



# Estadísticas longitudinales de supervivencia y longevidad en Andalucía, 2002-2016

## Metodología

<b>Índice</b> .....	<b>1</b>
<b>1. Introducción</b> .....	<b>2</b>
<b>2. Origen y relevancia de las estadísticas longitudinales de población</b> ....	<b>3</b>
<b>3. Fuente de información</b> .....	<b>4</b>
<b>4. Finalidad y diseño del estudio</b> .....	<b>5</b>
<b>5. Procesado y depuración de la información de base</b> .....	<b>6</b>
5.1 Integración de las fuentes en la BDLPA.....	6
5.2 Construcción de los episodios.....	7
5.3 Construcción del grid censal.....	8
<b>6. Contenidos de la publicación</b> .....	<b>8</b>
6.1 Tablas estadísticas con los datos de longevidad de las cohortes censales.....	9
6.2 Ficheros de microdatos de las cohortes censales.....	11
6.3 Indicadores y mapas de mortalidad por celda.....	12
<b>7. Control del riesgo de revelación y preservación del secreto estadístico</b> .....	<b>14</b>
<b>8. Anexos</b> .....	<b>14</b>
8.1 Algunos conceptos específicos de este tipo de estudios.....	14
8.1.1 Variables tiempo dependientes (VTD) y tiempo independientes (VTI).....	14
8.1.2 Concepto de episodio, estado, eventos y transiciones.....	15
8.1.3 Subdivisión de episodios basada en escalas de tiempo discretas.....	17
8.1.4 Precauciones en la interpretación de resultados.....	19
8.2 Anexo: Fuentes primarias.....	20
8.2.1 El padrón continuo.....	20
8.2.2 El Censo de Población de 2001.....	21
8.2.3 El Censo de Población de 2011.....	22
8.2.4 Estadísticas del MNP.....	22
8.3 Anexo: Lista abreviada de Causas.....	23

## 1. Introducción

La presente publicación, desarrollada por el Instituto de Estadística y Cartografía de Andalucía (IECA) con la colaboración del Departamento de Población del Consejo Superior de Investigaciones Científicas y cofinanciada por el Fondo Social Europeo, continúa la línea de producción estadística que inició en 2014 y amplió en 2016, basada en la integración y reutilización de información recopilada en otras operaciones estadísticas con el objeto de avanzar en el conocimiento demográfico de las funciones de supervivencia poblacional desde una perspectiva longitudinal.

Esta explotación se ha confeccionado con el objetivo de suministrar información adecuada para analizar las funciones de supervivencia y longevidad en combinación con la información sobre las características personales, familiares y del entorno recogidas en los Censos de Población y Viviendas de 2001 y 2011.

Las fuentes de información que ha sido preciso combinar e integrar para confeccionar esta publicación han sido las siguientes:

1. Los Censos de Población y Viviendas de Andalucía de los años 2001 y 2011.
2. Las estadísticas del Movimiento Natural de la Población, en concreto los registros correspondientes a los Boletines Estadísticos de Defunción correspondientes a los años 2002 a 2016.
3. Las variaciones residenciales de la población entre los años 2002 y 2016 y recogidas por el sistema de actualización permanente de los padrones municipales.

La integración de estas fuentes permite precisar si, con posterioridad a la operación censal de 2001 o, en su caso, a la operación censal de 2011, se ha producido el fallecimiento o la salida de Andalucía por emigración de alguno de los individuos recogidos en dichas operaciones, así como el momento exacto en que ocurrió dicho evento.

La integración de la información estadística de los eventos con la de los Censos de Población permite profundizar en las investigaciones sobre supervivencia poblacional en función de un amplio abanico de características personales, familiares, del entorno y con un detalle territorial inframunicipal, hasta ahora no disponible en ninguna otra fuente estadística en nuestro medio.

El sistema de información que permite realizar la integración de estas fuentes, y que se encuentra en permanente evolución y expansión, es la Base de Datos Longitudinal de Población de Andalucía (BDLPA). La ley 3/2013 del Plan Estadístico de Andalucía 2013-2020 recoge la importancia de esta nueva fuente al incluirla entre los “*ámbitos prioritarios de actuación en infraestructuras de información*” (Art 17.1.b) a desarrollar durante el periodo de vigencia del Plan.

En esta nueva edición de este producto de difusión derivado de la explotación de la BDLPA, la ventana de seguimiento se amplía tres años más, incluyendo la información de los eventos (defunciones y emigraciones exteriores) ocurridos durante los años 2014 a 2016 a la población censada en 2001. Asimismo se ha extendido el ámbito poblacional del estudio, al hacer también seguimiento entre los años 2012 y 2016 a la cohorte censal de 2011.

La investigación longitudinal de la supervivencia no agota el área temática de posibles explotaciones derivadas de la fuente BDLPA. Esta perspectiva analítica se vio complementada en el año 2017 con la publicación por parte de este organismo de las “*Estadísticas Longitudinales de Biografías Reproductivas en Andalucía, 2002-2013*”, en las que se centraba la atención en el fenómeno de la fecundidad. En los próximos años está previsto desarrollar nuevos productos estadísticos longitudinales, que continúen con los estudios de supervivencia y fecundidad, abarcando temas tales como la movilidad residencial de la población, las biografías de actividad económica o las de cualificación educativa.

## 2. Origen y relevancia de las estadísticas longitudinales de población

En Demografía existen dos acercamientos para analizar el comportamiento de una población con respecto a un determinado suceso: el análisis transversal clásico y el análisis longitudinal.

El análisis longitudinal implica el seguimiento de las unidades de investigación a lo largo de la duración de su existencia. La investigación longitudinal está directamente emparentada con conceptos como ciclo vital, desarrollo o envejecimiento. Para poder llevar a término este tipo de investigaciones es necesario disponer de sistemas que suministren información estadística desde una perspectiva biográfica.

En el análisis demográfico clásico, este tipo de estudios se desarrollaron inicialmente partiendo de datos agregados por variables longitudinales tales como la generación de nacimiento o las promociones de matrimonio y/o paridad. La disponibilidad de esta información en las series estadísticas transversales permitió las primeras investigaciones sobre los comportamientos que se daban en el seno de una misma generación o promoción, conforme ésta envejecía o transitaba por diferentes etapas de su ciclo vital.

Estos datos longitudinales se obtienen combinando tablas anuales publicadas incluyendo variables longitudinales (tales como la generación o la promoción de matrimonio o nacimiento). Tras combinar series temporales largas se puede obtener una reconstrucción, al menos parcial, de los comportamientos agregados de una generación o promoción concreta a lo largo de su ciclo de crecimiento / desarrollo / envejecimiento.

Sin embargo, dada la complejidad de las tablas que incluyen estas variables longitudinales, solo unos pocos países han publicado series adecuadas durante suficiente tiempo como para permitir las reconstrucciones longitudinales. Ahora bien, cuando las han publicado, éstas se han limitado a cruces relativamente sencillos y sin un nivel alto de detalle y desagregación con otras covariables.

Investigaciones longitudinales más sofisticadas comenzaron a desarrollarse a medida que se dispuso de fuentes estadísticas más complejas, tales como las proporcionadas por el seguimiento continuo de muestras representativas de las poblaciones como las del Estudio Longitudinal de Inglaterra y Gales, partiendo de una muestra de 1% de la población censada en 1970, o bien las obtenidas partir de grandes encuestas biográficas retrospectivas, como por ejemplo la encuesta francesa de la triple biografía o la encuesta socio-demográfica española de 1991. En la actualidad ya existen operaciones estadísticas europeas de base muestral con un claro componente longitudinal como la Encuesta Europea de Condiciones de Vida.

Una de las líneas de producción estadística que más ha contribuido a la proliferación de investigaciones longitudinales ha sido el desarrollo de sistemas de información estadística basados en el enlace de registros administrativos y estadísticos, tales como los que comenzaron a desarrollarse en los países nórdicos en los años 70. La recuperación sistemática de la información de utilidad estadística ha creado un nuevo y poderoso modelo de producción estadística, que ha sido fuertemente retroalimentado por la revolución en la tecnología del tratamiento de la información que paralelamente se ha producido en las cuatro últimas décadas. Paulatinamente, muchos países del resto de Europa han comenzado a acercarse con interés a este modelo, implantando en algunos casos sistemas similares, como es el caso de Holanda.

Una de las características fundamentales de estos nuevos sistemas de información basados en registros es la inclusión en su diseño de una perspectiva eminentemente biográfica, dado que las unidades de integración de la información dispersa a lo largo del tiempo y en múltiples registros sectoriales ya no son exclusivamente las categorías territoriales o los grupos sociales que subdividen la población, sino directamente los individuos concretos que la han protagonizado. Detalladas estadísticas territoriales o desagregaciones de colectivos poblacionales muy específicos se consiguen fácilmente con posterioridad, mediante la agregación de la información individual ya integrada.

El seguimiento individual de las unidades de observación a través del tiempo que dure el estudio es la característica más relevante de los estudios longitudinales. Si bien en condiciones ideales el objetivo es el seguimiento a lo largo de todo el curso de la vida, la mayoría de las veces hay que limitarlos a una estrecha "ventana de observación" durante la cual ha sido

posible registrar de manera continua aquellos acontecimientos relevantes y la fecha exacta en la que se han producido.

Mientras la perspectiva transversal prioriza la investigación de estructuras y relaciones de asociación entre determinadas características en un momento dado, el análisis longitudinal enfatiza los procesos y las relaciones temporales y de causalidad que se dan entre diferentes tipos de eventos y transiciones. En este sentido, la relevancia de la investigación causal en los estudios longitudinales hace a éstos de gran interés para la planificación y evaluación de políticas públicas, tal como lo expresa Frans Willekens<sup>1</sup>:

*“Si se consigue un conocimiento de la evolución de los procesos y se puede actuar sobre el contexto, se podrán dirigir los procesos. Ese es el fundamento de las intervenciones y la formulación de políticas. A una gran escala, la formulación de políticas es el seguimiento de los procesos en curso creando condiciones que los dirigen en la dirección deseada. La intervención no es eficaz a menudo porque las interacciones entre los procesos y su contexto no están suficientemente comprendidas.”*

### 3. Fuente de información

La fuente de información a partir de la cual se ha realizado este producto de difusión es una infraestructura de información estadística que bajo la denominación de “Base de Datos Longitudinal de Población de Andalucía” (BDLPA), está desarrollando el Instituto de Estadística y Cartografía de Andalucía (IECA). Este sistema es la columna vertebral de la actividad estadística “Registro de Población de Andalucía” (RPA). Las fuentes principales que la nutren provienen de las inscripciones en los padrones municipales de Andalucía, los boletines estadísticos del Movimiento Natural de Población y los censos de población y viviendas. La integración de esta información permite la incorporación de variables recopiladas en una fuente a las de otra, que o bien no disponía de dichas variables o bien las tenía recogidas con un menor nivel de detalle. Este proceso de integración y reutilización de información ya recopilada está propiciando la generación de una nueva gama de productos estadísticos que amplían y complementan la producción estadística clásica, con un coste menor que el que supondría diseñar nuevos circuitos de recogida de información.

