

## **Los enlaces de registros mejoran la calidad de las encuestas: la experiencia del Instituto de Estadística y Cartografía de Andalucía en la Encuesta Social 2010**

**Rubén Martín Gimeno  
María Macarena Muñoz Conde  
Rosa Bohórquez Baña**

Instituto de Estadística y Cartografía de Andalucía

Correos electrónicos: [ruben.martin.ext@juntadeandalucia.es](mailto:ruben.martin.ext@juntadeandalucia.es)  
[mmacarena.munoz@juntadeandalucia.es](mailto:mmacarena.munoz@juntadeandalucia.es)  
[rosa.bohorquez@juntadeandalucia.es](mailto:rosa.bohorquez@juntadeandalucia.es)

Dirección: Servicio de Estudios, Síntesis y Métodos Estadísticos  
Instituto de Estadística de Andalucía.  
Pabellón de Nueva Zelanda.  
Leonardo Da Vinci, 21  
Isla de la Cartuja  
41071-Sevilla

Los enlaces de registros mejoran la calidad de las encuestas: la experiencia del Instituto de Estadística y Cartografía de Andalucía en la Encuesta Social 2010.

**RESUMEN:** El interés por mejorar la calidad de los datos intentando evitar sesgos de distinta índole ha acompañado a las investigaciones realizadas mediante encuestas desde sus orígenes. A lo largo del progreso de la disciplina se han ido desarrollando y refinando acciones metodológicas para intentar conseguir datos de calidad. En las encuestas sociales del Instituto de Estadística y Cartografía de Andalucía se han incorporado estrategias para reducir los errores provocados por los problemas de no cobertura y de no respuesta. Estas acciones han sido posibles gracias al enlace del marco de población con sistemas de información gestionados por la administración pública. La primera experiencia de enlace, puesta en marcha en la “Encuesta Social 2010. Educación y Hogares en Andalucía”, ha sido muy positiva ya que se incrementaron considerablemente las tasas de respuesta situándose en el 72,7%. En este artículo se explican los procesos que se han llevado a cabo para obtener esta mejora.

**PALABRAS CLAVE:** Enlace de registro, Encuesta Social 2010, Instituto de Estadística y Cartografía de Andalucía, Errores de no respuesta, Tasa de respuesta.

Record linkages improve the quality of surveys: experience of the Institute of Statistics and Cartography of Andalusia in the Social Survey 2010.

**ABSTRACT:** Since the origins of survey research, the discipline has been interested in improving quality of data, trying to reduce different types of bias. Methodologies aiming at improving quality of data have been gradually incorporated over the progress of this discipline. New strategies to reduce errors caused by undercoverage and non-response in units sample have been used by the Institute of Statistics and Cartography of Andalusia in social surveys. These strategies have only been possible thanks to the record linkage between population frame and other information systems managed by Public Administration. The first experience of record linkage in the “Social Survey 2010: Education and Housing in Andalusia” has been very positive: the response rate has considerably increased standing at 72,7%. The aim of this article is to show the different processes leading to achieve these results.

**KEYWORDS:** Record linkage, Social Survey 2010, Institute of Statistics and Cartography of Andalusia, Non-response errors, Response rate.

**Recibido:** 28 de julio de 2014

**Revisado:** 18 de noviembre de 2014

**Aceptado:** 3 de diciembre de 2014

## 1. Introducción

El presente artículo tiene como objetivo principal explicar una serie de estrategias metodológicas que se pusieron en marcha en la “Encuesta Social 2010. Educación y Hogares en Andalucía” (en adelante Encuesta Social 2010) del Instituto de Estadística y Cartografía de Andalucía (IECA), y que ayudaron a mejorar la calidad de los datos de esta encuesta. Con ello se pretende mostrar una posible vía para mejorar las tasas de respuesta de las encuestas realizadas desde la estadística oficial en cualquier ámbito territorial.

Antes de adentrarse en el tema del artículo es conveniente describir brevemente las características generales de la Encuesta Social 2010<sup>2</sup>. El tema principal de esta encuesta ha sido conocer cuáles son los factores dentro de las familias que pueden influir en los distintos rendimientos educativos de los alumnos que estudian en Andalucía. La búsqueda de estos factores parte de una hipótesis central: existe relación entre los estilos educativos, el origen social de las familias y los resultados escolares de alumnos y alumnas. Con este objetivo se diseñó una muestra de 6.000 viviendas con alumnos nacidos en 1998 (11-12 años) y 1994 (15-16 años) en las que se les entrevistaba a ellos y a uno de sus progenitores. La elección de dos cohortes generacionales, cuyas edades corresponden escolarmente al final de la etapa primaria (los nacidos en 1998) y al final de la etapa secundaria (los nacidos en 1994), permite afrontar el objetivo de estudio desde una perspectiva longitudinal, tanto en el análisis de las trayectorias pasadas como en el seguimiento de estas cohortes en estudios futuros<sup>3</sup>.

Para el tema que nos ocupa en el presente artículo, la principal novedad metodológica de la encuesta ha sido la utilización de un sistema de información auxiliar para enlazar con los datos del Registro de Población de Andalucía (en adelante RPA)<sup>4</sup>, marco poblacional del que se extrajo la muestra. Este sistema de información denominado Séneca fue creado por la Consejería de Educación de la Junta de Andalucía<sup>5</sup> y cuenta con información incorporada por los distintos agentes educativos de los centros públicos y concertados (profesores, personal administrativo y personal directivo) relativos al seguimiento de los alumnos en distintos aspectos de su trayectoria educativa (rendimiento, comportamiento, competencias, etc.).

---

<sup>2</sup> La ficha técnica de la encuesta se puede consultar en el Anexo I.

<sup>3</sup> Para ver las ventajas de este tipo de estudios: Díaz de Rada, 2007: 137-139.

<sup>4</sup> El Registro de Población de Andalucía tiene como finalidad la comunicación de los diferentes órganos de la Administración de la Comunidad Autónoma de Andalucía con los interesados residentes en su territorio. Contiene los datos del nombre, apellidos, domicilios y fecha de nacimiento que constan en los padrones municipales de habitantes y en el censo electoral de todos los Ayuntamientos de la Comunidad Autónoma de Andalucía. Para la mejora de estos datos se realizan procesos de depuración y mejora mediante el uso de otras fuentes estadísticas y administrativas.

