Estadísticas longitudinales de biografías reproductivas en Andalucía, 2002-2019

Metodología

Índice

1. Introducción....1
2. Origen y relevancia de las estadísticas longitudinales de población....3
3. Fuente de información....4
4. Finalidad y diseño del estudio....5
   4.1 Diseño prospectivo....7
   4.2 Diseño retrospectivo....7
5. Procesado y depuración de la información de base....8
   5.1 Integración de las fuentes en la BDLPA....9
   5.2 Construcción de los episodios....9
   5.3 Construcción del grid censal....10
6. Contenidos de la publicación....11
   6.1 Tablas estadísticas con los datos de fecundidad y número de hijos de las cohortes censales....11
   6.2 Ficheros de microdatos de la cohorte censal....13
   6.3 Indicadores y mapas de fecundidad por celda....13
7. Control del riesgo de revelación y preservación del secreto estadístico....16
8. Anexos....17
   8.1 Algunos conceptos específicos de este tipo de estudios....17
      8.1.1 Variables tiempo dependientes (VTD) y tiempo independientes (VTI)....17
      8.1.2 Concepto de episodio, estado, eventos y transiciones....18
      8.1.3 Subdivisión de episodios basada en escalas de tiempo discretas....20
      8.1.4 Precauciones en la interpretación de resultados....21
   8.2 Anexo: Fuentes primarias....23
      8.2.1 El padrón continuo....23
      8.2.2 El Censo de Población de 2001....24
      8.2.3 El Censo de Población de 2011....25
      8.2.4 Estadísticas del MNP....25

1. Introducción

La presente publicación, desarrollada por el Instituto de Estadística y Cartografía de Andalucía (IECA), que en su origen contó con la colaboración del Departamento de Población del Consejo Superior de Investigaciones Científicas y fue cofinanciada por el Fondo Social Europeo, continúa la línea de producción estadística basada en la integración y reutilización de información recopilada en otras operaciones estadísticas. Esta publicación sigue avanzando en el conocimiento demográfico desde una perspectiva longitudinal que ya se inició con las "Estadísticas Longitudinales de Longevidad y Supervivencia" publicadas por este organismo, pero centrando ahora la atención sobre el fenómeno de la fecundidad, un suceso con características muy diferentes para su análisis.

Esta explotación se ha confeccionado con el objetivo de suministrar información adecuada para analizar los diferentes comportamientos de la población en relación al fenómeno demográfico de la fecundidad combinando la información del nacimiento de los hijos con la información sobre las características personales (de la madre y el padre), familiares y del entorno recogidas en los Censos de Población y Viviendas de 2001 y 2011.

Al menos desde el punto de vista del análisis demográfico, la continuación natural en la difusión desde el punto de vista longitudinal es la confección de un nuevo producto estadístico, pero esta vez centrado en los eventos relacionados con el origen de la vida, en vez de con su finalización: el área temática de la investigación en fecundidad y las dinámicas de crecimiento familiar. Evidentemente, la fecundidad tiene características diferenciales tales como que es un evento renovable desde el punto de vista de los progenitores, que se trata de un evento que involucra a varios individuos de forma simultánea o que puede estar influenciado por factores sociales.

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 1991, 2001 y 2011.
2. Las estadísticas del Movimiento Natural de la Población, en concreto los registros correspondientes a los Boletines Estadísticos de Parto correspondientes a los años 2002 a 2019.
3. Las variaciones residenciales de la población entre los años 2002 y 2019 y recogidas por el sistema de actualización permanente de los padrones municipales.

La integración de todas estas fuentes permite precisar el número de hijos de los individuos al inicio del estudio y si con posterioridad se ha producido un nacimiento en el que haya estado implicado como padre o madre. También se han considerado las entradas o salidas de Andalucía por migración, ya que las edades en las que se desarrolla el evento de interés coinciden con las edades de mayor movilidad poblacional. De los eventos migratorios y de los nacimientos se conoce en qué fecha suceden, de forma exacta en el caso de los nacimientos y de forma aproximada en el caso de las migraciones.

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

El sistema de información que se está construyendo para realizar la integración de estas fuentes se ha denominado Base de Datos Longitudinal de Población (BDLPA). El todavía vigente 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 los próximos años está previsto desarrollar nuevos productos estadísticos longitudinales que continúen con los estudios de supervivencia y fecundidad, abarcando temas tales como

movilidad residencial de la población, biografías de actividad económica o 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 acercase 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. De esta forma se consiguen detalladas estadísticas territoriales o desagregaciones de colectivos poblacionales muy específicos que 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 Frank Willekens[1]:

> *"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 "Estadística de itinerarios vitales de los residentes en Andalucía". 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 comenzando a generar 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 todavía vigente Plan Estadístico y Cartográfico de Andalucía 2013-2020, que en su artículo 17.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,

1 F. Willekens. El futuro del análisis longitudinal. Información Estadística y Cartográfica de Andalucía nº1 pg: 76-78.

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 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. 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 se realiza 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 continuación natural 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 fecundidad de la población andaluza que no están disponibles en las estadísticas anuales de fecundidad y natalidad que hasta ahora se han publicado.

