
II Metodología

1. Instrumentos y procedimiento de recogida de información

El cuestionario

PRINCIPALES FUENTES PARA SU ELABORACIÓN

Se analizaron los cuestionarios ya existentes que habían sido planteados desde marcos teóricos y con objetivos similares, aplicados en varios países afines o en el marco de proyectos muy sólidos metodológicamente, así como los cuestionarios que contenían preguntas sobre prácticas sexuales en España y por lo tanto era de gran interés el obtener información comparable:

a) La propuesta de la Acción Concertada Europea (1991 y 1996) para el estudio de las conductas sexuales relacionadas con la infección por el VIH/sida^{1;47}. Este cuestionario fue traducido ex profeso al español y se tuvo presente en todo el diseño. De él se incorporó al cuestionario de la ESHS prácticamente un módulo entero, el de nuevos encuentros o NEM (New Encounter Module) cuyo objetivo es conocer la frecuencia de nuevas parejas durante los últimos 12 meses así como su caracterización y la práctica de relaciones sexuales protegidas o no con las mismas. Esta información se consideró de especial importancia para diseñar programas de prevención de la infección por VIH.

b) El cuestionario utilizado en la *National Survey of Sexual Attitudes and Lifestyles - NATSAL (1990 y 2000)*^{11;17;21}, llevada a cabo en el Reino Unido, del que parecía especialmente interesante el abordaje epidemiológico de las prácticas de riesgo.

c) El cuestionario de la encuesta francesa *Analyse des Comportements Sexuels en France - ACSF (1991-92)*, no en su enfoque de estudio de la sexualidad humana, pero sí como estudio de gran magnitud que aporta sugerencias imprescindibles^{2;48-51}.

d) El proyecto del Programa Global de Acción de la Organización Mundial de la Salud GPA/OMS del cuestionario *Knowledge, Attitudes, Beliefs and Practices (1989)*, primera herramienta diseñada para su aplicación en distintos países^{52;53}.

e) El módulo del Center of Disease Control and Prevention *Standard Sexual Behavior Questionnaire (1998)* se utilizó principal-

mente por la claridad en la formulación de algunas preguntas y definición de las prácticas⁵⁴.

f) Las preguntas sobre prácticas sexuales de riesgo de la *Encuesta Domiciliaria sobre Abuso de Drogas en España (1996)*, por el interés de obtener información comparativa^{28;29}.

ESTRUCTURA Y PRINCIPALES VARIABLES

La estructura y organización del cuestionario se basó en numerosas consideraciones, entre las cuales destacan:

– Los objetivos del estudio, que definían las principales variables.

– La experiencia de la organización de los cuestionarios de estudios internacionales anteriormente mencionados y que se tomaron como principales modelos.

– El método mixto de administración: entrevista asistida por ordenador con entrevistador, pero mayoritariamente autoadministrada.

– La prioridad otorgada al establecimiento de una buena relación entre el entrevistador y el entrevistado, evitando las preguntas demasiado personales al comienzo, antes de haber logrado una cierta compenetración. Por este motivo se decidió comenzar por las preguntas sociodemográficas que además permitirían caracterizar a las personas en el caso de abandono prematuro.

– Introducir progresivamente las preguntas sexuales comenzando por aquéllas que parecían menos sensibles, como las fuentes de información sobre temas sexuales, lo que permitiría entrar paulatinamente en aspectos más delicados.

– Situar las preguntas sobre la experiencia sexual hacia la mitad de la encuesta, de forma que se minimice el impacto de los abandonos por cansancio, impedimentos de tiempo o rechazo.

– Organizar la secuencia de preguntas sobre la conducta sexual con orden cronológico, partiendo de la primera relación hasta los últimos treinta días para facilitar el recuerdo.

– Situar al final del cuestionario las preguntas sobre conocimientos, opiniones y actitudes, lo que permitiría responder a las preguntas sobre prácticas sexuales antes de solicitar emitir cualquier tipo de juicio u opinión valorativa de las mismas.

La estructura, con sus principales secciones, puede observarse resumida en la figura II.1.1. Asimismo, en el anexo 2 y en el servidor de internet del INE puede obtenerse la transcripción del cuestionario, aunque no el cuestionario electrónico que se empleaba en el ordenador. Debe tenerse en cuenta que todas las indicaciones correspondientes a filtros y saltos no eran perceptibles por entrevistador ni entrevistado al estar programados en la versión administrada.

Antes de comenzar con las preguntas que contenía el cuestionario, se presentaban dos pantallas de información para el encuestado. En la primera se le informaba sobre los cometidos del INE, la importancia y relevancia de la encuesta y su carácter voluntario. En la segunda se le informaba sobre las garantías de confidencialidad y las garantías del secreto estadístico establecidas por la Ley de la Función Estadística Pública.

El cuestionario constaba de las siguientes secciones:

A. Características sociodemográficas: además de las correspondientes al entrevistado (edad, sexo, nivel de estudios, estado civil, país de nacimiento y ocupación) incluye el nivel de estudios de los padres.

B. Estilos de vida: contiene preguntas sobre tipo de convivencia, frecuencia de salir por la noche y consumo de alcohol y drogas.

C. Información y experiencia sexual: consta de un apartado común y tres subsecciones.

C1: varones con relaciones heterosexuales

C2: varones con relaciones homosexuales

C3: mujeres

En el apartado común se indaga sobre las fuentes de información preferidas y reales sobre temas sexuales. Además se incluye la definición de *relaciones sexuales* a efectos

de contestar este cuestionario, y las preguntas sobre si se han tenido este tipo de relaciones y si ha sido con personas de su mismo o de diferente sexo o con ambos.

Los varones cumplimentan las subsecciones C1 y C2 en función del tipo de prácticas sexuales que pueden haber tenido alguna vez en la vida. En el caso de las mujeres, cumplimentan la subsección C3 si han tenido relaciones heterosexuales, mientras que si no han tenido dichas relaciones, pasan directamente a la sección de la prueba de VIH.

Las tres subsecciones tienen prácticamente la misma estructura, con los siguientes apartados y principales variables:

- La primera relación sexual, con variables sobre edad propia y de la pareja y uso del preservativo u otro método anticonceptivo.

- Relaciones sexuales a lo largo de la vida, incluyendo uso de prostitución en los varones, con preguntas sobre número de personas y uso del preservativo (alguna vez, en la última relación sexual y en relaciones de prostitución).

- Relaciones sexuales en los últimos 12 meses, diferenciado el tipo de pareja (estable, ocasional o a la que se ha pagado sólo en los varones), y el uso del preservativo con cada uno de ellos.

- Relaciones sexuales en los últimos 30 días, con preguntas sobre la frecuencia de relaciones.

- Relaciones sexuales con nuevas parejas: se indaga primero sobre la frecuencia de dichas relaciones y luego se incluye un amplio número de preguntas sobre las características de la persona y de la relación con la última nueva pareja.

D. Salud sexual: se indaga sobre las infecciones de transmisión sexual diagnosticadas, sobre cuál ha sido la última y cuándo ha sucedido.

E. Prueba de VIH: se pregunta sobre si se ha realizado la prueba al donar sangre, o si se trata de una prueba específica de detección, especificando si es con motivo de un embarazo o por otros motivos, así como el

lugar de realización y tiempo transcurrido desde la última prueba.