<sup>5</sup> Este enlace entre el marco de población y los datos de la encuesta con el Sistema de Información Séneca se ha realizado en el marco de colaboración institucional entre la Consejería de Educación de la Junta de Andalucía y el IECA. Por parte del IECA, los trabajos de enlace fueron realizados por técnicos del Servicio de Estadísticas Demográficas y Sociales con la colaboración de técnicos del Servicio de Estudios, Síntesis y Métodos Estadísticos.

El enlace de la información contenida en Séneca con la extraída del marco de población, configurando un único registro, ha permitido abrir interesantes vías de análisis de tipo longitudinal así como la posibilidad de establecer procedimientos metodológicos encaminados a mejorar la calidad de los datos. En este artículo se explicará detalladamente los procedimientos seguidos en este enlace y cómo se aprovechó la información de Séneca para reducir considerablemente las incidencias en los trabajos de campo.

## 2. Los intentos por reducir los errores en las encuestas

El estudio de las causas y los modos de reducir los distintos tipos de errores que se producen en las encuestas ha tenido un tratamiento desigual. Por ejemplo, la medición del efecto de los errores de observación<sup>6</sup> tuvo un temprano desarrollo en la literatura experta. Desde las primeras décadas del siglo pasado se iniciaron estudios experimentales en los que se trataba de medir los sesgos introducidos por los encuestadores y la incidencia de éstos dependiendo del tipo de encuestado (Cea D'Ancona, 2005:79-83). La mayoría de este tipo de experimentos apuntaba a que, efectivamente, se producen sesgos dependiendo de las características personales de encuestadores y encuestados y del rumbo que toma la interacción durante la situación de encuesta (Atienza y Noya, 1999:78-88). Y esto afecta especialmente a cierto tipo de cuestionarios o preguntas, principalmente cuando hay que posicionarse valorativa o ideológicamente (Alvira y Martínez, 1985:219-231).

También son numerosos los estudios que han puesto de relieve las diferencias en los datos atendiendo al modo en que se redacta la pregunta (Schuman y Presser, 1981) y a los tipos de escalas que se utilizan para medir los grados de intensidad en las acciones u opiniones de la población (Filgueira, 2001).

La realización de este tipo de estudios ha tenido efecto en las prácticas empleadas por investigadores e instituciones en distintos aspectos del diseño metodológico: realización de pre-test para mejorar los cuestionarios, formación de los encuestadores, supervisión y control de los encuestadores, etc. (Jäckle et al. 2011:1-4). En nuestro país hay pocos estudios dedicados a la labor de los encuestadores, por lo que gran parte de las acciones que se han desarrollado para mejorar su trabajo se han diseñado conforme a las experiencias y estudios internacionales (Díaz de Rada, 2006:241-242). Otro terreno en el que hay mucho que explorar es el de los errores provocados por factores que trascienden el momento de encuesta y que tienen su causa en la naturaleza del contexto en el que se realiza la investigación, como por ejemplo, las condiciones de trabajo de los encuestadores o las dificultades organizativas de una investigación (Fernández, 2003:57-62).

---

<sup>6</sup> Entendemos por errores de observación aquellos que tienen su origen en fallos causados por el instrumento de medida (Groves, 1989). Este error puede estar originado por cuatro motivos: errores producidos por los entrevistadores, por los entrevistados, errores debidos al cuestionario y errores producidos por el procedimiento de recogida de datos. Se distinguen de los errores de no observación, que son aquellos que surgen por no seleccionar correctamente una muestra de la población objeto de estudio.

Respecto a los errores de no observación, en la literatura especializada el protagonismo ha recaído en la medición del error muestral. En un segundo plano han quedado los estudios y los intentos por encontrar soluciones a los errores de no cobertura y no respuesta que afectan también de manera relevante a la representatividad de la muestra y a las posibilidades de inferencia de los datos (Cea D'Ancona, 2005:78). En nuestro país investigadores e instituciones han realizado distintos diagnósticos dependiendo del método de recogida. Por ejemplo, se han ido introduciendo análisis que profundizan en el intento de sistematizar el estudio de las incidencias que provocan este tipo de errores. Ejemplo de estas instituciones son el Instituto Nacional de Estadística (Ballano y Rebollo, 2006) o el Instituto de Estudios Sociales de Andalucía (Trujillo y Gutiérrez, 2006). También estudios en los que se comparan las incidencias y niveles de respuesta según el canal de recogida de datos en encuestas del CIS (Díaz de Rada, 2010: 75-122). A nivel internacional, por ejemplo, se pueden encontrar intentos sistemáticos de mejora en la metodología de las encuestas en las sucesivas oleadas de la Encuesta Social Europea (Cuxart y Riba, 2009) o en los esfuerzos del Institute for Social and Economic Research (ISER) de origen británico, por sistematizar la medición de los errores de no respuesta<sup>7</sup> (Lynn et al. 2001).

Los errores de no observación producidos por el marco no se pueden tratar de manera homogénea porque cada método de recogida presenta problemas distintos. En las encuestas telefónicas el problema principal ha sido la imposibilidad de encontrar un marco en el que estuviesen incluidos todos los sectores sociales: quedaban excluidos los que no tenían teléfono que a su vez correspondían con estratos muy concretos de la sociedad (Díaz de Rada, 2001: 140). A este problema se ha añadido otro en los últimos años: el crecimiento de la telefonía móvil. No hay listados específicos en nuestro país que recojan estos números, lo que dificulta la posibilidad de diseñar muestras fiables. No obstante, como se ha comprobado en distintos estudios empíricos, la generalización de este tipo de teléfonos también puede tener ventajas: acceden a la telefonía perfiles de población que anteriormente no lo hacían (Pasadas del Amo, 2006: 140-145).

Además, la incorporación de internet ha abierto nuevos retos en este sentido. Las encuestas vía web o por correo electrónico se han incrementado conforme lo ha hecho también el acceso de la población a la red. Este incremento, que se está produciendo con celeridad, no es óbice para que aún muchos sectores de la población estén al margen de esta tecnología, por lo que todavía quedan importantes errores de cobertura que no se pueden esquivar con este método de encuesta (Díaz de Rada, 2012:198-200). Por ello las encuestas vía web o por correo electrónico se están constituyendo como un modo muy útil de recogida, en combinación con otros métodos de encuesta. Este método mixto o multicanal, permite en muchos casos reducir costes, tiempo de ejecución y también, en situaciones concretas, errores de cobertura y no respuesta (Dillman et al. 2009: 302-306).