En concreto, con estas explotaciones de la BDLPA se pretenden generar nuevas estadísticas de fecundidad y dinámicas familiares hasta ahora no disponibles y que incluirán al menos estas líneas conductoras:

- Suministrar información a detalle geográfico infra-municipal.
- Reconstruir los episodios de paridad con objeto de poder estimar tasas de fecundidad de primera categoría, propensiones de ampliación de familiar y tablas de fecundidad.
- Investigar el efecto en la fecundidad y dinámica familiar de las características socio-económicas de los progenitores.
- Estudiar las diferencias que se producen entre generaciones, edades y sexo.

Si bien la maternidad o paternidad[2] son eventos renovables, pueden ser descompuestos en otros de tipo no renovable (como la mortalidad) bajo una serie de convenciones. Por ejemplo, tener un hijo es un evento renovable, pero tener el primer hijo es de tipo no renovable. De igual forma, se puede decir que es un evento no renovable el hecho de tener x hijos si se tienen x-1.

Otra característica del evento de la maternidad es que no es un fenómeno exhaustivo como sucedía con la mortalidad. Por evidente que resulte, hay que resaltar esta diferencia, ya que no todas las personas tienen hijos pero todas fallecen antes o después.

Otra apreciación importante es la limitación biológica del rango de edades en la que se producen los eventos. Así, si la mortalidad puede producirse a cualquier edad, la fecundidad no puede producirse en los primeros años de vida ni, en el caso de las mujeres, en edades avanzadas. Por lo general, la menopausia se produce entre los 45 y 55 años en países desarrollados, tal y como recogen diversos estudios. Tradicionalmente se ha tomado los 50 años como límite para la construcción de indicadores demográficos.

Por otro lado la delimitación de los “episodios” en relación a la fecundidad está determinada por el número de hijos al inicio y al final de un periodo de tiempo. Es evidente, que bajo la ventana temporal de 18 años considerada en el estudio, que un individuo concreto no haya tenido hijos no implica que no los tuviera antes de la fecha de inicio del estudio ni que no los vaya a tener después. Las herramientas conceptuales del “análisis de supervivencia”, pensadas en su inicio para estudios de tiempo de fallo o defunciones, han desarrollado una clara metodología que también es extrapolable para analizar este tipo de datos.

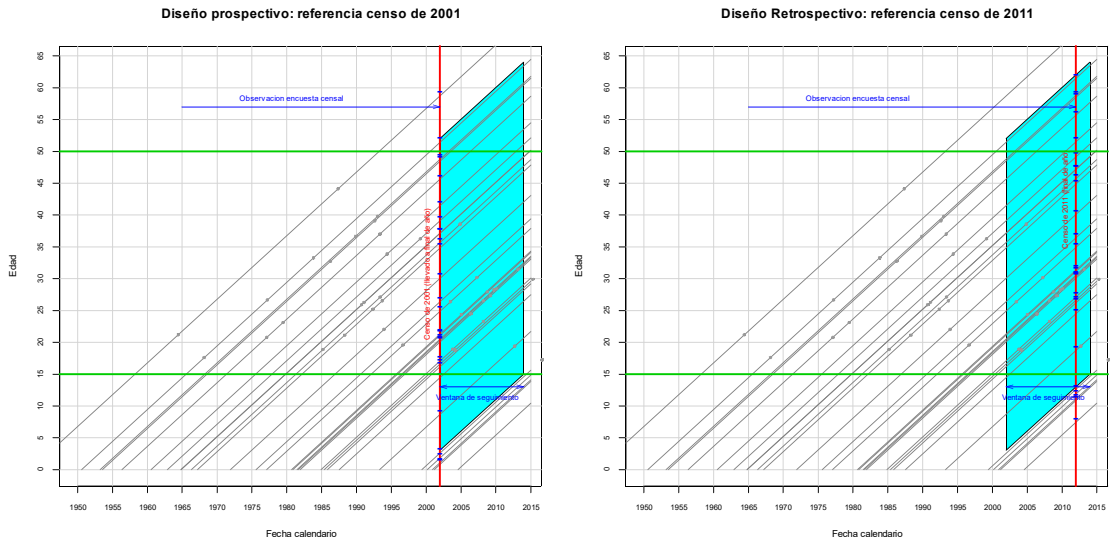
2 En lo sucesivo, cuando se hable de maternidad se hará extensivo al concepto paternidad cuando la información se refiera a hombres.