La actividad “Registro de Población de Andalucía”, se introdujo por primera vez en la planificación estadística en la Ley 8/2002 del Plan Estadístico de Andalucía con el siguiente objetivo:

*“Obtener la relación de las personas residentes en Andalucía y sus características principales, en base a los padrones municipales de habitantes como instrumento estadístico fundamental para la integración de las estadísticas demográficas y como fuente de nuevas actividades de interés (estadísticas de variaciones residenciales infra y supramunicipales, marco para estadísticas muestrales dirigidas a la población, trayectorias vitales, etc.)”*

En los años posteriores a su creación se ha ido perfilando el sistema de información de esta actividad bajo la denominación de “Base de Datos Longitudinal de Población de Andalucía”, quedando manifiesta su relevancia para el desarrollo de un nuevo sistema de producción estadística, y su trascendencia es explícitamente recogida en el Plan Estadístico y Cartográfico de Andalucía 2013-2020, que en su artículo 18.b la incluye como infraestructura de información dentro del ámbito prioritario de la población.

La BDLPA es la nueva fuente estadística donde se realiza la integración y reutilización de la información ya recopilada en otros registros estadísticos y administrativos de base poblacional. Esta fuente genera nuevos datos que no estaban disponibles en las fuentes originales explotadas sectorialmente. Por ejemplo, si para determinadas estadísticas, como las de mortalidad, se desea disponer de nuevas tablas que incluyan una variable como la educación, no disponible en los actuales formularios de recogida de información (el Boletín Estadístico de Defunción, en este caso), es posible conseguirla mediante el enlace de los registros de

---

1 F. Willekens. El futuro del análisis longitudinal. Información Estadística y Cartográfica de Andalucía nº1 pg:76-78.  
<http://www.juntadeandalucia.es/institutodeestadisticaycartografia/InformacionEstadisticayCartografica/RevistaFuturoPoblacion.pdf>

defunciones (que no dispone de la variable educación), con otros ficheros estadísticos que sí la tengan, como por ejemplo los censos de población o los padrones.

Esta integración permite, por un lado, la reutilización de la información recogida en una determinada actividad para comprobar, mejorar la calidad o ampliar la información de otras estadísticas sectoriales. Y por otro lado, permite diseñar nuevos productos de difusión, como es el caso de esta publicación, a partir de la información surgida de las interrelaciones de las fuentes originales. Todo ello, a un coste mucho menor del que implicaría la modificación de los circuitos de recogida preexistentes o el desarrollo de nuevas actividades estadísticas para la recogida de las nuevas variables.

#### **4. Finalidad y diseño del estudio**

Esta publicación es la actualización del primer producto de difusión relevante surgido de la actividad estadística Registro de Población de Andalucía y realizada a partir de la explotación de su banco de datos: la Base de Datos Longitudinal de Población de Andalucía. Se ha concebido con el objetivo de suministrar a los investigadores y al público en general, nuevos datos con información sobre las funciones de longevidad y mortalidad de la población andaluza que no están disponibles en las estadísticas anuales de mortalidad que hasta ahora se han publicado.

Existen claras razones por las que se comenzó la serie de productos estadísticos derivados de la BDLPA con investigaciones sobre supervivencia y longevidad de la población. Por un lado, históricamente las metodologías de la demografía se han desarrollado a partir de la investigación de la mortalidad. Sus características de evento irrepitable (no renovable, en terminología técnica) y de cobertura universal (todos terminaremos experimentándolo) permiten una conceptualización más sencilla que otros eventos demográficos. Por otro lado la delimitación de los “episodios vitales” en relación a la supervivencia es sencilla, estando enmarcados por dos claros eventos: el nacimiento y la muerte de los individuos. Aunque, para un individuo concreto, estos eventos no siempre ocurrirán en la ventana de observación que se define en el estudio. Las herramientas conceptuales del “análisis de supervivencia” han desarrollado una clara metodología para analizar este tipo de datos.

Además, los episodios vitales (nacimiento-muerte) enmarcan cualquier otro posible episodio en la vida de un individuo, tales como los residenciales, los de convivencia, o las variaciones en estructuras familiares. No es posible realizar investigaciones en ninguno de estos campos sin generar y validar previamente la calidad de los episodios vitales generados por la nueva fuente.

Por último, además de estas cuestiones prácticas asociadas con un lógico escalonamiento en los procesos de creación y explotación de la nueva fuente, la investigación sobre longevidad tiene sobrado interés en sí misma y supone un tema de gran actualidad. Por ejemplo, en relación al importante debate de la sostenibilidad del sistema de pensiones, o bien por la posibilidad de profundizar en la descripción y análisis de las desigualdades sociales en salud.

Con el objetivo de poder estimar la diferencia en mortalidad y longevidad en diferentes colectivos poblacionales definidos tanto por su lugar de residencia como por sus características socio-económicas, se diseñó un “estudio longitudinal de cohortes” basado en la información histórica almacenada en la BDLPA relativa a los padrones municipales, los Censos de Población y Viviendas de 2001 y 2011, y los Boletines Estadísticos de Defunción.

El diseño longitudinal planteado en esta publicación es de tipo prospectivo, fijando las características de las personas (salvo edad y residencia<sup>2</sup>) al inicio del estudio: para la cohorte censal de 2001, a través de la información recogida en el Censo de Población y Viviendas de 2001, y para la muestra que constituye la cohorte censal de 2011, a través de la información recogida en el Censo de Población y Viviendas de 2011. Se estudia así cómo evoluciona el estado residencial y vital de los individuos incluidos en cada cohorte durante las respectivas ventanas de observación (2002-2016 para la cohorte de 2001 y 2012-2016 para la cohorte de 2011).

---

<sup>2</sup> Edad y Residencia son las dos únicas variables personales tratadas como tiempo-dependientes en las explotaciones de esta publicación, mientras que el resto de las variables son observadas en una fecha temporal fijada: la censal.

Se parte por tanto de las características de la población recogidas en los cuadernillo censales en unas determinadas fechas de referencia (1/1/2002 y 1/12/2012), sobre las cuales se investiga los eventos relativos a defunciones o variaciones residenciales que le han podido ocurrir en años posteriores a las dos cohortes en estudio. Esta información ha sido seleccionada y combinada con objeto de obtener una base de datos de explotación para esta investigación, construida bajo las siguientes especificaciones:

- **Población de partida:** Individuos recogidos en el Censo de 2001, que pudieron ser localizados en algún alta padronal de un municipio de Andalucía (97,9% de los censados) y que residían en Andalucía el 1 de enero de 2002. Asimismo, se estudian los individuos recogidos en la muestra del Censo de 2011, que pudieron ser localizados en algún alta padronal de un municipio de Andalucía (98,1% de los censados encuestados) y que residían en Andalucía el 1 de enero de 2012.
- **Periodo de seguimiento (ventana de observación):** Sobre la población censada, residente en Andalucía el 1 de enero de 2002 (respectivamente, el 1 de enero de 2012) se comprueban las modificaciones de su situación de residencia o estado vital recogidas tanto en las altas y bajas padronales como en los Boletines de Defunción, desde el 1 de enero de 2002 (respectivamente, el 1 de enero de 2012) hasta el 31 de diciembre de 2016 (ambos inclusive).
- **Condiciones que determinan el fin del seguimiento:** El fin de seguimiento de cada uno de los individuos participantes se puede producir por una de estas tres causas:
  - *Fallecimiento:* registrado en MNP y/o baja padronal por defunción.
  - *Emigración fuera de Andalucía:* informada por una baja padronal. Estos casos son tratados como *episodios censurados* en la fecha de la primera salida de Andalucía. En un pequeño número de casos se produce la re-entrada en Andalucía con posterioridad a su salida. Esta información se suministra en los ficheros de microdatos dejando al usuario especializado la capacidad de decidir cómo tratar estos episodios: como truncamiento de intervalo o como censuras en la fecha de la primera salida.
  - *Censura por fin de estudio:* en la gran mayoría de los casos el fin de seguimiento se produce en personas que continúan residiendo en Andalucía en la fecha de cierre de estudio 31-12-2016.

## 5. Procesado y depuración de la información de base

El hecho de que la Base de Datos Longitudinal de Población de Andalucía posea datos de diversas fuentes de información conlleva una serie de ventajas mencionadas anteriormente, como la posibilidad de combinarlas. No obstante, existen problemas asociados al proceso de integración de la información.

### 5.1 Integración de las fuentes en la BDLPA

Para establecer una relación entre las fuentes de información se hace necesario vincular la información correspondiente a unas mismas unidades. Por ello, un paso previo es la identificación de registros padronales. Un proceso que no es sencillo, ya que si bien la información padronal dispone de una gran cantidad de identificadores personales directos, no es así en el caso del fichero censal, por el contrario este último es muy rico en información de interés estadístico. El enlace de ambas fuentes se realiza en fases sucesivas, donde se irán aplicando distintos criterios, desde los más estrictos y seguros a otros más laxos<sup>3</sup> que incluyen necesariamente mayor incertidumbre.

El criterio más restrictivo pasa por la coincidencia en ambas fuentes de todos los datos en un conjunto de identificadores estrictos. El enlace conseguido de esta forma se denomina determinístico y con él se establece la correspondencia entre fuentes por encima del 80% de los individuos censados.

3 Cánovas R, Montañés V, Poza E, Viciana F "Resultados del enlace de los registros del MNP con el RPA, 12 años de seguimiento: 1998-2009.", XVII JECAS, Cáceres, 2010.

<http://www.jecas.org/ponencias/miercoles/tarde/demograficas/resultadosdelenlacedelosregistros.pdf>

Para el resto se busca establecer un grado de correspondencia que si bien no sea perfecta, exista un alto grado de seguridad de que ambos registros son los mismos. Este grupo de enlace incluye necesariamente un componente probabilístico.

Por ejemplo, en fases posteriores se considerarían enlaces positivos cuando hay coincidencia de la hoja padronal, el año y el mes de nacimiento, aunque el día de nacimiento varíe. Aplicando criterios de este tipo, que siguen siendo estrictos, se llega a alcanzar un enlace cercano al 95%. En sucesivas aproximaciones, que incluyen criterios de búsqueda relacionados con los convivientes, parejas y familiares, se llega a alcanzar un nivel de enlace muy cercano al 99%.

Para asegurar la fiabilidad del sistema se toma una pequeña muestra de estos enlaces y se revisan manualmente para estimar si hay enlaces asimilados como verdaderos sin evidencia clara de que lo sean. Estos posibles enlaces erróneos son denominados falsos positivos. El análisis de estos pares revela que la tasa de falsos positivos es reducida (en todo caso inferior al 0,5%), lo que garantiza que la información emparejada es sobradamente fiable para su tratamiento estadístico.

Con la información del Movimiento Natural de Población se realiza un proceso similar, si bien esta información solo se emplea para cerrar episodios cuya baja por defunción no haya sido comunicada por los canales de información padronal en la fecha de cierre de este estudio.

Una vez obtenidos los enlaces, hay que establecer un orden de prioridad para resolver los posibles conflictos entre la información reportada por las distintas fuentes. La información censal se toma como base y se extraen de los censos de 2001 y 2011 todas las covariables de estudio que se consideran fijas en el tiempo a excepción del lugar de residencia. Para determinar los cambios en el lugar de residencia se toma la información procedente de las variaciones padronales con las que se construyen los episodios. Por último, del Movimiento Natural de Población se toman las fechas de defunción, en los casos en que no se disponga de dicha información en los ficheros padronales, y las causas de defunción.

En ocasiones, el municipio de residencia declarado en la defunción no coincide con el reportado en el padrón municipal de habitantes, optándose por considerar como válido el municipio padronal ya que la adecuación con los denominadores poblacionales es más apropiada.