Por último, en el caso de las encuestas presenciales el problema del marco muestral también ha estado presente. Esto ocurre principalmente en las encuestas a

---

<sup>7</sup> Con esta metodología se ha hecho el cálculo de la tasa de respuesta que se muestra en el apartado 4.

hogares en las que se estudia a la población en general. La ausencia de listados en los que se contemple con detalle las características de la población de estudio, provoca tener que recurrir a métodos de selección de la muestra que pueden reducir su representatividad.

No obstante, en las encuestas presenciales del IECA este problema está, hasta cierto punto, minimizado. Como organismo sometido al marco normativo de la Función Estadística Pública tiene acceso a un marco detallado, el Registro de Población de Andalucía (RPA) referido anteriormente, en el que consta información de las personas (nombre, sexo y edad) y sus direcciones. Esto permite realizar muestreos polietápicos que mejoran la representatividad de la muestra y que además facilitan la recogida de la información ya que permiten reducir considerablemente los costes materiales y de tiempo (Díaz de Rada, 2008: 213).

Sin embargo, la disponibilidad de este marco no evita que se sigan produciendo errores de cobertura y no respuesta<sup>8</sup>. Las razones pueden ser diversas. Una de las principales es que en numerosas ocasiones el padrón, fuente principal en la que se basa el RPA, no está actualizado, principalmente porque las personas no renuevan la información cuando realizan un cambio de domicilio; a pesar de que la información del RPA está enriquecida con otras fuentes además del Padrón Municipal de Habitantes, en muchas ocasiones no se puede conocer las variaciones en las pautas residenciales de la población. Otro factor que también tiene una incidencia importante es el que viene ocasionado por la dificultad de acceder a las viviendas de algunos sectores sociales, incluso diseñando estrategias específicas de contacto para este tipo de poblaciones<sup>9</sup>. Además de éstas hay otras causas debidas a defectos del marco: cambio en el nombre de las calles, direcciones registradas de manera incorrecta, etc.

Por tanto, disponer de un listado completo de direcciones y viviendas no soluciona por completo este tipo de errores. Como consecuencia de ello desde el IECA se han desplegado distintos procedimientos para mejorar las tasas de respuesta en las encuestas sociales. Además de ir perfeccionando los distintos procesos que afectan a la calidad de los datos (mejoras en la formación de los equipos de trabajo, en el diseño del cuestionario, en la inspección y el seguimiento de las encuestas, etc.), se instauró un procedimiento multicanal de encuestas para mejorar la cobertura a la población: aunque el procedimiento principal era la encuesta

---

<sup>8</sup> Esta denominación de los errores es la que estableció Groves y que reproduce Díaz de Rada (Díaz de Rada, 2004: 46-55). Los “errores de cobertura” son aquellos que vienen provocados cuando algunos elementos de la población objeto de estudio no están incluidos en el proceso de selección muestral. Los “errores de no respuesta” son los producidos por la ausencia de respuesta. En este apartado se pondrá atención en la no respuesta total, es decir, cuando una persona, por distintas circunstancias, no contesta ninguna pregunta del cuestionario; no se atenderá a la no respuesta parcial, que se produce cuando el entrevistado que estaba contestando el cuestionario decide no responder a determinadas cuestiones.

<sup>9</sup> En este sentido, las poblaciones con mayor dificultad para contactar detectadas en las encuestas sociales del IECA en Andalucía han sido las siguientes: a) temporeros, en buena medida de origen extranjero, que se empadronan por breves estancias de tiempo y b) residentes de temporada, de origen inglés y alemán principalmente, como ocurre en el caso de la Costa del Sol. Para estos y otros tipos de poblaciones (barrios marginales, zonas residenciales de clases altas con dificultades de acceso, zonas de veraneo, etc.) se llevan a cabo procesos especiales de localización que en muchas ocasiones son efectivos, pero no siempre.

presencial asistida por ordenador (CAPI), la población también podía hacer la encuesta vía web y en casos puntuales cuando así lo requiriesen, telefónicamente. En la Encuesta Social 2010 el impacto de las encuestas vía web sobre la muestra total fue del 7,7%, porcentaje elevado si se tiene en cuenta que era un método auxiliar de recogida y que era la primera experiencia web en las encuestas sociales del IECA.

Pero la principal novedad en la Encuesta Social 2010, como hemos comentado anteriormente, fue la posibilidad de tener otro marco alternativo con nueva información. La clave de todo ello fue la posibilidad de realizar un enlace entre ambos registros, RPA y Séneca, y disponer de teléfonos a través de Séneca para poder establecer contacto con la muestra antes de que los encuestadores comenzaran los trabajos de campo. A continuación se explicará cómo se llevaron a cabo estos trabajos y los efectos que tuvo en la reducción de las incidencias durante el campo.

### **3. Metodología: Cómo el enlace de registros puede mejorar la calidad de las encuestas**

Por enlace de registros se entiende la combinación de información que proviene de registros disponibles en bases de datos informatizadas (Winkler, 2006:1). La conciencia de la importancia del enlace de los registros para reconstruir la vida de la población a través de sus principales hitos vitales (nacimiento, trayectoria educativa, vida laboral, matrimonio, divorcios, fallecimientos, etc.) se remonta a las primeras décadas del siglo pasado (Dunn, 1946). Desde entonces se han depurado en gran medida los métodos para vincular fuentes de datos. Esto ha permitido realizar en muchos países estudios de gran calado analítico, especialmente aquellos que se llevan a cabo con enfoques longitudinales para estudios de índole económica, social o demográfica. En este sentido encontramos estudios realizados en Suecia (Hjelm, 2006), Canadá (Ryten, 2006), Inglaterra (Knies et al., 2011) o Estados Unidos (United States Census Bureau, 2012).

