lorena Sánchez Diosa
Yadira Efigenia Estrada Loaiza
Yuliana Castaño Isaza
Jorge Alberto Pulgarín Tobón

CAPÍTULO 3: RESULTADOS

Profesores

Carlos Alberto Tangarife Villa
Héctor Byron Agudelo García
Fabio León Rodríguez Ospina
Nora Adriana Montealegre Hernández
Doris Cardona Arango

Profesionales

Diego León Bastidas Lopera
Dillanelid Cañas Betancur
Diego Mauricio Montoya Restrepo.
Rosmary Aristizabal Guzmán

Estudiantes

Jhojandrelli Congote Espinosa
Cindy Maryei Crespo Sánchez
Yamile Andrea Baena Henao
María Alejandra Agudelo Martínez
Lucas Restrepo Molina
Lorena Sánchez Diosa
Yadira Efigenia Estrada Loaiza
Yuliana Castaño Isaza
Jorge Alberto Pulgarín Tobón

CAPÍTULO 4: ERRORES MUESTRALES

Profesores

Diana Marcela Marín Pineda
Lucelly López López
Hugo Grisales Romero

TABLA DE CONTENIDO

INTRODUCCIÓN	25
---------------------------	----

CAPÍTULO 1: MARCO CONCEPTUAL

Sistemas de información	29
Elementos que componen un sistema de información	31
Actividades del sistema de información	32
Tipos de sistemas de información	32
Dato	34
Bases de datos	36
Sistemas de bases de datos	37
Diccionario de datos	37
Metadatos	38
Bodega de datos	42
Calidad del dato	42
Dimensiones de la calidad del dato	43
Estándares de calidad del dato	44
Medición de la calidad de los datos	49
Errores o sesgos	50
<i>Medición de los errores</i>	54

Reporte de la calidad de los datos	56
Fuentes de datos para el proyecto ASIS, Colombia 2002-2007	56
Encuestas poblacionales	58
Encuesta Nacional de Salud (ENS) 2007	58
Encuesta Nacional de Demografía y Salud (ENDS) 2005	60
Encuesta de Calidad de Vida (ECV) 2003..	61
Encuesta Nacional de la Situación Nutricional de Colombia (ENSIN) 2005	62
Estudio Nacional de Salud Mental (ENSM) 2003	63
Segundo Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas (ENFREC) 1998	64
Tercer Estudio Nacional de Salud Bucal (ENSAB) 1998	65
Gran Encuesta Integrada de Hogares (GEIH) 2007	66
Registros continuos.	67
Estadísticas vitales (EEVV)	67
Registros Individuales de Prestación de Servicios (RIPS)	70
Sistema de Vigilancia en Salud Pública (SIVIGILA) 2003-2007..	71
Censo de Población y Vivienda (CENSO) 2005	72
OBJETIVOS	73
Objetivo general...	73
Objetivos específicos..	73

CAPÍTULO 2: DISEÑO METODOLÓGICO

Tipo de estudio	77
Población y fuentes de datos...	77
Fuentes de datos complementarias	78
Indicadores de calidad seleccionados	78
Descripción de los procesos.	82
Recursos	84
<i>Talento humano</i>	84
<i>Recurso físico</i>	84

<i>Recursos de datos</i>	84
<i>Recursos de software</i>	84
Consideraciones éticas	84

CAPÍTULO 3: RESULTADOS

Indicadores de documentación de la fuente	92
Indicador de cobertura	93
Indicadores de contenido	94
<i>Declaración de la edad</i>	94
<i>No respuesta</i>	96
<i>Sobre-respuesta y consistencia</i>	96
Indicadores de fiabilidad	97
Indicadores de representatividad	99

CAPÍTULO 4: ERRORES DE MUESTREO

INTRODUCCIÓN	103
Tipos de muestra	104
Muestras probabilísticas	104
De conglomerados	104
Estratificadas	105
Polietápicas	105
Cálculo de errores de muestreo	106
CONCLUSIONES	107
REFERENCIAS BIBLIOGRÁFICAS	108

ÍNDICE DE CUADROS

CAPÍTULO 1.

Cuadro 1.1	Definición de sistema de información desde el punto de vista de diferentes autores	30
Cuadro 1.2	Tipos de metadatos y sus funciones	40
Cuadro 1.3	Atributos y características de los metadatos.....	41
Cuadro 1.4	Dimensiones de calidad de datos según ISO/IEC 25012.....	45
Cuadro 1.5	Marco para evaluar la calidad de los datos: marco general	46
Cuadro 1.6	Clasificación de sesgos	50
Cuadro 1.7	Clasificación de errores según fuente	52
Cuadro 1.8	Métodos de medición según la fuente de datos	54
Cuadro 1.9	Caracterización de la Encuesta Nacional de Salud (Ministerio de la Protección Social, 2009)	58
Cuadro 1.10	Encuesta Nacional de Salud según muestra total, cantidad de muestra contratada y realizada, año 2007	59
Cuadro 1.11	Caracterización de la Encuesta Nacional de Demografía y Salud, año 2005	60
Cuadro 1.12	Caracterización de la Encuesta Nacional de Calidad de Vida, año 2003	62
Cuadro 1.13	Caracterización de la Encuesta Nacional de Situación Nutricional en Colombia, año 2005	63
Cuadro 1.14	Caracterización del Estudio Nacional de Salud Mental, año 2003	63
Cuadro 1.15	Caracterización del Segundo Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas, año 1998.....	64

Cuadro 1.16	Caracterización del Tercer Estudio Nacional de Salud Bucal, año 1998.....	65
Cuadro 1.17	Caracterización de la Gran Encuesta Integrada de Hogares, año 2007	67
Cuadro 1.18	Caracterización del certificado de nacido vivo.....	68
Cuadro 1.19	Caracterización del certificado de defunción	69
Cuadro 1.20	Caracterización del Registro Individual de Prestación de Servicios.....	70
Cuadro 1.21	Caracterización del Sistema de Vigilancia en Salud Pública	71
Cuadro 1.22	Caracterización del Censo de Población y Vivienda, año 2005.....	72

CAPÍTULO 2.

Cuadro 2.1	Indicadores seleccionados para evaluar la calidad de los datos.....	79
-------------------	---	----

CAPÍTULO 3.

Cuadro 3.1	Total de registros y atributos de las fuentes evaluadas. ASIS 2002-2007.....	92
Cuadro 3.2	Indicadores de cobertura por fuente de información. ASIS 2002-2007	93
Cuadro 3.3	Porcentaje de atributos con no respuesta por fuente. ASIS 2002-2007	96
Cuadro 3.4	Porcentaje de atributos con sobre-respuesta y consistencia por fuente. ASIS 2002-2007	97
Cuadro 3.5	Indicadores de existencia, duplicidad y completitud por fuente. ASIS 2002-2007	98

ÍNDICE DE GRÁFICOS

CAPÍTULO 2.

- Gráfico 2.1** Esquema de procesos para la caracterización de los datos disponibles para la realización del análisis de la situación de salud en Colombia, 2002-2007 82
- Gráfico 2.2** Diagrama de actividades para la caracterización de los datos disponibles para la realización del análisis de la situación de salud en Colombia, 2002-2007..... 83

CAPÍTULO 3.

- Gráfico 3.1** Calidad de los datos para las encuestas según fuente e indicador. ASIS 2002-2007. 90
- Gráfico 3.2** Calidad de los datos para los registros continuos de estadísticas vitales según fuente e indicador. ASIS 2002-2007..... 91
- Gráfico 3.3** Total poblaciones estudiadas de las fuentes según evaluación de la declaración de la edad. ASIS 2002-2007..... 95

INTRODUCCIÓN

Para llevar a cabo el análisis de la situación de la salud es necesario contar con metodologías objetivas de medición, que permitan conocer el impacto que factores diversos tienen en la salud de una población y con metodologías objetivas de medición sobre los datos para garantizar la calidad de la información y/o conocimiento obtenido sobre la situación de la salud al finalizar el proceso.

La información se convierte, entonces, en un insumo fundamental para la toma de decisiones, la orientación de políticas públicas, la formulación de estrategias, planes y programas que redunden en el bienestar de la población estudiada. Sin embargo, para que la información sea un recurso capaz de satisfacer las expectativas de los tomadores de decisiones, debe garantizarse la calidad –esto es, la confiabilidad y validez– la de los datos obtenidos y el conocimiento que de ellos se infiere.

Los avances actuales en el campo de la informática aumentan cada día la probabilidad de lograr la excelencia operacional en las actividades de recolección, procesamiento, almacenamiento y comunicación de datos para la generación de información, y su automatización contribuye en gran medida a garantizar la calidad.

En la búsqueda de la versión más objetiva posible de la situación de la salud en Colombia, el proyecto ASIS 2002-2007 tuvo como fuentes principales las siguientes: Registros Continuos de Estadísticas Vitales, que comprenden los registros de defunciones entre los años 1973 y 2006, los de defunciones fetales entre 1998 y 2006 y los de nacimientos entre los años 1998 y 2006; Registros Individuales de Prestación de Servicios (RIPS) del año 2007; los datos acopiados por el Sistema de Vigilancia en Salud Pública (SIVIGILA) durante los años 2003 a 2007; el Censo General de Población y Vivienda del año 2005 y las encuestas poblacionales que se enuncian a continuación: Encuesta Nacional de Salud (ENS) 2007, Encuesta Nacional de Demografía y Salud (ENDS) 2005, Encuesta de Calidad de Vida (ECV) 2003, Encuesta Nacional de la Situación Nutricional en Colombia (ENSIN) 2005, Estudio Nacional de Salud Mental en Colombia (ENSM) 2003, Segundo Estudio Nacional de

Factores de Riesgo de Enfermedades Crónicas (ENFREC) 1998, Tercer Estudio Nacional de Salud Bucal (ENSAB) 1998, Gran Encuesta Integrada de Hogares (GEIH) 2007 y la Encuesta Nacional de Hogares (ENH) realizada entre 1995 y 2000.

Con el fin de entregar medidas objetivas sobre la información acopiada, se lideró un proceso de evaluación de la calidad de los datos mediante el diseño de una metodología de trabajo que consideró actividades que van desde la documentación y validación de los datos hasta su disposición en una base de datos integrada para facilitar los procesos analíticos.

Todos los procesos fueron documentados y diagramados utilizando el estándar ANSI combinado, a su vez, con otras metodologías de documentación y estandarización de procesos.

Al final se obtuvo el informe descriptivo en dos versiones: ejecutivo y detallado, que permiten evidenciar el estado de los datos y servir de insumo para definir con objetividad las medidas a considerar en el análisis de la situación de la salud en Colombia.

CAPÍTULO 1:



**MARCO
CONCEPTUAL**

A continuación se describen los conceptos asociados a los procesos de evaluación de la calidad de los datos y los aspectos principales aspectos relacionados con la temática.

Sistemas de información

Para definir el concepto de sistema de información es preciso, en primera instancia, definir el concepto de sistema. “Un sistema es un conjunto delimitado de componentes que se relacionan e interactúan entre sí para lograr un objetivo común transformando entradas en salidas”. Así, por ejemplo, una organización es un sistema en el que cada uno de sus componentes puede ser, a su vez, un sistema que, mediante una actuación conjunta, transforma materias primas en bienes, productos y/o servicios. Ahora bien, “Todo sistema organizacional está íntimamente relacionado en mayor o menor medida con una entidad abstracta que debe alinearse con los objetivos corporativos denominada sistema de información”, cuyo objetivo, como su nombre lo indica, es proporcionar información sobre las operaciones y el desempeño organizacional para que la gerencia comprenda, mantenga y dirija la organización y oriente acciones que permitan el uso estratégico de los sistemas de información y, así, contribuir al mejoramiento continuo y el logro de las metas corporativas.

Un sistema de información tiene como finalidad la adecuada articulación de los componentes, tanto internos como externos, de la organización, de forma tal que produzca información útil para apoyar los procesos de planificación, seguimiento, control y evaluación corporativos, y puede ser usados con objetivos estratégicos que apunten a: lograr la excelencia operativa; la creación de nuevos productos; servicios y modelos de negocios; mantener las buenas relaciones con clientes y proveedores; apoyar una toma de decisiones mejorada; generar ventajas competitivas y, en muchos casos, contribuir a la supervivencia de las organizaciones.

El Cuadro 1.1 resume en orden cronológico las definiciones de sistema de información desde el punto de vista de diferentes autores.

Cuadro 1.1 Definición de sistema de información desde el punto de vista de diferentes autores

Autor	Definición de sistema de información
Langerfors [1973] y Goldkuhl y Lyytinen [1982]	"Es un medio implementado tecnológicamente, puesto en práctica con el objetivo de captar, almacenar y diseminar expresiones lingüísticas así como para apoyar la realización de procesos inferenciales. El sistema de información facilita la creación y el cambio de significados que sirven para definir propósitos tales como el control, la toma de decisiones y la argumentación" (Hirschheim <i>et al</i> , 1995:11).
Samuelson [1977]	"Es la combinación de recursos humanos y materiales que resultan de las operaciones de almacenar, recuperar y usar datos con el propósito de una gestión eficiente en las operaciones de las organizaciones" (Samuelson, 1977:3).
Lucey [1987]	"Es un sistema para convertir datos procedentes del exterior o del interior del mismo en información y para comunicar ésta en forma apropiada a los directivos de todos los niveles y de todas las funciones, para facilitar la toma de decisiones efectiva y oportuna para la planeación, dirección y control de las actividades de las cuales ellos son responsables" (Herrera y Salazar, 2008:25).
Patricia Burt [1990]	"Es una combinación de fuentes de información junto con una serie de mecanismos de recuperación, manipulación y uso. El objeto de cualquier Sistema de Información es conectar a un usuario con la fuente de información que necesita, para satisfacer sus necesidades" (Burt y Kinnucan, 1990:175-208).
Debons [1991]	"Es un conjunto de personas, maquinaria y procedimientos que integrados hacen posible a los individuos trabajar con entradas (inputs) y demandas que aparecen en el trabajo cotidiano" (López, 1991:26).
Codina [1996]	"Es un conjunto de elementos que interactúan unidos por una relación de influencia" (Codina, 1996:117-46).
Martínez Santana [1998]	"Es una organización compleja que relaciona todo el conjunto de métodos, datos, medios e instrumentos necesarios para satisfacer las necesidades de la dirección de un objeto determinado" (Herrera y Salazar, 2008:29).
Chain [2000]	"Es una agrupación de componentes que trabajan juntos para alcanzar un objetivo común". (Herrera y Salazar, 2008:27).
Gutiérrez J. [2006]	"Es un ordenamiento sistemático y permanente de los componentes responsables directamente de: recolección y elaboración de los datos requeridos (estadísticos, científicos, tecnológicos, históricos, etc.) de acuerdo con normas nacionales e internacionales, producción de las series estadísticas e indicadores necesarios, suministro de esos datos a los usuarios y colaboración para facilitar el proceso de información, diseño, puesta en operación y control de los registros estadísticos primarios, y de los subsistemas de informes, diseño y actualización de las bases de datos y archivos requeridos para el manejo del sistema, mantenimiento y análisis de las series cronológicas de población, estadísticas vitales y de recursos y servicios de salud, garantizar la calidad y oportunidad de los datos producidos, colaborar con otros sectores en la definición de las necesidades de indicadores y de otras estadísticas y en el aprovechamiento de los mismos dentro del proceso decisorio" (Gutierrez, 2006: 18).

Fuente: Adaptado de Herrera y Salazar, 2008: 4-8.

En el análisis desarrollado por Herrera y Salazar (2008:28), se diferencian dos perspectivas utilizadas históricamente para definir los sistemas de información: la primera corresponde a sus funciones u operaciones, y la segunda a su estructura o composición; también se hace referencia a posturas más recientes como las de O'Brien y Laudon, que definen dichos sistemas a partir de una combinación entre estructura y funciones.

Por su parte, Gutiérrez (2006:18) define los sistemas de información en salud como “un medio para describir y entender el estado de salud del individuo, de la familia y de la comunidad; para poder ejercer la vigilancia epidemiológica; para planear, dirigir, ejecutar, controlar y evaluar los servicios de salud; para asegurarse que las acciones de salud programadas para el individuo, la familia y la comunidad se están realizando; para reconocer las acciones de salud que el individuo, la familia y la comunidad ejecutan o deben ejecutar; para medir el impacto de las acciones sobre el estado de salud y las variaciones de éste en el tiempo y para tomar las decisiones pertinentes”.

De acuerdo a lo anterior, desde el punto de vista de diferentes autores y en el marco del proyecto de análisis de la situación de la salud en Colombia, se entenderá como sistema de información: “el conjunto formal de elementos –recurso humano, recursos tecnológicos (hardware, software y redes de comunicación), recursos de datos y procedimientos, estrategias, políticas, métodos y reglas– que operan en forma integrada y permiten el procesamiento de datos previamente recopilados y almacenados para producir, publicar y distribuir la información de apoyo a los procesos de evaluación, planificación, seguimiento y control y contribuir a una toma de decisiones mejorada” (Herrera y Salazar, 2008: 29)

Elementos que componen un sistema de información

Así pues, un sistema de información está compuesto por las herramientas tecnológicas o tecnologías de información-TI (hardware, software y redes), recursos de datos, recurso humano y procedimientos.

El hardware consiste en el equipamiento informático para llevar a cabo la entrada, proceso y salida de la información, en otras palabras, es el lugar físico donde el subsistema lógico realiza el tratamiento de datos. El software está constituido por los programas e instrucciones que se dan al ordenador, que procesa la información a través de ellos. Los datos son el conjunto de cifras o valores que representan hechos del mundo real mediante los cuales el sistema obtiene la información de salida una vez ésta haya sido procesada (más adelante, se amplía el concepto de dato y se describen sus características). Los recursos de red o la red es un conjunto de computadoras y/o dispositivos conectados por medio de cables, señales, ondas o cualquier otro método de transporte de datos, que comparten información de tipo archivos, recursos como impresoras y servicios como el de acceso a Internet, correo electrónico, entre otros. El recurso humano es el personal técnico y especializado –entre quienes los CIO– Gerentes de información o de sistemas

de información, jefes de proyecto, operadores informáticos, arquitectos, analistas de sistemas, programadores y administradores de bases de datos, entre otros- y los usuarios del sistema, tanto internos como externos. Los procedimientos están formados por las estrategias, políticas, métodos y reglas en torno a los sistemas, es decir, son las rutinas organizativas para el diseño y el desarrollo de nuevos proyectos, criterios de calidad, procedimientos relativos a cómo utilizar los sistemas, procedimientos de mantenimiento, programas de formalización de personal, entre otros. (Herrera y Salazar, 2008: 14)

Actividades del sistema de información

Además de los componentes que los caracterizan, los sistemas de información cuentan con actividades que producen la información para soportar la toma de decisiones. De acuerdo con las consideraciones de diferentes autores, estas actividades –de manera incluyente– son: la entrada, el almacenamiento y procesamiento de datos, la salida o generación de información y la retroalimentación, seguimiento y control del sistema.

La entrada de datos es el proceso mediante el cual el sistema de información toma los datos que requiere para procesarlos y transformarlos en información. El almacenamiento –una de las actividades más importantes de los sistemas de información– es la capacidad que tiene el sistema de guardar la información recolectada en la sección o proceso anterior para un uso posterior. El procesamiento es la posibilidad que ofrece el sistema de efectuar cálculos, ordenaciones, clasificaciones y comparaciones de datos –bien sea actuales o almacenados– de acuerdo con una secuencia de operaciones preestablecida. La salida o generación de información es la capacidad del sistema para transmitir y poner a disposición de los usuarios los datos de entrada o la información resultante, bien sea por medios impresos o a través de unidades de almacenamiento, plotters, etc. La retroalimentación se refiere a la salida que se devuelve al personal de la organización para ayudarlo a mejorar el desempeño del sistema de acuerdo con estándares preestablecidos. El control del desempeño del sistema comprende el monitoreo y la evaluación de las actividades previas para determinar si un sistema se está desplazando hacia el logro de su meta (Herrera y Salazar, 2008: 33-36).

Tipos de sistemas de información

Los sistemas de información se pueden clasificar de acuerdo a diferentes aspectos: 1. Según los niveles de decisión (sistemas de procesamiento de transacciones, de manejo de conocimiento y de apoyo a los trabajadores del conocimiento, de automatización de la oficina, de información administrativa, de información estratégica y de soporte a los ejecutivos); 2. Según las áreas fun-

cionales de la organización (sistemas de mercadeo y ventas, manufactura o producción, recursos humanos, contabilidad y finanzas); 3. Sistemas de información empresariales o sistemas de información para empresas digitales (sistemas de planeación de recursos empresariales-ERP, de manejo de las relaciones con los clientes-CRM, con los proveedores-SRM, con los socios-PRM y con los empleados-ERM, sistemas de manejo de la cadena de abastecimiento-SCM, de inteligencia de negocios-BI, de manejo del conocimiento KMS-KWS, de integración de aplicaciones empresariales-EAI, de manejo del recurso humano-HRM y sistemas de manejo del ciclo de vida de producto-PLM, entre otros). Otras clasificaciones consideran aspectos como: el grado de automatización (automatizados, semiautomatizados o manuales), el grado de formalidad (formales e informales) y el ámbito (locales, nacionales e internacionales).

Para Senn(1992: 25-29), los sistemas de información son de tres tipos:

1. Sistemas para el procesamiento de transacciones tienen como finalidad mejorar las actividades rutinarias de una empresa, pues de ellas depende toda la organización. En efecto, Senn define una transacción como “cualquier suceso o actividad que afecta a toda la organización”, y, por lo tanto, el procesamiento de transacciones es el conjunto de procedimientos para su manejo, el cual incluye las actividades de cálculo, clasificación, ordenamiento, almacenamiento, recuperación y generación de resúmenes.

2. Sistemas de información administrativa: que ayudan a los directivos a tomar decisiones estructuradas y a resolver problemas. El aspecto estructurado se refiere al hecho de que los administradores conozcan de antemano los factores que deben tenerse en cuenta para la toma de decisiones así como las variables con influencia más significativa sobre el resultado de una decisión.

3. Sistemas para el soporte de decisiones que permiten a los directivos tomar decisiones no muy estructuradas (lo que los diferencia de los sistemas de información administrativa) cuando no existen procedimientos claros y no es posible identificar con anticipación todos los factores que deben considerarse en la decisión.

Por su parte, O’Brien (2001: 13-14) clasifica los sistemas de información en sistemas de apoyo a las operaciones y sistemas de apoyo gerencial. Los primeros se subdividen en:

Sistemas de procesamiento de transacciones: los cuales registran y procesan datos resultantes de transacciones comerciales; los resultados de este procesamiento se utilizan para actualizar las bases de datos organizacionales.

Sistemas de control de procesos: mediante los cuales la computadora realiza de manera automática el ajuste toma de un proceso de producción física en tiempo real.

Sistemas de colaboración empresarial: los cuales permiten comunicar ideas, compartir recursos y coordinar esfuerzos de trabajo cooperativo.

Los sistemas de apoyo gerencial se subdividen en:

Sistemas de información gerencial: son aquellos que suministran a los usuarios finales productos de información que respaldan gran parte de sus necesidades de toma de decisiones diarias.