## **5.2 Construcción de los episodios**

Los episodios muestran la permanencia de un individuo en una localización geográfica determinada desde una fecha establecida como inicio del episodio hasta otra fecha en que esa etapa residencial concluye, bien por defunción, cambio de residencia a otro ámbito geográfico o bien porque esa fecha es considerada como fecha de finalización del estudio.

En esta publicación se tienen en cuenta los episodios residenciales en Andalucía que suceden durante el periodo 2002-2016 y se obtienen tras una depuración de los flujos de alta y baja recibidos en el Registro de Población de Andalucía, de manera que no se produzcan incongruencias y se cumpla la condición de que un individuo sólo puede estar asociado a un territorio en el mismo periodo de tiempo. Para ello se realizan las oportunas depuraciones e imputaciones, para suprimir los eventos repetidos y modificar, en su caso, las fechas de inicio o fin de episodio para garantizar la consistencia espacio-temporal en la secuencia de eventos.

Debido a modificaciones en la legislación que han afectado a las bajas administrativas de extranjeros, comunitarios y extracomunitarios<sup>4</sup>, se ha precisado imputar fechas de baja de individuos de este colectivo para los que se conocía que habían abandonado la comunidad autónoma pero se desconocía la fecha. Esta imputación<sup>5</sup> evita un descenso súbito de la cifra de efectivos poblacionales en las fechas en que comenzó a aplicarse la ley sobre los registros administrativos.

---

4 Ley Orgánica 14/2003 de 20 de Noviembre, sobre derechos y libertades de los extranjeros en España y su integración social, aplicable desde el 22 de Diciembre de 2005.

5 Montañés V., Cánovas R., Poza E., Viciana F. "Estimación de los efectivos y estructuras poblacionales de Andalucía a partir de la integración de fuentes estadísticas y administrativas". XVIII JECAS, Oviedo 2014.

La secuencia de eventos y los episodios se completan con información procedente del Movimiento Natural de Población en lo referente a defunciones, con objeto de establecer un correcto cierre de los episodios, así como la causa del fallecimiento.

### **5.3 Construcción del grid censal**

Uno de los productos ofrecidos en esta publicación consiste en la difusión de la información a través de mapas, utilizando las celdillas o cuadrículas resultantes de la construcción de una malla regular o grid sobre todo el territorio. Para generar la malla regular se siguen las indicaciones derivadas de las experiencias llevadas a cabo por el proyecto Geostat (ESSnet project Geostat) del Foro Europeo de Geografía y Estadística (EFGS), que desarrolla la generación de una malla formada por celdas de 1 km de lado, utilizando un mismo sistema de referencia espacial para la totalidad de Europa. Este proyecto realiza tareas de estimación de la distribución por celdas de la población para un conjunto amplio de países en todo el continente. Las celdas se codifican con un sistema estándar que sigue las indicaciones de la directiva Inspire.

La malla de referencia es la utilizada en la publicación “Distribución espacial de la población en Andalucía”. A partir de la malla de 250 metros completa para Andalucía es preciso situar cada una de las viviendas ocupadas en la fecha de referencia censal en la celda correspondiente. Para ello se realizó un trabajo de georreferenciación del Censo de 2001 utilizando como fuentes para la obtención de coordenadas la información territorial del Callejero Digital de Andalucía Unificado (CDAU) y el Censo de Edificios de 2011. A ellas se les unen como fuentes auxiliares los cambios de denominación de vía de los Callejeros del Censo Electoral y de la propia BDLPA. El sistema de referencia de coordenadas empleado es ETRS89 / UTM 30N (EPSG: 25830).

Este proceso se realiza a nivel de edificio para garantizar que todas las viviendas pertenecientes al mismo edificio sean situadas en el mismo lugar. Una vez representadas las coordenadas, se puede determinar su celda de pertenencia.

Inevitablemente, resta un conjunto de viviendas sin coordenadas cercano al 3,9%. Como el fin último es conocer la celdilla y no la propia georreferencia a nivel puntual, se opta por unas aproximaciones con alto grado de seguridad, utilizando como información básica el código de vía INE, el número y la manzana.

Para los restantes casos, situados fundamentalmente en núcleos secundarios y diseminados, se optó por una asignación aleatoria dentro del núcleo de pertenencia y entre los edificios existentes en CDAU y Censo de 2011.

Finalmente se realizó un proceso de contraste y reasignación en los casos en que la celdilla asignada inicialmente no pertenecía a la sección censal de residencia.

La asignación de la celdilla de 1 km se realiza mediante agregación de las de 250 metros. A estos elementos espaciales se les denominará indistintamente como celdas, celdillas, cuadrículas o secciones estadísticas.

## **6. Contenidos de la publicación**

A partir de las tablas extraídas de la BDLPA, tras su depuración estadística para suprimir inconsistencias, se han realizado tres productos para difusión:

- 1.) Un conjunto de tablas estadísticas sobre el conjunto de la población con seguimiento efectivo. Estas tablas combinan la información recogida en los censos (2001 y 2011, respectivamente) con los sucesos (fallecimiento o censura) y tiempos de exposición experimentados por esta población durante el periodo de seguimiento (2002-2016 y 2012-2016, respectivamente).
- 2.) Una serie de ficheros muestrales de microdatos, preparados para suministrar a usuarios especializados toda la información necesaria que pudieran necesitar en sus respectivos proyectos, que no haya sido satisfecha por las tablas estadísticas del plan de tabulación previo.



- 3.) La estimación de un indicador robusto de los niveles de mortalidad para área pequeña (celdillas de grid) y su representación sobre mapas del conjunto de Andalucía con los datos del Censo de 2001.

A continuación se describe con más detalle cada uno de estos tres productos.

### **6.1 Tablas estadísticas con los datos de longevidad de las cohortes censales**

Con objeto de proporcionar a los posibles usuarios una información fácilmente accesible sobre las características de supervivencia y longevidad de la población andaluza a comienzos del siglo XXI, se han definido una serie de tabulaciones que permiten estudiar la relación de la longevidad con un amplio abanico de características personales socio-demográficas, relativas al hogar de convivencia, al entorno físico o los patrones de movilidad recogidos por los Censos de Población y Viviendas de 2001 y 2011.

La tabulación de los ficheros estadísticos preparados para este trabajo se realiza cruzando un grupo de variables censales (tiempo-independientes) con otras relativas a la experiencia de seguimiento de los miembros de las cohortes. Las variables asociadas al seguimiento son tiempo-dependientes y pueden cambiar a lo largo de la duración del seguimiento. En este trabajo estas variables son la edad y periodo en el que se ha producido el suceso o ha existido una exposición; y la unidad territorial de residencia del suceso o la exposición (solo en algunas de las tablas). Los cruces de estas variables generan un hipercono multidimensional con un conjunto de celdillas o compartimentos donde se almacenan los datos o índices numéricos.

Estos compartimentos contienen información numérica referida a un subgrupo poblacional determinado (definido por las variables censales), con respecto a su experiencia de seguimiento mientras ha transitado por un grupo de edad y/o periodo calendario y/o unidad territorial. La información almacenada es básicamente el número de sucesos (fallecimientos) y el tiempo de seguimiento acumulado por todos los individuos que han transitado por el espacio de estados delimitado por las variables que definen la celdilla. A la suma de los tiempos de exposición a riesgo de todos los individuos que han transitado por ella lo denominamos "Personas-Año" (de exposición).

Así pues, cualquiera de las tabulaciones de esta fuente estadística produce un hipercono multidimensional de datos. Cada una de las dimensiones de este hipercono son o bien información censal o bien una característica que cambia durante el seguimiento (grupo de edad, periodo y residencia en algunas tablas). La información numérica almacenada en estos hiperconos ofrece datos bien sobre el "número de fallecimientos" en cada una de las subcategorías, en algún caso incluyendo además la causa de la defunción, o sobre el "tiempo de exposición en riesgo" en dichas subcategorías (personas-año). Además, se ofrece la tasa bruta de mortalidad (no ajustada) expresada en tanto por mil calculada para cada cruce, definida como el cociente entre el número de fallecimientos y el tiempo de exposición en riesgo.

El plan de tabulación ha definido un conjunto de cruces complejos que han generado una serie de hiperconos, los cuales han sido cargados en el Banco de Datos Estadístico de Andalucía (BADEA). BADEA es una herramienta de difusión que permite al usuario diseñar la consulta a medida de sus necesidades de información, a través de un interfaz intuitivo, posibilitando localizar y consultar datos de una manera muy flexible y exportarlos en múltiples formatos para uso posterior, o en su caso, generar gráficos y visualizar mapas con la información seleccionada. A partir de BADEA el número de posibles tablas estadísticas extraíbles de los hiperconos originales es enorme.

Los hiperconos originales también se suministran en formato PC-Axis para aquellos usuarios que prefieran usar sus propias herramientas análisis.

Este plan de tabulación se ha diseñado delimitando cada una de las tablas en base a tres criterios:

1. Criterio temático: determina el objetivo y secundariamente las variables (aunque no su nivel de detalle) a incluir en las tablas. Las *características demográficas sexo, edad y estado civil*, características transversales que se incluirán prácticamente en todos los hiperconos, son combinadas con algunas variables correspondientes a otras categorías temáticas: *características socio-económicas personales, estructuras de convivencia, calidad del entorno*,

*características relativas a la actividad económica, características relativas a la movilidad y de la causa de la defunción (solo para la cohorte censal 2001).*

**2. Criterio territorial:** se han definido tres agrupaciones con diferente nivel de detalle territorial: Andalucía y sus provincias, unidades territoriales del Plan de Ordenación del Territorio de Andalucía (POTA) y secciones censales<sup>6</sup>

**3. Colectivo sub-poblacional:** determina el subconjunto de la población a la que hace referencia una tabulación concreta. Dado que determinadas variables solo están disponibles en un determinado colectivo, la información de cada una de ellas está referida a la población filtrada para las categorías para las que se dispone información. Los tres criterios de filtrado usados para generar una tabla concreta son los siguientes:

- *Total de población censada en viviendas familiares y colectivas:* Dado que sobre los colectivos se recogió un cuestionario muy simplificado, los conjuntos temáticos que se podrán aplicar a este colectivo serán necesariamente limitados. En este colectivo solo se dispone de variables demográficas básicas y nacionalidad (solo para la cohorte censal 2001).
- *Población en viviendas familiares:* Sobre esta población se dispone del grueso de las variables contenidas en el cuestionario censal.
- *Población que había trabajado la semana anterior a la entrevista censal:* Una serie de preguntas relativas a la actividad económica de la población solo se aplicó sobre el colectivo que había trabajado la semana anterior. Por lo que solo es posible investigar estas características en este grupo.