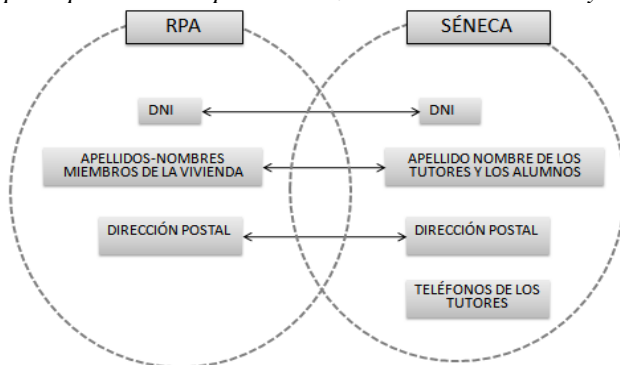
En España, la utilización de los registros se empezó a intensificar en los años 90 con distintos objetivos (Salaregui, 2006): diseño de marcos de población y de empresas (Porrás y Hernández, 2006), estadísticas derivadas de la gestión administrativa o como fuente auxiliar en encuestas por muestreo. En muchas de estas operaciones el enlace entre distintos registros ha sido fundamental. En nuestro país también se han realizado importantes operaciones desde la perspectiva longitudinal. A nivel autonómico, el Instituto de Estadística y Cartografía de Andalucía en las “Estadísticas de Trayectorias Vitales” reconstruye biografías integrando estadísticas demográficas de distintas fuentes (Viciano et al., 2005). Por la similitud con el enlace que se ha realizado en la Encuesta Social 2010, dentro de la experiencia española hay que destacar los trabajos desarrollados en la Encuesta Anual de Estructura Salarial, ya que integra información de origen muestral (Encuesta Trimestral de Coste Laboral) con información de origen administrativo (datos del Fichero General de Afiliación de la Seguridad Social y datos de la Agencia Tributaria) con el objetivo de evitar carga de respuesta en los encuestados y obtener series temporales para estudiar las evoluciones de estas personas en el tiempo (Moral-Arce y Martín, 2009: 488-500).

Por tanto, el uso de enlaces para la integración de la información con fines analíticos y como método para depurar y mejorar la información de marcos muestrales también se ha consolidado en nuestro país. En este sentido, el uso del enlace en la Encuesta Social 2010 sigue esta línea de acción, e incorpora como aspecto más novedoso la utilización de datos de contacto (teléfonos de los tutores legales del Sistema de Información Séneca) para mejorar las tasas de respuesta. A continuación se explica cómo se realizó el enlace para esta encuesta.

### 3.1 *Cómo se realizó el enlace*

En una primera fase de los trabajos, desarrollada antes de los trabajos de campo, los procesos se centraron en generar el enlace a través del conjunto de campos o variables que RPA y Séneca comparten. Para ello era fundamental la información nominal: el nombre, la fecha de nacimiento y el DNI de los hijos y de sus padres. En Séneca esta información estaba registrada ya que procedía de los datos que los tutores facilitan en el proceso de matriculación de sus hijos. La Consejería de Educación de la Junta de Andalucía remitió al IECA la información de todos los alumnos nacidos en 1994 y 1998<sup>10</sup> que constan en Séneca y de sus tutores<sup>11</sup>. Una vez se dispuso de esta información, desde el IECA se realizó el enlace con la muestra de la encuesta extraída del RPA: 6.000 unidades de muestra titular y 8.382 unidades de muestra reserva.

Gráfico 1.  
*Campos a partir de los que se realizó el enlace entre RPA y Séneca*



Para realizar estos trabajos se llevaron a cabo procesos de normalización de las variables de Séneca estableciendo como criterio el modo en que la información está

<sup>10</sup> Algunas de las personas nacidas en 1994 o 1998 que constaban en el RPA no estaban registradas en Séneca. La mayor parte de estos menores correspondía a alumnos que estaban matriculados en la enseñanza privada; aunque estos alumnos y sus familias no estaban registrados en Séneca también formaban parte de la muestra.

<sup>11</sup> Para la transmisión de la información entre ambos organismos se utilizaron herramientas de encriptación informática con el fin de asegurar los protocolos de protección de datos de carácter personal y salvaguarda del secreto estadístico.



registrada en el RPA. El objetivo era evitar discordancias provocadas por la forma en que está implementada la información en cada uno de los registros. En el caso de los nombres y apellidos se eliminaron caracteres extraños, tildes y partículas de los nombres y los apellidos. Con los documentos de identidad se comprobó si su longitud era la adecuada, ocho dígitos, completándose con ceros a la izquierda en los casos necesarios. Estas acciones incrementaban considerablemente las posibilidades de éxito del enlace desde el primer momento.

Además, en esta fase también se añadió en el enlace los números de teléfonos fijos y móviles que los tutores facilitan a los centros en los que estudian sus hijos. Esta información de contacto es muy fiable ya que se actualiza desde los centros para tener siempre la posibilidad de contactar con los tutores. Como se observa en la tabla 1 el proceso tuvo un alto porcentaje de éxito: después del enlace se disponía de algún teléfono de contacto en cerca del 90% de las viviendas. La disponibilidad de teléfonos permitió articular estrategias de contacto, que son las que se detallan en el siguiente apartado.

Tabla 1.

*Resultado del enlace de la muestra entre RPA y sistema de información Séneca*

	Total de viviendas		Viviendas Principales		Viviendas Reservas	
	Viviendas	%	Viviendas	%	Viviendas	%
<b>Total viviendas</b>	<b>14.382</b>	<b>100,0</b>	<b>6.000</b>	<b>100,0</b>	<b>8.382</b>	<b>100,0</b>
Enlazadas*	13.542	94,1	5.681	94,6	7.861	93,8
Enlazadas con tutor/a**	12.751	88,6	5.342	89,0	7.409	88,3
Enlazadas con algún teléfono***	12.725	88,5	5.332	88,8	7.393	88,2

\*Enlazadas: total de viviendas en las que coinciden los DNI de RPA y Séneca.

\*\*Enlazadas con tutor/a: viviendas en las que los componentes de la vivienda en RPA coinciden con los nombres de algún tutor en Séneca.

\*\*\*Enlazadas con algún teléfono: viviendas en las que después del enlace se tiene constancia de algún teléfono.

### 3.2 *Cómo se utilizó la información del enlace*

La posibilidad de contar con los teléfonos de los tutores permitió diseñar estrategias de contacto con la muestra. Para ello se creó una unidad de agentes telefónicos dentro del IECA que se encargó de realizar las llamadas a la muestra titular en la semana anterior a la que los encuestadores tuviesen que visitar la sección censal en la que estaba situada la vivienda.