F. Conocimientos, opiniones y actitudes: se pregunta sobre los mecanismos de transmisión del VIH, la opinión sobre la eficacia de diferentes medidas de prevención, con especial atención al uso del preservativo, las actitudes discriminatorias con las personas infectadas por el VIH, y la percepción del riesgo de infección.

G. Opiniones sobre el cuestionario y método de aplicación: esta sección no formaba parte propiamente del cuestionario de la encuesta, sino que constituía una evaluación sobre su contenido y su método de administración por parte del entrevistado.

El cuestionario, redactado inicialmente en español, se tradujo a las otras lenguas oficiales del Estado. En las comunidades autónomas con lengua propia se ofreció la posibilidad de elegir la lengua para contestar.

Procedimiento de administración del cuestionario

El método de recogida de información fue uno de los principales determinantes de los aspectos operativos del estudio. Por las razones mencionadas en la introducción^{17;55-64} se eligió un cuestionario asistido por ordenador mayoritariamente autoadministrado (CASI).

El cuestionario electrónico se contestaba en un pequeño ordenador portátil. Cada pregunta se presentaba en una pantalla diferente, debiendo anotarse las respuestas mediante un lápiz óptico. Se trataba de un programa que podía ser fácilmente manejado por todos los entrevistados después de aprender su funcionamiento al mismo tiempo que se iba desarrollando la primera parte de la entrevista. Con este fin, y tal como se muestra en la figura II.1.1, el entrevistado iba progresivamente aprendiendo y ganando autonomía en el manejo del ordenador.

1. La sección de características sociodemográficas se cumplimentaba de forma conjunta entre entrevistador y entrevistado, para que el entrevistado se familiarizase con el portátil y adquiriese habilidades para manejarlo correctamente. De este modo el entrevistado podía aprender sobre el procedimiento de respuesta a los diferentes tipos de preguntas, la forma de anotar los datos o el avance y retroceso de pantalla.

2. A partir de la sección de estilos de vida, el entrevistador entregaba el ordenador portátil al entrevistado para que continuase contestando él solo el resto del cuestionario. No obstante, como esta sección todavía contenía cuestiones menos íntimas que las siguientes, aún existía oportunidad de interacción entre entrevistado y entrevistador, realizando éste último una discreta supervisión del procedimiento.

3. Cuando comenzaban las preguntas relativas a fuentes de información sobre temas sexuales, el entrevistador se retiraba y ya no intervenía, aunque le indicaba al entrevistado que estaría disponible para cualquier aclaración o duda que se pudiera plantear. El entrevistador debía retirarse de forma que el entrevistado percibiera con claridad que no tenía posibilidad de conocer sus respuestas. Ahora bien, debía procurar que la contestación se realizara en la misma habitación.

Para dar las mayores garantías de privacidad, en el ordenador portátil no se incluía ningún dato personal de los informantes (nombres, apellidos, dirección, ...), disociando la información personal en origen. Los cuestionarios se identificaban únicamente con un número que permitía enlazar con el parte de trabajo, de forma que se pudiera conocer a quiénes se debían realizar las entrevistas así como poder controlar *a posteriori* cuáles eran las entrevistas que se habían realizado entre las personas previstas y cuáles no, pero sin posibilidad alguna de conocer a qué persona pertenecían las respuestas de un determinado cuestionario.

Figura II.1.1 Estructura y modo de administración del cuestionario

<p>A- Características sociodemográficas</p>	<p>Administrado por entrevistador</p> <ul style="list-style-type: none"> • el entrevistador formula las preguntas • el entrevistado responde, observa y aprende a utilizar el ordenador
<p>B- Estilos de vida</p>	<p>Autoadministrado con supervisión</p> <ul style="list-style-type: none"> • el entrevistado responde directamente en el ordenador • el entrevistador supervisa
<p>C – Información y experiencia sexual</p> <ul style="list-style-type: none"> • Apartado común • Pregunta sobre tipo de relaciones sexuales que se han tenido alguna vez en la vida • Apartados específicos según el tipo de relaciones sexuales que se hubieran mantenido. <p>D – Salud sexual</p> <p>E – Prueba de VIH</p> <p>F – Conocimientos, actitudes y opiniones sobre la infección por VIH</p>	<p>Autoadministrado</p> <ul style="list-style-type: none"> • el entrevistado contesta sin que el entrevistador pueda observar sus respuestas • el entrevistador disponible para responder a consultas
<p>G – Opiniones sobre cuestionario y método de aplicación</p>	<p>Autoadministrado</p>

2. Diseño muestral

Ámbito de la encuesta

- **Ámbito poblacional:** personas de 18 a 49 años de edad residentes en viviendas familiares principales.
- **Ámbito geográfico:** todo el territorio del Estado.
- **Ámbito temporal:** las entrevistas se llevaron a cabo entre el 13 de octubre y el 19 de diciembre de 2003.

Tipo de muestreo

Se diseñó una muestra independiente para cada comunidad autónoma con el fin de que determinadas variables pudieran ser analizadas con este nivel de desagregación.

Se utilizó un muestreo bietápico estratificado en la primera etapa, siendo las unidades de primera etapa las secciones censales y las unidades de segunda etapa las personas de 18 a 49 años residentes en viviendas familiares principales.

El marco utilizado para la selección de la muestra fue un marco formado por la relación de secciones censales existentes a 1 de enero de 2003. Para la selección de las unidades de segunda etapa se utilizó la relación de personas de 18 a 49 años empadronadas en cada una de las secciones seleccionadas para la muestra.

El criterio de estratificación utilizado en la primera etapa fue el tamaño del municipio al que pertenecía la sección. De acuerdo con este criterio se establecieron los siguientes estratos:

Estrato 0: municipios con más de 500.000 habitantes

Estrato 1: municipios capital de provincia (excepto los anteriores)

Estrato 2: municipios con más de 100.000 habitantes (excepto los anteriores)

Estrato 3: municipios de 50.000 a 100.000 habitantes (excepto los anteriores)

Estrato 4: municipios de 20.000 a 50.000 habitantes (excepto los anteriores)

Estrato 5: municipios de 10.000 a 20.000 habitantes

Estrato 6: municipios con menos de 10.000 habitantes

Tamaño de la muestra. Afijación

Para poder facilitar estimaciones con un determinado grado de fiabilidad a nivel nacional y por comunidades autónomas, se fijó una muestra de 13.600 personas distribuidas en 1.700 secciones censales.

Con objeto de garantizar un tamaño de muestra mínimo en cada comunidad autónoma, la muestra se distribuyó entre éstas asignando una parte fija y otra proporcional al tamaño de cada comunidad, considerando como tamaño el número de habitantes entre 18 y 49 años. El tamaño de muestra mínimo en cada comunidad se fijó en 300 individuos. El tamaño final de la muestra en las mismas, una vez sumada la parte proporcional correspondiente, puede observarse en la tabla II.2.1.

La distribución de la muestra entre estratos se realizó de forma proporcional al tamaño, pero se potenció la muestra en los

Tabla II.2.1 Muestra teórica de personas por comunidad autónoma

Total nacional	13.600
Andalucía	1.800
Aragón	544
Asturias (Ppdo. de)	488
Balears (Illes)	472
Canarias	672
Cantabria	392
Castilla y León	768
Castilla-La Mancha	624
Cataluña	1.552
Comunidad Valenciana	1.136
Extremadura	488
Galicia	784
Madrid (Com. de)	1.520
Murcia (Región de)	544
Navarra (Com. Foral de)	400
País Vasco	736
Rioja (La)	352
Ceuta y Melilla	328

estratos que contenían los municipios mayores, por esperar en éstos una mayor variabilidad en el comportamiento de la población, así como un mayor número de incidencias o de entrevistas fallidas.