La información básica sobre la que se ha trabajado es la siguiente:

<b>Con información relativa al total de la población residente</b>	
Características demográficas básicas	<ul style="list-style-type: none"> <li>• Seguimiento poblacional por sexo, edad, generación, periodo y grupo de nacionalidad. Provincia de residencia censal (todos los censados). (Tabulación disponible solo para la cohorte censal 2001)</li> </ul>
<b>Con información relativa a la población residente en viviendas familiares</b>	
Características demográficas básicas	<ul style="list-style-type: none"> <li>• Seguimiento poblacional por sexo, edad, generación, periodo, grupo de nacionalidad y estado civil. Provincia de residencia censal.</li> </ul>
Características socioeconómicas personales	<ul style="list-style-type: none"> <li>• Seguimiento poblacional por sexo, edad, generación, nacionalidad y estado civil. Provincia de residencia censal.</li> <li>• Seguimiento poblacional por estado civil, nacionalidad, estudios, sexo y edad. Áreas POTA de residencia censal.</li> <li>• Seguimiento poblacional por estado civil, nacionalidad, estudios, sexo y edad. Provincia de residencia censal.</li> <li>• Seguimiento poblacional por estado civil, nacionalidad, estudios, sexo y edad. Área POTA en fecha del evento.</li> <li>• Seguimiento poblacional por estado civil, nacionalidad, estudios, sexo y edad. Provincia en fecha del evento.</li> <li>• Seguimiento poblacional por nacionalidad, propiedad de la vivienda, estudios, sexo y edad. Sección censal en 2002 y 2012.</li> </ul>
Características del hogar	<ul style="list-style-type: none"> <li>• Seguimiento poblacional por estado civil, nacionalidad, propiedad de la vivienda, 2ª vivienda, disponibilidad de vehículo (solo para la cohorte censal 2001) y tipo de hogar, estudios, sexo y edad. Provincia de residencia censal.</li> <li>• Seguimiento poblacional por estado civil, nacionalidad, propiedad de la vivienda, 2ª vivienda, disponibilidad de vehículo (solo para la cohorte censal 2001), tamaño y características económicas del hogar, estudios, sexo y edad. Provincia de residencia censal.</li> <li>• Seguimiento poblacional por estado civil, nacionalidad, propiedad de la vivienda, disponibilidad de vehículo (solo para la cohorte censal 2001) y tipo de hogar, estudios, sexo y edad. Área POTA de residencia censal.</li> </ul>
Calidad del entorno	<ul style="list-style-type: none"> <li>• Seguimiento poblacional por nacionalidad, estado del edificio y variables de hacinamiento, estudios, sexo y edad. Área POTA de residencia censal.</li> </ul>

<sup>6</sup> Tanto el número como el nivel de detalle de las variables del campo temático se reducen al aumentar el nivel de detalle territorial para disminuir el riesgo de revelación.

	<ul style="list-style-type: none"> <li>• Seguimiento poblacional por valoración personal del entorno, propiedad de la vivienda, estudios, sexo y edad. Áreas POTA de residencia censal. (Tabulación disponible solo para la cohorte censal 2001)</li> <li>• Seguimiento poblacional por valoración vecinal del entorno, propiedad de la vivienda, estudios, sexo y edad. Conjunto de Andalucía. (Tabulación disponible solo para la cohorte censal 2001)</li> </ul>
Características en relación a la actividad económica	<ul style="list-style-type: none"> <li>• Seguimiento poblacional por clasificación ocupacional de los activos, situación profesional, rama de actividad, propiedad de la vivienda, estudios, sexo y edad. Provincia de residencia censal.</li> <li>• Seguimiento poblacional por clase socio-ocupacional, propiedad de la vivienda, estudios, sexo y edad. Áreas POTA de residencia censal (Población en viviendas familiares ocupada en fecha censal).</li> </ul>
Movilidad	<ul style="list-style-type: none"> <li>• Seguimiento poblacional por clase socio-ocupacional, tipo de desplazamiento, estudios, sexo y edad. Áreas POTA de residencia censal (Población en viviendas familiares ocupada en fecha censal).</li> <li>• Seguimiento poblacional por tipo desplazamiento, relación con la actividad, propiedad vivienda, estudios, sexo y edad. Conjunto de Andalucía.</li> </ul>
Causa de la defunción <sup>7</sup>	<ul style="list-style-type: none"> <li>• Seguimiento poblacional por periodo, nacionalidad, estado civil, estudios, propiedad de la vivienda, sexo y edad. Áreas POTA de residencia censal. (Tabulación disponible solo para la cohorte censal 2001)</li> <li>• Fallecidos por periodo, nacionalidad, estado civil, estudios, propiedad de la vivienda, causa de la defunción, sexo y edad. Áreas POTA de residencia censal. (Tabulación disponible solo para la cohorte censal 2001)</li> <li>• Seguimiento poblacional por periodo, nacionalidad, estado civil, estudios, propiedad de la vivienda, sexo y edad. Provincia de residencia censal. (Tabulación disponible solo para la cohorte censal 2001)</li> <li>• Fallecidos por periodo, nacionalidad, estado civil, estudios, propiedad de la vivienda, causa de la defunción, sexo y edad. Provincia de residencia censal. (Tabulación disponible solo para la cohorte censal 2001)</li> </ul>

## 6.2 Ficheros de microdatos de las cohortes censales

Para cada uno de los dos estudios incluidos en esta publicación electrónica, relativos a las cohortes censales de 2001 y 2011, se han preparado varias muestras de microdatos con una determinada selección de las variables extraída los ficheros estadísticos censales. Las muestras censales se han completado con la información del seguimiento vital y residencial incluida en la BDLPA.

Para la cohorte censal de 2001 se han extraído dos muestras desde los ficheros censales (una de personas y otra de viviendas familiares). El procedimiento de selección se ha realizado mediante muestreo estratificado utilizando como estratos las secciones censales en la muestra de personas y las unidades territoriales del POTA en la muestra de viviendas. En cada uno de los estratos se ha realizado un muestreo aleatorio simple de las respectivas unidades muestrales, con una fracción de muestreo del 10%. Para la cohorte censal de 2011 se ha utilizado la propia muestra censal elaborada por el INE para el censo de población de 2011.

Cada uno de los registros seleccionados en las muestras ha sido enlazado con la información contenida en la BDLPA relativa a fallecimientos, residencia y migraciones durante su seguimiento en la ventana de observación de finales del año de referencia censal hasta el 31-12-2016. En torno al 98% de los registros incluidos en las muestras han podido ser localizados en la BDLPA.

Para el estudio de la cohorte censal de 2001 se han preparado tres bloques de microdatos:

- Un primer bloque, sin detalle territorial, que incluye información censal de la muestra de viviendas enlazada con el tiempo de seguimiento hasta el fallecimiento, primera salida de Andalucía o fin de estudio, suministrado por la BDLPA. El fichero de seguimientos

<sup>7</sup> Las causas se adaptan a la lista abreviada basada en la lista de "Carga Global de Enfermedades Simplificada". (Ver Anexo Causas)

de este bloque solo incluye un registro (episodio) por registro muestral enlazado, con la duración de seguimiento del individuo y la causa de su finalización.

- Un segundo bloque, con información de migraciones interiores a nivel de zonas POTA de la misma muestra de viviendas anterior. Los registros censales que han sido enlazados con la BDLP tienen uno o mas registros en la tabla de seguimientos, uno por cada episodio residencial continuado en alguna de la demarcaciones territoriales empleadas.
- El tercer bloque de microdatos incluye información de sección censal de residencia en la fecha censal. Se ha confeccionado a partir de una muestra aleatoria de personas, no de viviendas, con objeto de reducir el riesgo de revelación. El fichero asociado con los microdatos de seguimiento tampoco incluye información de cambio residencial, por lo que solo incluye un episodio vital por persona enlazada.

En esta nueva actualización de la publicación de estadísticas longitudinales se ha optado por suprimir de los ficheros con seguimiento post-censal la información relativa a las re-entradas en Andalucía de aquellos individuos que en algún momento del seguimiento emigraron fuera de Andalucía. Su inclusión complicaba sobremanera la estructura de los ficheros y no aportaba información relevante para los estudios de supervivencia, así como se corría el riesgo de derivar indicadores de mortalidad sesgados por una incorrecta utilización de esta información.

Para el estudio de la cohorte censal de 2011 se han preparado igualmente tres bloques de microdatos:

- Un primer bloque, sin detalle territorial, incluye información censal de la muestra censal y, para los registros enlazados, la información relativa a la duración de seguimiento del individuo y la causa de su finalización (un solo episodio por registro muestral enlazado).
- Un segundo bloque, con información de migraciones interiores a nivel de zonas POTA, de la misma muestra anterior. Los registros censales que han sido enlazados con la BDLP tienen uno o mas registros en la tabla de seguimientos, uno por cada episodio residencial continuado en alguna de la demarcaciones territoriales empleadas.
- El tercer bloque de microdatos incluye información de sección censal de residencia en la fecha censal. El fichero de seguimientos asociado no contiene información de cambio, por lo que solo incluye un episodio por persona enlazada.

En ambos estudios, el primero de los bloques de microdatos se puede descargar sin restricciones desde la página de la publicación y permite a los usuarios especializados familiarizarse con la información del estudio y realizar los análisis que precisen para el conjunto de Andalucía. Los otros dos bloques de microdatos requieren una solicitud nominal y, tras su aprobación, la firma de un compromiso de uso y confidencialidad.

Los detalles de contenido, las estructuras y el diccionario de datos se encuentra en el documento "Ficheros de microdatos: estructura y diccionario de datos" que se puede descargar desde la página de la publicación.

### **6.3 Indicadores y mapas de mortalidad por celda**

Este producto estadístico ha sido concebido con el principal objetivo de poner a disposición de los potenciales usuarios una gran cantidad de información en bruto con objeto de que ellos mismos puedan crear y definir sus propios análisis. Sin embargo, como ejemplo de la potencia de la información disponible se ha incluido una estimación de indicadores de mortalidad por área pequeña.

La ampliación del detalle territorial de las estadísticas clásicas es uno de los añadidos más interesantes que la integración de fuentes realizada en la BDLP conlleva. Con objeto de mostrar su potencial se ha estimado el indicador de mortalidad "Razón de Mortalidad Estandarizada Suavizada" (RMES) por celda de residencia en el Censo de 2001, en diferentes subcolectivos de grupo de edad, sexo y causas de defunción. Se ha elegido este indicador por su relativa sencillez de interpretación y robustez que lo hacen adecuado para su uso como indicador de área pequeña.

La razón de mortalidad estandarizada (RME) es un indicador que no se ve afectado por las distintas estructuras poblacionales, lo que permite comparar la mortalidad entre poblaciones con estructuras muy distintas. La RME utiliza el método indirecto de estandarización, también conocido como de las tasas tipo. Para su cálculo se estima el número de casos (defunciones) que se producirían en una determinada zona (con estructura conocida) si ésta estuviera sometida a las intensidades recogidas por las tasas tipo. A este resultado se le denomina casos esperados (E). En este trabajo se han usado como tasas tipo las tasas de mortalidad por edad y sexo del conjunto de Andalucía. La razón entre los fallecidos realmente observados (O) y los esperados (E) es el indicador RME para una unidad de análisis determinada.

La reducción del ámbito territorial de análisis, que pasa de ser una sección censal (con unas poblaciones más o menos equilibradas) a una celda regular de 250 metros o 1 kilómetro de lado, trae como consecuencia que las observaciones en la unidad de interés pueden no ser suficientes para la estimación del indicador RME y su significatividad. Debido a esto, se hace necesario buscar nuevos métodos que reduzcan la variabilidad del indicador para poblaciones pequeñas y que permitan calcular su significatividad de forma más eficiente. De esta búsqueda surge el método de suavizado bayesiano local, que no sólo utiliza la información de la propia celda, sino que también utiliza la estimación del entorno de la celdilla analizada como apoyo para obtener una población mayor y considerar su influencia en la mortalidad de la propia celdilla. A partir de este entorno y mediante técnicas de remuestreo es posible obtener un indicador suavizado y su intervalo de credibilidad a un nivel de confianza establecido que permita clasificar la celdilla en uno de los 5 grupos de mortalidad considerados: baja, moderadamente baja, similar a la media de Andalucía, moderadamente alta y alta.