Sistemas de apoyo a decisiones: son sistemas de información interactivos que se basan en la computadora y utilizan modelos de decisión y bases de datos especializadas para apoyar los procesos de toma de decisiones de los usuarios finales.

Sistemas de información ejecutiva: son sistemas adaptados a las necesidades estratégicas de información de la alta gerencia, y su fin último consiste en proporcionar un acceso inmediato y fácil a información selectiva sobre factores clave que son fundamentales para el logro de los objetivos estratégicos de una organización.

Las clasificaciones descritas permiten ver los sistemas de información desde diferentes puntos de vista y de acuerdo con sus múltiples aspectos; revelan, además, la evolución de estos sistemas y la importancia de su valor agregado o producto final: la información y/o el conocimiento.

Dato

Si bien el concepto de dato ha sido abordado desde diferentes puntos de vista, su significado es similar en todos ellos. Desde una concepción general, el dato es una descripción elemental, una cifra, un valor o un hecho que se convierte en la unidad mínima de información; dicho de otra manera, el dato es la materia prima con la que se produce la información.

Desde el punto de vista de autores como Whitten, “el dato es una colección de hechos considerados de forma aislada, los cuales aportan un significado pero en general no son de utilidad por sí solos” (Whitten *et al*, 1998: 47).

En las diferentes ramas del conocimiento se tienen visiones más puntuales sobre lo que es el dato. En el mundo de la computación, el “dato es el término que se utiliza para describir la información representada por grupos de interruptores de encendido/apagado [...]. Los datos son las señales individuales en bruto y sin ningún significado que manipulan las computadoras para producir la información” (Norton *et al*, 2006: 185). Así pues, dato es el término que se usa para describir las señales con las cuales trabaja la computadora.

A diferencia de los datos, la información representa mayor valor y puede ser entendida como “un conjunto de datos relacionados, procesados y contextualizados que constituyen una estructura de mayor complejidad”, a su vez la información es el insumo requerido para la creación de conocimiento (Aja *et al*, 2002:1).

Para Whitten “La información es un dato o conjuntos de datos que han sido manipulados, con lo que resultan de utilidad para alguien”. Esto es, “la información debe tener valor, o en caso contrario sería un dato” (Whitten *et al*, 1998: 47).

Por su parte, Rojas plantea que “los datos, debidamente procesados y analizados, generan información y conforman la base para la formulación de políticas y la evaluación de las mismas” (Rojas, 2006: 7).

O’Brien (2001: 29) define información como “datos que han sido convertidos a un contexto significativo y útil para usuarios finales específicos”. Explica, además, que los datos son “sometidos a un procesamiento que se constituye en valor agregado donde se manipula y organiza su forma, se analiza y evalúa su contenido y se adapta al contexto adecuado para el usuario final”.

El análisis de la situación de la salud en Colombia ha adoptado esta definición de O’Brien y ha sometido los datos, compilados en las fuentes antes mencionadas, a los procesos de administración que exigen las actividades de gestión de información para agregarles valor. La gestión de información “comprende las actividades relacionadas con la obtención de la información adecuada, a un precio adecuado, en el tiempo y lugar adecuado, para la toma de decisiones” (Gutiérrez, 2006: 18).

En este sentido, Gutiérrez (2006: 18) elabora el concepto de organización o persona (usuario o cliente) que recibe el producto o servicio como resultado de la gestión de información. Ésta, a su vez, se contempla como parte esencial de la infraestructura para la gestión del conocimiento, debido a que los servicios de producción de información impulsan la generación del conocimiento y contribuyen en la búsqueda de soluciones a los problemas que enfrentan las organizaciones; dicho de otra manera, la gestión de la información se vincula con la generación y la aplicación de estrategias, el establecimiento de políticas, y el desarrollo de una cultura de carácter organizacional y social encaminada al uso racional, efectivo y eficiente de la información en función de los objetivos y metas de las compañías en materia de desempeño y de calidad (Gutiérrez, 2006: 18).

Por otra parte la gestión de información involucra actividades que permiten administrar su insumo principal: los datos que, según su estado, se convierten en la garantía de la calidad de la información.

Para resumir: existe una diferencia importante entre dato e información. Los datos individuales son hechos fuera de contexto y la mayoría de ellos tienen poco o ningún significado; pero cuando son agrupados, procesados y contextualizados en forma lógica, los datos conllevan significados especiales y se acercan más al concepto de información. Por otra parte, los datos deben ser administrables y la información, a su vez, debe ser sometida a procesos de gestión (administración de datos). Para facilitar estas tareas existen los sistemas administradores de bases de datos SGBD y las bases de datos.

Bases de datos

Para el procesamiento y utilización de los datos es preciso diseñar bases de datos. Según Scott (1992: 335), una base de datos es “un sistema de archivos de computadora que usa una organización de archivos particular para facilitar: la actualización rápida de registros aislados, la actualización simultánea de registros relacionados, el acceso de los programas de aplicaciones a todos los registros y el acceso rápido a todos los datos almacenados que deben unirse para satisfacer un informe o consulta particular de rutina o de propósito especial”.

Para O'Brien (2001:29), una base de datos es un “conjunto integrado de elementos de datos relacionados de manera lógica. Una base de datos consolida muchos registros almacenados de antemano en archivos separados de manera que una fuente común de datos sirve a muchas aplicaciones”; a su vez, el término registro “representa un conjunto de atributos que describen una entidad; un grupo de registros relacionados es un archivo de datos o una tabla”.

Este autor clasifica las bases de datos en dos tipos: operacionales y analíticas; las primeras se caracterizan porque almacenan detalladamente datos requeridos para respaldar las operaciones de toda la organización; las segundas se componen de los datos y la información resumida más necesaria para los gerentes de una organización y otros usuarios finales (O'Brien, 2001: 143-145).

Otros autores definen bases de datos -llamadas también bancos de datos- como “un conjunto de datos persistentes que es utilizado por los sistemas de aplicación de alguna empresa dada. Donde el término empresa es un término genérico conveniente para identificar a cualquier organización independiente de tipo comercial, técnico, científico u otro. Una empresa podría ser un individuo (con una pequeña base de datos personal), toda una corporación o un gran consorcio similar (con una gran base de datos compartida)” (Date, 2001: 10).

Para Prieto, una base de datos es “un sistema formado por un conjunto de datos y un paquete de software para la gestión de dicho conjunto de datos de tal modo que: se controla el almacenamiento de datos redundantes, los datos resultan independientes de los programas que los usan, se almacena las relaciones de los datos junto con éstos, y se puede acceder a los datos de diversas formas” (Prieto et al, 2006: 619). Plantea, además, como requisitos que debe cumplir una base de datos: el acceso y utilización múltiple, la flexibilidad, la confidencialidad y seguridad, la protección contra fallos, la redundancia controlada y la interfaz de alto nivel.

Para efectos del análisis de la situación de la salud de Colombia 2002-2007, se entenderá por base de datos el conjunto de datos debidamente estructurados, relacionados, organizados y almacenados que representan un fenómeno o problema del mundo real y permiten la centralización de las características asociadas al fenómeno, la disminución de la redundancia y su disposición para el uso de las aplicaciones que requieran su consulta de acuerdo con las expectativas de los usuarios en un Sistema Gestor de Bases de Datos-DBMS.

Sistemas de bases de datos

Los sistemas de bases de datos “son una colección de datos interrelacionados y un conjunto de programas que permiten a los usuarios tener acceso a éstos y modificarlos. La principal finalidad es simplificar la interacción de los usuarios con el sistema, pues ofrecen a los usuarios una visión abstracta de los datos, ocultando ciertos detalles del modo en que se almacenan y mantienen los datos” (Silberchatz et al, 2006: 3).

Según Silberchatz, esta abstracción de datos se da en tres niveles: físico, lógico y vistas. El primero es el más bajo: describe el modo como se almacenan los datos y detalla las estructuras de datos complejas de bajo nivel; el segundo ilustra qué datos se almacenan en la base de datos en términos de un número pequeño de estructuras simples; el tercero existe para simplificar la interacción con el sistema al proporcionar muchas vistas para la misma base de datos.

Los sistemas de bases de datos proveen un lenguaje de manipulación que permite a los usuarios tener acceso a los datos organizados, borrarlos, modificarlos y/o insertar nueva información en la base de datos. Así mismo proveen un lenguaje de definición de datos que especifica los esquemas de la base de datos y las propiedades de los datos (Silberchatz *et al*, 2006: 7).

Para Scott, un sistema de bases de datos incluye los siguientes componentes: los archivos de la base de datos que contienen los elementos individuales de los datos almacenados; un sistema de administración de bases de datos conformado por un conjunto de programas de software que, entre otras tareas, administra la base de datos, controla el acceso a ella y le proporciona la seguridad, un sistema de interfaz de lenguaje anfitrión, que es la parte del sistema de administración de bases de datos que se comunica con los programas de aplicaciones, los que, a su vez, realizan las mismas funciones de los programas convencionales, pero son independientes de los archivos de datos y usan definiciones estándar de los datos; un sistema de interfaz de lenguaje natural, que permite la actualización y las consultas en línea de los usuarios sin mayor conocimiento de los sistemas de cómputo, las terminales de acceso y actualización en línea, que pueden encontrarse adyacentes a la computadora o a miles de millas de distancia; el sistema de salida o generador de informes, que proporciona informes de trabajo de rutina, documentos e informes especiales; y el diccionario de datos, que es un depósito centralizado de información en forma computarizada sobre los datos en una base y una descripción y definición de sus atributos (Scott, 1992: 357-359).

Diccionario de datos

Los diccionarios de datos, “incluyen información acerca de la localización de datos en los archivos de la base y contienen reglas de acceso e información de seguridad y privacidad acerca de los datos” (Scott, 1992: 359).

Por su parte, Yourdon (1993:212) define el diccionario de datos como “un listado organizado de todos los datos pertinentes al sistema, con definiciones precisas y rigurosas para que tanto el usuario como el analista tengan un entendimiento común de todas las entradas, salidas, componentes de almacenes y cálculos intermedios”. En el marco de esta definición se desarrolló el análisis de la situación de la salud objeto de este estudio.

Yourdon plantea, además, que el diccionario de datos permite:

- Describir el significado de los flujos y almacenes que se muestran en los diagramas de flujo de datos.
- Describir la composición de agregados de paquetes de datos que se mueven a lo largo de los flujos, es decir, paquetes complejos (por ejemplo el domicilio de un cliente), que pueden descomponerse en unidades más elementales (como ciudad, estado y código postal).
- Describir la composición de los paquetes de datos en los almacenes
- Especificar los valores y unidades relevantes de piezas elementales de información de los flujos de datos y en los almacenes de datos.
- Describir los detalles de las relaciones entre almacenes que se enfatizan en un diagrama de entidad-relación (Yourdon, 1993: 212).

Para Luque un diccionario de datos es “Un catálogo o directorio computacional que contiene metadatos, es decir, datos sobre datos”. Y, agrega, “un diccionario de datos incluye un componente de software para manejar una base de definiciones de datos, es decir, metadatos sobre la estructura, los elementos de datos y otras características de las bases de datos como una organización” (2006: 9).

Metadatos

Los metadatos se generan a partir de un diccionario de datos, que en resumen se puede definir como “una lista de datos relacionados con las características, cualidades y contenido de un objeto de información” (Sánchez, 2003:1). Así pues, a partir de los diccionarios de datos se pueden determinar los metadatos que serán necesarios para la gestión de la información.

Los metadatos, definidos por lo general como “el conjunto de datos sobre los datos”, buscan establecer un lenguaje concreto que describa los datos y pueda ser entendido tanto por las personas que demandan su uso como por las que producen información (McLeod *et al*, 2000: 602).

Los metadatos presentan diferentes niveles de especificidad o estructura, su principal objetivo es describir, identificar y definir un recurso, con el fin de recuperar, filtrar e informar sobre condiciones de uso, autenticación y evaluación, preservación e interoperatividad (Lannela y Waugh (1997) Citado por Senso y Pinero, 2003: 98).

Según Sánchez, en sentido genérico los metadatos se definen como “datos de los datos, y surgieron de la necesidad de describir, identificar y caracterizar objetos de información (documentos impresos, sonoros, etc.) a través de campos que identifican cada dato, con el propósito de facilitar la gestión, búsqueda, acceso y localización de los documentos. Ejemplo: Título, autor, materia, resumen, editor” (Sánchez, 2003:3).

El mismo autor define también los metadatos en el contexto electrónico como “información que describe datos con respecto al contenido, la forma y las características técnicas y editoriales de la información electrónica, los metadatos son sometidos a procesos de generación, consulta, manipulación y distribución” (Sánchez, 2003:4).

Dadas las características de la información del proyecto ASIS Colombia 2002-2007, esta última definición resulta ser la más cercana a lo requerido, pues identifica tanto lo que deben incluir los metadatos en su estructura, como su funcionalidad.

Según Gilliland-Swetland (1998: 1), los metadatos son útiles para:

- Incrementar la accesibilidad, dado que la existencia de un conjunto de metadatos que describa correctamente uno o varios objetos aumenta la posibilidad de acceder a ellos.
- Hacer la búsqueda de información en múltiples colecciones a la vez: por medio del mapeo entre sistemas heterogéneos es posible consultar, con una única ecuación de búsqueda, bases de datos que utilicen diferentes sistemas de metadatos para describir sus objetos.
- Expandir el uso de la información, al facilitar la difusión de versiones digitales de un único objeto.
- Controlar las versiones, no solo en lo que se refiere a gestionar la vida de un objeto, sino también en lo que tiene que ver con su difusión; es decir, generar diferentes metadatos con distintas cantidades de información sobre un mismo objeto permite distribuirla a un público heterogéneo.
- Establecer claramente las restricciones de explotación, mediante la información sobre los derechos de autor, el control del uso de todo o una parte, del objeto, el método de pago por su disfrute y el control del acceso a información restringida.

Los metadatos pueden ser administrativos, descriptivos, de preservación, técnicos y de uso. En el Cuadro 1.2 se describe cada tipo con ejemplos de algunas de las funciones comunes para un sistema de información digital (Gilliland-Swetland, 1998: 3).

Cuadro 1.2 Tipos de metadatos y sus funciones

TIPO	DEFINICIÓN	EJEMPLOS
Administrativo	Usados en la gestión y administración de recursos de información	<ul style="list-style-type: none"> -Administración de información -Derechos y reproducción -Requerimientos legales para el acceso -Localización de información -Criterios de selección para la digitalización -Control de la versión
Descriptivo	Usados para representar recursos de información	<ul style="list-style-type: none"> -Registros catalográficos -Proporcionar ayuda en la búsqueda -Índices especializados -Hiperenlazar relaciones entre recursos -Anotaciones de los usuarios
Preservación	Para salvaguardar los recursos de información	<ul style="list-style-type: none"> -Informar sobre las condiciones de uso de los recursos físicos -Informar sobre las acciones llevadas a cabo para versiones físicas y digitales de recursos
Técnico	Relativos al modo como funcionan los sistemas o el comportamiento de los metadatos	<ul style="list-style-type: none"> -Documentación de hardware y software -Digitalización de la información (formato, ratio de compresión...) -Autenticación y datos de seguridad (encriptación, contraseñas...) -Control de tiempo de respuesta de sistemas
Uso	Relativos al nivel y tipo de uso que se hace con los recursos informativos	<ul style="list-style-type: none"> -Información sobre versiones -Reutilización del contenido del recurso

Fuente: Gilliland-Swetland, 1998:3

Es importante resaltar que existen diferentes formas de organizar y resumir las características de los metadatos. En el cuadro 1.3 se pueden observar las principales características de los metadatos con ejemplos (Baca, 1998:3).

Cuadro 1.3 Atributos y características de los metadatos

ATRIBUTO	CARACTERÍSTICAS	EJEMPLOS
Fuente	Metadatos internos generados por el agente creador con el propósito de informar sobre el momento de su creación	Nombres de ficheros Estructuras de directorios Formatos de ficheros y Algoritmia de compresión
	Metadatos externos relativos a una información que se modifica después de su creación	Registros catalográficos Información sobre derechos de autor
Método de creación	Metadato generado automáticamente por su ordenador	Índices de palabras clave
	Metadatos creados manualmente	Herramientas descriptivas
Naturaleza	Creados por el autor del documento objeto	Los utilizados en páginas HTML
	Generados por profesionales de la información, independientemente de quién sea el autor del documento objeto	Registros MARC Encabezamiento de materia
Estado	Estático: no cambian desde su creación	Título, fecha de creación
	Dinámico: varía con el uso del documento objeto	Estructuras de directorios
	A largo plazo: necesario para asegurarse de que el documento objeto será accesible en todo momento	Información de los derechos de (autor, de uso, de difusión...)
	A corto plazo: con clara vocación transaccional	Información sobre el uso
Semántica	Normalizados por medio de un vocabulario controlado	MARC AACR2
	No controlados	Etiquetas HTML
Nivel	Colecciones de metadatos relativos a colecciones de documentos objeto	Índices especializados
	Un metadato relativo a un documento objeto individual, fuera de cualquier colección	Información sobre el formato Leyenda de una imagen

Fuente: Baca, 1998:3

Los atributos de los metadatos contienen información fundamental (disponibilidad, significado y modo de obtención, entre otras) para usuarios, desarrolladores y administradores de las bases de datos, bodegas de datos y los almacenes de datos (Datawarehouse-Datamart); también describen la fuente para cada característica, especifican si ésta ha sido tomada directamente de un sistema fuente o si es calculada, documentan las transformaciones a las que debe someterse y la fecha de su última actualización en la base de datos o bodega de datos, entre otras informaciones.

Bodega de datos

Múltiples son las definiciones de bodega de datos dadas por diversos autores. A continuación se presentan algunas de ellas:

Para Luque (2006: 6), las bodegas de datos son “una construcción de grandes repositorios integrados de datos que pueden ser generados internamente o recibidos de fuentes externas, organizados de tal manera que facilitan el proceso de toma de decisiones”. Luego de lo cual agrega que en la actualidad “algunas organizaciones han comenzado a construir en vez de bodegas de datos, repositorios más pequeños, más fáciles y rápidos de implementar, conocidos como almacenes de datos, los cuales son organizados alrededor de aplicaciones específicas, como por ejemplo: los datos relevantes de un área de la organización almacenados por separado constituyen un almacén de datos”.

O’Brien (2006:145) afirma que “una bodega de datos almacena datos del año en curso y de años anteriores, que han sido extraídos de las diversas bases de datos operacionales de una organización”, por lo cual la define como “una fuente central de datos que ha sido seleccionada, editada, estandarizada e integrada, de manera que los gerentes y otros profesionales que son usuarios finales puedan utilizarla para una variedad de formas de análisis empresarial, investigación de mercado y respaldo a las decisiones”.

Según Anaya (1996: 96-8), la bodega de datos es un “proceso estructurado de métodos, técnicas y consideraciones organizacionales que permiten manejar datos con el propósito de obtener una vista consolidada y detallada de todo el negocio o parte de él”.

Y para Ramalho (1999:22-35), “una bodega de datos es una base de datos orientada a consultas, como resultado de un análisis extenso y de la transformación de datos de una organización. Una bodega de datos puede ayudar a las aplicaciones de apoyo de decisiones y OLAP (On-Line Analytical Processing en español procesamiento analítico en línea) porque provee datos que son consolidados y consistentes, orientados hacia un tema, históricos o solamente de lectura.” Agrega, Además, que una bodega de datos consolida los datos operacionales de varias fuentes, con convenciones de nombre, medidas, atributos físicos y semántica consistentes.

Para efectos del presente análisis se entiende como bodega de datos la colección de datos o hechos históricos, integrados y estables obtenidos de los sistemas transaccionales y de otras fuentes para representar fenómenos o problemas de potencial interés, que respondan a las preguntas clave formuladas por los tomadores de decisiones.

Calidad del dato

El concepto de calidad ha cambiado en los últimos años, y hoy se le considera como una filosofía que busca una mejora continua de los procesos, procedimientos, decisiones y productos o servicios.

La cultura de la calidad y la productividad asociadas, transforman la manera de hacer las cosas inercialmente, por una dinámica más integradora en donde el liderazgo es característico. Así, el proceso de cambio de lo tradicional a la calidad, propicia la aparición de la jerarquía que otorga la eficiencia y la eficacia con que se hagan y determinen las cosas.

ICONTEC, en la norma 1892, define la calidad como “un conjunto de propiedades o características de un producto o servicio con la habilidad para satisfacer necesidades implícitas o explícitas” (Londoño, 2009:2-3).

A su vez, el International Standard ISO 8402 define la calidad como “La totalidad de circunstancias y características de un producto o servicio que un usuario requiere para satisfacer sus necesidades” (Londoño, 2009: 2).

Los parámetros de calidad que se incorporan a este análisis se apoyan en las definiciones anteriores y, también, en las definiciones que consideran la calidad de un producto o servicio desde las expectativas del cliente, tal y como las formulan Armand V. Feigenbaum: “Satisfacción de las expectativas del cliente” o de William Edwards Deming: “Calidad es satisfacción del cliente” (Citado por Gopal, K. 2002: 55-56).

La calidad es aplicable a muchos conceptos, productos o servicios, y el dato no es ajeno a esta filosofía de mejora continua. Como ya se ha dicho, el dato es la colección de hechos que no han sido transformados en información, o lo que es lo mismo, es un hecho crudo cuyas conclusiones están por obtenerse. Estos hechos pueden describir características aisladas de personas, lugares, cosas, ideas, procesos o eventos, características que son útiles y de calidad si sirven para el propósito que se necesita. Se puede deducir, por lo tanto, que la calidad del dato es un concepto subjetivo: no solo depende del uso que se le vaya a dar, sino también de los distintos puntos de vista (multidimensional) del analista.

Dimensiones de la calidad del dato

Sin embargo, este carácter subjetivo del dato está sometido a objetivos concretos. En efecto, no solo debe servir para proveer información de calidad, sino también para apoyar los procesos de administración de los datos y hacer buen uso de la información que resulta de ellos. Varias son las dimensiones que caracterizan la calidad del dato, entre otras valen destacarse:

- La exactitud: mide el grado en que la información refleja lo que está pasando.
- La totalidad: refleja el grado en que las bases de datos cuentan con toda la información crítica.
- La oportunidad: mide que la información esté disponible cuando se requiere para tomar una decisión.
- La relevancia: expresa la utilidad de la información para la persona a la que se le proporciona.

- El nivel de detalle: esto es, que la información tenga todos los elementos requeridos dependiendo según el nivel organizacional y el tipo de decisión al cual está destinada.
- La consistencia: que la información sea la misma en cada uno de los usos que se le dé (Eugenio *et al*, 2007: 367).

Para poder garantizar la calidad de los datos se deben incorporar actividades de planificación, organización, control, mejora y aseguramiento; es decir, se debe gestionar la calidad de los datos y definir medidas para cada una de las dimensiones requeridas y para cada escenario en particular. La forma en la que se definen las medidas depende de los métodos de medición utilizados por lo tanto una misma dimensión de calidad puede tener tantas medidas diferentes como partes interesadas haya en un escenario (Eugenio *et al*, 2007: 367).