Con objeto de poder sustituir las entrevistas fallidas que se produjeran durante el trabajo de campo, en cada sección censal se seleccionaron 24 personas, 8 personas titulares y 16 personas reservas, agrupadas en cuatro grupos de edad de seis personas cada uno. En cada uno de estos grupos las dos primeras personas fueron las titulares y las cuatro restantes las reservas, pudiendo ser utilizadas estas últimas para sustituir indistintamente a cualquiera de las dos titulares.

Selección de la muestra

En la primera etapa, las secciones se muestrearon dentro de cada estrato con probabilidad proporcional a su tamaño, entendiéndose por tal el número de habitantes de 18 a 49 años. En la segunda etapa, en cada sección las personas se seleccionaron con igual probabilidad mediante muestreo sistemático con arranque aleatorio. Este procedimiento conduce a muestras autoponderadas en cada estrato.

Estimación, ponderación y elevación. Errores de muestreo

Para estimar todas las características de la muestra se han utilizado estimadores de razón a los que se han aplicado técnicas de reponderación, tomando como variables auxiliares los grupos de edad y sexo de la población de la comunidad autónoma.

Para ello se han seguido los siguientes pasos:

1. Obtención del estimador basado en el diseño.

$$\hat{Y}_d = \sum_h \sum_{i,j} \frac{1}{K_h \cdot \frac{8}{P_h^{(03)}}} y_{hij} = \sum_h \sum_{i,j} \frac{P_h^{(03)}}{p_h^t} y_{hij}$$

siendo:

h: estrato

i: sección

j: persona

K_h : número de secciones de la muestra en el estrato h

$P_h^{(03)}$: número de personas del estrato h según el marco de 2003

Y: variable objetivo

p_h^t : población muestral teórica del estrato h.

2. Corrección de la falta de respuesta. Con este fin se multiplicó el factor de elevación

anterior $\frac{P_h^{(03)}}{p_h^t}$ por la inversa de la probabi-

lidad de respuesta dentro del estrato, es decir:

$$\hat{Y}_2 = \sum_h \sum_{i,j} \frac{P_h^{(03)}}{p_h^t} \frac{p_h^t}{p_h^e} y_{hij} = \sum_h \sum_{i,j} \frac{P_h^{(03)}}{p_h^e} y_{hij}$$

donde p_h^e es la muestra efectiva de personas en el estrato h.

3. Estimador de razón, utilizando como variable auxiliar la proyección de población en el momento de la encuesta. Tiene por objetivo fundamental mejorar el estimador obtenido en los pasos anteriores, actualizando la población empleada en el momento de la selección de la muestra a la del momento de realización de la encuesta. Su expresión es:

$$\hat{Y}_3 = \sum_h \frac{\sum_{i,j} \frac{P_h^{(03)}}{p_h^e} y_{hij}}{\sum_{i,j} \frac{P_h^{(03)}}{p_h^e} p_{hij}} P_h = \sum_h \sum_{i,j} \frac{P_h}{p_h^e} y_{hij}$$

donde P_h es la proyección de población a mitad del periodo de realización de la encuesta para el estrato h.

4. Reponderación del factor anterior para ajustar la distribución estimada a la distribución de las poblaciones por comunidad autónoma, grupos de edad y sexo propor-

cionada por la Unidad de Proyecciones Demográficas del INE. Este calibrado se ha realizado mediante la macro CALMAR del Instituto Nacional de Estadística y Estudios Económicos francés (INSEE). Los grupos considerados han sido los doce siguientes: varones y mujeres entre 18 y 24, 25 a 29, 30 a 34, 35 a 39, 40 a 44 y 45 a 49 años.

Después de aplicar los pasos anteriores, se obtiene un factor de elevación final F_i para cada una de las personas de la muestra efectiva.

Así el estimador del total \hat{Y} de una característica Y es de la forma:

$$\hat{Y} = \sum F_i y_i$$

donde la suma se extiende a todas las personas de la muestra, y_i es el valor de la característica Y observada en el individuo i, y F_i es el factor de elevación final de dicho individuo.

Los estimadores de las proporciones $P = \frac{X}{Y}$

son de la forma $\hat{P} = \frac{\hat{X}}{\hat{Y}}$ donde las estima-

ciones \hat{X} e \hat{Y} se obtienen mediante la fórmula anterior.

Errores de muestreo

Para el cálculo de los errores de muestreo se ha aplicado el Método Jackknife, que proporciona una estimación del coeficiente de variación mediante la expresión:

$$CV(\hat{Y}) = \frac{\sqrt{\sum_h \frac{A_h}{A_h} \frac{1}{A_h} (\sum_{l_h} (\hat{Y}_{(lh)} - \hat{Y})^2)}}{\hat{Y}}$$

donde $\hat{Y}_{(lh)}$ es el estimador obtenido quitando el grupo de secciones l del estrato h.

Para obtener el estimador, y por simplicidad, en vez de recalculer los factores de elevación se multiplican los factores del

estrato donde se han quitado las secciones

por el factor: $\frac{n_h}{n_h \#(lh)}$.

De acuerdo a lo anterior:

$$\hat{Y}_{(lh)} = \sum_{i_h} F_i y_i + \sum_{i_{lh}} F_i \frac{n_h}{n_h \#(lh)} y_i$$

donde:

lh es un grupo de secciones del estrato h

n_h es el total de secciones del estrato h

A_h son los grupos de secciones del estrato h

$\#(l)$ es el número de secciones del grupo l

3. Evaluación de instrumentos y procedimientos

Tanto la estructura del cuestionario, como los métodos de administración y el diseño muestral que se acaban de describir hacen referencia al estudio definitivo. Sin embargo, antes de su diseño final hubo versiones previas que fueron evaluadas mediante tres tipos de procedimientos con el fin de analizar su adecuación y funcionamiento en las condiciones reales de aplicación: una encuesta piloto, la observación de los encuestadores de la encuesta piloto y estudios cualitativos. Cada uno de estos procedimientos ofrece información diferente y complementaria para aumentar la calidad de los instrumentos y procedimientos que van a ser empleados.

La encuesta piloto

Se llevó a cabo en el mes de junio de 2003 sobre una muestra de 320 personas de 18 a 49 años pertenecientes a 40 secciones censales de 10 provincias: Madrid, Valencia, Burgos, Vizcaya, Ávila, Cáceres, Sevilla, Murcia, Jaén y Toledo. Estas provincias no fueron elegidas aleatoriamente, sino que se intentó que representaran la variabilidad que puede darse en un estudio estatal de estas características con respecto a diferentes variables: situación geográfica, distribución de la población (grandes núcleos urbanos y población rural) ... También se tuvo en cuenta que en las mismas existiera disponibilidad de entrevistadores con mucha experiencia y con interés en participar en este estudio concreto.