**Ilustración 1. Esquema de vecindades para la estimación de la RMES de una celdilla A**

Para calcular la RMES en una celda A, se tienen en cuenta los valores de mortalidad observados y esperados en las celdillas de su entorno. Se toma como entorno mínimo 3 celdillas, comenzando desde la propia celda y expandiendo hacia afuera. En el primer caso, el entorno más inmediato consta de 8 celdillas. En el segundo, únicamente hay 2 celdillas, por lo que es necesario ampliar la zona de influencia para conseguir el mínimo número de celdas establecido. Si en un número máximo de ampliaciones prefijado no se llega al mínimo de celdillas (3) se considera como no estimable.

El proceso realizado para obtener cada una de las estimaciones puede resumirse en las siguientes fases:

1. Selección del entorno de influencia de la celdilla.
2. Cálculo de defunciones observadas y esperadas de la celda y su entorno, teniendo en cuenta la estructura poblacional.
3. Proceso de remuestreo bootstrap (con repetición) sobre la celda de interés y su entorno, asegurando la presencia de la propia celda en todas las muestras. Se toman 100 muestras por celda.

4. Cálculo de un estimador suavizado de la RME a partir de la estimación bayesiana local<sup>8</sup> de cada muestra.
5. Obtención del RMES de la celda, como media de sus 100 muestras bootstrap<sup>9</sup>, y de la distribución empírica del estimador que permite construir un intervalo de credibilidad para la misma.

El estimador bootstrap no precisa que los datos sean independientes e idénticamente distribuidos, no precisa de suposiciones paramétricas restrictivas y converge en probabilidad al verdadero valor del estimador. Además, la RMES es menos sensible a datos extremos que la RME.

El cálculo de la RMES se realiza con los tiempos de exposición a riesgo en el periodo 2002-2016 y se ha llevado a cabo en dos niveles de detalle:

- Utilizando celdillas de 250 metros de lado y desagregando por sexo para el total de población, para la población de 45 a 64 años y para la población de 65 y más años.
- Utilizando celdillas de 1 kilómetro de lado y desagregando por sexo para una serie de grupos de causas de muerte. Estos grupos se refieren a:
  - Enfermedades transmisibles
  - Neoplasias (tumores)
  - Enfermedades del sistema circulatorio
  - Causas externas
  - Todas las causas

En base a los niveles del indicador RMES se han representado en la herramienta MAPEA los mapas con los niveles de mortalidad por área pequeña para aquellas zonas pobladas en la fecha de referencia del censo de 2001.

La RMES estimada para una determinada celda se ha representado sobre el mapa con distinto gradiente de color. Las celdas con RMES que no difieren significativamente de la mortalidad media de Andalucía se han representado con una tonalidad neutra. Los grupos con RMES significativamente más altas o más bajas a la referencia de Andalucía, se han subdividido en dos subgrupos de efectivos similares y etiquetados respectivamente como superior (“moderadamente alta”) o muy superior (“alta”) en el caso de los superiores a la media; o inferior (“moderadamente baja”) o muy inferior (“baja”) en el caso de los inferiores al valor medio de Andalucía.

## **7. Control del riesgo de revelación y preservación del secreto estadístico**

La difusión de los resultados de toda actividad estadística se enfrenta siempre a un conflicto entre el derecho a la información y a la intimidad. El dilema se concreta entre el nivel de detalle con el que difundir la información y paralelamente la necesidad de preservar el secreto estadístico, o lo que es lo mismo, la identidad de los informantes. Dada la riqueza y detalle de la información contenida en la BDLPA, este conflicto es especialmente manifiesto.

En esta publicación se ha optado por una estrategia de difusión que maximice el nivel de detalle de la información bruta suministrada a los usuarios especializados con objeto de que éstos puedan realizar sus propios análisis, adecuados a sus objetivos e intereses. La principal limitación en cuanto al nivel de detalle y desagregación de las variables de esta publicación está motivada por el hecho de que el incremento del detalle incrementa proporcionalmente los riesgos de revelación. Por lo que, en último extremo, el máximo nivel detalle es una solución de compromiso entre el interés de la información y riesgo de revelación a él asociado.

---

<sup>8</sup> Basada en la metodología propuesta en “Mapping disease and mortality rates using Empirical Bayes Estimators (Applied Statistics, 40)”, Marshall R M (1991); “Interactive Spatial Data Analysis (Harlow: Longman)” Bailey T, Gatrell A (1995).

<sup>9</sup> Otros usos de la simulación bootstrap para estimaciones en áreas pequeñas se pueden encontrar en “Estimadores de áreas pequeñas basados en modelos para la Encuesta de Población Activa (Estadística Española Vol.51)”, Herrador M y otros (2009).

Se han tomado tres medidas tendentes a garantizar la necesaria preservación del secreto estadístico:

- Introducir una pequeña cantidad de perturbaciones aleatorias en los ficheros de explotación estadística, a partir de los cuales se han realizado las tabulaciones y la extracción de los ficheros de microdatos. Estas perturbaciones introducen un muy bajo nivel de ruido aleatorio que no alteran los resultados de los análisis estadísticos.
- Realizar una estrategia de agregación de las categorías de las variables en función de su poder de identificación y de las combinaciones con otras variables identificadoras en una misma tabla. Cuando se combinan dos o más variables con alta capacidad identificadora se utiliza en al menos una de estas variables una agregación que reduzca apreciablemente el nivel de detalle. Por ejemplo, no se incluyen tabulaciones que mezclen detalle territorial y temporal, o en los datos con gran detalle territorial se restringe el nivel de detalle de otras variables.
- Por último, los ficheros de microdatos con una muestra de los registros individuales, que tienen asociado un mayor riesgo de revelación, se distribuyen tras petición específica para un objetivo estadístico justificado y previa firma de un compromiso de uso y no redistribución de los datos brutos proporcionados.

## 8. Anexos

### 8.1 Algunos conceptos específicos de este tipo de estudios

#### 8.1.1 Variables tiempo dependientes (VTD) y tiempo independientes (VTI)

Una de las principales ventajas de las estadísticas derivadas de la nueva fuente BDLPA es la posibilidad de añadir información adicional no disponible en las fuentes primarias. En concreto, en este estudio se ha incluido gran cantidad de información recogida en los censos de 2001 y 2011, lo que permite investigar las relaciones entre esta y la longevidad de los distintos subgrupos poblacionales delimitados por las covariables censales.

Las variables censales, como nivel educativo, actividad económica o características de la vivienda, tienen que ser tratadas en este estudio como **variables tiempo independientes (VTI)**, es decir, referidas a sus características en la fecha censal ya que no disponemos de los cambios que en ellas han podido producirse a lo largo de los periodos de seguimiento. En aquellas variables de tipo estructural este hecho no tiene especial relevancia, algunas son claramente fijas como la generación o lugar de nacimiento. Otras, como por ejemplo el máximo nivel educativo alcanzado en los mayores 30, presumiblemente no cambiarán significativamente a lo largo del periodo de seguimiento, por lo que es lícito dividir la población participante en compartimentos estancos que no sufrirán intercambio entre ellos durante el seguimiento. En estos casos es posible calcular indicadores de duraciones medias, tales como la *esperanza de vida por nivel educativo* (en población mayor de 30) o por *lugar de nacimiento*.

No ocurre lo mismo con otras características, como por ejemplo la actividad económica, que varía tanto en función del ciclo económico como del vital: es más probable que un activo sufra un episodio de desempleo durante una recesión que en épocas de crecimiento, o es muy probable que una persona pase a ser jubilada al cumplir 65 años. En estos casos no es adecuado presuponer que la característica en la fecha censal se mantenga estable en el futuro. Para el cálculo de las duraciones en estos casos se precisaría disponer de los cambios (transiciones) en el *espacio de estados* de la actividad económica producidos durante el seguimiento para analizarlos mediante modelos multiestado. Aunque se está trabajando para obtener esta información, aún no se dispone del nivel de detalle suficiente para incluir estas variables (la actividad económica, el estado civil...), como variables tiempo-dependientes.

Aun así, variables coyunturales observadas al comienzo del seguimiento se utilizan ampliamente en muchos estudios de cohortes, ya que sin duda son útiles predictores. Por ejemplo, con los datos de este estudio es posible comprobar que las personas con estudios universitarios presentan distintos índices de supervivencia en los años posteriores de seguimiento que aquellas con estudios básicos. Sin embargo, con la información que suministra este diseño, no se pueden estimar cuestiones como la esperanza de vida de las “personas desempleadas” o de los “trabajadores en activo”, aunque sí de los que tuvieron un

episodio de desempleo en la fecha de la entrevista censal o estaban trabajando la semana previa a la entrevista.

Las **Variables Tiempo Dependientes** (VTD), son aquellas para las que se necesita precisar su estado en distintos tiempos del periodo de seguimiento. Por ejemplo el *lugar actual de residencia* sería una variable VTD, mientras que el *lugar de residencia en la fecha censal* es una variable de tipo VTI. En este estudio las únicas variables VTD que se han incluido son la de *residencia actual* (categorizada como provincia y unidad territorial definida en el POTA) y las escalas de tiempo *edad* y *periodo* una vez discretizadas. Por ello, excluyendo las escalas de tiempo, el espacio de estados más complejo que podemos estudiar con estos datos está formado por los estados vivo y fallecido y las 34 unidades territoriales del POTA.

Si bien este estudio incluye muchas variables con gran detalle de información territorial y de características personales o familiares, hay que tener en cuenta que éstas son solo observadas en las fechas censales y por lo tanto tienen que ser tratadas como VTI en el análisis.

### **8.1.2 Concepto de episodio, estado, eventos y transiciones**

Las poblaciones de partida de este estudio quedan definidas por una parte como los individuos entrevistados en la operación de recuento censal de 2001, que se encontraron inscritos y en alta en alguno de los padrones municipales de Andalucía a fecha del 1-1-2002 (el 97,9% de la población censada), y por otra parte como los individuos recogidos en la muestra del Censo de 2011, que pudieron ser localizados en algún alta padronal de un municipio de Andalucía (98,1% de los censados encuestados) y que residían en Andalucía el 1-1-2012. Estas poblaciones cerradas a nuevas incorporaciones, es decir, que no incorporan nacimientos ni entradas por inmigración, van reduciéndose progresivamente con el paso de los años por los fallecimientos y las emigraciones.

De los participantes originales que comenzaron su seguimiento al inicio de los respectivos periodos de observación, la gran mayoría mantendrá su residencia en Andalucía hasta el final del seguimiento el 31-12-2016. Sin embargo algunos saldrán de Andalucía (emigración) en algún momento del periodo de observación, e incluso unos pocos de los que salieron retornarán después de haber pasado un periodo fuera de Andalucía. Por último, y lo que es más relevante para el objetivo de esta investigación, muchos de ellos fallecerán durante este periodo de seguimiento.

Así pues, el número de participantes no se mantiene fijo a lo largo del periodo de seguimiento, ya que este efectivo poblacional está modificándose continuamente por los flujos de salida y, en su caso, de re-entrada.

Tampoco se mantienen invariables sus características (es decir, su estructura). Por ejemplo muchas de las variables recogidas en el momento de la entrada en el estudio (la información de cuadernillo censal en nuestro caso), no se mantienen invariables durante el seguimiento. Determinadas características relevantes, como por ejemplo la edad, van cambiando a lo largo del periodo de seguimiento: los participantes envejecen durante el estudio y este dato es relevante para sus funciones de supervivencia. También algunos participantes cambian de residencia, lo cual puede también ser relevante para determinado tipo de investigaciones.