Las medidas de calidad de los conjuntos de datos pueden clasificarse en: orientadas al modelo de datos –medidas, por lo general, objetivas–, y orientadas a los propios datos– las cuales requieren metadatos que complementen el significado del dato con respecto a la dimensión de calidad y según la parte interesada (Eugenio *et al*, 2007: 367). Para efectos de facilitar el estudio de la calidad de datos, las medidas de calidad han sido agrupadas en dimensiones o calidades menores y se han adoptado como estándares de calidad de datos los reconocidos internacionalmente.

Estándares de calidad del dato

Según el Estándar ISO 25012, la mejor forma de abordar el estudio de datos para un contexto específico es dividirla en partes más pequeñas llamadas “calidades menores” y conocidas como dimensiones de calidad de datos (Guerra, 2009: 6-7). Estas “calidades menores” se convierten en un conjunto usable en un contexto y determinan los denominados “Modelos de calidad de datos”.

El Estándar ISO 25012 propone un modelo que categoriza los atributos de calidad de datos en 15 características o dimensiones consideradas desde dos puntos de vista: inherente y dependiente del sistema (Guerra, 2009: 6-7).

Inherente al sistema: se refiere al grado en el cual las características de calidad de datos tienen el potencial intrínseco para satisfacer las necesidades implicadas cuando el dato es usado bajo condiciones específicas.

Dependiente del sistema: se refiere al grado en el cual la calidad de los datos es enriquecida y preservada dentro de un sistema de cómputo cuando el dato es usado bajo condiciones específicas.

Como se muestra en el Cuadro 1.4, la medición de las dimensiones definidas en el Estándar ISO 25012 permite valorar las características esenciales de los datos para garantizar su calidad (Hernández, 2008: 28).

Cuadro 1.4 Dimensiones de calidad de datos según ISO/IEC 25012

Características de calidad	Definición	Inhe-rente	Depen-diente del sistema
Exactitud (Accuracy)	El grado en el que el dato tiene atributos que representan correctamente el valor verdadero del atributo instanciado en un concepto o evento en un contexto específico de uso.	X	
Complejidad (Completeness)	El grado en el que el dato asociado a un sujeto con una entidad tiene todos los valores esperados para el atributo y para las instancias de la entidad relacionadas en un contexto específico de uso.	X	
Consistencia (Consistency)	El grado en el que el dato tiene atributos que son libres de contradicción y son coherente con otros datos en un contexto específico de uso.	X	
Credibilidad (Credibility)	El grado en el que el dato tiene atributos que son considerados como verdaderos y creíbles por usuarios en un contexto específico de uso.	X	
Actualidad (Currentness)	El grado en el que el dato tiene atributos que son de la edad correcta en un contexto específico de uso.	X	
Accesibilidad (Accessibility)	El grado en el que el dato puede ser accesado en un contexto específico de uso, particularmente por la gente que necesita el soporte de tecnología o una configuración especial porque tiene alguna indisponibilidad.	X	X
Conformidad (Compliance)	El grado en el que el dato tiene atributos que se adhieren a las normas, convenciones o regulaciones vigentes y reglas similares relacionadas con la calidad de datos en un contexto específico de uso.	X	X
Confidencialidad (Confidentiality)	El grado en el que el dato tiene atributos que aseguran que éste es solo accesible e interpretable por usuarios autorizados en un contexto específico de uso.	X	X
Eficiencia (Efficiency)	El grado en el que el dato tiene atributos que pueden ser procesados y proporciona los niveles esperados de desempeño al utilizar las cantidades y los tipos de recursos apropiados en un contexto específico de uso.	X	X
Precisión (Precision)	El grado en el que el dato tiene atributos que son exactos o que proporcionan la discriminación en un contexto específico de uso.	X	X
Trazabilidad (Traceability)	El grado en el que el dato tiene atributos que proporcionan un rastro de auditoría al acceso a los datos y de cualquier cambio realizado a los datos en un contexto específico de uso.	X	X
Entendibilidad (Understandability)	El grado en el que el dato tiene atributos que le permiten ser leído e interpretado por usuarios, y es expresado en lenguajes apropiados, símbolos y unidades en un contexto específico de uso.	X	X
Disponibilidad (Availability)	El grado en el que el dato tiene atributos que le permiten ser recuperados por usuarios autorizados y/o aplicaciones en un contexto específico de uso.		X
Portabilidad (Portability)	El grado en el que el dato tiene los atributos que le permiten ser instalado, substituido o movido de un sistema a otro conservando la calidad existente en un contexto específico de uso.		X
Recuperabilidad (Recoverability)	El grado en el que el dato tiene atributos que le permiten mantener y conservar un nivel especificado de operaciones y calidad, aún en caso de falla, en un contexto específico de uso.		X

Fuente: ISO. ISO/IEC FDIS 25012: Software engineering - software product quality requirements and evaluation (SQuaRE) - data quality model2008.

Otra modelo o marco de referencia para evaluar la calidad de los datos es el propuesto por S. Carson (2001: 20-23. Cuadro 1.5).

Cuadro 1.5 Marco para evaluar la calidad de los datos: marco general

Dimensiones de la calidad	Elementos	Indicadores
0. Condiciones previas de la calidad.	0.1 Entorno jurídico e institucional – El entorno jurídico respalda la labor estadística.	0.1.1 Se asignan claramente las tareas de recopilación, procesamiento y divulgación de estadísticas. 0.1.2 El intercambio de información y la coordinación entre las entidades que laboran datos son adecuados. 0.1.3 La confidencialidad de los datos que proporcionan los declarantes individuales está garantizada y éstos se utilizan solo con fines estadísticos. 0.1.4 La declaración de datos se efectúa por ley y/o gracias a disposiciones que alientan la declaración voluntaria.
	0.2 Recursos – Los recursos guardan relación con las necesidades de los programas estadísticos.	0.2.1 Los recursos humanos, instalaciones, recursos informáticos y recursos financieros guardan relación con los programas estadísticos. 0.2.2 Se han adoptado medidas para garantizar un uso eficiente de los recursos.
	0.3 Carácter pertinente de las estadísticas – Las estadísticas comprenden información pertinente en el campo correspondiente.	0.3.1 Se evalúa el carácter pertinente y la utilidad práctica de las estadísticas en relación con las necesidades de los usuarios.
	0.4 Otra gestión de la calidad – La calidad es una piedra angular de la labor estadística.	0.4.1 Se han establecido procedimientos para centrar la atención en la calidad. 0.4.2 Se han establecido procedimientos para llevar a cabo un seguimiento de la calidad del programa de estadísticas. 0.4.3 Se han establecido procedimientos para abordar los aspectos de calidad en la planificación de los programas de estadísticas.
1. Garantías de Integridad Riguroso respeto del principio de objetividad en la recopilación, procesamiento y divulgación de las estadísticas.	1.1 Profesionalismo – El profesionalismo es un principio rector de las políticas y prácticas estadísticas.	1.1.1 Las estadísticas se elaboran en forma imparcial. 1.1.2 La elección de fuentes y métodos estadísticos así como las decisiones sobre divulgación, obedecen únicamente a consideraciones estadísticas. 1.1.3 La entidad estadística pertinente puede formular comentarios si ha habido una interpretación errónea o uso indebido de las estadísticas.

Dimensiones de la calidad	Elementos	Indicadores
<p>1. Garantías de Integridad Riguroso respeto del principio de objetividad en la recopilación, procesamiento y divulgación de las estadísticas.</p>	<p>1.2 Transparencia – Las políticas y prácticas estadísticas son transparentes.</p>	<p>1.2.1 Se dan a conocer al público los términos y las condiciones en que se recopilan, procesan y divulgan las estadísticas. 1.2.2 Se indican públicamente los casos en que las autoridades del gobierno tienen acceso a las estadísticas antes de su divulgación. 1.2.3 Se identifican claramente los productos de las entidades o unidades estadísticas. 1.2.4 Se anuncia en forma anticipada toda modificación sustancial de la metodología, los datos fuente o las técnicas estadísticas.</p>
	<p>1.3 Normas éticas – Las políticas y prácticas estadísticas se rigen por normas éticas.</p>	<p>1.3.1 Se han establecido normas de conducta para el personal de las que éste tiene un conocimiento cabal.</p>
<p>2. Rigor metodológico - La base conceptual de las estadísticas se apoya en normas, directrices o buenas prácticas aceptadas internacionalmente.</p>	<p>2.1 Conceptos y definiciones – Los conceptos y definiciones corresponden a marcos estadísticos de aceptación internacional.</p>	<p>2.1.1 La estructura global de los conceptos y definiciones se ajusta a las normas, directrices o buenas prácticas de aceptación internacional.</p>
	<p>2.2 Alcance – El alcance de los datos se ajusta a las normas, directrices o buenas prácticas de aceptación internacional.</p>	<p>2.2.1 El alcance se ajusta a las normas, directrices o buenas prácticas de aceptación internacional.</p>
	<p>2.3 Clasificación /sectorización – Los sistemas de clasificación y sectorización utilizados se ajustan a las normas, directrices o buenas prácticas de aceptación internacional.</p>	<p>2.3.1 Los sistemas de clasificación/sectorización utilizados se ajustan en general a las normas, directrices o buenas prácticas de aceptación internacional.</p>
	<p>2.4 Base de registro – El registro y la valoración de los flujos y saldos se ajustan a las normas, directrices o buenas prácticas de aceptación internacional.</p>	<p>2.4.1 Se utilizan precios de mercado para valorar los flujos y saldos. 2.4.2 La contabilidad se prepara en valores devengados. 2.4.3 Los procedimientos de registro en cifras netas/brutas se ajustan en general a las normas, directrices o buenas prácticas de aceptación internacional.</p>
<p>3. Exactitud y fiabilidad Los datos fuente y las técnicas estadísticas son sólidos y los productos estadísticos reflejan adecuadamente la realidad.</p>	<p>3.1 Datos fuente – Los datos fuente disponibles constituyen una base adecuada para la compilación de estadísticas.</p>	<p>3.1.1 Los datos fuente se obtienen de programas de recopilación de gran alcance en los que se tiene en cuenta las condiciones particulares del país. 3.1.2 Los datos fuente se ajustan en grado razonable a las definiciones, el alcance, la clasificación, la valoración y el momento de registro que se requieren. 3.1.3 Los datos fuente son puntuales.</p>
	<p>3.2 Evaluación de los datos fuente – Los datos fuente se evalúan periódicamente.</p>	<p>3.2.1 Los datos fuente –incluidos los censos, las encuestas muestrales y los registros administrativos– se evalúan regularmente, por ejemplo, para examinar la cobertura, el error muestral, errores de respuesta y errores no muestrales; los resultados de estas evaluaciones son objeto de seguimiento y orientan los procesos estadísticos.</p>

Dimensiones de la calidad	Elementos	Indicadores
<p>3. Exactitud y fiabilidad</p> <p>Los datos fuente y las técnicas estadísticas son sólidos y los productos estadísticos reflejan adecuadamente la realidad.</p>	<p>3.3 Técnicas estadísticas</p> <p>– Las técnicas estadísticas utilizadas se basan en procedimientos estadísticos sólidos.</p>	<p>3.3.1 El proceso de compilación de datos emplea técnicas estadísticas sólidas para ajustar las fuentes de datos.</p> <p>3.3.2 Los demás procedimientos estadísticos (por ejemplo, los ajustes y conversiones de los datos así como el análisis estadístico) también se basan en técnicas estadísticas sólidas.</p>
	<p>3.4 Evaluación y validación de datos intermedios y productos estadísticos</p> <p>– Los resultados intermedios y productos estadísticos se evalúan y validan periódicamente.</p>	<p>3.4.1 Cuando procede, los resultados intermedios se validan comparándolos con otra información.</p> <p>3.4.2 Las discrepancias estadísticas de los datos intermedios se evalúan e investigan.</p> <p>3.4.3 Se investigan las discrepancias estadísticas y otros indicadores de posibles problemas de los productos estadísticos.</p>
	<p>3.5 Estudios de revisión</p> <p>– A fin de evaluar la fiabilidad, se efectúa un seguimiento de las revisiones para aprovechar la información que proporcionan.</p>	<p>3.5.1 Se efectúan regularmente estudios y análisis de las revisiones y éstos se utilizan para mejorar los procesos estadísticos (véase la sección 4.3.3).</p>
<p>4. Utilidad de las estadísticas para el usuario</p> <p>Las estadísticas que poseen una adecuada periodicidad y puntualidad, son coherentes y están sujetas a una política de revisión previsible.</p>	<p>4.1 Periodicidad y puntualidad</p> <p>– La periodicidad y puntualidad de las estadísticas se ajustan a normas de divulgación de datos de aceptación internacional.</p>	<p>4.1.1 La periodicidad de los datos se ajusta a las normas de divulgación.</p> <p>4.1.2 La puntualidad de los datos se ajusta a las normas de divulgación.</p>
	<p>4.2 Coherencia</p> <p>– Las estadísticas son coherentes dentro del conjunto de datos al que pertenecen y a lo largo del tiempo, y son compatibles con otros conjuntos de datos.</p>	<p>4.2.1 Las estadísticas son coherentes dentro del conjunto de datos al que pertenecen.</p> <p>4.2.2 Las estadísticas son coherentes o pueden conciliarse en un período razonable.</p> <p>4.2.3 Las estadísticas son coherentes o pueden conciliarse con los datos de otras fuentes o marcos estadísticos.</p>
	<p>4.3 Política y procedimientos de revisión</p> <p>– Las revisiones de los datos se efectúan de acuerdo con un procedimiento regular y de conocimiento público.</p>	<p>4.3.1 Las revisiones se efectúan de acuerdo con un calendario regular y transparente.</p> <p>4.3.2 Los datos preliminares y/o revisados se identifican claramente.</p> <p>4.3.3 Se dan a conocer al público los estudios y análisis de las revisiones (véase la sección 3.5.1).</p>
<p>5. Acceso</p> <p>Se dispone fácilmente de datos y metadatos y se brinda asistencia adecuada a los usuarios.</p>	<p>5.1 Acceso a los datos</p> <p>– Las estadísticas se presentan en forma clara y comprensible, las formas de divulgación son adecuadas y las estadísticas se dan a conocer con imparcialidad.</p>	<p>5.1.1 La presentación de las estadísticas facilita una interpretación adecuada y permite efectuar comparaciones válidas (formato y claridad del texto, cuadros y gráficos).</p> <p>5.1.2 Los medios y los formatos utilizados en la divulgación de datos son adecuados.</p> <p>5.1.3 Las estadísticas se divulgan de acuerdo con el calendario previamente anunciado.</p> <p>5.1.4 Las estadísticas se divulgan simultáneamente a todos los usuarios.</p> <p>5.1.5 Los datos que no se divulgan regularmente se proporcionan a solicitud de los interesados.</p>

Dimensiones de la calidad	Elementos	Indicadores
5. Acceso Se dispone fácilmente de datos y metadatos y se brinda asistencia adecuada a los usuarios.	5.2 Acceso a los metadatos – Se divulgan metadatos pertinentes y actualizados.	5.2.1 Se dispone de documentación sobre los conceptos, el alcance, las clasificaciones, la base de registro, las fuentes de datos y las técnicas estadísticas, y se señalan las discrepancias con respecto a las normas, directrices o buenas prácticas de aceptación internacional. 5.2.2 Se ofrecen diversos niveles de desglose de la información según las necesidades del público al que se desea llegar.
	5.3 Asistencia a los usuarios – Se ofrece un respaldo oportuno y competente.	5.3.1 Se da a conocer al público un punto de contacto para cada área. 5.3.2 Pueden obtenerse fácilmente catálogos de publicaciones, documentos y otros servicios, así como información sobre sus precios.

Fuente: International Monetary Fund, Carol S. Toward a Framework for Assessing Data Quality: Annex I. Data Quality Assessment Framework -generic framework. 2001:20-3

Independiente del modelo de calidad de los datos o del marco de referencia utilizado, su abordaje estandarizado y documentado se convierten en un referente fundamental al momento de estudiar la calidad de los datos en diferentes contextos.

Medición de la calidad de los datos

En la mayoría de los casos, los problemas de investigación científica implican mediciones sobre los conceptos que intervienen en ella; estas mediciones deben describir las características o las relaciones existentes según el problema estudiado (Anderson *et al*, 2003: 823).

Los conceptos de medición y medida, se pueden definir de una manera intrínseca medir algo se puede definir como: determinar el número de veces que una unidad o patrón de medida cabe en el objeto que se pretende medir. En otras palabras, medir es comparar el objeto con el patrón de medición para poder determinar el número de unidades y fracciones que éste incluye (Anderson *et al*, 2003: 823).

La medición, puede definirse como “la asignación de valores a la dimensión de una variable empleando un sistema de reglas y escala preestablecidos, siendo el dato, el valor que se le asigna a la variable” (Anderson *et al*, 2003: 823).

Si lo que se desea es evaluar la calidad de los datos, éstos se convierten en el objeto de estudio y de la medición de sus características a partir de métricas o indicadores se obtiene una valoración de su calidad. Pero como la calidad de datos es un concepto multidimensional, para medirla es necesario descomponerla en características observables o, lo que es lo mismo, en dimensiones de calidad de datos (ver supra). Es importante resaltar que cuando se efectúa la descripción y el análisis de un conjunto de datos y de sus indicadores derivados, es necesario tener en cuenta los

posibles errores y las diversas fuentes de error que pueden afectar los productos de información obtenidos.

Las fuentes de error se derivan principalmente de factores como: las limitaciones del instrumento de medida, la naturaleza de la magnitud a medir, los problemas que surgen durante la aplicación del instrumento y el proceso de registro de datos. Estas fuentes o causas de error dan origen a su vez a los errores o sesgos en las mediciones (Abraira: 1).

Errores o sesgos

Tradicionalmente se distingue entre dos tipos de error, el error debido a la precisión limitada del instrumento que atenta contra la reproducibilidad de la medición e introduce en ella un error aleatorio; y el error debido a la validez, también limitada, que introduce un error sistemático o sesgo. La validez depende del instrumento y se refiere a si éste mide lo que debe medir; la precisión depende tanto del instrumento como del proceso de medición y tiene que ver con cuánto se aproxima la medida al valor real de la magnitud. El error aleatorio determina la precisión y el error sistemático la validez. Existen instrumentos más precisos y/o válidos que otros, pero ninguno es infinitamente preciso y válido; por otra parte, como los errores pueden ocurrir en cualquier fase de la investigación, es prácticamente imposible eliminarlos en su totalidad (Abraira: 1).

Para disciplinas como la estadística y la epidemiología, el sesgo se puede clasificar en sistemático y no sistemático (no necesariamente error aleatorio), y se define como un error que aparece en los resultados de un estudio generado por factores que dependen del proceso realizado a los datos (recolección, procesamiento, almacenamiento, análisis, interpretación, publicación o revisión), el cual puede conducir a conclusiones que son sistemáticamente diferentes de la verdad o incorrectas con respecto a los objetivos de una investigación (Hernández *et al*, 2000: 439).

Existen diferentes clasificaciones de los errores que varían su comportamiento de acuerdo al tipo de estudio realizado, al instrumento utilizado o a los procesos de gestión de los datos. Según Sackett (1979: 51), existen más de 35 tipos posibles de sesgos provenientes del muestreo y la medición pero, de manera general, pueden agruparse en dos o tres grandes categorías: en el primer caso se trata de sesgos de información o clasificación –también llamados de observación– y sesgos de selección; en el segundo caso de sesgos de selección, información o medición y confusión. En el Cuadro 1.6 se puede ver una breve descripción de los tipos de sesgo.

Cuadro 1.6 Clasificación de sesgos

Tipo de sesgo	Descripción
Selección	<p>Se presentan durante el proceso para escoger la muestra o los sujetos. Son errores sistemáticos que se introducen al inicio del estudio, durante la selección o el seguimiento de la población y propician una conclusión equivocada sobre la hipótesis en evaluación. Los errores de selección pueden ser originados por el mismo investigador o ser el resultado de relaciones complejas en la población que pueden no ser evidentes para el investigador y pasar desapercibidas (Hernández <i>et al</i>, 2000: 440).</p> <p>Ocurren cuando hay errores en los diseños muestrales por una deficiente definición del marco de muestreo, al incluir voluntarios, pacientes u otros elementos que no son representativos de una población y se busca la generalización de resultados (Hernández <i>et al</i>, 2000: 440).</p> <p>Dentro de los sesgos de selección se pueden destacar: sesgos de autoselección, sesgos por pérdidas de seguimiento, por muestras no representativas (paradoja de berkson), por supervivencia selectiva (paradoja de neyman) y sesgos de diagnóstico o de detección entre otros.</p>
Información	<p>Se presentan después de tener los individuos de la muestra. Se definen como “errores que se introducen durante la medición de las variables en la población en estudio y se presentan de manera diferencial entre los grupos que se comparan, y ocasionan una conclusión errónea respecto de la hipótesis de investigación” (Hernández <i>et al</i>, 2000: 441).</p> <p>Ocurren cuando un valor de la variable difiere sistemáticamente de su verdadero valor por causa de calibración en el instrumento o la apreciación diferencial del observador y oportunidad en la exigencia del dato observado (Hernández <i>et al</i>, 2000: 441-443).</p> <p>Pueden ser no diferenciales (aleatorios) cuando el grado de error del instrumento o técnica empleada es el mismo para los grupos que se comparan y, diferenciales (no aleatorios) cuando el grado de error es diferente para los grupos estudiados, el sesgo de información se refiere articuladamente a este último tipo (Hernández <i>et al</i>, 2000: 441-443).</p> <p>Dentro de los sesgos de selección se pueden destacar: sesgos del observado (de recuerdo o de memoria), del observador (detección), de medición, de atención, de respuesta invariable, de notificación, de vigilancia, de falseamiento y del instrumento (instrumento defectuoso) entre otros.</p>
Confusión	<p>Se presentan cuando los resultados obtenidos en la población en estudio apoyan una conclusión falsa o ilegítima sobre la hipótesis en evaluación debido a la influencia de otras variables que no fueron controladas adecuadamente ya sea durante la fase de diseño o de análisis, por lo cual puede resultar en una sobre o subestimación de la asociación real. Todos los resultados derivados de estudios observacionales están potencialmente influenciados por este tipo de sesgo. Existe sesgo de confusión cuando se observa una asociación no causal entre la exposición y el evento en estudio o cuando no se observa una asociación real entre la exposición y el evento en estudio por la acción de una tercera variable que no es controlada (Hernández <i>et al</i>, 2000: 445-446).</p>

Es importante resaltar que las estrategias utilizadas para el control de los sesgos y errores determinarán la calidad de los datos y la información generada a partir de ellos.

Los errores presentes en un conjunto de datos también se pueden clasificar como errores de cobertura y errores de contenido, los primeros ocurren por la insuficiencia en la contabilidad en

las personas, hechos o cosas, motivo por el cual se presentan omisiones o sobre-registros (errores por defecto o por exceso); en la práctica, es más común que ocurra lo primero, es decir que los casos registrados sean menores que los reales. Los errores de contenido son aquellos derivados de insuficiencias en la información recogida, en este caso el individuo y sus datos han sido incluidos pero en forma incorrecta (Rodríguez *et al*, 1997: 44).