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 fecundidad tiene sobrado interés en sí misma y supone un tema de gran actualidad. Por ejemplo, los niveles de fecundidad actuales pueden influir en el futuro del sistema de pensiones, el número de potenciales trabajadores a medio plazo o los indicadores de envejecimiento. También es de interés conocer el calendario biológico en el que las personas tienen sus hijos para poder predecir futuras necesidades de equipamientos educativos en el tiempo y el espacio.

Con el objetivo de poder estimar la diferencia en fecundidad en diferentes colectivos poblacionales definidos tanto por su lugar de residencia como por sus características socio-económicas, se ha diseñado 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 Parto.

Esta publicación plantea dos tipos de diseño longitudinal. El primero de ellos es de tipo **prospectivo**, fijando las características de las personas (salvo edad y residencia[3]) al inicio del estudio a través de la información recogida en el censo de población y viviendas de 2001. El *segundo* es de tipo **retrospectivo**, es decir, observa las características de las personas al final del estudio (con la información recopilada por el censo de 2011). En ambos (Ilustración 1) se estudia cómo evoluciona la historia reproductiva de los individuos incluidos en cada estudio, durante la ventana de observación 2002-2019, pero en el prospectivo los eventos reproductivos ocurren después de observar las características personales recogidas en el censo de 2001 y el diseño retrospectivo lo eventos ocurren, mayoritariamente, antes de observar las características personales recogidas en el censo de 2011. Estas diferencias de diseño tienen importantes implicaciones en cuanto a cómo analizar, estimar indicadores y plantear hipótesis causales en cada uno de los dos tipos de estudio.

Ilustración 1. Esquema comparativo de estudio prospectivo y retrospectivo



A diferencia de la anterior publicación de supervivencia y longevidad en Andalucía, para la que se seleccionó toda la población censada en 2001, en esta ocasión sólo se han seleccionado a las personas de la generaciones nacidas entre 1950 y 2005[4] (ambos inclusive). Dos criterios fundamentales han determinado esta decisión. En primer lugar, se contemplan cinco décadas al completo, permitiendo divisiones por tramos quinquenales y decenales. En segundo lugar, por razones relativas al rango biológico en el que suceden la mayoría de los partos y el rango de edad habitualmente usada en demografía para el cálculo de los indicadores de fecundidad entre 15 y 49 años. Fuera de este rango, los partos suelen ser eventos poco frecuentes, suponiendo menos de un 0,1% del total de maternidades y un 3% de las paternidades. La

3 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. En el caso del nivel estudios alcanzado, hay una observación censal y otra final, pero no hay transiciones intermedias.

4 En el caso de la cohorte censal de 2001 el periodo de seguimiento se restringe a los nacidos entre 1950 y 2001.

elección de las fechas 1950 y 2005 se traduce finalmente en que las personas nacidas en 1950 tenían 51 años al iniciarse el estudio (63 al final) y los nacidos en 2005 tenían 14 años al finalizar el periodo de observación.

4.1 Diseño prospectivo

Se parte de las características de la población recogidas en los cuadernillo censales en una determinada fecha de referencia (1/1/2002[5]), sobre la cual se investiga los eventos relativos a partos, defunciones o variaciones residenciales que le han podido ocurrir en años posteriores. En concreto, el estudio se plantea bajo las siguientes especificaciones:

- **Población de partida**: Individuos recogidos en el Censo de 2001 nacidos entre 1950 y 2001 (5.355.108), que pudieron ser localizados en algún alta padronal de un municipio de Andalucía (5.241.395, 97,9% de los censados) y que residían en Andalucía el 1 de enero de 2002.

- **Periodo de seguimiento (ventana de observación)**: Sobre la población censada, residente en Andalucía el 1 de enero de 2002 se comprueban las modificaciones de su situación de residencia, estado vital y número de hijos recogidos tanto en las altas y bajas padronales como en los Boletines de Defunción y Parto, desde el 1 enero de 2002 hasta el 31 de diciembre de 2019 (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 31-12-2019.

Este diseño tiene la ventaja de disponer de una batería de variables socioeconómicas censales para la práctica totalidad de la población de referencia, permitiendo el conocimiento de la población en áreas pequeñas. Por el contrario, tiene la limitación de que algunas variables, como el nivel de estudios o la relación con la actividad cambian mucho para la población de seguimiento durante los 18 años del estudio para los menores de 36 años. Además, este tipo de estudio no recoge la influencia del boom migratorio de inicios de siglo, ya que no se considera a aquellas personas que entraron en Andalucía después del 1 de enero de 2002[6].