Este tipo de estudios se pueden analizar en base a modelos conceptuales diferentes. En un primer acercamiento, directamente derivado del análisis de supervivencia, las unidades de investigación son los individuos, los cuales están categorizados por un conjunto de covariables invariables en el tiempo y una variable dependiente: el tiempo hasta la muerte o censura (por cierre de estudio o por salida, emigración, antes de la fecha de cierre). Este esquema tiene la ventaja de su relativa sencillez y algunas limitaciones como la no fácil inclusión de variables tiempo dependientes y la imposibilidad de utilizar más de una escala temporal continua como variable dependiente. A pesar de estas limitaciones el acercamiento permite una rica variedad de análisis, en concreto, el fichero de microdatos de la muestra de personas que se distribuye con esta publicación ha sido diseñado para derivar fundamentalmente este tipo de modelos. En este tipo de esquema de análisis cada uno de los individuos participantes interviene con un único episodio de seguimiento que comienza con el inicio del estudio y termina con el primer evento de censura (fallecimiento, emigración o cierre del estudio).

Si se precisa incluir variables tiempo-dependientes o escalas de tiempo discretas hay que subdividir los seguimientos individuales en varios episodios definidos por los distintos valores



que toman las variables tiempo-dependientes en cada uno de ellos. En este caso, las unidades de investigación no son ya los individuos participantes, si no los distintos **episodios de seguimiento** que cada uno de los participantes generan a lo largo de su seguimiento individual.

Así pues, se define **episodio** como una *duración de seguimiento* (es decir, un intervalo temporal), durante el cual un individuo concreto no ha modificado su *estado*. Los *episodios* están delimitados por dos fechas: la de “*inicio*” y la de “*fin*” del episodio. Cada uno de estos episodios queda caracterizado por las variables tiempo-independientes propias de cada participante, además del valor de las variables tiempo-dependientes específicas del episodio y que pueden cambiar en un mismo individuo a lo largo de su seguimiento, y por lo tanto, generar varios episodios en un mismo individuo.

La fecha (o momento) en que se produce un nuevo episodio se asocia con un evento que es el responsable del cambio en el **estado** de un individuo (por ejemplo vital o residencial: un fallecimiento, una migración) y por ello a este tipo de eventos se les denomina **transiciones**.

Por ejemplo, si definimos como relevante para la investigación la edad y esta variable continua es discretizada en años cumplidos, un participante que cumpliera 40 años el 1-1-2002 y que siguiera vivo y residiendo en Andalucía el 1-1-2017, generaría 15 episodios distintos de un año de duración, con una edad cumplida diferente y con distinta propensión a fallecer en cada uno de ellos.

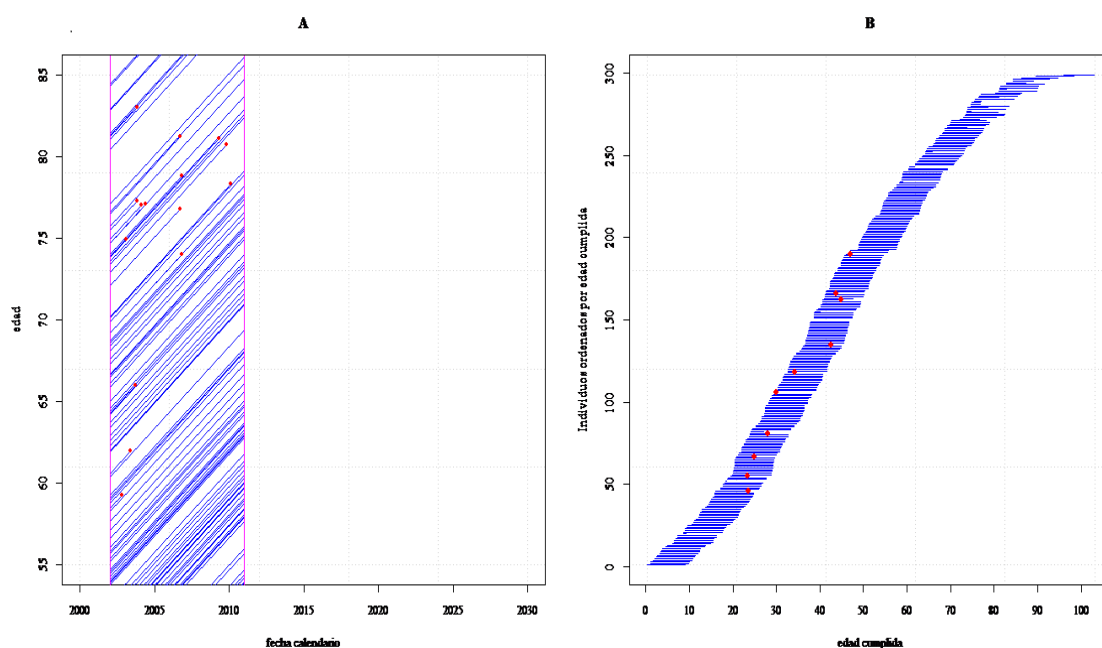
Los *episodios* se definen en función de un **espacio de estados** explícito y definido tanto en base a los objetivos de la investigación en curso como a la calidad y nivel de detalle de la información de la fuente que se utilice. En este estudio, se ha definido un sencillo *espacio de estados* en torno a estas dimensiones:

1. El estado vital de los participantes: vivo o fallecido.
2. Su lugar de residencia: en Andalucía o fuera de Andalucía o la unidad territorial que le corresponda.
3. Dos escalas de tiempo: el periodo y la edad.

Estas características son las únicas sometidas a vigilancia durante el periodo de seguimiento del estudio y por lo tanto susceptibles de investigación como *variables tiempo dependientes*. Conforme se vayan desarrollando los sistemas de información asociados a la BDLPA será posible incluir nuevas variables tiempo dependientes, como el nivel educativo, la actividad económica o el estado civil.

En el gráfico 1.A se ha representado una pequeña muestra de los episodios de este estudio definidos sobre la dimensión estado vital (vivo o muerto) sobre un recuadro de un diagrama de Lexis (de 20 años de lado). En él, se aprecia un subconjunto de episodios vitales observados durante varios años de seguimiento. Todas las *líneas de vida* comienzan el 1-1-2002, fecha de inicio del seguimiento y la mayoría acaban al final del periodo de estudio. Unos pocos episodios finalizan previamente a la fecha de estudio, bien por fallecimiento, evento marcado como un punto rojo en el diagrama, bien por emigración hacia fuera de Andalucía, sin una marca específica.

**Gráfico 1. Representación de una muestra de 300 seguimientos de la cohorte censal sobre un diagrama de Lexis y una representación alternativa con una sola escala de tiempo**



Si del gráfico 1.A suprimimos las escalas temporales de fecha calendario, dejando solo la edad y ordenando por edad al comienzo del estudio, tenemos el esquema de seguimientos del gráfico 1.B que es útil en análisis de supervivencia. En él se muestran los episodios vitales solo en la escala de tiempo edad. Como se aprecia en este gráfico, los episodios vitales se encuentra censurados por la derecha (censura tipo I no aleatoria por fin de estudio o censura tipo II aleatoria por salida de Andalucía) y truncados por la izquierda a la edad al comienzo del estudio el 1-1-2002.

### 8.1.3 Subdivisión de episodios basada en escalas de tiempo discretas

Las *escalas de tiempo edad y periodo* son habituales predictores en los modelos demográficos. Si bien es posible trabajar con ellas como variables continuas si se suministra en forma de microdatos, lo habitual es transformarlas en unidades discretas, tales como años cumplidos o periodos anuales, para facilitar la agregación de los episodios y la producción de tablas y resúmenes numéricos con eventos y tiempos de exposición a riesgo.

Un *episodio* definido en una determinada dimensión (por ejemplo estado vital o residencial) permanece invariable en la dimensión por la que ha sido definido (por ejemplo la residencia), pero no ocurre lo mismo con las escalas de tiempo durante las cuales dicho episodio existe. Por ejemplo, la edad o la fecha calendario cambian de un modo continuo desde el inicio al fin del episodio. La discretización de las variables continuas temporales es la única manera de poder generar tablas resumen agregadas, de eventos y tiempos de exposición, con el conjunto de la información de la población en seguimiento.

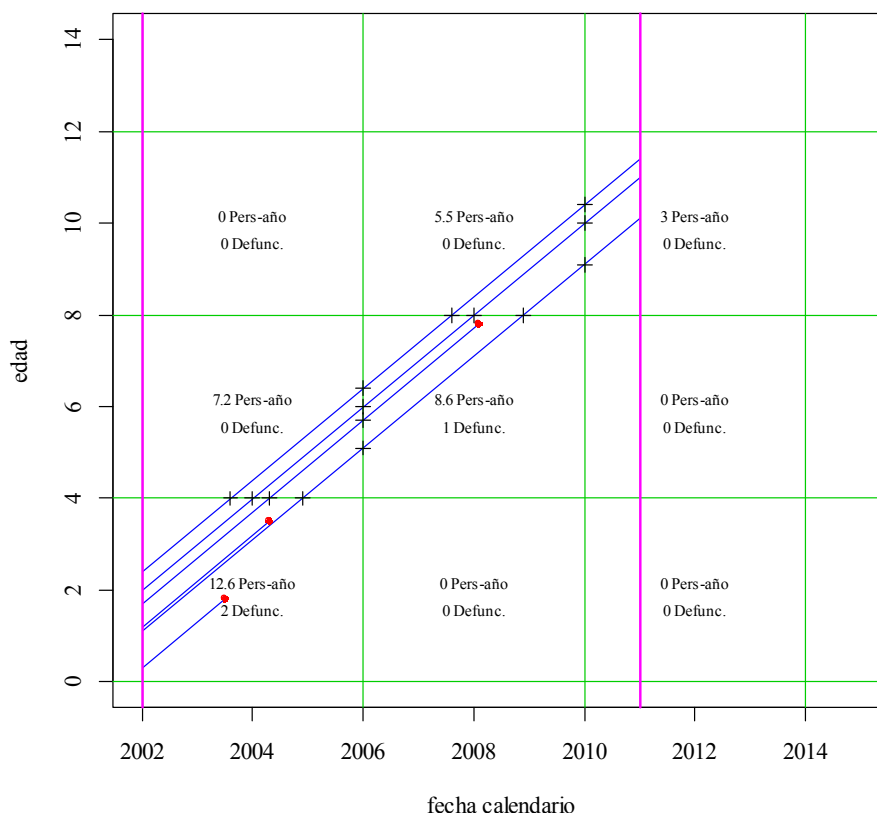
Para discretizar las escalas de tiempo se definen una serie de *fronteras temporales* en cada una de las escalas, a partir de las cuales se establece la partición discreta de éstas.

A nivel práctico, este paso consiste en transformar las variables continuas (por ejemplo, la edad) en funciones escalonadas, con la implicación de asumir que el efecto de la variable temporal es constante en todo el intervalo que se ha discretizado. Por ejemplo, el efecto de la edad sobre la supervivencia: una persona con 40 años de edad cumplida, se considera con igual propensión a fallecer el día que cumple esta edad, como 364 días después. Estas fronteras, definidas sobre las escalas edad y periodo, dibujan un malla sobre el diagrama de Lexis que permite dividir los episodios en sub-episodios con edad y periodos ya invariables (al haber sido transformados a una escala discreta).