La estimación, el conocimiento y la claridad de los errores se convierten en aspectos fundamentales para la evaluación de la calidad de los datos y la información y/o conocimiento que éstos generan. Los errores presentes varían también según el tipo de fuente –censos, encuestas, registros continuos entre otros–; a continuación se describen algunos de ellos:

Cuadro 1.7 Clasificación de errores según fuente

Tipo de fuente	Errores
Censos	<p>En los censos los errores que afectan la calidad de la información básica censal de la población son tanto errores de cobertura -también llamados de omisión o de sobre-enumeración- como errores de contenido.</p> <p>Por lo general, en los errores de cobertura “la cantidad de personas omitidas es mayor a la cantidad de censadas” y las cifras varían de acuerdo al país y sus zonas geográficas, también existen omisiones en variables como la edad y el estrato social entre otras, y a veces se presenta la omisión completa de áreas geográficas, viviendas y personas (Rodríguez <i>et al</i>, 1997: 45).</p> <p>Los errores de contenido tienen su origen en diversas etapas de la operación censal; pueden originarse en deficiencias del trabajo cartográfico, errores en la capacitación de los empadronadores o inadecuada coordinación y supervisión de la recolección de datos (Rodríguez <i>et al</i>, 1997:44).</p> <p>Uno de los errores de contenido más comunes en los censos de población son los referentes a la edad, debido al desconocimiento y equivocaciones de los informantes, dejando al descubierto la mala declaración de la edad y la preferencia de dígitos (Rodríguez <i>et al</i>, 1997: 45).</p> <p>Otro problema resaltado es el del número de hijos, debido a que las informantes en edad avanzada tienden a olvidar el número exacto de hijos tenidos, y además en algunos casos son omitidos los hijos muertos; a menudo en otros casos a menudo las respuestas son contabilizadas como mujeres sin hijos, o las mujeres que no declaran hijos se registran por error del empadronador, como no respuesta, Dado a que estos dos últimos errores son usuales, se han desarrollado procedimientos de ajuste basados en ecuaciones de regresión sobre datos empíricos de censos en diferentes lugares del mundo (Rodríguez <i>et al</i>, 1997: 45).</p> <p>También son comunes los errores en las respuestas a las preguntas relacionadas con la actividad económica, lo que en algunos censos prácticamente ha impedido la utilización de los datos recogidos sobre este tema (Rodríguez <i>et al</i>, 1997:46).</p> <p>Mientras que en el error de contenido se ven afectadas prácticamente todas las preguntas en diversas etapas de la operación censal, pueden suceder otro tipo de errores en la aplicación de los instrumentos causados por: una mala redacción en las preguntas y su inadecuada selección y/o distribución inconveniente; estos errores se pueden presentar al momento de la recolección de los datos ya sea por desconocimiento del informante o por una incompleta preparación de los entrevistadores. Por último, es importante resaltar que es posible que se presenten errores adicionales en los procesos de codificación y/o digitación de datos (Rodríguez <i>et al</i>, 1997: 46).</p>

Tipo de fuente	Errores
Registros continuos	<p>En los registros continuos los errores pueden ser originados por: falta de recursos del sistema (personal, infraestructura e información), omisiones de la legislación, falta de empoderamiento y comprensión sobre la utilidad del registro de los hechos, problemas de coordinación dentro de la administración pública, deficiencias en los informes (incapacidad o falta de compromiso de los informantes o de los encargados para llenarlos a cabalidad) (Rodriguez <i>et al</i>, 1997: 46).</p> <p>Los errores de cobertura son comunes y se refieren a que los hechos registrados sean menores a los ocurridos, a esto se le suele llamar subregistro, y cuando los casos registrados superan a los ocurridos se habla de sobre-registro. Estos aspectos están ligados directamente a los conceptos de alcance y de cabalidad. La cabalidad es "la medida en que se da cuenta de todos hechos o eventos allí donde se aplica el sistema de inscripción" (Rodriguez <i>et al</i>, 1997: 46).</p> <p>Otros error típico de cobertura de los registros continuos se relaciona con: la oportunidad, o más concretamente con la inscripción tardía, ya sea por ignorancia, comodidad o razones de fuerza mayor. Lo anterior puede alterar los indicadores demográficos, sobre todo si la tendencia al registro tardío tiene fluctuaciones marcadas año a año (Rodriguez <i>et al</i>, 1997: 46).</p> <p>Los registros continuos también se ven afectados por la duplicación de inscripciones, aunque los avances informáticos y los progresos en la legislación pertinente tienden a reducir drásticamente su ocurrencia (Rodriguez <i>et al</i>, 1997: 46).</p> <p>Los errores de contenido indican la deficiencia extendida en la información contenida en los registros que puede ir desde la no respuesta hasta la certificación no especializada (los certificados carecen de confiabilidad) (Rodriguez <i>et al</i>, 1997: 46).</p> <p>Algunos errores son atribuibles a deficiencias administrativas del sistema de registro de hechos reflejadas principalmente en la gestión de los datos por distintas dependencias gubernamentales, con diferentes criterios y enfoques (Rodriguez <i>et al</i>, 1997: 46).</p>
Encuestas	<p>En un estudio muestral, como es el caso de las encuestas, la muestra debe seleccionarse a partir de la población objetivo o de estudio, procurando que sea representativa de la misma y controlando que sus características sean una aproximación de las de la población, con un margen de error tolerable y conocido. Se hace énfasis en este criterio porque en este tipo de investigación es más importante la representatividad de la muestra que la preocupación por el tamaño de la misma.</p> <p>De acuerdo con lo anterior, entre los errores presentes en las encuestas se destacan aquéllos debidos al uso de muestras, las cuales, por lo general, son utilizadas en los estudios para minimizar los costos, hacer un número mayor de preguntas que las de un censo y/o lograr una mayor profundidad (Rodriguez <i>et al</i>, 1997: 46).</p> <p>La variabilidad que se observaría entre todas las muestras posibles constituye el error de muestreo, el cual no se conoce pero puede ser estimado a partir de los datos suministrados por la muestra realmente seleccionada.</p> <p>Dada la complejidad de las muestras de las fuentes utilizadas para el proyecto ASIS, resulta pertinente diferenciar e identificar los tipos de errores de muestreo o de cobertura y los errores de contenido (Rodriguez <i>et al</i>, 1997: 44).</p> <p>Los errores de muestreo o de cobertura que se presentan en las encuestas, pueden encontrarse en un diseño mal definido o en equivocadas reglas de asociación entre unidades de observación y unidades muestrales y evidencian problemas como: la falta de cobertura de todos los elementos seleccionados en la muestra, errores en la formulación de las preguntas y en el registro de las respuestas, confusión o incapacidad de los entrevistados para dar la información y errores de codificación o de procesamiento. Para controlar el error por cobertura se parte del intervalo de confianza (Rodriguez <i>et al</i>, 1997: 44).</p>

Tipo de fuente	Errores
Encuestas	<p>Algunos de los errores se pueden reducir mediante el uso de procedimientos como: diseño cuidadoso del cuestionario, pruebas piloto, intensa capacitación de los entrevistadores, supervisión intensa y permanente del trabajo de campo, revisión y validación de los cuestionarios en el campo, supervisión apropiada en la etapa de codificación y procesamiento de los datos, limpieza cuidadosa de archivos de datos con informes de retroalimentación a los supervisores, críticos y entrevistadores (Rodríguez <i>et al</i>, 1997: 45).</p> <p>El error de contenido, al igual que en el censo, afecta prácticamente todas las preguntas de la encuesta, y tiene su origen en diversas etapas de la operación, puede surgir por mala redacción de las preguntas o su selección y distribución inadecuada; también puede presentarse durante la recolección de los datos, bien sea por una incompleta preparación de los entrevistadores o bien desconocimiento del informante. Las deficiencias de información en la edad o el número de hijos pueden producirse durante los procesos de digitación y codificación. Se presenta también la no respuesta, la cual ocurre cuando algunas de las unidades seleccionadas para ser medidas no responde una o varias de las preguntas del cuestionario diseñado para la investigación, fenómeno que causa pérdida de representatividad de la población en las variables no contestadas o de los individuos en el caso de la no respuesta total (Rodríguez <i>et al</i>, 1997: 45).</p> <p>Los errores que no son atribuidos al proceso de muestreo pueden ser aleatorios y no aleatorios, y son originados por varias causas como: errores en la definición del problema o de las escalas, el diseño del cuestionario, los métodos de entrevista y la preparación y análisis de datos, entre otras (Instituto Nacional de Estadística, 2004: 303-305).</p>

Todos los defectos o errores mencionados se deben medir y asociar a las dimensiones de la calidad para, así, dar cuenta de la calidad de un conjunto de datos en un contexto determinado.

Medición de los errores

Debido a que el concepto de calidad es multidimensional, deben definirse los indicadores o métricas asociadas a cada defecto o error reconocido en el conjunto de datos; por otra parte, los defectos deben asociarse previamente con cada dimensión de la calidad de los datos para, así, valorarla de manera integral. A continuación se describen algunas de las métricas y métodos de medición según la fuente de datos.

Cuadro 1.8 Métodos de medición según la fuente de datos

Tipo de fuente	Métodos de medición
Censo	<p>Naciones Unidas: omisión diferencial de individuos en ciertas edades e irregularidades de los datos por sexo y edad.</p> <p>Método de Myers: evalúa la atracción o rechazo de declaración de la edad en los dígitos finales.</p> <p>Método de Whipple: muestra la mala declaración de la edad respecto a la preferencia de los dígitos 0 y 5 (Rodríguez <i>et al</i>, 1997:46)</p>

Tipo de fuente	Métodos de medición
Registros continuos	<p>Las Naciones Unidas proponen que la calidad de los datos puede medirse con respecto a su completitud, exactitud, disponibilidad y puntualidad, como se dispone a continuación (Naciones Unidas, 2003: 36).</p> <p>Cuando la preparación de los datos requiere gran número de operaciones administrativas, se recomienda emplear “técnicas de control de calidad por muestreo” para comprobar la precisión de los procesos administrativos sin necesidad de hacer una nueva comprobación de todo el archivo.</p> <p>Es preciso seguir de cerca los informes estadísticos remitidos por los registradores locales, para descubrir posibles problemas en la elaboración de los informes estadísticos (Naciones Unidas, 2003:20).</p> <p>Existen diversas técnicas para evaluar la exactitud cualitativa de los datos de las estadísticas vitales, que suponen un examen de la coherencia interna de los datos.</p> <p>Al igual que para los censos, en las estadísticas vitales se pueden analizar las edades declaradas para hallar la preferencia de edades terminadas en ciertos dígitos (en 0 o en 5, o en números pares) mediante el uso de diferentes técnicas (Rodríguez <i>et al</i>, 1997:100).</p>
Encuestas	<p>Técnicas para mantener bajo el nivel de no respuesta como: revistas (Call-backs), recordatorios (Follow-ups), submuestreo de no respondientes y aleatorización de la respuesta.</p> <p>Corrección de datos faltantes: análisis de casos completos, análisis de casos disponibles, procedimientos ponderados y procedimientos basados en imputación (DANE, 2002: 3-8).</p> <p>Lo importante a la hora de utilizar las estrategias para el manejo de datos faltantes para efectos de evaluación de la calidad, es tener el registro documentado de los procedimientos utilizados y el número de registros afectados, para que esto a su vez se pueda utilizar para el cálculo de una métrica que represente dicho aspecto.</p> <p>Uso de métricas de inconsistencia, ausencia, incoherencia e inhabitualidad de los datos.</p> <p>La crítica estadística es usada para la aplicación de reglas predeterminadas al conjunto de datos. Si un registro no cumple con alguna de las reglas establecidas, debe ser señalado para ser corregido o cambiado; una vez definido el plan de crítica, se deben verificar, validar los datos y calcular las métricas respectivas para dar cuenta los de datos modificados (DANE, 2002: 3-8).</p>

La calidad de un conjunto de datos puede ser representada por uno o más resultados agregados de las medidas de calidad de los datos combinadas sobre diferentes elementos, subelementos o dimensiones estandarizadas.

En este contexto los indicadores de la calidad de los datos son medidas que describen el cumplimiento de los estándares y, para hacerlo, se debe gestionar la obtención de información sobre las dimensiones que se observen mediante las métricas previamente definidas con respecto a los contenidos de cada característica o variable de interés (México. Universidad Autónoma. Estándares e indicadores de calidad. <http://www.tuobra.unam.mx/publicadas/040803214240.htm>).

Las mediciones o métricas deben ser reportadas mediante el uso de guías o plantillas predefinidas, que estandaricen la presentación de los informes para su posterior divulgación en un reporte de calidad de datos.

Reporte de la calidad de los datos

El reporte de evaluación de la calidad de los datos debe dar cuenta clara de la calidad del conjunto de datos y de los valores de las métricas asociadas. Por lo general, este reporte se presenta bien como metadatos o bien como un informe que tiene importancia cuando los resultados plasmados en los metadatos son reportados en términos de conformidad (conforme/no conforme), o cuando se generan resultados agregados.

El informe debe contener información referente a las fuentes, funciones de análisis utilizadas en los distintos pasos, estimaciones de la validez de los datos, métodos de transformación, procedimientos de control de la calidad utilizados y otros aspectos que se consideren relevantes para los usuarios según el escenario.

Los informes de calidad deben ser de interés tanto para los usuarios como para los productores. A éstos les permite conocer su línea de producción real, comprobar que los datos siguen los procesos establecidos por la organización y establecer especificaciones claras de la calidad para sus productos, que luego podrá validar.

Desde el punto de vista del usuario, los informes recuperan una información fundamental para entender la calidad de los datos, favorecen la posibilidad de derivar nuevos productos y aplicaciones de acuerdo a sus necesidades y permiten confrontarlos con otras bases de datos. Dicho en otras palabras, para los usuarios los datos en general son útiles en tanto que pueden transformarse en información y estén relacionados con una actividad predeterminada que será desarrollada con posterioridad.

En conclusión, la información de la calidad comprende un conjunto de actividades dirigidas a la documentación sobre el comportamiento de los principales indicadores o métricas que dan cuenta de su calidad.

Fuentes de datos para el proyecto ASIS, Colombia 2002-2007

Las fuentes de datos para la obtención de la información y/o conocimiento se clasifican en primarias y secundarias. Son primarias cuando la información se obtiene como producto del contacto directo con el sujeto u objeto de estudio, es el caso, por ejemplo, de la entrevista, la observación y el cuestionario. Las fuentes secundarias son las colecciones de datos obtenidos de procesos de registro previos tales como archivos, historias clínicas o censos. Vale resaltar que esta clasificación es relativa, pues un determinado dato puede ser primario para ciertos estudios y secundario para otros, sin embargo su importancia radica en la vinculación entre los datos y el proceso de explotación.

Una segunda clasificación de las fuentes, discrimina entre los datos directos –los que se recolectan de una manera simple, esto es cuando no existen mayores supuestos de recolección– e

indirectos –los que se obtienen por medio de operaciones matemáticas y su aplicación requiere de fuertes supuestos–. Otra clasificación de las fuentes las divide en fuentes de datos brutos o representados sin desagregaciones, ajustes o correcciones, y fuentes de datos corregidos, cuando se han sometido a técnicas matemáticas que permiten suavizar su comportamiento y ajustarlos a patrones lógicos o refinados.

En salud, algunas de las fuentes de datos utilizadas son: los censos, las encuestas y los registros continuos; estas fuentes proporcionan datos relevantes, centralizados y confiables para describir y analizar fenómenos como la morbilidad, mortalidad, fecundidad, migración y los fenómenos poblacionales, entre otros.

Los censos “constituyen la fuente primaria más importante y amplia de información estadística. Sus resultados suministran antecedentes básicos referentes a la población, en todas sus unidades muestrales, los cuales son necesarios para el diseño e implementación de políticas y programas públicos y para la decisión de inversiones y estudios privados” (Rodríguez *et al*, 1997: 41). Las principales características de los censos son: el secreto estadístico, el auspicio oficial, un territorio bien definido, la universalidad, la unidad censal y la simultaneidad.

El grado de exactitud del recuento de la población total de un país depende de la precisión con que se hayan llevado a cabo todas las operaciones relativas al levantamiento del censo: si las operaciones censales han sido realizadas con poco cuidado y las características de la población no se han determinado con exactitud, es poco probable que el recuento general sea correcto. Para evaluar los resultados censales existen dos métodos: directo e indirecto. El primero consiste en volver a censar por lo menos una parte de la población (encuestas post-censales), haciéndolo de forma tal que pueda obtenerse el número de personas que fueron omitidas o incluidas equivocadamente en el recuento censal; el segundo consiste en la aplicación de técnicas que permitan el análisis demográfico de la información recolectada (Rodríguez *et al*, 1997: 42-43).

Las encuestas poblacionales “son utilizadas como fuentes de datos en análisis referidos, básicamente, a los componentes del cambio de la población. En general permiten profundizar el conocimiento sobre morbilidad, fecundidad, mortalidad y migración. Éstas habitualmente recogen información a nivel de cada individuo, pero también consideran criterios para definir otros grupos como hogares o viviendas” (Rodríguez *et al*, 1997: 43). Dichas encuestas pueden ser de dos tipos: prospectivas que registran los movimientos y hechos demográficos de la población bajo observación durante un periodo de tiempo; y retrospectivas, aquellas que aplican un cuestionario en una sola ocasión.

Los registros continuos de estadísticas vitales “son registros oficiales sobre los principales hechos vitales que ocurren en una población, tales como nacimientos defunciones, matrimonios, separaciones, adopciones, etc.” (Rodríguez *et al*, 1997: 44).

Para garantizar la calidad de la información de los registros continuos es imprescindible que se adopten técnicas, métodos y herramientas que permitan reducir los diferentes tipos de errores

y sesgos, y además de establecer métodos de medición y métricas que garanticen la calidad de los datos durante todo el proceso de producción de la información.

A continuación se describen los aspectos más relevantes de cada una de las fuentes de información suministradas por el Ministerio de la Protección Social para desarrollar el proyecto “Análisis de la Situación de Salud-ASIS Colombia 2002-2007”. Para efectos de la caracterización, las fuentes fueron clasificadas en encuestas poblacionales y registros continuos.

Encuestas poblacionales

Las encuestas poblacionales suministradas fueron: la Encuesta Nacional de Salud(ENS) 2007, Encuesta Nacional de Demografía y Salud (ENDS) 2005, Encuesta de Calidad de Vida (ECV) 2003, Encuesta Nacional de la Situación Nutricional de Colombia (ENSIN) 2005, Estudio Nacional de Salud Mental (ENSM) 2003, Segundo Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas (ENFREC) 1998, Tercer Estudio Nacional de Salud Bucal (ENSAB) 1998 y la Gran Encuesta Integrada de Hogares (GEIH) 2007. A continuación se describirá cada una.

Encuesta Nacional de Salud (ENS) 2007

Cuadro 1.9 Caracterización de la Encuesta Nacional de Salud (Ministerio de la Protección Social, 2009)

ENS-2007	
Objetivo	Determinar la situación de salud de Colombia por medio de la información que brindaron hogares, usuarios e instituciones del sistema de salud colombiano, con el propósito de identificar las necesidades y prioridades en salud; de esta forma, aportar elementos de análisis para la toma de decisiones y futuras investigaciones de la situación de salud del país.
Muestra	Se tomó información de 41.147 hogares, en los que se entrevistaron efectivamente 224 entidades territoriales municipales, 1.168 Instituciones Prestadoras de Servicios de Salud (IPS), y en ellas se realizaron 123.964 encuestas a usuarios de servicios de urgencias, consulta externa y hospitalización. Además, en 1.035 IPS se validaron, recogieron y procesaron de manera prospectiva los Registros Individuales de Prestación de Servicios (RIPS) correspondientes a un mes de cada institución.
Unidad de análisis	Hogares, IPS, usuarios de consulta externa , hospitalización y urgencias, entidades territoriales.
Módulos o componentes	*Hogares Módulo 1: Encuesta hogares Módulo 2: Encuesta para personas de 6 y más años Módulo 3: Encuesta para niños menores de 6 años Módulo 4: Encuesta para personas de 18 años y más

ENS-2007

Módulos o componentes	Ficha para la caracterización del segmento
	*Encuestas IPS
	Módulo 2: Servicio de consulta ambulatoria
	Módulo 3: Verificación de consulta ambulatoria
	Módulo 4: Verificación de servicios de obstetricia
	*Usuarios
	Módulo 1: Consulta Externa
	Módulo 2: Hospitalización y urgencias
	*Entidades Territoriales
	Encuesta para administraciones municipales

En el Cuadro 1.10 se presentan aspectos relevantes del muestreo, tales como la muestra total del estudio, la muestra total contratada y la muestra ejecutada (Ministerio de la Protección Social, 2007).

Cuadro 1.10 Encuesta Nacional de Salud según muestra total, cantidad de muestra contratada y realizada, año 2007

Encuesta de hogares	Según contrato	Realizadas
• Módulo 1: encuesta de hogares	40.000	41.147
• Módulo 2: encuesta para personas de 6 y más años		79.596
• Módulo 3: encuesta para niños menores de 6 años		18.835
• Módulo 4: encuesta para personas de 18 y más años		29.136
Encuestas IPS		
• Módulo 1: encuesta de IPS	1.200	1.168
• Módulo 2: servicios consulta ambulatoria		7.413
• Módulo 3: verificación consulta ambulatoria		7.014
• Módulo 4: verificación servicios obstetricia		840
Encuesta usuarios de IPS	120.000	123.964
• Módulo 1: consulta externa		93.975
• Módulo 2: hospitalización y urgencias		30.169
Encuesta Administraciones Municipales	231	224
• Ficha caracterización de Barrio		5.216

Fuente: Encuesta Nacional de Salud, Colombia 2007

Dado el número representativo de hogares, usuarios, IPS y entidades territoriales que hicieron parte de este proyecto, debe resaltarse la importancia del factor de expansión para inferir los resultados al total de la población colombiana. Por otra parte, los resultados derivados están afectados por los errores –muestrales y no muestrales– propios de las muestras complejas (Ministerio de la Protección Social, 2007).

Por lo general, los errores no muestrales de la ENS-2007 fueron generados durante el proceso de recolección y procesamiento de los datos, y se minimizaron mediante: el diseño de la encuesta, la capacitación de los entrevistadores y la supervisión exhaustiva durante los procesos de recolección, codificación, depuración y procesamiento de los datos. Los errores muestrales fueron estimados mediante un algoritmo que permitió obtener aproximaciones bastante razonables, aunque no exactas dada la complejidad del diseño muestral (Ministerio de la Protección Social, 2007).

Esta encuesta es considerada la más completa de los últimos 18 años, y a partir de ella la comunidad científica, académica y los distintos actores del Sistema General de Seguridad Social en Salud, han iniciado sus análisis para generar conocimiento de utilidad para la toma de decisiones en materia de salud (Ministerio de la Protección Social, 2007).