El estudio pretendía evaluar diferentes aspectos del funcionamiento de los instrumentos y procedimientos en su contexto real de aplicación, y según algunas variables sociodemográficas relevantes: edad, sexo, nivel de estudios, ámbito rural o urbano ... En primer lugar se estudiaba la factibilidad de utilizar un diseño muestral de reciente introducción en la rutina del INE: el uso del padrón continuo para seleccionar desde el inicio a la persona concreta que debía contestar el cuestionario. Hasta ese momento se habían empleado diseños que muestreaban los hogares y una vez en ellos se entrevistaba a todos los miembros o bien se seleccionaba a las

personas concretas que deben contestar. El hecho de buscar a una persona concreta para realizar la encuesta planteaba nuevas consideraciones de accesibilidad, pues cabía suponer que algunas plantearían reticencias: ¿por qué yo?, además de la dificultad añadida de contactar con la persona seleccionada.

La encuesta piloto pretendía también valorar la posibilidad de que pudieran producirse reacciones de rechazo en amplios sectores sociales al hecho de que una institución pública preguntara sobre comportamientos tan íntimos. Esta situación se había producido en la década anterior en algunos países de nuestro entorno, obligando a retrasar algunas encuestas^{11;65;66}. En España esta reacción se creía improbable, pero no totalmente descartable.

El estudio preliminar pretendía también analizar el porcentaje de personas que no aceptaba realizar la entrevista una vez informado de su contenido. Era igualmente de enorme interés conocer si existía rechazo por el hecho de utilizar un ordenador, pues todavía constituía un método novedoso en este contexto. Se trataba también de conocer la carga de trabajo para el entrevistador y para el entrevistado, y si existía un porcentaje importante de personas que abandonaban la entrevista una vez comenzada, y en qué momento y por qué causas se producían dichos abandonos: duración, complejidad de la circulación, dificultad para manejar el ordenador o entender las preguntas, incomodidad con el contenido o la formulación de las preguntas ...

La falta de sinceridad en las respuestas y la posible influencia de la deseabilidad social son dos amenazas constantes a la validez de las respuestas proporcionadas, por lo que se estimó necesario controlar en lo posible la incidencia de estos factores en algunas preguntas cuyos contenidos pudieran resultar más comprometedores, tales como el consumo de alcohol y drogas, las relaciones sexuales o la infidelidad. Por ello se recurrió a un diseño experimental introduciendo en esta encuesta piloto dos versiones alternativas del cuestionario. Una de ellas planteaba directamente las preguntas (versión 1), y otra (versión 2), incorporaba

dos estrategias pensadas para controlar el efecto de la falta de sinceridad y de la deseabilidad social en las respuestas a preguntas comprometedoras: la estrategia de *aproximaciones sucesivas* y la introducción de *expresiones de descarga emocional*.

– La primera consistió en aproximar progresivamente al encuestado con preguntas relacionadas con el tema, pero poco o nada comprometedoras para él, hasta llegar a la pregunta o preguntas objetivo. Se supone que de esta forma el encuestado proporcionaría respuestas más sinceras y le resultaría más difícil abandonar el proceso de respuestas al bloque de preguntas comprometedoras.

– La segunda consistió en presentar antes de la pregunta algún enunciado indicativo de que el comportamiento objeto de la pregunta no es censurable, ni infrecuente, facilitando así la descarga emocional del entrevistado.

La observación por los encuestadores del estudio piloto

La encuesta piloto no fue realizada por entrevistadores contratados como los que llevaron a cabo el estudio definitivo. Por el contrario, fue realizada por entrevistadores pertenecientes a la plantilla del INE, con amplia experiencia y que manifestaron su interés en participar. Asistieron a una reunión de formación de dos días de duración en los servicios centrales del Instituto con un contenido similar al que luego se refiere en el apartado de formación de encuestadores. Se recurrió a este tipo de entrevistadores expertos para que pudieran efectuar aportaciones relevantes derivadas de su experiencia en el proceso de establecimiento de contacto y de persuasión para conseguir la participación, así como de la observación del comportamiento del encuestado durante el proceso de respuesta, del registro de sus comentarios, preguntas y dudas, de su valoración de las condiciones y el contexto de realización de la entrevista. En la reunión de formación se explicó la relevancia de su función, detallando lo que se esperaba de ellos y dotán-

doles de instrumentos que permitieran recoger sus aportaciones. Estas aportaciones de los encuestadores se recogieron de varias formas:

– Mediante la *Hoja de valoración*, en la que los encuestadores valoraban la disponibilidad, comprensión, sinceridad e intimidad, detectadas en las situaciones de entrevista.

– A través de una reunión entre encuestadores y técnicos participantes en la encuesta piloto y los responsables de la encuesta en los servicios centrales. Esta reunión se llevó a cabo una vez que hubo finalizado la encuesta piloto.

– Formando parte de un grupo específico de discusión en el estudio cualitativo que se describe más adelante.

Los estudios cualitativos o pruebas en profundidad

Estos procedimientos, frecuentes en otros ámbitos, no se habían empleado de forma habitual en el INE como herramienta para probar los cuestionarios antes del diseño final. Tienen como objetivo obtener evidencias que permitan conocer el proceso de respuesta de los encuestados a las preguntas, identificando así posibles fuentes de error por dificultades en la comprensión de las preguntas o en la elección de las alternativas de respuesta, las estrategias de recuerdo utilizadas o las reacciones a preguntas sensibles. El estudio cualitativo fue diseñado y realizado por el Departamento de Psicología Social y Metodología de las Ciencias del Comportamiento de la Facultad de Psicología de la Universidad de Granada.

Las pruebas en profundidad consistieron en entrevistas cognitivas y grupos de discusión. Las entrevistas cognitivas permiten identificar posibles fuentes de error relacionadas con problemas en la comprensión de las preguntas, en la percepción de la tarea a realizar o en la elección de las alternativas de respuesta. Los grupos de discusión permiten identificar problemas en la comprensión e interpretación de conceptos clave

y determinar las perspectivas que utilizan los encuestados para contestar las preguntas del cuestionario.

Participaron 50 personas de 18 a 49 años. De estos 50 participantes, un subgrupo estaba compuesto por hombres homosexuales, dada la relevancia de sus posibles aportaciones para el objetivo general del estudio. Se formaron otros cuatro grupos homogéneos según las variables sexo y edad. Además, se realizó otro grupo específico con los entrevistadores que habían llevado a cabo la encuesta piloto.

Resultados de la evaluación previa

A partir de los resultados del estudio piloto, del análisis de las entrevistas en profundidad y grupos de discusión, y de las observaciones de los encuestadores, se extrajeron conclusiones para mejorar el diseño de todos los instrumentos y procesos de la encuesta: diseño muestral, estructura y formulación final del cuestionario, diseño del cuestionario electrónico, elaboración de la documentación definitiva y formación de los encuestadores. Los cambios fueron menores, pues los estudios mostraron que no existía un nivel elevado de rechazo, que el método de recogida de la información era adecuado a los objetivos de la encuesta y compatible con el cuestionario. Su contenido, conjuntamente con el método de administración, la convertía en una encuesta difícil pero factible, que atraía, interesaba, y prácticamente no se producían abandonos una vez comenzada. Asimismo mostró que la metodología CASI facilitaba notablemente el proceso, que el porcentaje de encuestados con dificultades para manejar el programa de ordenador era muy bajo y que era infundado el temor a que el uso del ordenador fuera interpretado como un *gran controlador*, siendo por el contrario entendido como una garantía de confidencialidad y un instrumento más aceptado que el cuestionario tradicional. La duración de la entrevista osciló entre 30 y 45 minutos, tiempo que había sido estimado de antemano como un rango aceptable.