En el diagrama de Lexis adjunto (gráfico 2) se han representado seis episodios vitales de este tipo de estudio: tres de ellos han sufrido un evento defunción, representado con un punto rojo y

los otros tres han sobrevivido hasta el fin de periodo de seguimiento. Sobre este conjunto de episodios se muestra la malla (edad, periodo) utilizada para discretizar las escalas temporales, con objeto de derivar una tabla de *eventos y personas-año*. Con objeto de simplificar la representación se han elegido grupos de edad y periodo cuatrienales que no son habitualmente usados en las estadísticas de población.

**Gráfico 2 Recorte de episodios al discretizar las escalas de tiempo edad y periodo**



Las líneas verticales violeta marcan el comienzo y el fin del periodo de seguimiento en este ejemplo, la rejilla verde representa los *tiempos frontera* que se han usado en cada *escala de tiempo* para discretizarla. Las cruces negras sobre las *líneas de vida* azules representan las intersecciones de las líneas de vida con las *fronteras* de las categorías temporales discretas (grupos de edad y de periodo). En este sentido las “cruces” son transiciones entre los estados definidos por las escala de tiempo discretas (por ejemplo grupo de edad). Los episodios vitales originales, quedan divididos por estas marcas en *subepisodios* durante los cuales se considera que los efectos de la edad y del periodo no van a cambiar significativamente (a nivel práctico se toman como constante a lo largo de toda la duración del subepisodio).

En el centro de los cuadrados delimitados por esta rejilla, se han colocado las sumas de las duraciones de los episodios (personas-año) y el número de eventos producidos en los intervalos temporales discretos definidos en cada una de las celdillas de la malla.

Es posible diseñar mallas con más de dos escalas de tiempo, introduciendo por ejemplo la dimensión *generación* de nacimiento, o mallas más sencillas, por ejemplo definiendo solo grupos de edad (y no haciendo relevante el efecto periodo).

#### 8.1.4 Precauciones en la interpretación de resultados

Los estudios longitudinales, son especialmente valiosos para realizar investigación causal: qué relación hay entre poseer determinadas características, habitualmente denominadas factores

de exposición en estos estudios, y la posterior aparición de determinado suceso o evento, entre la población que comporte dicho factor.

Esta publicación ha sido específicamente diseñada para la investigación de las relaciones causales entre las características personales y del entorno, recogidas en los censos de 2001 y 2011<sup>10</sup>, y la frecuencia o cronología con la que, en años posteriores, ocurren defunciones<sup>11</sup> en dicha población.

La relativa sencillez para estimar indicadores de supervivencia y mortalidad en subpoblaciones definidas en base a múltiples características socio-económicas, geográficas o demográficas, recogidas en los cuadernillos censales, conlleva un cierto riesgo al permitir derivar conclusiones erróneas a partir de estrategias de análisis no planteadas correctamente. El gran número de variables analizables y la diversidad de estrategias que pueden plantearse a partir de distintas combinaciones y agregaciones de las variables disponibles obliga a ser especialmente cauteloso a la hora de explicar determinados resultados.

Las relaciones causales exigen la precedencia temporal y la asociación estadística entre causa y efecto. Pero ambos criterios, aunque necesarios, no son suficientes para asegurar la existencia de una "relación causal". En particular, una asociación estadística entre la exposición y el efecto está condicionada por la existencia y la magnitud de dos tipos de error: el "error aleatorio" y el "error sistemático". El error aleatorio es connatural con la variabilidad espontánea de los procesos sociales; su existencia no actúa tanto sobre la validez de determinadas estimaciones, si no sobre la precisión de estas. El sesgo o error sistemático, por el contrario, sí afecta a la validez de los resultados de un análisis, es decir a la generalización de unos resultados en unas determinadas conclusiones sobre la existencia o magnitud de un efecto.

Con respecto al "error aleatorio", este estudio, al suministrar información sobre un colectivo de gran tamaño, produce indicadores agregados con muy bajo error aleatorio. Sin embargo la posibilidad de estratificar la población combinando múltiples variables hace que rápidamente aparezcan subpoblaciones de muy pequeño tamaño en determinados estratos donde el error aleatorio asociado a una determinada estimación puede ser muy superior al efecto que se pretende medir. Afortunadamente la teoría estadística suministra abundantes herramientas para manejar y estimar este tipo de error.

Más sutil y complejo de manejar es la aparición de sesgos de confusión y la derivación de conclusiones a partir de ellos. La estadística define a un estimador como sesgado cuando sus valores se diferencian sistemáticamente en una misma dirección sobre el valor teórico poblacional. En los estudios donde se investiga una relación causal, como es el caso de los estudios longitudinales, los sesgos de confusión se manifiestan por la aparición de asociaciones estadísticas entre las variables de exposición y las de efecto cuando realmente no hay asociación causal, o por el contrario la inexistencia de asociación estadística cuando hay relación causal.

Es frecuente la aparición de asociaciones estadísticas que no responden a relaciones causales reales sino a la existencia de sesgos, de confusión o de selección, en una determinada estrategia de análisis planificada. Por ejemplo, la aparición de una asociación estadística positiva entre la mortalidad y la propiedad de una segunda residencia, en un análisis donde no se haya tenido incluido la edad de las personas de referencia del hogar, puede estar ocasionado fundamentalmente por esta variable omitida. La no inclusión en el análisis de un predictor como la edad, que está asociado tanto con la causa como con el efecto (la propiedad de la vivienda y la mortalidad aumenta con la edad), produce la aparición de asociación estadística espuria, que puede ser fácilmente mal interpretada. Desafortunadamente no existe un procedimiento seguro para evitar estos errores.

La interpretación de los resultados derivados de una fuente con la riqueza y complejidad de la suministrada plantea la posibilidad de derivar conclusiones no correctamente fundamentadas, por la posible introducción de un componente de error no valorado.

Por otro lado, hay que tener presente que los indicadores de supervivencia o mortalidad derivados de esta fuente, no pueden coincidir con los indicadores anuales estimados a partir de

10 De esta información se extraerán las variables independientes del modelo estadístico (causas)

11 El evento defunción, su aparición o no y el momento (tiempo) en que aparece es el efecto (la variable dependiente de los modelos estadísticos)

las fuentes transversales habitualmente publicadas. Varias razones justifican estas diferencias. Por un lado, la población de la cohorte censal de 2001, si bien es muy similar a las estimaciones de población en fechas cercanas a la de referencia censal, se va distanciando conforme avanza el seguimiento de la cohorte. La población residente en Andalucía está continuamente incorporando población por nacimientos o inmigraciones, sin embargo ambas entradas no son incluidas en la cohorte censal, que se mantiene como una población cerrada a entradas. La razón para no incorporar nuevas entradas es por no disponer, para los nuevos nacimientos e inmigraciones, del conjunto de variables de exposición recogidas durante la operación censal de 2001.

Los eventos recopilados para las cohortes censales (defunciones y emigraciones exteriores), tampoco coinciden con los eventos producidos en el conjunto de la población andaluza en los años posteriores a los censo de 2001 y 2011, dado que las defunciones anuales y emigraciones de la población de Andalucía incluyen los eventos ocurridos en las nuevas entradas por migración y nacimiento después de 2001 y 2011, pero no en la cohorte censal. Otra razón de posible origen de discrepancia entre las fuentes puede deberse a errores en los enlaces de los registros y los eventos posteriores. Se ha estimado que esta fuente de subregistro se sitúa en torno al 1%.

Ambas diferencias en numeradores (eventos) y denominadores (poblaciones medias expuestas a riesgo) impiden la coincidencia perfecta de los indicadores derivados de esta fuente longitudinal con las fuentes transversales hasta ahora usadas para los estudios de mortalidad y el análisis de supervivencia. Aun así, la evolución y correlación de los indicadores derivados de ambas fuentes es notable.

## **8.2 Anexo: Fuentes primarias**

El estudio sobre mortalidad en Andalucía basado en las cohortes censales de 2001 y 2011 se apoya en las tres fuentes clásicas de información demográfica de España: el censo, el padrón y el movimiento natural de población. A continuación se describe el origen y las principales características de estas fuentes.

### **8.2.1 El padrón continuo**

El padrón municipal es el registro administrativo donde constan todos los vecinos de un municipio. Este registro administrativo sirve de base para una gran cantidad de actividades administrativas y estadísticas, ya que se usa, por ejemplo, para obtener las cifras oficiales de población, la estadística de variaciones residenciales o la modificación del registro del censo electoral.

Hasta 1996, el padrón era un registro con renovaciones quinquenales a través de un sistema de visitas a los domicilios de los administrados de cada ayuntamiento. Cada 5 años se desechaba el padrón y se creaba uno nuevo en cada municipio, exigiendo una operación de campo exhaustiva.

Esta situación cambió radicalmente con la reforma legislativa de 1996, la Ley 4/1996, que modificaba a la Ley 7/1985, de 2 de abril, Reguladora de las Bases del Régimen Local y su posterior desarrollo reglamentariamente por el Real Decreto 2612/1996, de 20 de diciembre, por el que se modificaba el Reglamento de Población y Demarcación Territorial de las Entidades Locales. Con este cambio legislativo el padrón pasa a actualizarse de manera continua, en lugar de cada cinco años como venía realizándose hasta entonces. La información más relevante se encuentra entre los artículos 15 y 17 de la citada ley. Las ideas fundamentales son:

- Toda persona que viva en España está obligada a inscribirse en el Padrón del municipio en el que resida habitualmente.
- La inscripción deberá ser objeto de renovación periódica cada dos años cuando se trate de la inscripción de extranjeros no comunitarios sin autorización de residencia permanente<sup>12</sup> (ENCSARP).
- La inscripción contendrá obligatoriamente: nombre, apellidos, sexo, domicilio, nacionalidad, documento nacional de identidad o similar, titulación académica y lugar y fecha de nacimiento.

<sup>12</sup> Ley Orgánica 14/2003 sobre derechos y libertades de los extranjeros en España y su integración social que modifica el artículo 16.1 de la Ley 7/1985.

- Los datos del Padrón Municipal se cederán a otras Administraciones Públicas que lo soliciten sin consentimiento previo del afectado solamente cuando les sean necesarios para el ejercicio de sus respectivas competencias, y exclusivamente para asuntos en los que la residencia o el domicilio sean datos relevantes. También pueden servir para elaborar estadísticas oficiales sometidas al secreto estadístico.
- Corresponde a los ayuntamientos la formación, mantenimiento, revisión y custodia del Padrón y al Instituto Nacional de Estadística (INE) su coordinación.
- Los distintos órganos y Organismos de la Administración General del Estado competentes por razón de la materia remitirán periódicamente a cada ayuntamiento información sobre las variaciones de los datos de sus vecinos que con carácter obligatorio deben figurar en el padrón municipal, a fin de que puedan mantenerse debidamente comprobados y actualizados dichos datos. En particular, las Oficinas del Registro Civil, el Ministerio del Interior y el Ministerio de Educación.

En resumen, la importancia estadística de esta fuente reside en que con esta reforma legislativa se posibilitó la existencia de un registro de población exhaustivo y actualizado continuamente que, aunque es gestionado descentralizadamente por los ayuntamientos, es coordinado centralizadamente por el INE.

### **8.2.2 El Censo de Población de 2001**

Los Censos Demográficos han sido tradicionalmente los **proyectos estadísticos de mayor envergadura** que periódicamente debe acometer la oficina de estadística de cualquier país. Un censo es un conjunto de operaciones que consisten en recopilar, resumir, valorar, analizar y publicar los datos de carácter demográfico, cultural, económico y social de todos los habitantes del país y de sus divisiones político-administrativas, referidos a un momento o período dado. Esta operación va dirigida a todas las personas que residen en viviendas, ya sean familiares o colectivas.