La planificación de estas llamadas tuvo que realizarse intentando evitar un efecto de saturación en la población. Había que tener presente que se había realizado un primer contacto vía correo ordinario dos semanas antes de comenzar los trabajos de campo. Por tanto, la llamada telefónica supondría el segundo contacto con las personas de la muestra titular. En esta llamada era importante lograr un equilibrio que permitiera, por un lado, obtener el mayor número de información posible para

facilitar el trabajo de los encuestadores y por otro, establecer una comunicación lo más breve posible. Con este propósito se fijaron únicamente dos objetivos como fundamentales: a) confirmar las direcciones del marco muestral y b) concertar una cita con los entrevistados<sup>12</sup>.

La confirmación de las direcciones consistía en asegurar que la persona en muestra efectivamente vivía en la dirección que constaba en el RPA. Si como resultado de la comunicación telefónica se confirmaba que la persona no vivía en esa dirección, se sustituía la unidad muestral, excepto en casos concretos en los que la vivienda se localizase en algún lugar que no afectase a los criterios de estratificación territorial fijados en el diseño muestral. También se podía detectar si el teléfono no correspondía a un hogar (comercios, oficinas, etc.), si la vivienda estaba en ese momento sin habitar, si el nombre de la calle u otros aspectos de la dirección postal habían cambiado, etc. El cotejo de esta información, además de depurar el marco y reducir los costes producidos por los desplazamientos de los encuestadores, también permitió reducir el número de incidencias al localizar y acceder a la vivienda. Por ejemplo, si la dirección de la vivienda correspondía a lugares que no estaban situados en núcleos urbanos (parajes, cortijos, viviendas situadas en zonas de carretera, etc.) se pedían especificaciones para poder localizarla. Toda esta información, registrada por los agentes telefónicos en aplicaciones informáticas, quedaba automáticamente grabada en los ordenadores que utilizaban los encuestadores para recoger la información.

Cuando se confirmaba que la dirección del marco muestral era la correcta, se procedía a intentar concertar una cita. El momento lo decidían las personas en muestra y como único requisito desde el IECA se establecía que el encuentro tenía que producirse dentro de la semana en que los encuestadores visitaban la sección censal a la que pertenecía la vivienda. El establecimiento de estas citas se convirtió en un proceso fundamental para reducir las incidencias en campo que vienen provocadas por las dificultades para localizar a las personas en la vivienda, ya que el encuentro se producía cuando las personas encuestadas lo consideraban oportuno<sup>13</sup>.

Como resultado de estas acciones, se mejoraron considerablemente las tasas de respuesta. Esta mejora incidió principalmente en la reducción de las incidencias producidas por no contacto o ausencia. En el siguiente apartado comprobaremos estos y otros aspectos comparando los datos relativos a la población de padres de la Encuesta Social 2010 con los de la anterior encuesta social realizada por el IECA: la

---

<sup>12</sup> El contacto telefónico podría ser contraproducente ya que suele ser más sencillo rechazar una encuesta por teléfono que de manera presencial. Por ello, se diseñaron fórmulas de comunicación con argumentos concretos, con las que intentar evitar este problema. Además, aunque la encuesta se rechazase telefónicamente, los encuestadores debían realizar una visita a la vivienda para intentar realizar la encuesta.

<sup>13</sup> Muchas de las negativas que se producen en las encuestas vienen provocadas porque los encuestadores realizan los intentos por encuestar en momentos inoportunos para las personas en muestra. Un análisis, realizado a través de grupos de discusión a encuestadores, que permite observar algunos matices interesantes de las circunstancias que rodean a los intentos de contacto que los encuestadores realizan en campo, lo encontramos en Camacho et al. (2000).

“Encuesta Social 2008. Hogares y medio ambiente en Andalucía<sup>14</sup>” en la que no se pusieron en marcha estos procedimientos, porque no se dispuso de un marco auxiliar que tuviese información de contacto telefónico.

#### 4. Resultados de los trabajos de campo

Como primer dato que sintetiza la calidad de la muestra, se ha calculado la proporción de encuestas realizadas en viviendas titulares y reservas en ambas operaciones. Como se puede observar en la tabla 2, el porcentaje de viviendas titulares en la Encuesta Social 2010 (72,4%) aumentó sensiblemente respecto a la Encuesta Social 2008 (59,7%).

Este primer dato nos habla de una mejoría clara en los niveles de respuesta. El siguiente paso será analizar las causas de esta mejora. Para ello hay que profundizar en las incidencias de los trabajos de campo. Para su clasificación y análisis se ha tomado como referencia la clasificación y cálculos del ISER (Institute for Social and Economic Research), instituto británico perteneciente a la Universidad de Essex. La metodología que sigue, formulada en Lynn et al. (2001), nació con la finalidad de establecer unos estándares comunes a nivel europeo para la denominación de las incidencias de campo y el cálculo de los niveles de respuesta.

Tabla 2.  
*Comparación de viviendas principales y reservas encuestadas en la Encuesta Social 2008 y en la Encuesta Social 2010*

	Principal	Reserva	Total (muestra efectiva)
<b>ESOC-2008</b>			
Valores absolutos	3.391	2.293	<b>5.684</b>
Porcentajes	59,7	40,3	<b>100,0</b>
<b>ESOC-2010</b>			
Valores absolutos	3.955	1.506	<b>5.461</b>
Porcentajes	72,4	27,6	<b>100,0</b>

Para comenzar con el análisis, se analiza cuáles fueron las incidencias que evitaron la realización de la entrevista en la muestra principal en una y otra encuesta. Para ello, en primer lugar, se puede observar que los porcentajes de viviendas

<sup>14</sup> En el Anexo II se muestra la ficha técnica de la Encuesta Social 2008. La diferencia más importante con respecto a la Encuesta Social 2010 es la población de estudio. En la Encuesta Social 2008 es toda la población mayor de 16 años, mientras que en la de 2010 los datos que se muestran son relativos a un sector muy específico de la población: las madres y padres con algún hijo nacido en el 1998 o en 1994. Por tanto, los datos de las tablas que se muestran en el apartado 4 tienen que tomarse con precaución; si bien, como se verá más adelante, la comparación entre ambas encuestas sirve para poner en contexto las mejoras metodológicas de la Encuesta Social 2010.

“encuestables” y “no encuestables” son muy parecidos entre ambas encuestas (tabla 3).