4. El trabajo de campo

La formación del personal

En el desarrollo de las actividades de recogida de información y supervisión de las mismas participó personal de tres tipos:

a) Los inspectores de encuesta: fueron los responsables del cumplimiento de las normas de la encuesta, formación del personal a su cargo, planificación de la inspección, y organización, seguimiento y control de los trabajos de campo en cada provincia.

b) Los inspectores de entrevistadores: 25 personas contratadas que se encargaron de la revisión e inspección directa de la labor de los entrevistadores, apoyando también el trabajo y gestión de los inspectores de encuesta.

c) Los entrevistadores: 137 personas contratadas (73,5 % mujeres y 26,5 % hombres) que recogieron la información y cumplieron el parte de trabajo.

Todos los inspectores de encuesta asistieron a una reunión de formación teórico-práctica de 2 días de duración en los servicios centrales del INE. En ellos se justificó la necesidad del estudio, se explicaron los conceptos metodológicos y teóricos relevantes, se expuso detalladamente la estructura y el contenido del cuestionario, se indicaron las normas de cumplimentación del mismo y se analizaron y precisaron las definiciones fundamentales. Expertos en infección por VIH/sida explicaron el interés y la utilidad de la encuesta desde el punto de vista de la salud pública. También se les instruyó en la utilización del portátil, el procedimiento de administración del cuestionario, las normas de realización de la entrevista y del trabajo de campo, las incidencias en la recogida y su tratamiento, así como otras consideraciones de orden práctico.

Igualmente se explicó de forma detallada el contenido y funcionamiento de la aplicación de seguimiento y control de los trabajos de campo (ESHS-WIN) que se había desarrollado ad hoc para este estudio.

Todos los aspectos relevantes para la formación del personal de campo del proyecto quedaron recogidos en detalle en

el documento de trabajo *Manual del entrevistador*.

A su vez los inspectores de encuesta formaron a los inspectores de entrevistadores y a los entrevistadores en cada delegación provincial de INE. Además se incidió en la importancia de su misión y la relevancia que tiene en la recogida de datos la correcta realización de la entrevista y el manejo de aquellos factores que influyen en la colaboración de las personas y en la calidad de las respuestas.

El trabajo de campo

El Área de Estudios y Programación de los servicios centrales del INE se encargó del diseño y preparación de los documentos e instrumentos de la encuesta, de la programación de los trabajos de campo y de su seguimiento y control en todas las delegaciones provinciales del INE. En las delegaciones se efectuaron los trabajos de recogida, inspección, seguimiento y control de la información de cada provincia a cargo de los inspectores de encuesta, inspectores de entrevistadores y entrevistadores.

Al menos una semana antes de iniciar la recogida de datos, se remitió por correo una carta de presentación del estudio firmada por el Director General de Productos Estadísticos del INE a las personas titulares y a las reservas seleccionadas. En ella se les anunciaba la visita del entrevistador, se les ofrecía una sucinta explicación de los objetivos e importancia de la ESHS, se informaba sobre la confidencialidad de los datos, se les proporcionaba un número de teléfono gratuito 900 para efectuar las consultas necesarias y se solicitaba la colaboración de la persona, si era titular, o se indicaba su posible colaboración, si era reserva.

El tratamiento de las incidencias se reguló según el tipo. En el caso de incidencias en el marco (viviendas y personas):

a) En los casos de viviendas vacías, ilocalizables o destinadas a otros fines se sustituía la persona titular seleccionada por una reserva del mismo grupo de edad.

b) Cuando se trataba de viviendas inaccesibles sólo se sustituía la persona seleccionada si no desaparecía la causa de la inaccesibilidad durante todo el tiempo que duraba el trabajo de campo en la sección.

c) Las personas fuera del ámbito de estudio (mayores de 49 y menores de 18 años) y las ilocalizables, se sustituyeron por otras personas reservas del mismo grupo.

En el caso de incidencias en personas:

a) En el grupo humano residente, las negativas e incapacidades para contestar se sustituyeron por reservas pertenecientes al mismo grupo de edad. Las ausencias sólo eran sustituidas cuando se confirmaba esta situación durante todo el tiempo que duraba el trabajo de campo en la sección.

b) En la persona seleccionada. Las personas que tenían alguna incapacidad para contestar, las negativas totales, las negativas parciales y las negativas por el uso del ordenador se sustituían por otras personas reservas del mismo grupo de edad. Las ausentes sólo se sustituían cuando se confirmaba que esta situación se iba a prolongar durante todo el periodo de tiempo que duraba el trabajo de campo en la sección.

La recogida de información se realizó durante 9 semanas (entre el 13 de octubre y el 19 de diciembre de 2003).

Seguimiento y control de calidad

Para comprobar sobre el terreno la marcha de la recogida de la información se realizaron periódicamente inspecciones de la visitas realizadas por los entrevistadores. Además se hicieron otras inspecciones ocasionales en los lugares donde hubo problemas en la recogida, o dudas para afrontar alguna situación difícil de tipo metodológico. La principal finalidad de la inspección fue comprobar que los entrevistadores realizaban correctamente su trabajo en la entrevista original, siguiendo las normas establecidas, comprobar y en su caso corregir la asignación de incidencias.

El inspector de la encuesta elaboró un plan de inspección, especificando las secciones a inspeccionar por orden cronológico. Los inspectores de entrevistadores debían inspeccionar al menos cuatro secciones por entrevistador durante todo el período de trabajo de recogida de datos. El inspector de la encuesta decidía las inspecciones que efectuaría personalmente, según las necesidades que surgieran durante los trabajos de campo. El número de inspecciones que debía realizar en esta situación era de tres secciones en total, en todo el período de recogida de datos. Si no disponía de inspector de entrevistadores, el inspector de la encuesta debía inspeccionar cinco secciones en total en todo el período para comprobar cómo habían realizado el trabajo los entrevistadores a su cargo. Una vez finalizado el plan de inspección se remitía a los servicios centrales.

Para comprobar si el entrevistador había entendido bien el mecanismo de la encuesta, se inspeccionaron tanto las encuestas realizadas como las sustituciones. Las personas seleccionadas para inspección eran las que figuraban en el parte de trabajo como encuestas terminadas. Los resultados de la inspección se anotaban en el correspondiente parte de inspección.

Con el objetivo de facilitar la realización de determinadas funciones de control y seguimiento de los trabajos de campo se elaboró una aplicación informática (ESHS-WIN). Esta aplicación permitió cargar las secciones seleccionadas en la muestra de cada provincia, determinar a qué grupo de semanas pertenecía cada una de ellas, asignar entrevistador a cada grupo de semanas, imprimir los partes de trabajo y los de inspección, grabar los partes de costes y efectuar su envío a servicios centrales, obtener resúmenes de la marcha de los trabajos, e imprimir o consultar los calendarios. Se elaboró un manual de la aplicación (ESHS-WIN) para conocer su uso y resolver dudas. La aplicación, como instrumento de seguimiento y control, fue gestionada y controlada por el inspector de la encuesta.