Bajo la denominación *Censos Demográficos* se englobaron en realidad tres censos diferentes en el año 2001: el Censo de Población, el Censo de Viviendas y el Censo de Edificios. Los recuentos de las distintas unidades censales se referían a una única fecha censal, que en este caso se fijó en el 1 de noviembre de 2001.

A grandes rasgos, en el Censo de Población se incluyen únicamente las personas, de cualquier nacionalidad, que tienen fijada su residencia habitual en el territorio nacional, aunque estén temporalmente en el extranjero.

Por lo que se refiere al Censo de Viviendas, su ámbito poblacional abarca a las viviendas familiares (habitadas o no) y a las viviendas colectivas. Se consideran viviendas todos los recintos destinados a habitación humana, y aquellos otros que sin tener esa finalidad están efectivamente habitados en la fecha del censo, denominados alojamientos (tales como cuevas, caravanas...).

Uno de los objetivos de los censos de 2001 es servir de base para estudios demográficos. Tal y como indica la metodología de la operación censal del INE *“El Censo de Población permite obtener datos que son indispensables para el estudio de la evolución de la población”*. Asimismo, sirve para calcular tasas específicas de las características investigadas en las estadísticas de flujos.

En el censo de 2001 desapareció la doble definición de población (de hecho y de derecho) de anteriores censos, unificándola bajo el concepto población. No obstante, se introduce el concepto de población vinculada como el conjunto de personas censales (es decir, con residencia habitual en España) que tienen algún tipo de vinculación habitual con un municipio concreto, ya sea porque residen, trabajan o estudian allí, o porque, no siendo su residencia habitual, suelen pasar allí ciertos periodos de tiempo (veraneos, puentes, fines de semana...).

#### **8.2.2.1 Relación entre el Censo de 2001 y el padrón continuo**

La principal innovación metodológica del Censo de 2001 fue el aprovechamiento de los registros administrativos, en concreto de los Padrones Municipales de Habitantes y de la base de datos del Catastro Urbano, que fueron por primera vez utilizados para simplificar la operación de campo.

La opción que se consideró óptima para la realización de los censos españoles del 2001 fue una operación basada en un recorrido exhaustivo del territorio, fuertemente apoyada en el padrón continuo, y con un cuestionario también exhaustivo y único.

Los artículos 78 y 79 del Reglamento de Población, preservan perfectamente la relación de mutuo beneficio que ha venido dándose hasta ahora entre ambos documentos.

- El INE puede llevar a cabo operaciones de control de la precisión de los Padrones municipales, informando del resultado a los correspondientes Ayuntamientos.
- La formación del Censo de Población, que constituye una competencia exclusiva del Instituto Nacional de Estadística, se apoyará en los datos de los Padrones municipales, y servirá para controlar la precisión de los datos padronales y, en su caso, para introducir en ellos las rectificaciones pertinentes.

Inciendo en esta interrelación, se optó por hacer llegar a cada vivienda, junto con el cuestionario censal, los datos padronales<sup>13</sup> vigentes en ese momento, con objeto de que las personas empadronadas en la misma tuvieran ocasión de comprobar su exactitud y de introducir, en su caso, las variaciones pertinentes.

### **8.2.3 El Censo de Población de 2011**

El Censo de Población de 2011 fue el primer censo realizado en España de una forma no tradicional. Los avances tecnológicos, la disposición en España de un padrón municipal de habitantes con una trayectoria de 15 años y otros factores, llevaron a cambiar el modelo de realización del Censo en el año 2011. Otros países de la Unión Europea optaron también por cambiar el modelo tradicional de Censo, aunque con diferentes estrategias adaptadas a las especificidades de cada país. En España se optó por una operación de tipo muestral para personas y viviendas, con una fracción de muestreo variable en función del volumen de habitantes de los municipios. Simultáneamente, se realizó un censo exhaustivo sobre los edificios.

Este cambio de estrategia en la realización supone una gran reducción del coste económico de la operación censal, pero por otro limita el ámbito geográfico en el que los datos son estadísticamente significativos. La fecha de referencia se fijó a 1 de noviembre de 2011, si bien los trabajos de recogida se extendieron al primer trimestre de 2012.

A grandes rasgos, en el Censo de Población se incluyen únicamente las personas, de cualquier nacionalidad, que tienen fijada su residencia habitual en el territorio nacional, aunque estén temporalmente en el extranjero en la fecha de referencia.

### **8.2.4 Estadísticas del MNP**

Las estadísticas del Movimiento Natural de la Población (MNP), que se refieren básicamente a los nacimientos, matrimonios y defunciones ocurridos en territorio español, constituyen uno de los trabajos de mayor tradición en el organismo estadístico nacional. El primer volumen se publicó en el año 1863 por la Junta General de Estadísticas del Reino conteniendo datos relativos al periodo 1858 a 1861, obtenidos a partir de los registros parroquiales. Desde entonces se viene publicando en España información de los fenómenos demográficos sin más interrupción que durante el periodo 1871 a 1885, época de implantación del Registro Civil, de donde se obtendrían, en lo sucesivo, datos para estas estadísticas.

Las estadísticas del Movimiento Natural de la Población recogen los datos de los boletines estadísticos cumplimentados en el momento de inscribir estos hechos demográficos en el Registro Civil.

El artículo 20 del Reglamento de la Ley del Registro Civil, dice que los Encargados del Registro Civil remitirán al Instituto Nacional de Estadística, a través de sus Delegaciones, los boletines de nacimientos, matrimonios y defunciones.

En lo relativo a Andalucía, en 1992 se estableció un acuerdo de colaboración entre IECA e INE para la elaboración de las Estadísticas de defunciones según causa de muerte. Desde que en 1996 entró en vigor un nuevo convenio de colaboración para la realización de las Estadísticas

---

<sup>13</sup> La información estadística contenida en el Padrón, única que sería exhaustiva, es muy escasa: sexo, nacionalidad, y fecha y lugar de nacimiento.

del Movimiento Natural de la Población (MNP) y Defunciones según la causa de muerte, el Sistema Estadístico de nuestra comunidad está participando activamente en los circuitos de producción y distribución de la información estadística del MNP, incluyendo en este intercambio, a partir de entonces, partos y matrimonios además de las defunciones.

#### 8.2.4.1 Relación entre MNP y Padrón Continuo

Además de esta vertiente de uso de la información de los boletines para su aprovechamiento estadístico inmediato, existe una vertiente administrativa. Así, los nacimientos y las defunciones se comunican a cada Ayuntamiento para que procedan a dar de baja a las defunciones y de alta a los nacimientos en el Padrón Municipal. En el caso de los nacidos, la inscripción requiere de la localización en los padrones de sus progenitores, lo cual puede producir ciertos retrasos en las inscripciones.

### 8.3 Anexo: Lista abreviada de Causas

DESCRIPCIÓN	CIE10
<b>Signos, síntomas y afecciones mal definidas</b>	
Signos, síntomas y afecciones mal definidas	R00-R99
<b>Enfermedades transmisibles</b>	
Enfermedades infecciosas intestinales	A00-A09
Tuberculosis	A15-A19
Ciertas enfermedades transmisibles por vectores y rabia	A20, A44, A75-A79, A82-A84, A85.2, A90-A98, B50-B57
Ciertas enfermedades inmunoprevenibles	A33-A37, A80, B05, B06, B16, B17.0 B18.0-B18.1, B26
Meningitis	A39, A87, G00-G03
Septicemia, excepto neonatal	A40-A41
Enfermedad por el VIH (SIDA)	B20-B24
Infecciones respiratorias agudas	J00-J22
Resto de ciertas enfermedades infecciosas y parasitarias	Resto de A00-B99
<b>Neoplasias (Tumores)</b>	
Tumor maligno del estómago	C16
Tumor maligno del colon y de la unión rectosigmoidea	C18-C19
Tumor maligno de los órganos digestivos y del peritoneo excepto estómago y colon	C15, C17, C20-C26, C48
Tumor maligno de la tráquea, los bronquios y el pulmón	C33-C34
Tumor maligno de los órganos respiratorios e intratorácicos, excepto tráquea, bronquios y pulmón	C30-C32, C37-C39
Tumor maligno de la mama	C50
Tumor maligno del cuello del útero	C53
Tumor maligno del cuerpo del útero	C54
Tumor maligno del útero, parte no especificada	C55
Tumor maligno de la próstata	C61
Tumor maligno de otros órganos genitourinarios	C51-C52, C56-C57, C60, C62-C68
Leucemia	C91-C95
Tumor maligno del tejido linfático, de otros órganos hematopoyéticos y de tejidos afines	C81-C90, C96
Tumores malignos de otras localizaciones y de las no especificadas	Resto de C00-C97
Tumores in situ, benignos y los de comportamiento incierto o desconocido	D00-D48
<b>Enfermedades del sistema circulatorio</b>	
Fiebre reumática aguda y enfermedades reumáticas crónicas	I00-I09
Enfermedades hipertensivas	I10-I15



Enfermedades isquémicas del corazón	I20-I25
Enfermedad cardiopulmonar, enfermedades de la circulación pulmonar y otras formas de enfermedad del corazón	I26-I45, I47-I49, I51
Paro cardíaco	I46
Insuficiencia cardíaca	I50
Enfermedades cerebrovasculares	I60-I69
Aterosclerosis	I70
Resto enfermedades del sistema circulatorio	I71-I99
<b>Ciertas afecciones originadas en el período perinatal</b>	
Ciertas afecciones originadas en el período perinatal	P00-P96
<b>Causas externas</b>	
Accidentes de transporte terrestre	V01-V89
Los demás accidentes de transporte y los no especificados	V90-V99
Caídas	W00-W19
Accidentes por disparo de arma de fuego	W32-W34
Ahogamiento y sumersión accidentales	W65-W74
Accidentes que obstruyen la respiración	W75-W84
Exposición a la corriente eléctrica	W85-W87
Exposición al humo, fuego y llamas	X00-X09
Los demás accidentes	W20-W31, W35-W64, W88-W99, X10-X39, X50-X59, Y40-Y84, X40-X49
Lesiones autoinfligidas intencionalmente (suicidios)	X60-X84
Agresiones (homicidios)	X85-Y09
Eventos de intención no determinada	Y10-Y34
Las demás causas externas	Y35-Y36, Y85-Y89
<b>Todas las demás enfermedades</b>	
Diabetes mellitus	E10-E14
Deficiencias nutricionales y anemias nutricionales	E40-E64, D50-D53
Trastornos mentales y del comportamiento	F00-F99
Enfermedades del sistema nervioso, excepto meningitis	G04-G99
Enfermedades crónicas de las vías respiratorias inferiores	J40-J47
Resto de enfermedades del sistema respiratorio	J30-J39, J60-J98
Apendicitis, hernia de la cavidad abdominal y obstrucción intestinal	K35-K46, K56
Cirrosis y ciertas otras enfermedades crónicas del hígado	K70, K73, K74, K76
Resto de enfermedades del sistema digestivo	Resto de K00-K93
Enfermedades del sistema urinario	N00-N39
Hiperplasia de la próstata	N40
Embarazo, parto y puerperio	O00-O99
Malformaciones congénitas, deformidades y anomalías cromosómicas	Q00-Q99
Resto de las enfermedades	Resto de A00-Q99