Encuesta Nacional de Demografía y Salud (ENDS) 2005

La Encuesta Nacional de Demografía y Salud –proyecto quinquenal, desarrollado desde 1990 por profamilia desde 1990, con el apoyo técnico de Macro International Inc. de Calverton, Maryland– es un estudio epidemiológico de carácter descriptivo y cualitativo que ofrece información sobre la salud de las mujeres en edad fértil, su historia reproductiva y su entorno familiar.

En el Cuadro 1.11 se describen algunas de sus características principales (PROFAMILIA, 2005: 1-10).

Cuadro 1.11 Caracterización de la Encuesta Nacional de Demografía y Salud, año 2005

ENDS-2005	
Objetivo	Identificar y describir los problemas, características y situaciones relacionadas con la salud sexual y reproductiva en Colombia, sus regiones, subregiones y departamentos.
Muestra	3935 segmentos distribuidos en 200 municipios para un total de 37211 hogares
Unidad de análisis	<ul style="list-style-type: none"> • Hogares • Mujeres 13-49 años, información sobre esposo y niños menores de 5 años • Citología y mamografía, mujeres 18-69 años • Medidas antropométricas miembros de hogar de 18-64 años
Módulos o componentes	<ul style="list-style-type: none"> • Cuestionario hogares • Cuestionario individual

ENDS-2005

Módulos o componentes	<ul style="list-style-type: none">•Cuestionario mamografía y citología•Cuestionario peso y talla
Cobertura geográfica y niveles de desagregación	Fue a nivel nacional con representatividad urbana y rural. Se dividió al país en 6 regiones: Atlántica, Oriental, Bogotá, Central, Pacífica y Amazonía y Orinoquía y 16 subregiones: 1. Guajira, Cesar, Magdalena. 2. Barranquilla y su Área Metropolitana 3. Atlántico, San Andrés, Bolívar Norte 4. Bolívar Sur, Sucre, Córdoba. 5. Santanderes 6. Boyacá, Cundinamarca, Meta 7. Bogotá. 8. Medellín y su Área Metropolitana 9. Antioquia sin Medellín. 10. Caldas, Risaralda, Quindío. 11. Tolima, Huila, Caquetá. 12 Cali y su Área Metropolitana 13. Valle sin Cali ni litoral Pacífico. 14. Cauca y Nariño sin Litoral Pacífico. 15. Litoral Pacífico. 16. Orinoquia-Amazonía.

Encuesta de Calidad de Vida (ECV) 2003

La articulación bienestar económico / felicidad constituye el núcleo central del fenómeno de "calidad de vida". Según Blanco Abarca (1985) "la relación entre felicidad y bienestar económico concede a la calidad de vida un profundo sentido psicosocial, dentro de la tradición psicosocial del carácter social" (Citado por Bobes *et al*, 1993:5).

Después de concluir la encuesta sobre calidad de vida del año 1979, España propuso definir este concepto como una suma, esto es, como el grado en que las condiciones de una sociedad determinada permiten a sus miembros realizarse de acuerdo con los valores ideológicos establecidos, proporcionándoles una experiencia subjetiva y satisfactoria de su existencia (Citado por Bobes *et al*, 1993:5).

En los años 2003 y 2008 Colombia realizó encuestas de calidad de vida de carácter nacional, con el propósito de desarrollar una política social coherente con las necesidades del país mediante la realización de diversos programas y acciones por parte de los ministerios y de los institutos oficiales, cuya finalidad esencial es lograr el mejoramiento de las condiciones de vida de los hogares colombianos, en especial de los sectores de la población más afectados por la crisis, la inflación y las políticas de ajuste macroeconómico (DANE, 2003).

Cuadro 1.12 Caracterización de la Encuesta Nacional de Calidad de Vida, año 2003

ECV-2003	
Objetivo	Medir en forma actualizada las condiciones socioeconómicas de la sociedad colombiana y permitir la descripción y análisis de la estructura social.
Muestra	24090 hogares
Unidad de análisis	Hogares y personas
Módulos o componentes	A. Identificación y control B. Datos de la vivienda (diligencie este capítulo sólo para el hogar 01 de la vivienda) C. Servicios del hogar (diligencie este capítulo para todos los hogares de la vivienda) D. Tenencia y financiación de la vivienda que ocupa el hogar E. Características y composición del hogar (para todas las personas del hogar) F. Salud (para todas las personas del hogar) G. Cuidado de los niños y niñas menores de cinco años H. Labores de los niños y niñas de 5 a 11 años I. Educación (para todas las personas de 5 años y más) J. Preferencias y hábitos sobre televisión (para todas las personas de 5 años y más) K. Capacitación para el trabajo (para todas las personas de 12 años y más) L. Fuerza de trabajo (para todas las personas de 12 años y más) M. Fuerza de trabajo (para todas las personas de 12 años y más) N. Gasto de los hogares
Cobertura geográfica y niveles de desagregación	Las regiones y los departamentos que la conforman son las siguientes: Antioquia; Atlántica (Guajira, Cesar, Magdalena, Atlántico, Bolívar, Sucre y Córdoba); Valle; Pacífica (Chocó, Cauca y Nariño); Central (Caldas, Quindío, Risaralda, Tolima, Huila y Caquetá); Oriental (Norte de Santander; Santander, Boyacá, Cundinamarca y Meta); Bogotá D.C.; Orinoquia - Amazonía (Arauca, Casanare, Vichada, Guainía, Guaviare, Vaupés, Amazonas y Putumayo); San Andrés y Providencia.

**Encuesta Nacional de la Situación Nutricional de Colombia (ENSIN) 2005
(Instituto Colombiano de Bienestar Familiar, 2006: 27-46)**

Las características relevantes de la Encuesta Nacional de la Situación Nutricional se enuncian en el Cuadro 1.13.

Cuadro 1.13 Caracterización de la Encuesta Nacional de Situación Nutricional en Colombia, año 2005

ENSIN-2005	
Objetivo	Establecer una línea de base de la situación nutricional de la población colombiana, con el fin de plantear políticas y programas de intervención.
Muestra	17.740 hogares de los 32 departamentos del país y de Bogotá, D.C.
Unidad de análisis	La encuesta se realizó a nivel de hogares
Módulos o componentes	<ol style="list-style-type: none"> 1. Valoración del estado nutricional por indicadores antropométricos 2. Valoración del estado nutricional por indicadores bioquímicos <ul style="list-style-type: none"> • Módulo de bioquímica 3. Lactancia materna y alimentación complementaria 4. Ingesta dietética <ul style="list-style-type: none"> • Recordatorio de 24 horas, módulo de consumo de alimentos 5. Evaluación cualitativa de la seguridad alimentaria en el hogar <ul style="list-style-type: none"> • Escala de percepción de seguridad alimentaria en el hogar 6. Actividad física <ul style="list-style-type: none"> • Módulo de salud y actividad física 7. Tiempo dedicado a ver televisión y a jugar con videojuegos 8. Autopercepción del peso corporal y conductas asociadas 9. Autoreporte de hipertensión arterial y diabetes mellitus
Cobertura geográfica y niveles de desagregación	Para la encuesta se estudiaron seis regiones y catorce subregiones

Estudio Nacional de Salud Mental (ENSM) 2003

Cuadro 1.14 Caracterización del Estudio Nacional de Salud Mental, año 2003

ENSM-2003	
Objetivo	Estimar la prevalencia y la edad de aparición de los trastornos mentales, e identificar sus factores asociados. Específicamente, dar respuesta a qué tan extendida está la enfermedad mental en el país, cuáles son los trastornos más importantes y quiénes los padecen, si existen variaciones regionales, cómo se hace uso de los servicios y cuánta discapacidad está asociada con los trastornos mentales.
Muestra	4544 adultos (18-65 años) provenientes de 5.526 hogares urbanos; concentrados en 420 manzanas de 60 municipios escogidos al azar de una muestra del Ministerio (1.070 municipios de 25 departamentos).

ENSM-2003

Unidad de análisis	La población civil, no institucionalizada, de las zonas urbanas, con edades comprendidas entre 18 y 65 años.
Módulos o componentes	Instrumento para adultos.
Cobertura geográfica y niveles de desagregación	Se llevó a cabo un estudio de prevalencia con una diseño probabilístico, multietápico, y estratificado a nivel nacional. Para tal fin el país se dividió en 5 regiones (Bogotá, Central, Atlántica, Oriental y Pacífica).

Segundo Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas (ENFREC) 1998

La situación de la salud en Colombia se encuentra atravesando una fase intermedia de transición, con predominio claro de las enfermedades crónicas y degenerativas, seguidas por las lesiones, las enfermedades transmisibles, las deficiencias nutricionales y las enfermedades del grupo materno infantil. Es importante, por lo tanto, conocer la distribución y tendencia de los factores de riesgo y los comportamientos y conocimientos que la población tiene sobre ellos, para incidir de manera efectiva sobre enfermedades como el cáncer y las cardio-cerebro-vasculares.

El propósito de esta encuesta es brindar una amplia información sobre los factores de riesgo de enfermedades crónicas. A continuación se describen sus principales características (Ministerio de la Protección Social, 1998: 19. ENFREC Tomo I módulo general).

Cuadro 1.15 Caracterización del Segundo Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas, año 1998

ENFREC-1998

Objetivo	Conocer la distribución y tendencia de los factores de riesgo, comportamientos y conocimientos que la población tiene de ellos.
Muestra	Las muestras de hogares descritos son submuestras aleatorias de la muestra maestra de la población de investigación de salud, disponible en el Ministerio de Salud desde 1998. Esta muestra estuvo conformada por 27.060 hogares.
Unidad de análisis	El universo de estudio estuvo constituido por población civil, no institucional y con edades entre 12 a 69 años, residentes en todos los departamentos del país, con excepción de las zonas rurales y las localidades no documentada cartográficamente de la Orinoquia y la Amazonía.

ENFREC-1998

Módulos/ Componentes	<ul style="list-style-type: none"> Formulario del hogar. Formulario individual par hombres de 18 a 69 años. Formulario individual para mujeres de 18 a 69 años. Formulario ampliado para adultos hombres y mujeres de 18 a 69 años. Formulario individual para adolescentes de 12 a 17 años. Formulario de evaluación clínica para niños de 5, 6,7 y 12 años. Formulario de listado de viviendas y hogares de segmento urbano. Formatos para la selección de personas.
Cobertura geográfica	Nivel Nacional
Niveles de desagregación	Departamental

Tercer Estudio Nacional de Salud Bucal (ENSAB) 1998

En 1997, el Ministerio de Salud lideró el III Estudio Nacional de Salud Bucal (ENSAB III), con el cual se buscaba establecer de manera certera las condiciones de salud bucal de la población y los avances alcanzados en materia de prevención y educación. En el marco del Sistema General de Seguridad Social en Salud, ENSAB III permitió identificar las necesidades de atención odontológica y, por ende, orientar la prestación de los servicios asistenciales; asimismo brindó información valiosa sobre las necesidades de recurso humano y del nivel de representaciones, conocimientos y prácticas de la población respecto a la salud bucal y sus determinantes para, así, orientar las acciones de promoción y prevención. El Cuadro 1.16 presenta la descripción de los aspectos relevantes de dicho estudio.

Cuadro 1.16 Caracterización del Tercer Estudio Nacional de Salud Bucal, año 1998

ENSAB III-1998	
Objetivo	Identificar la prevalencia y los conocimientos actitudes y prácticas de la población nacional con respecto a las enfermedades bucodentales, como caries, enfermedad periodontal y fluorosis, así como la demanda y utilización de servicios y las necesidades de atención relacionadas con la salud bucal de la población.
Muestra	4400 niños(as) y 8448 adultos
Unidad de análisis	Niños(as) de 5 a 7 y 12 años. Adultos de 15 a 44 y de 55 a 74 años.
Módulos/Componentes	Evaluación clínica para niños(as). Evaluación clínica para mayores de 14 a 74 años.
Cobertura Geográfica	Se desagregó en 5 regiones, 11 subregiones y 32 departamentos
Niveles de desagregación	Regiones-Subregiones: Central (Medellín, Resto), Atlántica (Barranquilla, San Andrés y Providencia, Resto), Oriental (Amazonía, Orinoquia y Resto), Occidental (Cali, Anden pacifico, resto), Bogotá.

Gran Encuesta Integrada de Hogares (GEIH) 2007

El Departamento Administrativo Nacional de Estadísticas (DANE) realizó durante varios años esta encuesta con el fin de obtener información de los hogares, la fuerza de trabajo, la inserción laboral y las oportunidades de empleo. Pero en 2007 surgió la necesidad de tener conocimiento de las personas ocupadas, desocupadas, entre otras, por lo que el DANE llevó a cabo la Gran Encuesta Integrada de Hogares que, con una serie de mejoras en la ejecución de la recolección de información, tuvo como finalidad proporcionar al país y a los diferentes entes investigativos, nacionales e internacionales, una información coherente y eficaz en materia de mercado laboral, base para la toma de decisiones.

Cuadro 1.17 Caracterización de la Gran Encuesta Integrada de Hogares, año 2007

GEIH-2007	
Objetivo	Proporcionar información básica sobre el tamaño y estructura de la fuerza de trabajo (empleo, desempleo e inactividad) de la población del país, así como de las características sociodemográficas de la población colombiana.
Muestra	Las 24 ciudades capitales con sus áreas metropolitanas.
Unidad de análisis	La población civil no institucional, residente en todo el territorio nacional.
Módulos/ Componentes	<ul style="list-style-type: none"> • Datos de la vivienda. • Datos del hogar. • Características generales. • Seguridad Social en Salud. • Educación. • Fuerza de trabajo. • Ocupados. <ul style="list-style-type: none"> • Asalariados. • Independientes • Asalariados e independientes. • Trabajo secundario. • Empleo con insuficiencia de horas y situaciones de empleo inadecuado. • Calidad del empleo principal • J. Desocupados. • K. Inactivos. • L. Otras actividades y ayudas en la semana. • M. Otros ingresos (para todas las • Personas de 10 años y mas) mes pasado.
Cobertura geográfica	Nivel Nacional
Niveles de desagregación	Departamental.

Registros continuos

Estadísticas vitales (EEVV)

Los registros vitales son estadísticas continuas que recogen información sobre nacimientos, defunciones, adopciones, matrimonios y separaciones, y permiten contar con un flujo de información continuo que revela los cambios ocurridos en los niveles y patrones de mortalidad, fecundidad y nupcialidad, a la vez que proporcionan una visión dinámica de la población, como complemento al enfoque estático que proveen los censos (DANE).

Los datos son recolectados por el DANE, con la asesoría de medicina legal, en formatos diseñados según normas internacionales y avalados por el Ministerio de la Protección Social. Definidos como fuentes de información fundamentales, son utilizados por el estado colombiano para establecer las políticas y determinar las prioridades en salud pública (Agudelo *et al*, 2005: 85-88).

El sistema de estadísticas vitales colombiano incluye los nacimientos y las defunciones fetales y no fetales, y estos eventos son diligenciados en los certificados de nacido vivo y de defunción. Se espera que en el futuro cercano se incorporen los demás hechos vitales (DANE).

El certificado de nacido vivo, o certificado médico de nacimiento, se expide para acreditar el hecho de un individuo nacido vivo (Ministerio de la Protección Social, 1997: 2-3), y en ellos se identifican variables importantes como: sexo, peso y talla al nacer, fecha y lugar de nacimiento, tipo y persona que atendió el parto y datos de los padres.

Los encargados de diligenciar este documento son: a nivel institucional el médico que asistió el parto o la enfermera; si el parto fue asistido en zonas rurales por parteras u otro tipo de personas, el certificado puede ser diligenciado, en orden jerárquico, desde el máximo nivel por el director de las IPS y en un nivel inferior por un promotor de la salud (Ministerio de la Protección Social, 1997: 2-3)

El certificado de nacido vivo se divide en tres partes o componentes, a saber: datos propios del nacimiento, datos de los padres del nacido vivo y datos de la persona que expide el certificado. A continuación se describen sus características principales (Ministerio de la Protección Social, 1997: 2-3)

Cuadro 1.18 Caracterización del certificado de nacido vivo

Certificado de nacido vivo	
Objetivo	Acreditar el hecho del individuo nacido vivo
Muestra	Para esta fuente se utilizan todos los certificados de nacido vivo a nivel nacional, suministrados por el Ministerio de la Protección Social de Colombia
Unidad de análisis	Bases de datos a nivel nacional suministradas por el Ministerio de la Protección Social de Colombia
Módulos/ Componentes	I. Datos del nacimiento II. Datos de los padres del nacido vivo - Datos de la madre del nacido vivo en el momento del parto - Datos del padre del nacido vivo III. Datos de quien certifica el nacimiento
Cobertura geográfica y niveles de desagregación	Todo el territorio colombiano

El certificado de defunción sirve como certificado oficial de la muerte del individuo –define, por lo tanto, la terminación de sus derechos civiles– y como fuente de información demográfica y soporte estadístico para el establecimiento de las políticas en salud pública.

En el certificado de defunción se incluyen variables como: sexo, edad, lugar y fecha, tipo de defunción (fetal o no fetal) y causas de la muerte (Agudelo *et al*, 2005: 85-88), y los encargados de diligenciarlo son: a nivel institucional el médico tratante o no tratante, pero si la defunción fue en un hecho externo (muerte violenta), el certificado es diligenciado por medicina legal en cabeza del médico forense.

El certificado de defunción es considerado un documento médico con información estadística de singular valor epidemiológico, médico-legal, jurídico y administrativo. Se divide en cinco capítulos, a saber: información general, defunciones fetales o de menores de un año, defunciones de mujeres en edad fértil, muertes violentas y causas de defunción. Es importante resaltar que la información general y las causas de defunción se diligencian para todas las defunciones. A continuación se describen los aspectos relevantes de este registro (Agudelo *et al*, 2005: 85-88).

Cuadro 1.19 Caracterización del certificado de defunción

Certificado de defunción	
Objetivo	Oficializar la muerte del individuo y definir la terminación de sus derechos civiles Ayudar a establecer políticas en salud pública brindando información demográfica y soporte estadístico
Muestra	Para esta fuente se utilizan todos los certificados de defunción a nivel nacional, suministrados por el Ministerio de la Protección Social de Colombia
Unidad de análisis	Bases de datos a nivel nacional suministradas por el Ministerio de la Protección Social de Colombia
Módulos/ Componentes	Capítulo I: Información general Capítulo II: Defunciones fetales o de menores de un año Capítulo III: Defunciones de mujeres en edad fértil Capítulo IV: Muertes violentas Capítulo V: Causas de defunción para todos los casos
Cobertura geográfica y niveles de desagregación	Todo el territorio colombiano

La calidad del dato en el sistema de estadísticas vitales está influenciada por la claridad de los formularios utilizados, el entrenamiento y supervisión de las personas encargadas de llenarlos y las estrategias de control de calidad establecidas en los procesos de digitación en las bases de datos (Álvarez y López, 2006: 19).

Para garantizar y apoyar la calidad de los datos que se requieren en estos certificados, el gobierno colombiano ha expedido, entre otras, la siguiente legislación: Ley 23 de 1981, Decreto número 1171 del 28 de abril de 1997, por el cual se reglamentaron los artículos 50 y 51 de la Ley 23 de 1981 y la Resolución número 1346 de 1997 (mayo 5). El 13 de noviembre del 2007 se expidió la Circular Externa Conjunta No.0081, que plantea algunos cambios en el certificado, entre los que vale la pena resaltar:

- La eliminación del uso del papel carbón.
- El original de cada certificado se denominará “certificado de nacido vivo antecedente para el registro civil” y “certificado de defunción antecedente para el registro civil”.
- El número del certificado se presenta mediante un rótulo adhesivo.

- Se incrementó el número de preguntas de cada certificado.
- Se modificó el orden de las preguntas.
- La redacción de las preguntas se cambió, lo mismo que la presentación de las opciones de respuesta. (Ministerio de la prorección social, 2007: 2)

Registros Individuales de Prestación de Servicios (RIPS)

Es el conjunto de datos mínimos y básicos que el Sistema General de Seguridad Social en Salud requiere para los procesos de dirección, regulación y control, y como soporte de la venta de servicio, cuya denominación, estructura y características se ha unificado y estandarizado para todas las entidades de salud. Los datos de este registro se refieren a la identificación del prestador del servicio de salud, del usuario que lo recibe, de la prestación del servicio y el motivo que originó su prestación: diagnóstico y causa externa (Ministerio de la Protección Social, 2000: 2).

Cuadro 1.20 Caracterización del Registro Individual de Prestación de Servicios

RIPS	
Objetivo	Suministrar los datos mínimos y básicos que el Sistema General de Seguridad Social en Salud, requiere para los procesos de dirección, regulación y control, y como soporte de la venta de servicio.
Muestra	Muestra de RIPS de la Encuesta Nacional de Salud y RIPS nacionales entre 2001 y 2007
Unidad de análisis	Bases de datos suministradas por el Ministerio de la Protección Social de Colombia
Módulos/ Componentes	CT=Archivo de control AF=Archivo de transacciones US=Archivo de usuarios AD=Archivo de descripción agrupada de los servicios de salud prestados AC=Archivo de consulta AP=Archivo de procedimientos AH=Archivo de hospitalización AU=Archivo de urgencias AN=Archivo de recién nacidos AM=Archivo de medicamentos AT=Archivo de otros servicios
Cobertura geográfica y niveles de desagregación	Muestra de la ENS y total nacional de los años 2001 a 2007

Sistema de Vigilancia en Salud Pública (SIVIGILA) 2003-2007

Con el fin de cubrir los vacíos de información en vigilancia y control en salud pública, tanto en el orden departamental como municipal, el Ministerio de la Protección Social creó el Sistema de Vigilancia Epidemiológica (SIVIGILA) cuyo objetivo es facilitar la generación de datos y el intercambio de información entre las Unidades Primarias Generadoras del Dato (UPGD) –municipios, departamentos, distritos y el Instituto Nacional de Salud (INS)– sobre los eventos de interés en salud pública para su continua vigilancia y control.

Las UPGD (esto es, toda persona natural o jurídica relacionada con la prestación de servicios de salud en la atención a enfermedades inmuno o no inmuno-prevenibles) son responsables del reporte inicial al ente territorial, bien sea de manera colectiva –cuando se presenta el mismo evento en el mismo periodo y semana epidemiológica y en diferentes personas– o básica –cuando en el periodo y semana epidemiológica solo una persona notifica el evento ante el ente territorial–. Existen eventos que exigen ser notificados de manera inmediata, los que no deben seguir el siguiente ciclo: la UPGD notificara la presencia de eventos a la Dirección Local de Salud (DLS), ésta a la Dirección Departamental de Salud la que notifica al Instituto Nacional de Salud, una vez a la semana.