Las viviendas encuestables son aquellas en las que a priori se podía localizar a la persona seleccionada en la muestra, independientemente de que posteriormente se pudiera realizar la encuesta o no. De manera más concreta, el número total de viviendas encuestables es el resultado de la suma de encuestas que a) se realizaron de manera completa, b) se realizaron de manera parcial y c) viviendas en las que no se pudo realizar la encuesta por motivos no relacionados con el marco: no se pudo localizar al encuestado (Ausencia), no quiso realizar la encuestas (Negativa) y otros motivos de no colaboración (Incapacidad para contestar, etc.).

Tabla 3.

*Comparación de viviendas principales que eran “encuestables” y “no encuestables” en la Encuesta Social 2008 y la Encuesta Social 2010*

	Absolutos		Porcentajes	
	Encuesta Social 2008	Encuesta Social 2010	Encuesta Social 2008	Encuesta Social 2010
<b>Muestra titular</b>	<b>6.000</b>	<b>6.000</b>	<b>100,0%</b>	<b>100,0%</b>
Encuestables	5.078	5.169	84,6%	86,2%
No encuestables	922	831	15,4%	13,8%

Las viviendas no encuestables son las que agrupan los motivos que imposibilitan la realización de la encuesta por: a) defectos del marco (viviendas no elegibles que estaban vacías, destinadas a otros fines, etc.) y b) por tener una elegibilidad dudosa, es decir, viviendas que no pudieron ser localizadas por los encuestadores pero en las que no se pudo confirmar totalmente que fuera a causa de un error de marco.

Tabla 4.

*Comparación de resultados en las viviendas principales “no encuestables” en la Encuesta Social 2008 y la Encuesta Social 2010*

	Absolutos		Porcentajes	
	Encuesta Social 2008	Encuesta Social 2010	Encuesta Social 2008	Encuesta Social 2010
<b>No encuestables</b>	<b>922</b>	<b>831</b>	<b>100,0%</b>	<b>100,0%</b>
No elegibles	580	517	62,9%	62,2%
Elegibilidad dudosa sin contacto	342	314	37,1%	37,8%

Los datos de la tabla 3 muestran porcentajes de viviendas muy parecidos entre ambas encuestas, lo que permite establecer que la calidad del marco en ambas

encuestas es muy similar. Por tanto, como primera conclusión podemos establecer que la mejora en el porcentajes de viviendas titulares en la Encuesta Social 2010 no se explica por la mejora en el marco. Esto se puede corroborar de manera más exacta en la tabla 4 en la que se muestra que sólo hay una leve mejoría en el 2010 en los distintos motivos por los que las viviendas están en el grupo de no encuestables.

Por el contrario, si el análisis se centra en las viviendas encuestables, se observa que la mejora de la calidad de los datos de la Encuesta Social 2010 respecto a la Encuesta Social 2008 se debe a la reducción de la falta de respuesta provocada por la Ausencia de la persona que había que encuestar: en el 2008 supone el 17% de las situaciones que se producen en las viviendas encuestables, y desciende al 3,8% en el 2010.

Tabla 5.  
*Comparación de resultados en las viviendas principales “encuestables” en la Encuesta Social 2008 y la Encuesta Social 2010*

	Absolutos		Porcentajes	
	Encuesta Social 2008	Encuesta Social 2010	Encuesta Social 2008	Encuesta Social 2010
<b>Total encuestables</b>	<b>5.078</b>	<b>5.169</b>	<b>100,0%</b>	<b>100,0%</b>
Encuestadas	3.391	3.967	56,5%	66,1%
Falta de respuesta	1.687	1.202	28,1%	20,0%
- Ausencia	861	197	17,0%	3,8%
- Negativas	730	856	14,4%	16,6%
- Otros	96	149	1,9%	2,9%

En las encuestas sociales del IECA antes de que una vivienda pueda ser sustituida por no contacto los encuestadores tienen que realizar como mínimo cuatro visitas a la vivienda en al menos dos días y en tramos horarios distintos (por ejemplo: miércoles primera y última hora de la tarde y jueves primera y última hora de la mañana). A pesar de este protocolo, como se puede observar en la Encuesta Social 2008, se produce un volumen importante de incidencias por este motivo.

Es precisamente en este tipo de situaciones en las que los contactos telefónicos previos a la visita de los encuestadores son más efectivos. Los encuestados ponían fecha y hora para realizar la encuesta, reduciendo de este modo las situaciones en las que los encuestadores acuden a la vivienda y no encuentran a la persona que tiene que realizar la encuesta. Durante los trabajos de la Encuesta Social 2010 se realizaron más de seis mil llamadas telefónicas de las cuales se obtuvieron alrededor de mil citas telefónicas. Esta es la causa principal para la reducción de este tipo de incidencias y, consecuentemente, de la mejora de los niveles de respuesta.

Como dato final de síntesis que corrobore esta mejora se muestra el cálculo de la Tasa de Respuesta siguiendo también la metodología ISER. Para calcular esta tasa se divide el número de encuestas principales realizadas (parciales y completas), entre el número de unidades de la muestra elegibles. Son elegibles aquellas

viviendas en las que a priori podría realizarse la encuesta porque el marco proporciona una información fiable y completa, es decir, a efectos de cálculo, incluye las encuestas realizadas, las negativas, las viviendas con personas incapacitadas para contestar y los no contactados. Para explicar con más detalle el cálculo se detalla la fórmula concreta:

$$RR = \frac{I + P}{(I + P) + (R + NC + O) + (e_c UC + e_n UN)}$$

- Entrevista completa (I)
- Entrevista parcial (P)
- Ausencia o no contacto (NC)
- Negativa o Rechazo (R)
- Otros tipos de no respuesta: incapacidad para contestar, seleccionada anteriormente y otra no respuesta (O)
- Elegibilidad dudosa con contacto (UC)
- Elegibilidad dudosa con no contacto: inaccesible, ilocalizable y menor residente en otra vivienda (UN)
- $e_c$  = Proporción de los casos contactados de elegibilidad dudosa que son elegibles.
- $e_n$  = Proporción de los casos no contactados de elegibilidad dudosa que son elegibles; Encuestables / (Encuestables + No elegibles)

Utilizando esta fórmula, y teniendo presente que en ninguna de las dos encuestas hubo ni “Entrevistas parciales” ni viviendas de “Elegibilidad dudosa con contacto”, a continuación se muestran los resultados de las Tasas de Respuesta:

-Encuesta Social 2008:

$$RR = \frac{3391}{(3391) + (1687) + (0,90 * 342)} \times 100 = 63,0\%$$

-Encuesta Social 2010:

$$RR = \frac{3967}{(3967) + (1202) + (0,91 * 314)} \times 100 = 72,7\%$$

Por tanto, esta mejora en la tasa de respuesta en casi diez puntos permite concluir que efectivamente la metodología ha sido efectiva. Además, la reducción de estas incidencias también puede permitir el ahorro de costes. No es el objetivo de este artículo evaluar este ahorro con exactitud. Sin embargo, se puede establecer una estimación a este respecto, siempre teniendo presente que ambas encuestas tienen diseños muestrales distintos. En la Encuesta Social 2008 el número medio de visitas por encuesta realizada fue de 2,8 mientras que en la Encuesta Social 2010 fue de 2,3

visitas por encuesta. Los resultados de estas cifras se deben principalmente a los trabajos de contacto telefónico con las personas en muestra, tal y como se muestra en la tabla 5. En ambas operaciones el coste estimado de cada dieta-desplazamiento que realizaban los encuestadores en cada intento de realizar una encuesta, se presupuestó en 15,4 euros. De estas cifras se obtiene que el coste en desplazamientos por encuesta en la Encuesta Social 2008 fue de 43,1 euros (15,4 euros de cada dieta-desplazamiento por 2,8 visitas por encuesta) y el de la Encuesta Social 2010 fue de 35,4 euros. Esto supondría un ahorro de 7,7 euros por encuesta<sup>15</sup>.

Sin embargo, todavía quedan distintas líneas en las que se pueden establecer mejoras metodológicas que permitan utilizar la información de los registros con mayor eficacia. En este sentido, se ha puesto de manifiesto que el contacto telefónico con la muestra no se ha traducido en un menor número de negativas. Sin duda, en próximas experiencias éste será un aspecto importante en el que reflexionar: cómo utilizar la información administrativa de contacto para reducir los niveles de negativas.

## 5. Conclusiones

La literatura científica poco a poco ha ido nutriéndose de estudios que han tenido como objetivo la mejora de la calidad de los datos en las investigaciones mediante encuesta. Las estrategias y protocolos que se recomiendan se han ido incorporando de manera generalizada en la estadística oficial en nuestro país. Además, los organismos que forman parte de la Función de la Estadística Pública disponen de ventajas en las encuestas a hogares respecto a otras instituciones. Por ejemplo, este tipo de organismos utilizan marcos de población detallados a partir de los cuales se pueden realizar diseños muestrales complejos que aseguran una mayor representatividad.

Sin embargo, todavía no se aprovecha de manera generalizada otro tipo de información que también podría mejorar la calidad de los datos. Nos referimos a la utilización de los datos de los registros administrativos que disponen los distintos organismos que pertenecen a la Administración Pública. La utilización de enlaces entre distintos registros se señala en numerosos foros como una clara línea de futuro para aprovechar información existente, no duplicar esfuerzos, y consecuentemente,

---

<sup>15</sup> A esta estimación del ahorro habría que aplicarle los costes que supusieron los trabajos de contacto telefónico y los trabajos de enlace de registros. Ambos son complicados de estimar. En primer lugar, es difícil definir el coste que supuso el trabajo de contacto telefónicos porque esta función era sólo una de las que tenían que realizar los agentes telefónicos en su jornada de trabajo: las otras eran la validación telefónica de los cuestionarios y la revisión de la codificación de variables sociodemográficas a partir de la información literal que recogían los encuestadores durante los trabajos de campo. Más complejo es aún estimar el coste de los trabajos realizados por los técnicos que hicieron el enlace de registro. El enlace no se realizó de manera intensiva, sino que fue fruto de un proceso que se alargó en el tiempo ya que los técnicos implicados lo realizaron en combinación con el resto de funciones que tenían que desempeñar. Además participaron técnicos de distintos servicios dentro del IECA y en la Consejería de Educación. Por tanto, los autores no están en disposición de realizar una estimación fiable del coste de estos trabajos, cuestión que habría que definir mejor para próximas operaciones estableciendo criterios de contabilidad analítica. No obstante, aun teniendo presente estos gastos, los autores de este artículo consideran que la eficiencia económica de los trabajos de contacto y enlace han supuesto un ahorro considerable.

reducir costes. De hecho, sus ventajas se han puesto de manifiesto en distintas experiencias en el ámbito internacional y nacional, principalmente con el objetivo de mejorar análisis sobre elementos concretos de la realidad social.

En el presente artículo se ha descrito la primera experiencia realizada en el IECA como parte de la metodología de sus encuestas sociales. El objetivo del enlace del marco muestral con el sistema de información Séneca no era sólo el aprovechamiento de su información para profundizar en el conocimiento de la educación en Andalucía. Además, y lo que supone quizás el aspecto más novedoso, la información del enlace se utilizó para desplegar procedimientos que permitiesen reducir incidencias, facilitando la labor de los encuestadores antes del comienzo de los trabajos de campo.

Este método de trabajo se puede utilizar en otros contextos, especialmente en aquellas operaciones que se desarrollen desde la estadística oficial. Para ello es necesario establecer marcos de colaboración entre instituciones. Los institutos de estadística, tanto a nivel nacional como autonómico, podrían utilizar registros administrativos que cuenten con información de contacto con la ciudadanía. Por poner un ejemplo, este tipo de información está disponible, con un alto grado de actualización, en las bases de datos donde están registrados los usuarios del sistema sanitario público.

Como conclusión final, las acciones que se llevaron a cabo tuvieron un efecto claro en el incremento de algunos indicadores de calidad de la muestra respecto a encuestas similares desarrolladas en el IECA y también en comparación con los datos de otros organismos. No obstante, habrá que seguir comprobando la efectividad de estas acciones en otras operaciones estadísticas. Además, todavía quedan distintas líneas en las que se pueden establecer mejoras metodológicas que permitan utilizar la información de los registros con mayor eficacia (por ejemplo, para reducir el número de negativas). Por tanto, queda camino que transitar para avanzar en la mejora en la calidad de los datos. Para ello es fundamental establecer mejoras en los trabajos de campo. Seguro que la utilización de información administrativa tendrá mucho que aportar en este sentido.