5. Respuesta a la encuesta

La tasa de respuesta

La tasa de respuesta de una encuesta y su distribución según diferentes variables relevantes son medidas importantes de la representatividad de la muestra efectiva. Una de las principales aportaciones del diseño de las encuestas con muestra probabilística, como es el caso de la ESHS, es la posibilidad de poder calcular tasas de respuesta teniendo en cuenta las diferentes situaciones que pueden haber llevado a que las personas concretas seleccionadas a priori para contestar la encuesta no lo hagan finalmente. Esta posibilidad está mucho más limitada o es incluso inexistente en otro tipo de diseños como las encuestas telefónicas o las encuestas con cuotas y rutas aleatorias.

Para lograr las 10.980 entrevistas válidas finales hubo que visitar 30.887 viviendas que constaban en el listado como viviendas familiares principales en las que residía alguna persona entre 18 y 49 años. De esas 10.980 entrevistas válidas, 5.263 (47,9%) se realizaron en personas que habían sido designadas como titulares en la muestra teórica inicial y 5.717 (52,1%) en personas señaladas como reservas (Tabla II.5.1). Esto supuso alcanzar un 80,7% de muestra efectiva con respecto a la muestra teórica que se había planteado inicialmente (13.600). Del total de titulares sólo se logró encuestar al 39%, cifra realmente baja si la comparamos con las de otras encuestas llevadas a cabo por el INE. Ahora bien, hay que tener en cuenta que esta encuesta iba dirigida a una persona concreta y únicamente podía ser contestada por dicha persona seleccionada a priori, diseño que no ha sido habitual hasta el momento.

El porcentaje de muestra efectiva lograda no fue homogéneo por comunidades autó-

nomas. Así, tres de ellas alcanzaron porcentajes superiores al 90% (Región de Murcia, Principado de Asturias y Cantabria), sobresaliendo Murcia, con más de un 96%. En el extremo opuesto, con porcentajes de muestra efectiva total inferiores al 70%, destacan Madrid con un 66% y Cataluña con un 69% (Tabla II.5.2).

No existe una forma consensuada de calcular la tasa de respuesta y la selección de una única tasa dificulta en ocasiones la comparación con otras encuestas que optan por definiciones no idénticas, por lo que se ha preferido realizar cuatro tasas diferentes indicando cómo se ha calculado cada una de ellas (Figura II.5.1). Una definición detallada de los diferentes tipos de incidencias puede encontrarse en el Manual del entrevistador. Todas las tasas están calculadas respecto del total de personas titulares y reservas.

La tasa menos ventajosa, calculada como el cociente entre las entrevistas realizadas y el número total de viviendas visitadas, fue del 35,5%. Sin embargo, es importante tener en cuenta que de las 30.887 viviendas visitadas, se registraron incidencias en el marco muestral en 6.464. Estas incidencias fueron debidas a diferentes razones: bien la vivienda estaba destinada a otros fines, era ilocalizable, inaccesible o estaba vacía, o bien la persona seleccionada tenía una edad fuera del intervalo seleccionado o, lo que era más frecuente, estaba ilocalizable. Esta última situación, en que la persona ya no residía en la dirección que figuraba en la relación de personas de la muestra, supuso un 80% del total de las incidencias de marco. No debemos olvidar que la población objeto de estudio, personas de 18 a 49 años, es la que tiene más movilidad territorial, ya que puede efectuar frecuentes cambios de residencia por motivos laborales,

Tabla II.5.1 Distribución de las viviendas visitadas y de la muestra efectiva

	<i>Total</i>	<i>Titulares</i>	<i>Reservas</i>
<i>Viviendas visitadas</i>	30.887	13.600	17.287
<i>Incidencias</i>	19.907	8.337	11.570
<i>Entrevistas realizadas</i>	10.980	5.263	5.717

Tabla II.5.2 Muestra teórica y muestra efectiva por comunidad autónoma

	Muestra teórica	Muestra efectiva			
		Total (titulares + reservas)		Personas titulares	
		Personas	%	Personas	%
Total nacional	13.600	10.980	80,74	5.263	38,70
Andalucía	1.800	1.561	86,72	820	45,56
Aragón	544	434	79,78	195	35,85
Asturias (Ppdo. de)	488	450	92,21	219	44,88
Baleares (Illes)	472	356	75,42	168	35,59
Canarias	672	535	79,61	230	34,23
Cantabria	392	354	90,31	158	40,31
Castilla y León	768	661	86,07	333	43,36
Castilla-La Mancha	624	544	87,18	262	41,99
Cataluña	1.552	1.071	69,01	475	30,61
Comunidad Valenciana	1.136	960	84,51	456	40,14
Extremadura	488	423	86,68	223	45,70
Galicia	784	692	88,27	325	41,45
Madrid (Com. de)	1.520	1.007	66,25	433	28,49
Murcia (Región de)	544	525	96,51	298	54,78
Navarra (Com. Foral de)	400	306	76,50	151	37,75
País Vasco	736	594	80,71	271	36,82
Rioja (La)	352	249	70,74	121	34,38
Ceuta y Melilla	328	258	78,66	125	38,11

de estudio o personales, no llevando siempre asociado este cambio la correspondiente modificación en la inscripción padronal. Si se eliminan las incidencias de marco del denominador, teniendo en cuenta que la mayoría se debió a una falta de notificación de las personas sobre su cambio de domicilio, la tasa de respuesta asciende al 45%. Podría considerarse también otra tasa basada en el supuesto de que todas las ausencias de grupo fueran en realidad negativas encubiertas a contestar a la encuesta la persona seleccionada. Bajo esta asunción, y eliminando del denominador todas las ausencias de grupo, la tasa de respuesta ascendería al 49,6%.

Otra forma menos conservadora de calcular la tasa de respuesta es restringir el denominador de potenciales entrevistados a aquéllos en los que se pudo identificar que la persona seleccionada efectivamente vivía allí y podía teóricamente ser entrevistada o negarse a realizar la entrevista. Es decir, excluir del denominador, además de las incidencias de marco, las ausencias de todo