Cuadro 1.21 Caracterización del Sistema de Vigilancia en Salud Pública

SIVIGILA, 2003-2007	
Objetivo	Facilitar la generación de datos y el intercambio de información entre las unidades generadoras del dato (UPGD), los municipios, departamentos y distritos y el Instituto Nacional de Salud (INS) acerca del reporte de eventos de interés en salud pública, para su continua vigilancia y control con el propósito de satisfacer las necesidades de información del SIVIGILA.
Muestra	Total de notificaciones hechas por las Unidades Primarias Generadoras del Dato (UPGD), los municipios, departamentos, distritos y el Instituto Nacional de Salud (INS) en un periodo epidemiológico determinado.
Unidad de análisis	Notificadoras de eventos infecciosos y/o parasitarios que ponen en riesgo la salud pública.
Módulos/Componentes	<ul style="list-style-type: none"> • FICHA INDIVIDUAL Información general. Identificación del paciente. Notificación. Espacios territoriales. <ul style="list-style-type: none"> • FICHA COLECTIVA Información general. Caracterización. Información complementaria.
Cobertura geográfica	Nivel Nacional.
Niveles de desagregación	Unidad Primaria Generadora del Dato (UPGD).

Censo de Población y Vivienda (CENSO) 2005

Los Censos Generales de Población constituyen una actividad de investigación dirigida a recabar, procesar, analizar y difundir información básica para contribuir al desarrollo económico, social, científico y tecnológico de un país.

Los censos son la fuente primaria más importante y amplia de información estadística. Sus resultados suministran antecedentes básicos referentes a la población, a las viviendas, a los hogares, y a las personas, necesarios para el diseño e implementación de políticas y programas públicos y para la decisión de inversiones y estudios privados (Tacla, 2006: 7).

El Censo General realizado en Colombia en 2005 ha hecho importantes innovaciones que han permitido mejorar la calidad de la información recolectada, reducir los costos y lograr una mayor oportunidad en la publicación de los resultados. La actualización de la cartografía, la introducción de nuevas tecnologías para el levantamiento de la información, la implantación de nuevos modelos de gestión de personal y el desarrollo de programas automatizados de edición e imputación han resultado en una mejora considerable de la calidad y oportunidad de la información sobre la población colombiana (DANE, 2005: 2).

Cuadro 1.22 Caracterización de Censo de Población y Vivienda, año 2005

CENSO 2005	
Objetivo	Realizar un censo de tipo general que permita disponer de información precisa, oportuna, confiable e integrada sobre el volumen y composición de la población los hogares y las viviendas, así como los marcos censales básicos de los establecimientos económicos y las unidades agropecuarias.
Muestra	Aproximadamente 1.300.000 hogares en el país.
Unidad de análisis	Todas las viviendas, todos los hogares, las personas y unidades económicas que residen u operan en el territorio nacional. También se tuvieron en cuenta las zonas rurales, los predios con vivienda asociadas a ellos.
Módulos/Componentes	Entorno Urbanístico, unidades económicas: Módulo de hogares, Módulo de personas, Módulo de viviendas, Módulo de unidades agropecuarias. Lugares especiales de alojamiento: Módulo de instituciones y el Módulo de personas.
Cobertura geográfica y niveles de desagregación	Alcance nacional con información desagregada hasta el nivel de predio.

Con base en los resultados arrojados por el DANE, puede afirmarse que el censo de 2005 tuvo una calidad satisfactoria de información y muy superior cuando se la compara con los censos anteriores de población y vivienda que no cubrían espacios de información valiosa para el país. En efecto, de acuerdo con los resultados de la conciliación final, el Censo de 2005 tuvo una no cobertura de 3,7% y de este porcentaje, el 1,4% se debió principalmente a dificultades de acceso

a los hogares por razones de violencia; el 2,3% restante corresponde a la no cobertura final de unidades censales.

OBJETIVOS

Objetivo general

Gestionar la calidad de los datos de las fuentes proporcionadas por el Ministerio de la Protección Social (ENDS-2005, ENS-2007, ECV-2003, GEIH 2007, ENH 1995 y 2000, ENSIN-2005, ENSM-2003, ENFREC-1998, ENSAB-1998, Censo General 2005, RIPS, SIVIGILA y Estadísticas Vitales) para el análisis de la situación de salud en Colombia, 2002-2007.

Objetivos específicos

- Documentar los metadatos de todas las fuentes de información utilizadas en el análisis de la situación de salud en Colombia (ASIS), 2002-2007.
- Validar la calidad de los microdatos contenidos en las fuentes de información disponibles, con el fin de determinar su consistencia, presencia y cobertura, entre otros.
- Estimar métricas de calidad de los datos a partir de los errores e inconsistencias, mediante la distribución de frecuencias, medidas de tendencia y de variabilidad.
- Realizar la limpieza y depuración de los microdatos de las fuentes de información utilizadas para el ASIS 2002-2007.
- Transformar los datos de las variables necesarias mediante procesos de clasificación, codificación, recodificación y/o postcodificación.
- Documentar todos los procesos y procedimientos utilizados en la ejecución del proyecto.
- Describir el estado de la calidad de los microdatos de todas las fuentes de información utilizadas en el Análisis de la Situación de Salud en Colombia (ASIS), 2002-2007.

CAPÍTULO 2:



**DISEÑO
METODOLÓGICO**

Tipo de estudio

Se realizó un estudio descriptivo, que permitió evaluar la calidad de los datos recolectados en cada una de las fuentes disponibles para el Análisis de la Situación de Salud en Colombia entre los años 2002 y 2007.

Población y fuentes de datos

Para el análisis de la situación de salud del país, el Ministerio de la Protección Social suministró fuentes de datos secundarias que se pueden clasificar en registros continuos y encuestas poblacionales. Estas fuentes fueron:

Encuestas poblacionales:

- Encuesta Nacional de Salud (ENS), año 2007
- Encuesta Nacional de Demografía y Salud (ENDS), año 2005
- Encuesta de Calidad de Vida (ECV), año 2003
- Encuesta Nacional de Situación Nutricional en Colombia (ENSIN), año 2005
- Estudio Nacional de Salud Mental en Colombia(ENSM), año 2003
- Segundo Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas (ENFREC), año 1998
- Tercer Estudio Nacional de Salud Bucal (ENSAB), año 1998
- Gran Encuesta Integrada de Hogares (GEIH), año 2007
- Encuesta Nacional de Hogares (ENH), entre 1995 y 2000

Registros continuos:

- Registros de Estadísticas Vitales: defunciones entre 1973 y 2006; defunciones fetales entre 1998 y 2006; nacimientos entre 1998 y 2006

- Registro Individual de Prestación de Servicios (RIPS) a nivel nacional, año 2007
- Sistema de Vigilancia en Salud Pública (SIVIGILA), años 2003 a 2007
- Censo General de Población y Vivienda, año 2005

La población de estudio estuvo conformada por el conjunto de datos asociado a cada componente o módulo de las fuentes anteriormente mencionadas; o lo que es lo mismo, la población de estudio corresponde al número de registros por componente de la fuente de datos asociado a cada unidad de análisis.

Fuentes de datos complementarias

Los microdatos de las fuentes anteriores fueron objeto de estudio, análisis y evaluación para asimilar sus contenidos; durante este proceso fue necesario acceder a otras fuentes de información complementaria, tales como: documentos metodológicos, instrumentos, instructivos, informes de los operativos de campo y de los resultados generados por cada encuesta o proceso investigativo, informes sobre errores en las mediciones y de evaluación de las limitaciones identificadas en la información obtenida, documentación informática correspondiente a los algoritmos, diccionarios de datos y metadatos, entrevistas y reuniones con los participantes en los proyectos, revisión de resúmenes en video –en particular, la ENDS– y la exploración de la encuesta co-censal (para el Censo).

Indicadores de calidad seleccionados

Para la medición de la calidad de los datos se hizo una selección exhaustiva de las dimensiones y atributos de la calidad referenciados por el Fondo Monetario Internacional y la Organización Internacional de Estándares (ISO). Los criterios de selección de estos indicadores fueron la exigibilidad, la necesidad técnica y la factibilidad. A continuación se documenta la información básica sobre cada indicador incluido.

Cuadro 2.1 Indicadores seleccionados para evaluar la calidad de los datos

Dimensión	Nombre del indicador	Descripción	Unidad de análisis	Indicador
Documentación de la fuente	Documento de metadatos	Se dispone fácilmente de diccionarios de datos y/o metadatos y se brinda asistencia adecuada a los usuarios. Se dispone de documentación sobre los conceptos, el alcance, las clasificaciones, la base de registro, las fuentes de datos y las técnicas estadísticas, y se señalan las discrepancias con respecto a las normas, directrices o buenas prácticas de aceptación internacional. *	Fuente	$\frac{\text{No. de fuentes con metadatos}}{\text{Total de fuentes}}$
	Documento de diccionarios de datos			$\frac{\text{No. de fuentes con diccionarios de datos}}{\text{Total de fuentes}}$
Cobertura	Cobertura	Es el grado de observación y/o registro de datos obtenido y/o almacenado de una unidad de análisis de un estudio de una población o muestra determinada *****	Hogares Viviendas Personas IPS Usuarios Entidades territoriales	$\frac{\text{Total unidad de análisis-UA estudiada} \times 100}{\text{No. UA calculada en la muestra}}$
	Error de cobertura	Los errores de cobertura son los provenientes de la insuficiencia en la contabilidad de las personas, hechos o cosas. Caso en el cual las personas o los hechos que se desean investigar son omitidos o sobre-registrados (errores por defecto o por exceso). *****	Hogares Viviendas Personas	$\frac{\text{UA calculada en la muestra} - \text{UA estudiada} \times 100}{\text{UA calculada en la muestra}}$
	Declaración de la edad	Método de Whipple: Preferencia de dígitos 0 y 5. Método de Myers: Atracción de los dígitos finales	Personas	$\begin{aligned} & * \text{ Método Whipple} \\ & * \text{ Método Myers} \end{aligned}$
	No respuesta **	Algunos de los datos de los atributos carecen de valores	Atributos	$\frac{\text{Nº de datos con no respuesta para cada atributo}}{\text{Total de datos válidos para el atributo}}$
Contenido	Sobre respuesta **	El total de datos es superior al número de datos válidos para el atributo **	Atributos	$\frac{\text{Nº de atributos con ausencia de alguno de sus datos}}{\text{Total de atributos}}$
				$\frac{\text{Nº de datos que supera los datos válidos del atributo}}{\text{Total de datos válidos del atributo}}$
				$\frac{\text{Nº de atributos con ausencia de alguno de sus datos}}{\text{Total de atributos}}$

Dimensión	Nombre del indicador	Descripción	Unidad de análisis	Indicador
Contenido	Consistencia	Indica la presencia de datos aceptables para el campo al que corresponden y estén registrados, de acuerdo a la definición hecha por la fuente establecida previamente y especificada en los diccionarios de datos. Es importante anotar que un valor válido de un campo no necesariamente significa un valor exacto o correcto **	Atributos	$\frac{\text{Nº de datos consistentes}}{\text{Total de datos válidos para el atributo}}$
	Presencia	Grado de coincidencia entre cada uno de los atributos registrados con respecto a cada uno de los atributos solicitados en los instrumentos de las fuentes. ****	Atributos	$\frac{\text{Nº de atributos registrados}}{\text{Total de atributos en el Instrumento}}$
Relevancia		Los atributos críticos para el análisis deben tener algún valor ***	Atributos	$\frac{\text{Nº de atributos con datos}}{\text{Total de atributos}}$
				$\frac{\text{Nº de registros con valores duplicados}}{\text{Total de registros}}$
Fiabilidad	Duplicidad	El registro tiene representación duplicada del mismo objeto ***	Registros	$\frac{\text{Nº de registros con valores duplicados iguales}}{\text{Total de registros}}$
				$\frac{\text{Nº de registros con valores duplicados diferentes}}{\text{Total de registros}}$
Complejidad		El grado en el que el dato asociado a un sujeto con una entidad tiene todos los valores esperados para el atributo y para las instancias de la entidad relacionadas en un contexto específico de uso ***	Atributos	$\frac{\text{Nº de atributos con la totalidad de sus datos}}{\text{Total de atributos}}$
	Casos ponderados	Es la frecuencia absoluta observada y corregida en la muestra en cada categoría de la variable estudiada, tomando en cuenta los casos representados de la población por cada individuo estudiado, es decir, toma en consideración el factor de expansión	Atributos críticos para el análisis	fi corregido

Dimensión	Nombre del indicador	Descripción	Unidad de análisis	Indicador
	Casos sin ponderar	Es la frecuencia absoluta observada en la muestra en cada categoría de la variable estudiada		f_i
	Valor estimado	Según la naturaleza de la variable y el interés del observador puede ser una proporción, una media, una varianza		$p = \frac{f_i}{n}$ $x = \frac{\sum w_i f_i}{n}$
	Error estándar (EE)	El error estándar de un promedio, porcentaje, diferencia o cualquier otra estadística calculada con los datos de la muestra se define como la raíz cuadrada de la varianza ($\text{var}(r)$) de la estadística, y es una medida de su variación en todas las muestras posibles		$EE = \sqrt{\text{var}(r)}$
Representatividad	Efecto de diseño (EDIS)	Es la razón entre el error estándar (EE) correspondiente al diseño empleado y el error estándar que se obtiene tratando la muestra como si hubiera sido aleatoria simple (EE _{mas})	Atributos críticos para el análisis	$EDIS = \frac{EE}{EE_{mas}}$
	Error relativo (ER)	Para valorar o calificar la magnitud de un error se piensa, tácita o explícitamente, en términos relativos. Se calcula dividiendo el error estándar entre el error del parámetro (estadístico poblacional)		$er = \frac{e}{p}$ (parámetro)
	Intervalo de confianza	Intervalos dentro de los cuales hay una determinada confianza de que se encuentra el valor poblacional		$p \pm z^* \sqrt{p^*(1-p)/n}$
	Factor de expansión (Fv)	Es el recíproco de la probabilidad total, donde ésta, es el producto de las probabilidades parciales de cada una de las etapas del proceso de selección		$Fv = \frac{1}{P_t}$

*Definición fondo monetario

**Definición adaptada para el proyecto

***Definición ISO

****Exigencia contractual

Para efectos de estandarizar las operaciones, se elaboró una guía metodológica con las instrucciones para el cálculo de cada uno de los indicadores (ver Anexo 2), la cual contiene la documentación de los procesos de validación de la calidad de los microdatos y la estimación de las métricas de calidad, depuración y limpieza y de transformación de datos.

Descripción de los procesos

El Gráfico 2.1 presenta el esquema de los procesos ejecutados para caracterizar los datos disponibles según las fuentes de datos proporcionadas por el Ministerio de la Protección Social para la realización del análisis de la situación de salud en Colombia entre los años 2002 y 2007.

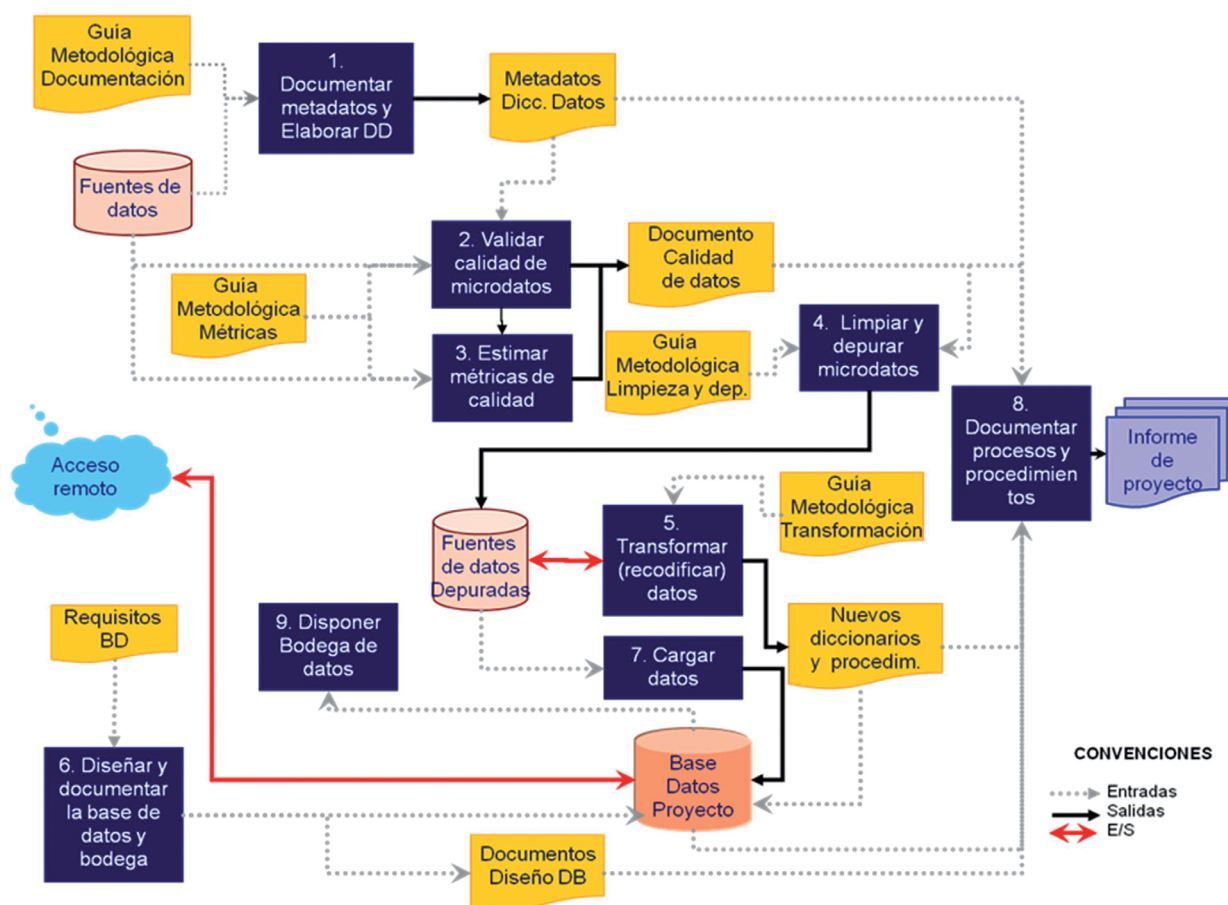


Gráfico 2.1. Esquema de procesos para la caracterización de los datos disponibles para la realización del análisis de la situación de salud en Colombia, 2002-2007

Los procesos se manejaron de manera industrializada. Al inicio el equipo de trabajo desarrolló dos guías de trabajo: una para la documentación de metadatos y diccionarios y otra para las

métricas de evaluación de la calidad. Estas guías se convirtieron en el insumo principal para el desarrollo de las actividades de evaluación de la calidad de los microdatos del proyecto.

En el siguiente grafico se ilustra la sincronización de actividades requerida para los procesos de evaluación de la calidad de los microdatos.

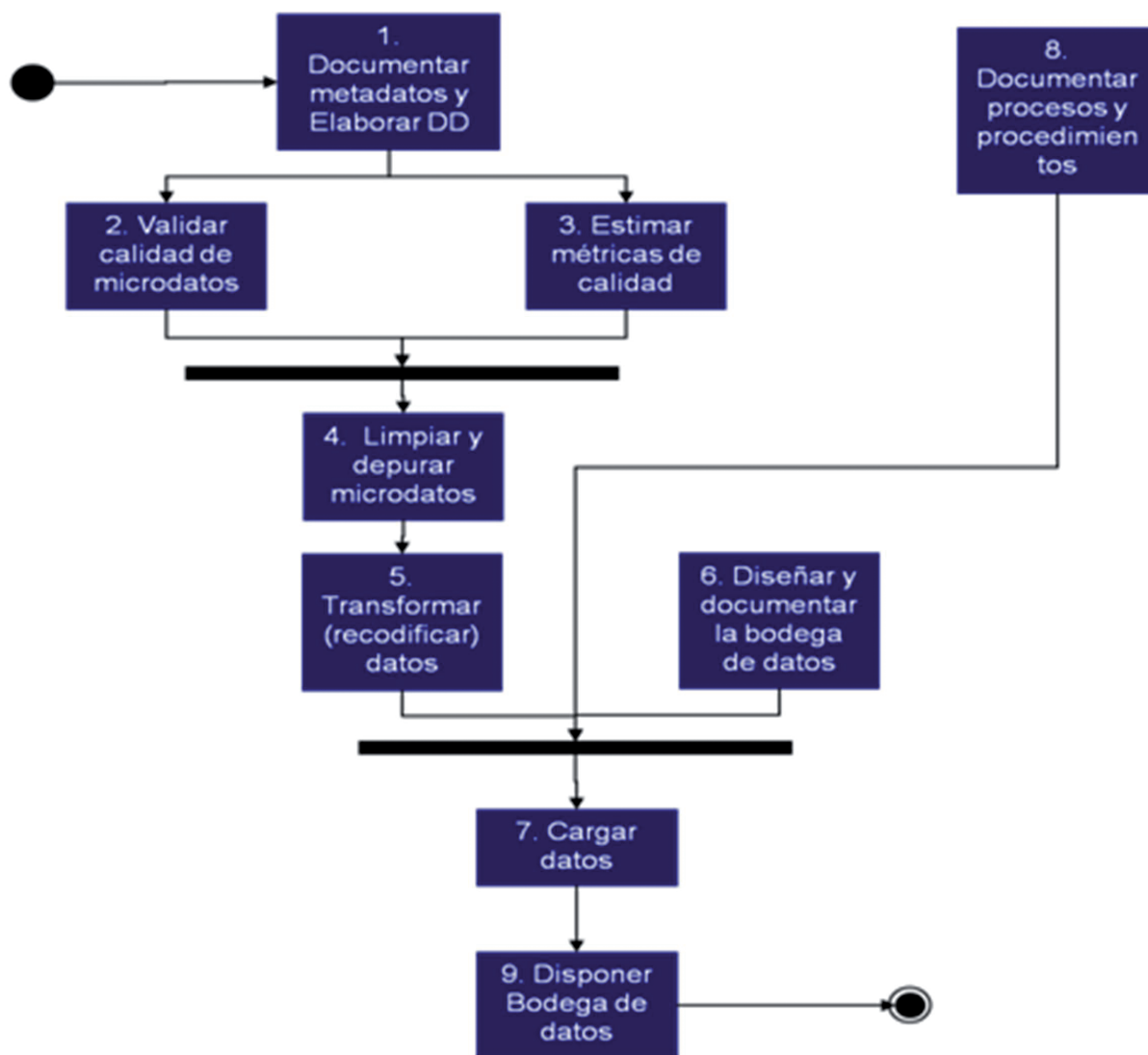


Gráfico 2.2. Diagrama de actividades para la caracterización de los datos disponibles para la realización del análisis de la situación de salud en Colombia, 2002-2007

Todos los procesos referidos (documentar los metadatos y elaborar diccionarios de datos, validar la calidad de los microdatos, estimar las métricas de calidad, limpiar y depurar microdatos, transformar datos, diseñar y documentar la base de datos integrada, cargar datos, documentar procesos y procedimientos y disponer la bodega de datos) fueron documentados y diagramados

siguiendo el estándar ANSI (American National Standards Institute/ Instituto Nacional Estadounidense de Estandarización), lo que permitió examinar y estandarizar al detalle las actividades realizadas, especificándolas a nivel de tareas (en el Anexo 1 se puede ver un ejemplo de cómo se documentaron los procesos).