## Bibliografía

- Alvira, F. y Martínez, E. (1985). El efecto de los entrevistadores sobre las respuestas de los entrevistados. *Revista Española de Investigaciones Sociológicas*, 29, 219-256.
- Atienza, J. y Noya, F.C. (1999). La encuesta como interacción social. Una aproximación empírica. *Empiria*, 2, 73-92.
- Ballano, C. y Rebollo, M.D. (2006). Las incidencias en las encuestas del Instituto Nacional de Estadística. *Metodología de Encuesta*, 8, 23-32.
- Camacho, C.; Romero, B.; Valero, S. y Camacho, C. (2000). ¿Cómo hacen encuestas los encuestadores? *Metodología de Encuesta*, 2 (1), 117-133.
- Cea D`Ancona, M.A. (2005). La senda tortuosa de la «calidad» de la encuesta. *Revista Española de Investigaciones Sociológicas*, 111, 75-103.



- Cuxart, A. y Riba, C. (2009). Mejorando a partir de la experiencia: la implementación de la tercera ola de la ESE en España. *Revista Española de Investigaciones Sociológicas*, 125, 147-168.
- Díaz de Rada, V. (2001). Problemas de cobertura en las encuestas telefónicas. *Revista Española de Investigaciones Sociológicas*, 93, 133-164.
- Díaz de Rada, V. (2004). Problemas de representatividad en las encuestas con muestreos probabilísticos. *Papers*, 74, 45-66.
- Díaz de Rada, V. (2006). Cómo mejorar la preparación de los encuestadores. *Revista Española de Investigaciones Sociológicas*, 114, 241-254.
- Díaz de Rada, V. (2007). Tipos de encuestas considerando la dimensión temporal. *Papers*, 86, 131-145.
- Díaz de Rada, V. (2008). La selección de los entrevistados últimos en encuestas presenciales: un análisis de la utilización conjunta del método de rutas y el método de cuotas. *Revista Española de Investigaciones Sociológicas*, 123, 209-247.
- Díaz de Rada, V. (2010). *Comparación entre los resultados proporcionados por encuestas telefónicas y personales: el caso de un estudio electoral*. Madrid: Centro de Investigaciones Sociológicas, Colección "Opiniones y Actitudes", n°. 66.
- Díaz de Rada, V. (2012). Ventajas e inconvenientes de la encuesta por Internet. *Papers*, 97, 193-223.
- Dillman, D.A.; Smyth, J.D. y Christian, L. M. (2009). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. New Jersey: Wiley.
- Dunn, H.L. (1946). Record Linkage. *American Journal of Public Health*, 36 (12), 1412-1416.
- Fernández, M. (2003). Criterios de calidad en la investigación social. *Empiria*, 6, 47-77.
- Filgueira, E. (2001). La calidad de la medición frente al error estadístico. *Empiria*, 4, 193-210.
- Groves, R. M. (1989). *Survey error and survey cost*. Nueva York: Wiley.
- Hjelm, C.G. (2006). Enlarging Statistical System and Improving Economic and Social Analysis; The Role Of Administrative Records. Ponencia en el Seminario *El papel de los registros administrativos en relación con el análisis social y económico y el desarrollo del sistema estadístico nacional*. Instituto de Estudios Fiscales. Madrid: enero de 2006.
- Jäckle, A.; Peter, L.; Sinibaldi, J. y Tipping, S. (2011). The Effect of Interviewer Personality, Skills and Attitudes On Respondent Co-Operation with Face-To-Face Surveys. ISER. *Working Papers Series*, 14.
- Knies, G.; Sala, E. y Burton, J. (2011). Consenting to Health Record Linkage: Evidence from the British Household. Panel Study. *Working Papers*, 27, Colchester, University of Essex: Institute for Social and Economic Research.
- Lynn, P.; Beerten, R.; Laiho, J. y Martin, J. (2001). Recommended Standard final Outcome Categories and Standard Definitions of Response Rate for Social Surveys. *Working Papers*, 23, Colchester, University of Essex: Institute for Social and Economic Research.

- Moral-Arce, J. y Martín, E. (2009). Integración de información administrativa y muestral en estadísticas económicas estructurales. La Encuesta Anual de Estructura Salarial. *Revista Estadística Española*, 51(172), 487-504.
- Porras, J. y Hernández, F. (2006). Aprovechamiento de la información administrativa en la formación del marco para las Encuestas de Hogares en el INE. Ponencia en el seminario *El papel de los registros administrativos en relación con el análisis social y económico y el desarrollo del sistema estadístico nacional*. Instituto de Estudios Fiscales. Instituto de Estudios Fiscales. Madrid: enero de 2006.
- Pasadas del Amo, S. (2006). El impacto de la telefonía móvil en la cobertura de las encuestas telefónicas. *Metodología de Encuesta*, 8 (2), 49-59.
- Ryten, J. (2006). El papel de la estadística de base administrativa en las estadísticas económicas. La historia de su evolución en Canadá. Ponencia en el seminario *El papel de los registros administrativos en relación con el análisis social y económico y el desarrollo del sistema estadístico nacional*. Instituto de Estudios Fiscales. Instituto de Estudios Fiscales. Madrid: enero de 2006.
- Saralegui, J. (2006). El uso estadístico de las fuentes administrativas como factor de cohesión territorial. Ponencia en el seminario *El papel de los registros administrativos en relación con el análisis social y económico y el desarrollo del sistema estadístico nacional*. Instituto de Estudios Fiscales. Instituto de Estudios Fiscales. Madrid: enero de 2006.
- Schuman, H. y Presser, S. (1981). *Questions and Answers in Attitude Surveys: Experiments on Questions, Forms, Wording and Context*. California: Academic Press inc.
- Trujillo, M. y Gutiérrez C. (2006). Incidencias durante el trabajo de campo en encuestas personales realizadas por el IESA. *Metodología de Encuesta*, 8, 49-60.
- U.S. Census Bureau (2012). *National Longitudinal Mortality Study. Reference Manual*.
- Viciano, F.; Montañés, V.; Ruiz, E.; Ríos, J. y González, A. (2005). Estadísticas de Trayectorias Vitales: un nuevo producto estadístico derivado de la gestión y mantenimiento del Registro de Población. Ponencia en *Jornadas de Estadísticas de las Comunidades Autónomas (JECAS)*. Córdoba: noviembre de 2005.
- Winkler, W.E. (2006). *Overview of Record Linkage and Current Research Directions*. Bureau of the Census. Research Report Series.