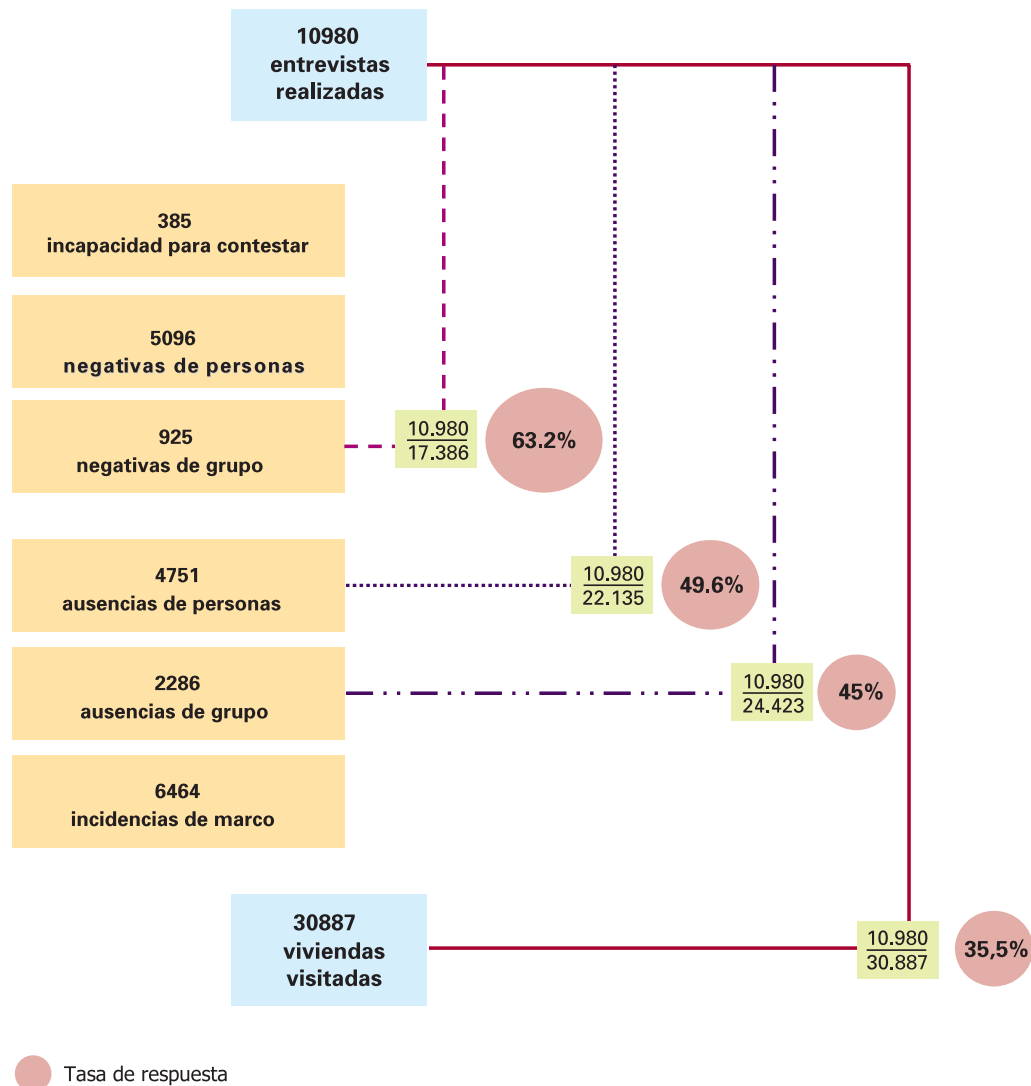
el grupo de convivientes y las de la persona concreta seleccionada. En este caso la tasa de aceptación se elevaría al 63,2%. Posiblemente sea este cálculo el que es más fácilmente comparable con la forma en que se ha calculado la tasa de respuesta en muchas de las encuestas sobre comportamientos sexuales en relación con el VIH realizadas en Europa^{1;11;21}.

La tasa de respuesta más utilizada en el ámbito de la estadística oficial se obtiene dividiendo la muestra efectiva de personas titulares (muestra finalmente obtenida de la lista original de titulares, es decir, 5.263) por la muestra teórica de titulares (13.600). Esta definición arroja una tasa de respuesta de 38,7%.

Análisis de la no respuesta

La figura II.5.1 permite analizar también la distribución de los diferentes tipos de incidencias que llevaron a que algunas personas inicialmente seleccionadas no contes-

Figura II.5.1 Estimación de la tasa de respuesta bajo diferentes asunciones



DEFINICIONES

- **Incapacidad para contestar:** cuando el seleccionado está incapacitado para responder, por discapacidad, enfermedad, desconocimiento del idioma, analfabetismo u otro motivo.
- **Negativa de la persona:** cuando el seleccionado se niega a cooperar parcial o totalmente. Se han distinguido tres tipos:
 1. *Negativa total:* se niega a facilitar cualquier tipo de información.
 2. *Negativa parcial:* contesta sólo a una parte del cuestionario, sin llegar a finalizarlo. Se considera que se produce esta incidencia cuando la parte de cuestionario contestada no llega a un mínimo fijado.
 3. *Negativa por empleo del ordenador:* cuando el seleccionado se niega a dar información debido al uso del ordenador, por reticencia, miedo o incapacidad para utilizarlo.
- **Negativa de grupo:** cuando en la vivienda del seleccionado hay un rechazo inicial a colaborar en la encuesta, de forma que no hay posibilidad de que la persona seleccionada se plantee colaborar
- **Ausencia de la persona:** cuando el seleccionado está ausente y va a continuar estándolo durante todo el periodo del trabajo de campo en la sección.
- **Ausencia de grupo:** cuando en la vivienda del seleccionado, no hay nadie en el momento de ir a realizar la entrevista, porque están ausentes todos sus ocupantes y van a seguir estándolo durante todo el periodo del trabajo de campo en la sección.
- **Incidencias de marco muestral:** cuando la vivienda es ilocalizable, está destinada a otros fines distintos de residencia familiar, es inaccesible o está vacía, o bien la persona seleccionada tiene una edad fuera del intervalo seleccionado o no reside ya allí, aunque formalmente continúe empadronada.

taran finalmente la encuesta. La incapacidad para contestar constituye un porcentaje muy bajo del total de incidencias (1,9%), mientras que las otras categorías podrían agruparse en tres grandes grupos con porcentajes muy similares: las incidencias de marco (32,5%), las ausencias (35,4%) y las negativas (30,3%). De esta información se dispone tanto para titulares como para reservas, aunque la distribución del tipo de incidencia no mostró diferencias relevantes entre ambos grupos.

Con el fin de profundizar en el análisis de la no respuesta, se diseñó un cuestionario de evaluación que se cumplimentó únicamente para los titulares que no colaboraron en la encuesta, pero no para los reservas. Se trata de un documento con el que se pretendía obtener información sobre las características sociodemográficas básicas de estas personas. Para el análisis de la falta de respuesta se comparó la distribución de los diferentes tipos de incidencias en las personas (no se consideraron las incidencias de marco) según las principales variables sociodemográficas con la distribución de la población española en el censo de 2001 con objeto de analizar posibles sesgos en la información obtenida en la ESHS. Un análisis detallado puede encontrarse en el documento *Evaluación de la falta de respuesta en la Encuesta de Salud y Hábitos Sexuales*⁶⁷. Los datos más relevantes fruto de la comparación de estas dos distribuciones indican que las ausencias se produjeron fundamentalmente en varones mientras las negativas se concentraron más en las mujeres. Además, la mayor proporción de negativas se observó entre los casados y la de ausencias entre los solteros. Con respecto al nivel de estudios, las negativas predominaron en las personas con estudios primarios, mientras las ausencias no mostraron una predominancia tan clara en uno y a veces dos de los niveles de estudios recogidos.

6. Preparación de datos para el análisis

La realización de la entrevista por ordenador permitió un volcado directo de los cuestionarios al programa de depuración y análisis, obviando así posibles errores cometidos en el proceso de informatización de los cuestionarios.

La administración por ordenador, incorporando saltos y filtros automáticos, así como la restricción en las posibles respuestas a cada pregunta, evitando la grabación de valores fuera de rango, incrementa el trabajo en la fase de diseño de instrumentos, pero disminuye sustancialmente los errores de cumplimentación y, por lo tanto, el trabajo de detección y corrección de los mismos. Sin embargo, en la programación del cuestionario se optó por incluir sólo una parte de las consistencias lógicas que un análisis detallado del proceso habría sugerido. Esto se hizo por un lado por el enlentecimiento que habría supuesto para el programa de grabación del portátil, y por otro, por las repercusiones negativas que podría haber planteado para algunos entrevistados. De hecho adoptar la estrategia de buscar la concordancia total entre las respuestas hubiera obligado a muchos individuos a volver varias veces hacia atrás en el cuestionario y repetir secciones enteras del mismo hasta lograr respuestas totalmente coherentes, lo que quizá se hubiera traducido en una percepción de control externo excesivo que podría haber provocado rechazo.