Recursos

Talento humano

Para caracterizar los datos disponibles se contó con un comité técnico conformado por docentes y profesionales en Gerencia de Sistemas de Información en Salud; se contó también con un grupo de estudiantes en práctica de la Tecnología en Sistemas de Información en Salud, que fue capacitado en el manejo básico del software estadístico SPSS, herramienta seleccionada para la exploración y medición de las características de calidad de los datos.

Recurso físico

Se dispuso un espacio de trabajo acondicionado con computadores personales conectados en red, y de un servidor para el desarrollo y pruebas de las aplicaciones de software.

Recursos de datos

El Ministerio de la Protección Social entregó matrices de datos de las fuentes de información a caracterizar; sus diferentes formatos se describen en los metadatos de cada fuente.

Recursos de software

Se contó con las licencias y el software requerido para la evaluación de la calidad de los datos, entre ellos cabe destacar: SPSS, STATA, SQL Server, Epidat, Visual studio 2008 y herramientas de ofimática.

Consideraciones éticas

Para la utilización de las fuentes de información proporcionadas por el Ministerio de la Protección Social se suscribió con el Centro de Investigación de la Facultad Nacional de Salud Pública, un acta de confidencialidad donde se asumieron los siguientes compromisos:

- Garantizar la confidencialidad y sigilo de los datos.
- Hacerse responsable del acceso y manejo de las bases de datos.
- No realizar ninguna copia adicional de las bases de datos.
- Hacer devolución de las bases de datos usadas, al Centro de Investigación al término del proyecto.
- Informar oportunamente al Centro de Investigación y al coordinador del proyecto, la pérdida o daño de las bases de datos usadas.
- No divulgar, sin autorización previa del coordinador del proyecto, ninguna información que se derive del análisis.
- No hacer uso de la información con propósitos distintos al aprobado en el contrato.
- Hacer explícito reconocimiento al Ministerio de la Protección Social en la presentación de los resultados y en cualquier publicación derivada del análisis.

El acta de compromiso fue firmada por todos los participantes en el proyecto que requirieron el acceso directo a los datos.

CAPÍTULO 3:

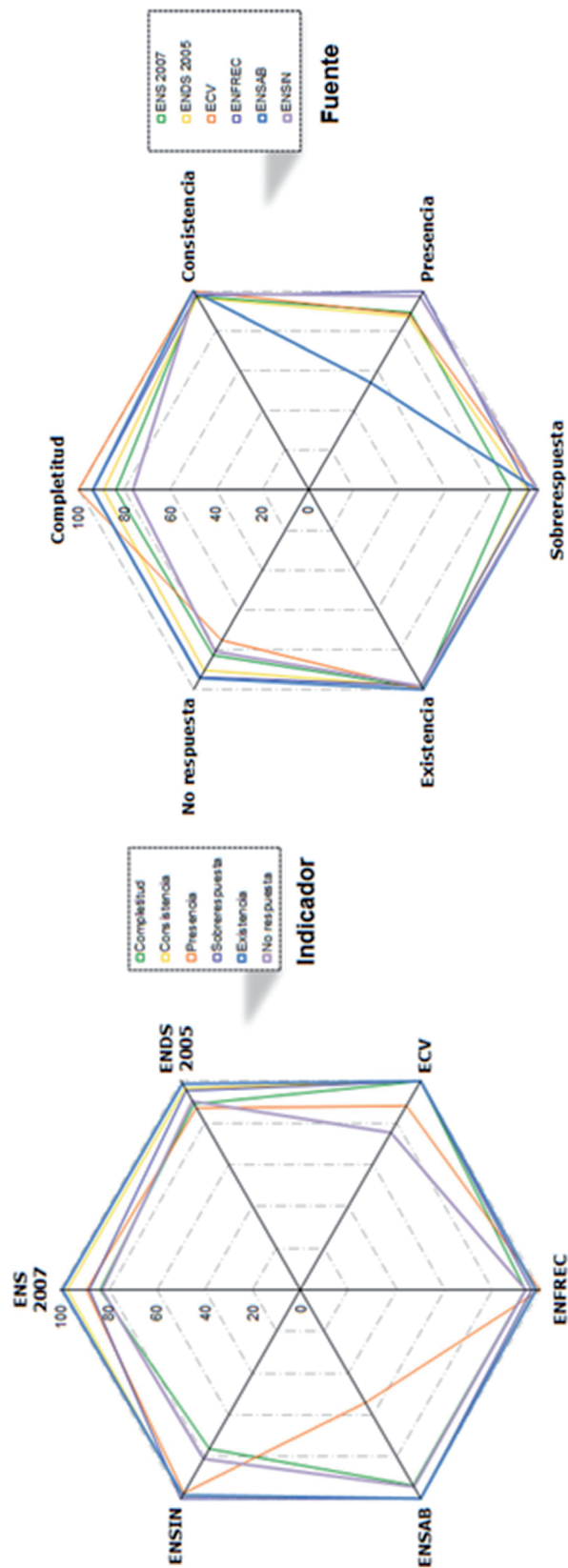
RESULTADOS

A continuación se describen los principales hallazgos, obtenidos al evaluar la calidad de los datos de todas las fuentes de información proporcionadas por el Ministerio de la Protección Social para el análisis de la situación de salud del país. Es importante resaltar que los informes detallados de cada fuente se publicarán en el sitio web del análisis de la situación de la salud en Colombia. A continuación se presentan, de manera resumida, los principales resultados del proceso de evaluación de la calidad de datos.

Para las encuestas se encontró que a pesar de presentar los datos algunos defectos de calidad, pueden ser utilizados para producir la información requerida para el análisis de la situación de salud en Colombia, o para cualquier otro tipo de análisis. Es importante resaltar que los valores obtenidos para los indicadores se consultaron y validaron previamente, por lo cual se convierten en el insumo fundamental para definir los procesos de transformación y ajuste de acuerdo con los supuestos y condiciones exigidos para la aplicación de diferentes técnicas de análisis y explotación de datos.

Para la evaluación de la calidad desde el punto de vista de sus dimensiones, se estimaron los siguientes indicadores: acceso a la documentación de la fuente, cobertura, contenido, fiabilidad y representatividad. En el Gráfico 3.1 se presenta un análisis general para los indicadores de contenido y fiabilidad; en él se puede observar que los indicadores más afectados son los de presencia, completitud y no respuesta para las fuentes ENSAB, ENSIN y ECV, y los mejor valorados son la consistencia y la existencia; también se evidencia que la fuente con mayor calidad, de acuerdo con los valores obtenidos para sus indicadores, es la ENFREC.

Para los registros continuos, la fuente más afectada fue Nacimientos; y al hacer el análisis por indicador, los problemas más comunes son de completitud y no respuesta. En el Gráfico 3.2 se puede observar el comportamiento general por fuente y por cada indicador de calidad.



Convencciones: ENS: Encuesta Nacional de Salud, ENDS: Encuesta Nacional de Demografía y Salud, ECV: Encuesta de Calidad de Vida, ENFREC: Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas, ENSAB: Encuesta Nacional de Salud Bucal, ENSIN: Encuesta Nacional de la Situación Nutricional en Colombia-Módulo ingesta.

Gráfico 3.1. Calidad de los datos para las encuestas según fuente e indicador. ASIS 2002-2007

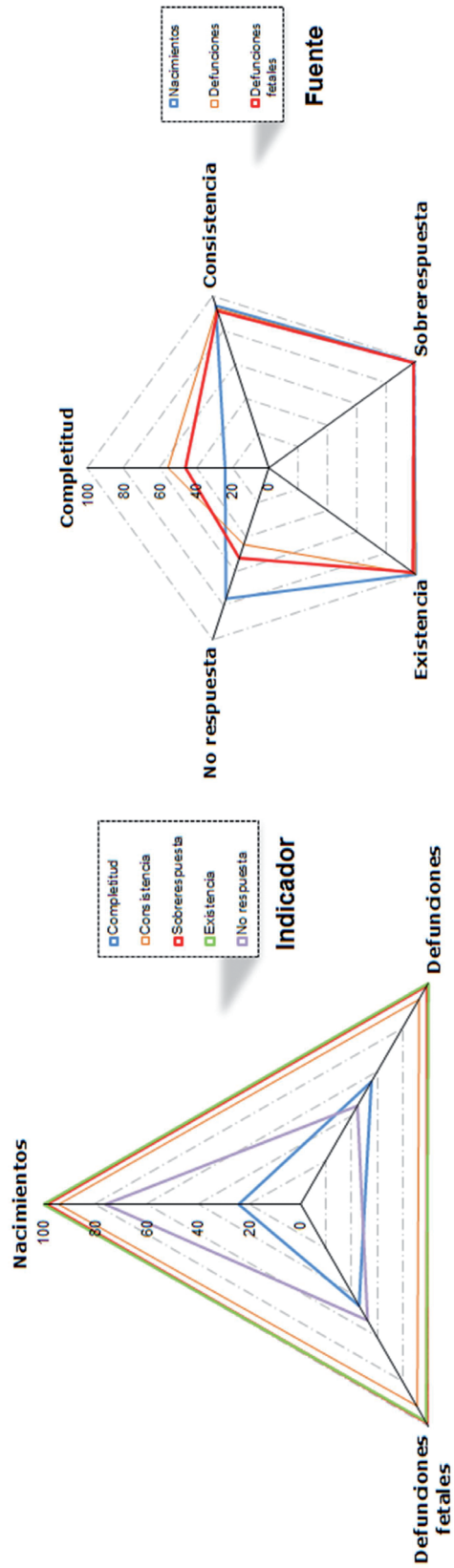


Gráfico 3.2. Calidad de los datos para los registros continuos de estadísticas vitales según fuente e indicador. ASIS 2002-2007.

Según los resultados, es importante destacar que para las encuestas y los registros continuos de estadísticas vitales, los indicadores más afectados fueron no respuesta y completitud y, de manera general, las fuentes más afectadas fueron Nacimientos y la ENSAB.

A continuación se presentan los resultados obtenidos para los indicadores en cada dimensión de la calidad evaluada y para cada una de las fuentes sometida a la evaluación; es de anotar que no todos los indicadores se estimaron para todas fuentes e, inclusive, que fuentes como ENSIN, ENH y GEIH no se evaluaron completamente.

Indicadores de documentación de la fuente

En el Cuadro 3.1 se presenta un resumen del total de registros, atributos y atributos coincidentes con los instrumento sometidos a la evaluación de calidad. Cabe observar que el número de registros representa la sumatoria de registros para todas las unidades de análisis de cada fuente, también que el indicador de acceso a los metadatos fue del 28,5%, es decir de 14 fuentes de información proporcionadas por el Ministerio de la Protección Social, 4 tenían como documentación anexa metadatos; por otra parte, el indicador de acceso a los diccionarios de datos y a otras fuentes de documentación fue de 78,5%. El hecho de que una fuente no incluya toda la documentación requerida, dificulta la comprensión sobre su contenido y no permite evidenciar los procesos de control que se llevaron a cabo para poblarla y garantizar la calidad de sus datos.

Cuadro 3.1 Total de registros y atributos de las fuentes evaluadas. ASIS 2002-2007

Fuente	Número total		Atributos coincidentes con el instrumento*	Indicador acceso a metadatos, diccionarios de datos y otra documentación asociada		
	Registros	Atributos		Metadatos	Diccionario de datos	Otra documentación asociada**
Encuestas						
ENS 2007	518.831	4.326	4.067	✓	✓	✓
ENDS 2005	372.801	1.968	1.602		✓	✓
ECV	2.000.325	741	676		✓	✓
ENFREC	36.849	555	490	✓	✓	✓
ENSAB	9747	233	126	✓	✓	✓
ENH	512.117	766	518			
GEIH					✓	✓
ENSIN				✓	✓	✓
ENSM						

Fuente	Número total		Atributos coincidentes con el instrumento*	Indicador acceso a metadatos, diccionarios de datos y otra documentación asociada		
	Registros	Atributos		Metadatos	Diccionario de datos	Otra documentación asociada**
Registros Continuos						
SIVIGILA	438.555	228	217			
Nacimientos	6.513.005	357	353		✓	✓
Defunciones	5.357.442	1.685	914		✓	✓
Defunciones fetales	168.331	536	503		✓	✓
RIPS						
Indicador de Acceso				28.5%	78.5%	78.5%

*Los atributos coincidentes con el instrumento de captura fueron los sometidos al proceso de evaluación de calidad de los datos.

**Hace referencia a los instructivos, formularios, documentos del operativo de campo, documentos de auditoría y la documentación del modelo de datos entre otros.

Convenciones: ENS: Encuesta Nacional de Salud, ENDS: Encuesta Nacional de Demografía y Salud, ECV: Encuesta de Calidad de Vida, ENFREC: Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas, ENSAB: Estudio Nacional de Salud Bucal, ENH: Encuesta Nacional de Hogares, GEIH: Gran Encuesta Integrada de Hogares, ENSIN: Encuesta Nacional de la Situación Nutricional en Colombia, ENSM: Estudio Nacional de Salud Mental, SIVIGILA: Software del sistema de vigilancia en salud pública, RIPS: Registros Individuales de Prestación de Servicios.

Indicador de cobertura

Con respecto a la cobertura o rendimiento de la muestra, la mayoría de las fuentes superaron el 80%. Las fuentes que tuvieron un rendimiento más bajo fueron la ENFREC y la ENSAB, con un valor mínimo de 75,7% para los adultos de 18 a 69 años en la ENFREC y de 72,83% para los adultos de la ENSAB.

Cuadro 3.2 Indicadores de cobertura por fuente de información. ASIS 2002-2007

Fuente	Unidad de análisis	Población			Indicador de cobertura (%)	
		Calculada	Estudiada	Registrada	Estudiado / Calculado*	Registrado / Calculado**
	Personas	164.474	166.474	166.474	101,2	101,2
	Hogares	41.543	41.833	41.543	100,7	100
ENS	IPS	1.472	1.222	1.170	83	79,5
	Usuarios IPS	123.917	124.141	123.917	100,2	100
	Administraciones municipales	238	225	238	94,5	100

Fuente	Unidad de análisis	Población			Indicador de cobertura (%)	
		Calculada	Estudiada	Registrada	Estudiado / Calculado*	Registrado / Calculado**
ENSIN	Ingesta	42.887	39.413	43.386	91,9	101,1
ENDS	Hogares	42.033	37.211	37.865	88,5	90,1
	Mujeres	41.552	38.143	42.156	91,8	101,5
ENFREC	Adultos 18 a 69 años	5.280	3.999	3.999	75,7	75,7
	Adolescentes 12 a 17 años	1.056	823	823	77,9	77,9
	Hogares	5841	9302	5781	159,3	99,0
ENSAB	Niños	4.400	3.594	3.594	81,6	81,6
	Adultos	8.448	6.153	6.153	72,8	72,8

*Corresponde el cálculo de la población que referencian los informes publicados dividida por la población del diseño muestral

** Es el cálculo entre la población obtenida de los archivos de datos dividida la población del diseño muestral

Convenciones: ENS: Encuesta Nacional de Salud, ENDS: Encuesta Nacional de Demografía y Salud, ENFREC: Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas, ENSAB: Estudio Nacional de Salud Bucal, ENSIN: Encuesta Nacional de la Situación Nutricional en Colombia

Es importante destacar que para los registros continuos no aplica este indicador y por lo tanto, no fue estimado.

Indicadores de contenido

A continuación se presentan los resultados obtenidos para los indicadores de cuatro de los aspectos que permiten la medición del contenido de los conjuntos de datos almacenados en las fuentes; estos aspectos son: declaración de la edad, no respuesta, sobre-respuesta y consistencia.

Declaración de la edad

Este indicador fue estimado para las encuestas. En el Gráfico 3.3 se observa que tanto en los hombres como en las mujeres –sobre todo en éstas– hay una menor preferencia por los dígitos 1 y 9, y que el dígito preferido fue el 5. Por otra parte, y según el índice de Whipple, se observa que los datos de la mayoría de las encuestas se consideran “datos aproximados”, excepto para la Encuesta Nacional de Salud en la que pueden considerarse como “datos relativamente precisos”.

El índice de Myers permitió evaluar la atracción o rechazo, presente en la declaración de la edad, en los dígitos finales y, en la mayoría de los casos, hizo evidente una atracción “mediana” por los dígitos finales para ambos sexos, excepto para los hombres de la ENDS cuyo índice fue de 4,9, lo que permite evidenciar una atracción “baja” por los dígitos finales.

Fuente	Myers	Whipple	
		Hombres	Mujeres
ECV 112 8.9-7.7			
ENDS 111 4.9-8.5			
ENFREC 111 7.4-7.8			
ENS HOGAR 109 6.2-6.6			
ENSIN IN- GESTA 111 11.7-12			

Convenciones: ENS: Encuesta Nacional de Salud, ENDS: Encuesta Nacional de Demografía y Salud, ECV: Encuesta de Calidad de Vida, ENFREC: Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas, ENSIN: Encuesta Nacional de la Situación Nutricional en Colombia, ENSM: Estudio Nacional de Salud Mental

Gráfico 3.3. Total poblaciones estudiadas de las fuentes según evaluación de la declaración de la edad. ASIS 2002-2007

Para los registros continuos –RIPS, vigilancia en salud pública, nacimientos, defunciones y defunciones fetales– no aplica este indicador, y por lo tanto no fue estimado; y para el Estudio Nacional de Salud Bucal no se pudo calcular este indicador debido a que se aplicó a población estaba con edades de 5, 6, 7, 12, 15 a 44 años y 55 a 74 años y los métodos utilizados Myers y Whipple requieren la información de las edades simples en forma ininterrumpida para realizar el cálculo del indicador.

No respuesta

Para la evaluación del contenido también se estimó el indicador de no respuesta. El porcentaje más alto de este indicador se obtuvo para los registros continuos: cerca del 76% para los nacimientos, seguidos de las defunciones generales y fetales con un 52%; para las encuestas, el porcentaje más alto lo obtuvo la ENS-2007, con cerca del 12% de no respuesta.

Cuadro 3.3 Porcentaje de atributos con no respuesta por fuente. ASIS 2002-2007

Fuente	Atributos	% de no respuesta*	Rangos de no respuesta*		
			Menor del 5%	Entre 5% y 10%	Mayor del 10%
Encuestas					
ENS 2007	4.067	11,9	9,8	0,4	1,7
ENDS 2005	1.602	9,8	5,3	0,6	3,9
ECV	676	0,4	0,3	0,1	
ENFREC	490	6,7	4,9	1	0,8
ENSAB	126	5,6	4,8	0,8	
ENH	518	1,7	1,7		
ENSIN	138	4,0	4,0	0,0	3,0
Registros Continuos					
SIVIGILA	217	1,8	1,8		
Nacimientos	353	75,6	47,3	10,5	17,6
Defunciones	914	52,1	27,9	5,8	18,4
Defunciones fetales	503	52,3	16,9	7,6	27,8

*No respuesta se presenta cuando algunos de los datos de los archivos carecen de valores

Convenciones: ENS: Encuesta Nacional de Salud, ENDS: Encuesta Nacional de Demografía y Salud, ECV: Encuesta de Calidad de Vida, ENFREC: Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas, ENSAB: Estudio Nacional de Salud Bucal, ENH: Encuesta Nacional de Hogares, GEIH: Gran Encuesta Integrada de Hogares, ENSIN: Encuesta Nacional de la Situación Nutricional en Colombia, ENSM: Estudio Nacional de Salud Mental, SIVIGILA: Software del sistema de vigilancia en salud pública, RIPS: Registros Individuales de Prestación de Servicios.

Sobre-respuesta y consistencia

En general se evidenció una mayor sobre-respuesta en las encuestas, mientras que con respecto a la consistencia los más afectados son los registros continuos. Para las encuestas, el porcentaje más alto de sobre-respuesta fue del 16,2%, y la consistencia más baja fue para la ENDS-2005 con cerca del 95%. En el caso de los registros continuos la consistencia estuvo por encima del 95%, y el mayor porcentaje de sobre-respuesta fue para las defunciones con cerca del 2%. Debe resaltarse

que las únicas fuentes sin sobre-repuesta fueron la ECV, ENSAB y el SIVIGILA. Los valores detallados para estos indicadores se pueden observar en el Cuadro 3.4.

Cuadro 3.4 Porcentaje de atributos con sobre-respuesta y consistencia por fuente.
ASIS 2002-2007

Fuente	Atributos	Porcentaje de atributos	
		Sobre-respuesta*	Consistencia**
Encuestas			
ENS 2007	4.067	16,2	97,2
ENDS 2005	1.602	8,2	95,2
ECV	676	0	100
ENFREC	490	2,9	95,1
ENSAB	126	0	100
ENH	518	0,6	99
ENSIN-INGESTA	138	0	98,8
Registros continuos			
SIVIGILA	217	0	99,5
Nacimientos	353	1,1	94,1
Defunciones	914	2,3	91,8
Defunciones fetales	503	1,2	90,3

*Sobre-respuesta se presenta cuando el total de datos es superior al número de datos válidos para el atributo

**Consistencia indica la presencia de datos aceptables para el campo al que corresponden y que estén registrados de acuerdo a la definición hecha por la fuente establecida previamente

Convenciones: ENS: Encuesta Nacional de Salud, ENDS: Encuesta Nacional de Demografía y Salud, ECV: Encuesta de Calidad de Vida, ENFREC: Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas, ENSAB: Estudio Nacional de Salud Bucal, ENH: Encuesta Nacional de Hogares, ENSIN: Encuesta Nacional de la Situación Nutricional en Colombia, ENSM: Estudio Nacional de Salud Mental, SIVIGILA: Software del sistema de vigilancia en salud pública, RIPS: Registros Individuales de Prestación de Servicios.

Indicadores de fiabilidad

Los indicadores de fiabilidad hacen referencia a la presencia, existencia, duplicidad y completitud de los conjuntos de datos almacenados en las fuentes.

Con respecto al indicador de presencia es importante resaltar que solo los atributos de los conjuntos de datos de la ECV coinciden totalmente con los instrumentos utilizados; las demás fuentes no consideran el total de los atributos de los instrumentos. El valor más sobresaliente es el reportado por las defunciones –con un indicador de presencia del 81%– debido a que, históricamente, el registro no se ha poblado con todos los datos en formato digital (este hecho ha

mejorado considerablemente en los últimos años, pero al evaluar la fuente de manera general para todos los años el problema es evidente).

La existencia de los atributos críticos para el análisis reportó valores satisfactorios: el valor más bajo fue para las defunciones fetales con cerca de un 98%, y la mitad de las fuentes reportaron una existencia del 100%.

El indicador de duplicidad fue el más favorable para todas las fuentes, debido a que ninguna de ellas reportó registros duplicados. Debe resaltarse que este indicador estaría afectado si se hubiesen incluido los RIPS en el análisis.

Por último, el indicador de completitud, es decir la presencia de todos los valores esperados por atributo, permite evidenciar que las fuentes más afectadas son los registros de estadísticas vitales (nacimientos, defunciones y defunciones fetales) con valores por debajo del 49%. A continuación se presentan los resultados.

Cuadro 3.5 Indicadores de existencia, duplicidad y completitud por fuente. ASIS 2002-2007

Fuente	Atributos	Porcentaje de atributos			
		Presencia *	Existencia **	Duplicidad ***	Completitud ****
Encuestas					
ENS 2007	4.067	98,4	98,9	0	87,3
ENDS 2005	1.602	97	99,8	0	89,1
ECV	676	100	100	0	99,5
ENFREC	490	99,8	99,4	0	92,9
ENSAB	126	98,4	100	0	92,9
ENH	518	-	100	0	98,3
ENSIN-INGESTA	138	97,4	98,4	0	76,2
Registros continuos					
SIVIGILA	217	-	100	0	98,2
Nacimientos	353	84,8	100	0	24,4
Defunciones	914	81	99,6	0	49,1
Defunciones fetales	503	93,4	98,2	0	45,9

*Presencia es el grado de coincidencia entre cada uno de los atributos registrados con respecto a cada uno de los atributos solicitados en los instrumentos de las fuentes

**La Existencia, se evalúa bajo la consideración de que los atributos críticos para el análisis deben tener algún valor

***Duplicidad se presenta cuando el registro tiene representación duplicada del mismo objeto

****Completitud es el grado en el que el dato asociado a un sujeto con una entidad tiene todos los valores esperados para el atributo y para las instancias de la entidad relacionadas en un contexto específico de uso.

Convenciones: ENS: Encuesta Nacional de Salud, ENDS: Encuesta Nacional de Demografía y Salud, ECV: Encuesta de Calidad de Vida, ENFREC: Estudio Nacional de Factores de Riesgo de Enfermedades Crónicas, ENSAB: Estudio Nacional de Salud Bucal, ENH: Encuesta Nacional de Hogares, GEIH: Gran Encuesta Integrada de Hogares, ENSIN: Encuesta Nacional de la Situación Nutricional en Colombia, SIVIGILA: Software del sistema de vigilancia en salud pública.

Indicadores de representatividad

Con respecto a los indicadores de representatividad los resultados se dispondrán en el sitio web del Análisis de la Situación de Salud en Colombia, donde se podrá consultar los errores, factores de expansión, intervalos de confianza, frecuencias simples y medidas de tendencia estimadas para cada una de las variables de las fuentes de información proporcionadas por el Ministerio de la Protección Social.

CAPÍTULO 4:

ERRORES DE MUESTREO

INTRODUCCIÓN

Conocer los aspectos estadísticos considerados en una investigación es relevante para entender el alcance de los resultados que se exponen. En el ASIS 2002-2007 se contó con fuentes secundarias y, por lo tanto, los análisis estuvieron condicionados por aspectos del muestreo tales como diseño, tipo, método de selección y cálculos de factores de expansión previamente considerados en las distintas encuestas poblacionales.

Las bases de datos entregadas por el Ministerio de la Protección Social y utilizadas para el ASIS 2002-2007 no contenían la información completa de los factores de expansión a la población, razón por la cual, con excepción de la ENS-2007, no pudieron estimarse con la precisión deseada los resultados de las encuestas ENDS-2005, ENSIN-2005, ECV-2003 y ENFREC-1998.

Con respecto a la ENS-2007, el análisis incluyó diferentes unidades de observación, y su nivel de desagregación estuvo sujeto a la prevalencia de los problemas de salud priorizados por grupo etéreo. Los hogares y sus habitantes formaron parte de muestras probabilísticas, y para el análisis se aplicaron los correspondientes factores de expansión y el cálculo de los errores estándar. Las IPS's y los usuarios de los servicios de consulta externa, urgencias y hospitalización estuvieron circunscritos a los entes territoriales que hicieron parte de esta encuesta y, por lo tanto, dichas muestras no fueron representativas de sus respectivos universos.

Para el análisis también se utilizaron todas las estadísticas vitales proporcionadas por el DANE, el Censo General de 2005 y la información proveniente de los RIPS del año 2007.

Este capítulo recopila los aspectos relacionados con el muestreo e ilustra el modo como se calcularon los errores estándar de las estimaciones de la ENS-2007.

TIPOS DE MUESTRA

Las encuestas ENDS-2005, ENSIN-2005 y ENS-2007 provienen de la muestra maestra de Profamilia, pero la última encuesta fue actualizada con una muestra complementaria de hogares y personas y de IPS's circunscritas a los entes territoriales. Las tres encuestas poblacionales mencionadas, la ENFREC-1998 y la ECV-2003 se caracterizan por provenir de muestras probabilísticas, estratificadas y polietápicas, con alguna diferencia en el nivel de representación. A continuación se explica cada uno de estos aspectos de muestreo referenciados metodológicamente en las respectivas encuestas nacionales (Gómez y Sánchez, 2007:4).

Muestras probabilísticas

Cada unidad de los universos estudiados tuvo una probabilidad de selección conocida y superior a cero. Las muestras probabilísticas tienen la ventaja de permitir establecer a priori, y de manera aproximada, la precisión requerida en las estimaciones, y de facilitar a posteriori el cálculo de la precisión observada en las estimaciones obtenidas. La probabilidad final de cada sujeto de estudio (hogar y sus miembros, instituciones de salud y sus usuarios) estuvo dada por el producto de las probabilidades de cada fase del proceso de selección. El efecto de esta operación fue el de probabilidades finales desiguales, lo cual conllevó la aplicación de ponderaciones correctoras para evitar sesgos en la estimación de los resultados para los universos estudiados.

De conglomerados

Para las encuestas que utilizaron la muestra maestra de Profamilia, los conglomerados fueron de varios tipos. En particular, para la ENS-2007 las Unidades Primarias de Muestreo (UPM) estuvieron constituidas por municipios o combinación de dos o más de ellos (si eran muy pequeños). Las Unidades Secundarias de Muestreo (USM) variaron según el tipo de universo: para el universo de hogares y población general, estuvieron constituidas por manzanas en la zona urbana, y por secciones censales en la zona rural; para el universo de usuarios por las IPS's. Para el estudio de las variables institucionales de éstas, no hubo un segundo tipo de conglomerados. Las Unidades Terciarias de Muestreo (UTM) del universo de población general fueron los segmentos de hogares contiguos (12 en promedio), que se seleccionaron de las USM escogidas. La conglomeración de la muestra –es decir, su concentración en un número limitado de conglomerados– tiene grandes ventajas en el estudio de universos grandes y dispersos, pues reduce los costos unitarios de la recolección a cambio de un moderado incremento en el error estándar de las estimaciones. A su vez, esta pérdida de precisión se compensa con un pequeño incremento del tamaño total de la

muestra. El efecto de la conglomeración (deff: design effect) depende del grado de homogeneidad interna de los conglomerados, la cual se mide por la correlación intraclase.

Estratificadas

Antes de la selección y para aumentar la precisión de las estimaciones, la ENS-2007 llevó a cabo la clasificación o estratificación de las unidades de los universos de estudio en grupos o estratos de unidades similares. La estratificación de las UPM (municipios) se hizo con algunas de las mejores variables de estratificación disponibles en Colombia a nivel municipal: región, departamento y subregiones económicas y geográficas intra-departamentales; nivel de urbanización de las cabeceras municipales según su tamaño; composición urbano – rural de la población municipal e Índice de Necesidades Básicas Insatisfechas (NBI). En las UPM, el estrato socioeconómico predominante en las manzanas, cuando estaba disponible, fue un buen criterio de estratificación para el universo de población general; y en las zonas rurales, la distancia a los centros poblados o a las cabeceras municipales. En el universo de instituciones de salud, la estratificación se realizó en función de su naturaleza jurídica (pública, mixta y privada) y el tamaño. En las instituciones, los usuarios fueron preclasificados según el tipo de servicio recibido; debe tenerse en cuenta, sin embargo, que la selección de las IPS's y los usuarios estuvo circunscrita a los entes territoriales seleccionados en la ENS-2007. La estratificación es la técnica de optimización por excelencia en el muestreo probabilístico. La ganancia en precisión está asociada con el grado de homogeneidad interna de las unidades de los estratos, y el grado de heterogeneidad entre un estrato y otro. Para una mayor complementariedad, las variables de estratificación deben estar altamente correlacionadas con las variables centrales de estudio (variables dependientes) y mínimamente correlacionadas entre sí.

Polietápicas

Para la ENS-2007, la selección de los hogares se efectuó en un proceso de cuatro etapas sucesivas:

- *Primera:* las UPM con probabilidad proporcional al tamaño (PPT), en la que los 59 municipios más grandes del país fueron de inclusión forzosa y tuvieron probabilidad 1.
- *Segunda:* las USM con PPT.
- *Tercera:* las UTM con probabilidad igual en el trabajo inicial de Profamilia; los hogares de las UTM (segmentos), con unas pocas excepciones, fueron encuestados en su totalidad.
- *Cuarta:* se submuestrearon los adultos de hogares de numerosos miembros (3 o más adultos) y las personas para los exámenes clínicos.

En total, se seleccionaron 231 UPM y 4272 USM y UTM. Las IPS's (USM) se obtuvieron en una segunda etapa, a nivel central, con probabilidad igual en los estratos conformados. En una tercera etapa se seleccionaron, a nivel institucional, los usuarios de los servicios por tipo de atención (consulta, urgencias y hospitalización), también con probabilidad igual.

CÁLCULO DE ERRORES DE MUESTREO

Para obtener indicadores precisos de las características poblacionales es necesario calcular una probabilidad final de selección, cuyo resultado es el producto de las probabilidades de selección de cada una de las etapas del proceso. De todas las encuestas poblacionales analizadas, únicamente para la ENS-2007 fue posible obtener dicha probabilidad final de selección, corregida por un factor de ajuste de estructura y de cobertura, y por ende el "factor de expansión", recíproco de dicha probabilidad. La importancia de este último valor es que su aplicación genera estimaciones no sesgadas de los parámetros de investigación (Gómez y Sánchez, 2007:12).

Los factores de expansión de cada unidad final de observación (hogares o personas según modulo estudiado), fueron utilizados para el cálculo de los errores de muestreo, el intervalo de confianza del 95%, el coeficiente de variación, el recuento no ponderado y el efecto de diseño de cada una de las preguntas de los módulos 1: Hogares, 2: Personas de 6 a 69 años, 3: Personas menores de 6 años y 4: Submuestreo de mayores de 18 años (Anexo, Cuadros 4.1 a 4.553).

A diferencia de los resultados presentados por la Unión Temporal SEI y la Universidad Javeriana, que utilizaron el comando de Stata "svyset" para los cálculos de los errores estándar, la Universidad de Antioquia –con la información suministrada– debió utilizar el módulo de muestras complejas de SPSS 17.0 especificando lo siguiente:

- Estrato del diseño muestral: Departamento
- Unidad Primaria de Muestreo: Municipio
- Factor Final de Expansión
- Probabilidad de selección del hogar

CONCLUSIONES

El proceso exhaustivo, riguroso, industrializado y metodológicamente innovador de evaluación de la calidad de datos permite afirmar que, si bien persisten defectos en la calidad de los datos disponibles para el análisis de la situación de salud de Colombia –en particular, para los indicadores de presencia, completitud y no respuesta–, los datos se pueden explotar para obtener productos de información y/o conocimiento en apoyo de la toma de decisiones y la retroalimentación de las políticas en materia de salud; para hacerlo, tan solo es necesario consultar previamente los resultados de la evaluación de la calidad y, así, justificar el uso de diferentes técnicas de procesamiento, ajuste y transformación de los datos, según el tipo de fuente y la revisión exhaustiva de los indicadores estimados para cada dimensión de la calidad de datos. Se debe tener en cuenta, también, que la selección de los indicadores incluidos se hizo de manera exhaustiva de acuerdo con los términos contractuales y las necesidades explícitas del proyecto, y se sustentaron de acuerdo con criterios de exigibilidad, necesidad técnica y factibilidad. Por otra parte, los indicadores seleccionados se agruparon en dimensiones de la calidad de los datos referenciadas por estándares, instituciones y publicaciones científicas mundialmente aceptados: ISO, Fondo Monetario Internacional y otras fuentes adicionales; estas dimensiones y la valoración de sus indicadores permitieron concebir una visión integral de la calidad de los conjuntos de datos de las fuentes disponibles a nivel nacional para Colombia.

BIBLIOGRAFÍA

Abraira, V. "Errores en las mediciones y clasificaciones clínicas: precisión y validez". Madrid. [Sitio disponible en]: http://www.hrc.es/investigacion/bioest/Intro_errores.html Consultado el 24 de febrero de 2009.

Agudelo, A, A Henao y A Méndez (2005). "El certificado de defunción: un instrumento para la vigilancia en salud pública". [Sitio disponible en]: <http://www.nacer.udea.edu.co/pdf/libros/libro1/elcertificado-dedefuncion.pdf> Consultado el 20 de febrero de 2009.

Aja, L (2002). "Gestión de información, gestión del conocimiento y gestión de la calidad de las organizaciones". ACIMED. [Sitio disponible en]: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-9435200200500004&ing=es&nrm=iso Consultado el 19 de febrero de 2009.

Álvarez, A y A López (2006). *Evaluación de la calidad de los datos recolectados en los instrumentos de registro del nacido vivo y la defunción en dos empresas sociales del estado*. Medellín.

Anaya, R (1996). "Las bodegas de datos como apoyo a sistemas de información acerca del negocio". *Revista universidad EAFIT* 104:96-98.

Anderson, D, D Sweeney y T Williams (2003). "Estadística para administración y economía". [Sitio disponible en]: <http://books.google.com.co/books?hl=es&lr=&id=rT7WtNM2rOQC&oi=fnd&pg=PA4&q=ESTADISTICA%2B+MEDICION+DE+DATOS&ots=MeHtI7Tg6c&sig=paSUKqmw51KRskFYmy7blQpThXQ#PPT102,M1> Consultado el 20 de febrero de 2009

Baca, M (1998). *Introduction to metadata: pathways to digital information*. Los Angeles: Getty Information Institute.

Bobes, J, M Bousoño y E Suárez (1993). "Desarrollo histórico del concepto de calidad de vida". [Sitio disponible en]: http://www.uniovi.es/psiquiatria/publicaciones/documentos/1993/1993_Bobes_De-sarrollo.pdf Consultado el 21 de febrero de 2009.

Burt, P y T Kinnucan (1990). "Information models and modeling techniques for Information Systems". En: *Annual Review of Information Science and Technology*, 25(1):175-208.

Carson, S (2001). *Toward a Framework for Assessing Data Quality: Annex I. Data Quality Assessment Framework -generic framework*. Washington: International Monetary Fund.

Chaín, C (2000). *La Gestión de Información en la Organizaciones*. Murcia: Universidad de Murcia.

Codina, L (1996). "La investigación en sistemas de información". En: *Tramullas. Actas del Seminario Tendencias de Investigación en Documentación*. Zaragoza: Universidad de Zaragoza. pp.117-146.

Colombia, Ministerio de la Protección Social (1997). Decreto 1171 de 1997, Por el cual se reglamentan los artículos 50, 51 de la Ley 23 de 1981. Bogotá.

_____ (1998). III Estudio Nacional de Salud Bucal y II Estudio Nacional de Factores de Riesgo de las Enfermedades Crónicas. Colombia.

_____ 2000). Resolución 3374 de 2000. [Sitio disponible en]: <http://www.fosyga.gov.co/LinkClick.aspx?link=MarcoNormativo%2FECAT%2FResolucion+3374+de+2000.pdf&tabid=310&mid=598> Consultado el 21 de agosto de 2009.

_____ 2007). Nuevo certificado de defunción y nacido vivo. [Sitio disponible en]: <http://www.minproteccion-social.gov.co/VBeContent/NewsDetail.asp?ID=16856&IDCompany=3> Consultado el 2 de marzo de 2009.

_____ (2007). Encuesta Nacional de Salud 2007. [Sitio disponible en]: <http://www.minproteccion-social.gov.co/VBeContent/NewsDetail.asp?ID=18361&IDCompany=3> Consultado el 18 de febrero de 2009.

DANE: Departamento Administrativo Nacional de Estadísticas (2002). "Metodología y tratamiento de la no respuesta en encuestas estadísticas". En: *Estadísticas al día*, 3:3-8.

_____ (2003). "Encuesta nacional de calidad de vida 2003: documentos metodológicos y formularios". [Sitio disponible en]: http://www.dane.gov.co/files/investigaciones/fichas/ficha_calidad_vida_2003.pdf Consultado el 21 de febrero de 2009.

_____ (2005). Censo general 2005: conciliación demográfica. [Sitio disponible en]: <http://190.25.231.242/cgibin/RpWebEngine.exe/PortalAction?&MODE=MAIN&BASE=CG2005BASICO&MAIN=WebServerMain.inl> Consultado el 15 de febrero de 2009.

_____ (2009) Estadísticas vitales. [Sitio disponible en]: http://www.dane.gov.co/index.php?option=com_content&task=category§ionid=16&id=36&Itemid=148 Consultado el 20 de febrero de 2009.

Date, C (2001). *Introducción a los sistemas de bases de datos*. México: Prentice Hall. Séptima edición.

Eugenio, V, C Ismael, C Coral y P Mario (2007). "Documentos xml según criterios de calidad de datos". [Sitio disponible en]: http://www.iadis.net/dl/final_uploads/200713C056.pdf Consultado el 20 de febrero de 2009.

Gilliland-Swetland, A (1998). *Defining metadata. Introduction to metadata: pathways to digital information*. Los Angeles: Getty Information Institute.

Gómez, LC y H Sánchez (2007). "Muestreo probabilístico de hogares, instituciones de salud y usuarios". Documento técnico presentado por SEI S.A. ENS-2007.

Guerra, C (2009). "Calidad y medición de sistemas de información". [Sitio disponible en]: <http://alarcos.inf-cr.uclm.es/doc/cmsi/trabajos/Cesar%20Guerra.pdf> Consultado el 20 de febrero de 2009.

Gutiérrez, J (2006). "Generación del Conocimiento como estrategia del desarrollo territorial: dentro la incertidumbre y el determinismo". En: *Revista del Centro de Altos Estudios CANDANE*, 1(2): 18. ISSN 1909-2466.

Hernández, A (2008). *Características de calidad de datos de los almacenes de datos*. España.

Hernández, M, F Garrido y E Salazar (2000). "Sesgos en estudios epidemiológicos". En: *Salud Pública de México*, 42(5):438-446.

Herrera, A y M Salazar (2008). "Caracterización de los sistemas de información de IPS publicas de II nivel. Área metropolitana de Medellín". Trabajo de grado (Gerente de Sistemas de Información en Salud) Medellín: Universidad de Antioquia.

Hirschheim, R, H Klein, y K Lyytinen (1995). *Information systems development and data modelin*. Australia: Cambridge University Press.

Instituto Colombiano de Bienestar Familiar (2006). Encuesta Nacional de la situación Nutricional en Colombia 2005. Bogotá: Imprenta Nacional de Colombia.

Instituto Nacional de Estadística y Ministerio de Salud y Deportes (2004). *Los errores de muestreo. Encuesta Nacional de Demografía y Salud, 2003*. Bolivia.

Londoño, M, G Morera y P Laverde. (2003). Auditoria En Salud, Editorial Médica Panamericana. Segunda edición.

López, J (1991). "El desarrollo de los sistemas de información y documentación". En: *Cuadernos EUBD*: 23-33

Luque, A (2006). "Bodega de Datos: consideraciones, seguridad y control". [Sitio disponible en]: <http://www.acis.org.co/fileadmin/Conferencias/ConfAlfonsoLuqueNov30-06.pdf> Consultado el 20 de febrero de 2009.

Mcleod, R, E García, S Alberto y N Fernández (2000). "Sistemas de información gerencial". [Sitio disponible en]: http://books.google.com.co/books?id=zmnjBpmufKIC&pg=PA273&lpg=PA273&dq=bodega+de+datos&source=bl&ots=TuDf3PjBca&sig=-mlk5plIEglgUcerlCwUivauNnM&hl=es&ei=U4a2SdCDPKKBtwfVntm1CQ&sa=X&oi=book_result&resnum=6&ct=result#PPA273,M1 Consultado el 10 de marzo de 2009.

Naciones Unidas (2003). *Principios y recomendaciones para un sistema de estadísticas vitales*. Nueva York: Naciones Unidas.

Norton, P (2006). *Introducción a la computación*. México: Mc Graw-Hill. Sexta edición.

O'Brien, J (2001). *Sistemas de Información Gerencial. Manejo de la tecnología de información en la empresa interconectada en red*. Bogotá: Mc Graw-Hill, Cuarta edición.

_____ (2006). *Sistemas de Información Gerencial*. México: Mc Graw-Hill. Séptima Edición.

Prieto, A, A Lloris y J Torres (2006). *Introducción a la informática*. España: Mc Graw-Hill. Cuarta edición.

Profamilia (2005). [Sitio disponible en]: http://www.profamilia.org.co/encuestas/01encuestas/2005resultados_generales.htm Consultado el 12 de febrero de 2009.

Quiroz, H (2001). *Fundamentos y legislación en calidad y auditoria en salud*. Colombia: Zuluaga.

Ramalho, J (1999). *SQL SERVER 7*. Capítulo 14: Iniciación y referencia. Colombia: McGraw-Hill Interamericana S.A.

Rodriguez, J, J Chackiel, J Guzmán, D Jaspers-Faijjer, G Macció, J Pujol et al (1997). *Demografía I*. México: PROLAP-IISUNAM.

- Rojas, E (2006). "Realismo Estadístico". En: *Revista del Centro de Altos Estudios CANDANE*, 1(2): 7. ISSN 1909-2466.
- Sackett, D (1979). *Bias in analytic research*. Gran Bretaña: Pergamon Press Ltd.
- Samuelson, K (1977). "Information Systems and networks". En: *Book Reviews*, 12(3): 3.
- Sánchez, J (2003). "¿Qué son los metadatos?" [Sitio disponible en]: www.uportal.cl/siel/siel_docs/estandarizacion/metadatos_SIEL.pdf Consultado el 18 de febrero de 2009.
- Scott, G (1992). *Principios de sistemas de información*. México: Mc Graw-Hill.
- Senn, J (1992). *Análisis y diseño de sistemas de información*. México: Mc Graw-Hill. Segunda edición.
- Senso, J y A Pinero (2003). "El concepto de metadato. Algo más que descripción de recursos electrónicos". [Sitio disponible en]: <http://www.scielo.br/pdf/ci/v32n2/17038.pdf> Consultado el 23 de febrero de 2009.
- Silberchatz, A, H Korth y S. Sudarshan (2006). *Fundamentos de bases de datos*. España: Mc Graw-Hill / Interamericana de España, S.A.U. Quinta Edición.
- Tacla, O (2006). *La omisión censal en América Latina 1930-2000*. Santiago de Chile: Naciones Unidas.
- Whitten, J, L Bentley y V Barlow (1998). *Análisis y diseño de sistemas de información*. Madrid: Mc Graw-Hill. Tercera edición.
- Yourdon, E (1993). *Análisis estructurado moderno*. México: Prentice Hall Inc.



ASIS
Análisis de Situación en Salud



Libertad y Orden

Ministerio de la Protección Social
República de Colombia



**UNIVERSIDAD
DE ANTIOQUIA**
1803

Facultad Nacional de Salud Pública
"Héctor Abad Gómez"