Una vez incorporados los cuestionarios a la aplicación de detección y depuración, se realizó en primer lugar un detallado análisis descriptivo de frecuencias de todas las variables para el conjunto de la muestra, así como estratificando por sexo, con objeto de realizar una depuración.

La depuración del fichero con todos los registros constó de tres fases:

1. Eliminación de registros con insuficiente información

El cuestionario electrónico incorporaba controles para evitar que las personas que no quiesieran aportar información contaminaran la encuesta con respuestas *No sabe o no contesta*. Para ello, el cuestionario electrónico exigía, por un lado, que la pre-

gunta inicial de los bloques más importantes de la encuesta tuvieran respuesta válida. Por otro lado, iba contabilizando el número de respuestas *No sabe o no contesta*, de forma que si se superaba un porcentaje previamente determinado, en el momento en que finalizaba la entrevista el ordenador identificaba el cuestionario como no válido. Los casos en que no se lograba recabar esta información mínima, se contabilizaban como incidencia de negativa parcial, que daba lugar a sustitución. Esta situación se dio en un total de 49 ocasiones.

No obstante, estos controles para evitar cuestionarios inválidos no eran excesivamente exigentes, por un lado para mantener la operatividad del programa de grabación, y por otro para evitar la sensación de control excesivo. De ahí que en la fase de depuración se detectaran otros 142 cuestionarios inservibles para el análisis. Éstos fueron eliminados del fichero, quedando la muestra efectiva reducida de 10.980 a 10.838 registros.

2. Detección y depuración de inconsistencias (errores aleatorios)

El equipo de trabajo de la Secretaría del Plan Nacional sobre el Sida y el Instituto Nacional de Estadística elaboró una exhaustiva lista de relaciones entre variables poco probables o imposibles que, en caso de ser observadas en los cuestionarios, debían ser revisadas. A continuación se comprobó, en la medida de lo posible, su veracidad mediante el análisis del resto de variables del cuestionario implicado. Este estudio se hizo para cada uno de los registros que presentó alguna inconsistencia.

3. Imputación de errores determinísticos

Durante el análisis de las frecuencias de las variables del cuestionario y de las inconsistencias definidas en el paso anterior se detectaron algunos errores no aleatorios, que seguían un patrón. Estos casos se corrigieron mediante programas que asignaban a las variables erróneas un valor válido coherente con el resto del cuestionario.

Esta situación se detectó en la variable *motivo de realización de la última prueba*

del VIH. Se encontraron nueve hombres que contestaron que se hicieron la prueba porque *Estaba embarazada*. Estos casos se imputaron a la categoría de respuesta *otro motivo*. También se consideró que las 70 mujeres y 15 hombres que habían señalado un inicio de las relaciones sexuales inferior a los 10 años podrían haber tenido un error de cumplimentación con el lápiz óptico y estas respuestas se reemplazaron por *no consta*.

La fase de depuración afectó a variables aisladas de aproximadamente un 4% de los registros del fichero.

Se decidió no realizar ningún tipo de imputación para los valores perdidos de las variables, sino calcular los porcentajes sobre el total de las personas con información para la variable que se evaluaba en cada caso. Este criterio para el cálculo de indicadores, utilizado por muchos institutos de estadística y recomendado por Eurostat (la Oficina de Estadística de la Comunidad Europea), equivale a suponer que el comportamiento de los registros con valores perdidos es similar al de aquéllos con valores válidos.

La tasa de falta de respuesta parcial o por pregunta resultó ser inferior al 2% para la mayoría de las variables, sin embargo fue superior al 5% en algunas, como el uso de preservativos con las parejas ocasionales o el número de copas con alcohol que se acostumbra a beber al día, e incluso superó el 10% en el número de parejas tenidas en los últimos doce meses, el número de días con relaciones sexuales en los últimos 30 días o la edad a la que se iniciaron las relaciones sexuales. La variable con mayor porcentaje de falta de respuesta se encuentra en la sección de hombres con relaciones homosexuales, en la que se desconoce un 19% de las respuestas acerca de la edad que tenía el hombre con el que el entrevistado tuvo su primera relación homosexual.

7. Análisis y presentación de datos

Una vez realizados todos los procesos anteriores se elaboró un amplio plan de explotación que incluyó las siguientes tareas. Primero, definir nuevas variables de análisis sobre comportamientos, actitudes, opiniones o conocimientos que fueran relevantes por su significado para los objetivos principales del estudio. En segundo lugar, seleccionar cuáles iban a ser las variables de clasificación sociodemográficas o de estilos de vida elegidas para cruzar con las variables de análisis. En tercer lugar, realizar una agregación de las categorías de muchas variables en función de las frecuencias muestrales de sus cruces. Este proceso fue especialmente relevante para algunas de las variables sociodemográficas que iban a emplearse como variables fundamentales de clasificación. Así, por ejemplo, la edad fue agregada inicialmente en siete grupos para ser finalmente reducida a los tres que se observan en todas las tablas de resultados. Un proceso similar se planteó para el nivel de estudios o el tamaño del hábitat. En cuarto lugar, decidir el nivel máximo de estratificación que parecía oportuno publicar teniendo en cuenta las distribuciones de frecuencias de forma que se dispusiera de suficientes efectivos en todos los estratos.

Todos los análisis se han realizado estratificando por sexo. Además, teniendo en cuenta el último criterio mencionado, en este informe se presentan únicamente algunos de los cruces de las variables de análisis más relevantes con variables sociodemográficas o de estilos de vida. Los indicadores seleccionados son los de la comunicación con los padres, las fuentes de información preferidas y reales sobre temas sexuales, el tipo de práctica sexual (homo, hetero o ambas), la edad de inicio de las relaciones sexuales, el uso del preservativo en la primera y en la última relación sexual, el número de parejas sexuales,

el uso de prostitución, las medidas de prevención con las nuevas parejas sexuales, la realización de la prueba del VIH, el conocimiento de los mecanismos de transmisión, la actitud discriminatoria, el conocimiento de las medidas de prevención y las opiniones sobre el preservativo. Los cruces de estas variables se realizan únicamente con:

- Sexo y edad
 - Sexo, edad y nivel de estudios, aunque en algunos indicadores se suprimió la estratificación por edad limitándose a sexo y nivel de estudios
 - Sexo y tamaño de municipio
 - Sexo y tipo de convivencia
- Sexo y estilos de vida: salir de noche en los últimos doce meses, dormir fuera de casa por razones de trabajo o de estudios en los últimos doce meses y la frecuencia con la que el encuestado estuvo bebido el último mes
- Sexo y comunidad autónoma, aunque sólo para un número reducido de indicadores

El análisis que se ha realizado es únicamente descriptivo, intentando mostrar los resultados más interesantes. No se han planteado para este informe otro tipo de análisis multivariantes más complejos y de interés para analizar los factores asociados a determinadas prácticas, opiniones o actitudes. Todos los resultados se muestran con porcentajes estimados en base a los datos elevados de la población y sin incluir en su cálculo los valores perdidos, pues como ya se ha señalado, no se han realizado imputaciones para éstos.

En www.ine.es se puede encontrar una versión completa de las tablas descriptivas realizadas de acuerdo con el plan de explotación.