4.2 Diseño retrospectivo

El diseño retrospectivo[7] se construye de forma inversa al prospectivo. Se parte de la información de las características de la población observada al final del estudio (en este caso de forma aproximada, ya que el estudio se cierra el 31/12/2019 y las variables censales se recogen a finales de 2011) y se trata de estudiar diferencia en la biografía reproductiva de las personas según las características notificadas en el último censo. Hay que recordar que la operación censal de 2011 dejó de ser un censo tradicional y se basó en una muestra inferior al 10% del total poblacional.

Este segundo diseño se construye bajo las siguientes especificaciones:

---

5 Se optó por el 1/1/2002, en este caso dado que los trabajos de campo del censo de 2001 se extendieron de noviembre de 2001 a Abril de 2002

6 En las estimaciones publicadas en mapas sí se consideran reentradas e incluso individuos que no fueron censados en 2001 ni residían en Andalucía a 1 de enero de 2002.

7 El diseño no es retrospectivo en sentido estricto, ya que después de la fecha de referencia se siguen estudiando los individuos durante 2 años

- **Población participante**: Individuos recogidos en la muestra del censo de 2011 nacidos entre 1950 y 2005 (475.351), que pudieron ser localizados en algún alta padronal de un municipio de Andalucía (465.467, 97,9% de los censados encuestados) y que hayan residido en Andalucía durante el periodo de estudio.

- **Periodo de seguimiento (ventana de observación)**: Sobre la población censada se comprueban las modificaciones de su situación de residencia, estado vital y número de hijos recogidos tanto en las altas y bajas padronales como en los Boletines de Defunción y Parto, desde la última entrada en Andalucía posterior al 1 enero de 2002 y anterior al 1 de Enero de 2012, hasta el 31 de diciembre de 2019 (ambos inclusive). El comienzo de seguimiento de los participantes es bien el 1/1/2002 si ha residido de manera continuada en Andalucía desde dicha fecha, bien la fecha de la entrada del episodio residencial activo durante la operación censal de 2011

- **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:

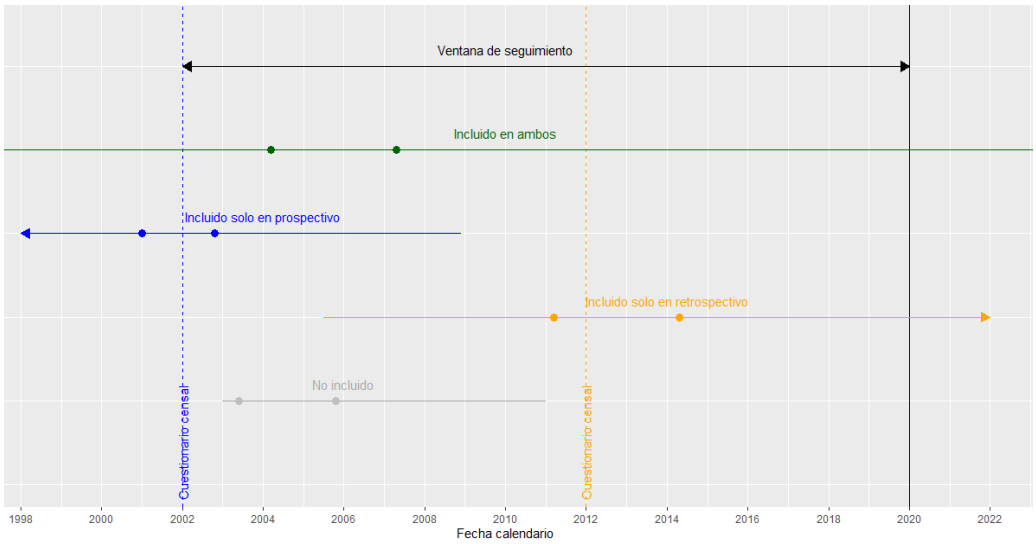
    - *Fallecimiento:* registrado en MNP y/o baja padronal por defunción, evidentemente con posterioridad a la fecha censal.

    - *Emigración fuera de Andalucía*: informada por una baja padronal, con posterioridad a la fecha censal.

    - *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 31-12-2019.

En este diseño, sólo se perderá una pequeña parte de la población por fin de seguimiento, ya que sólo transcurren 8 años desde la observación censal, en lugar de los 18 años del caso prospectivo. Este diseño tiene la ventaja de disponer de una batería de variables socioeconómicas cercanas al final del periodo de estudio. Se recoge a las personas que llegaron a Andalucía con el boom migratorio de inicios de siglo y que permanecieron posteriormente en la comunidad autónoma (*Ilustración 2*). Por el contrario, al tratarse de una operación muestral, no se puede abordar con garantías el cálculo de indicadores detallados en área pequeña.

Ilustración 2. Esquema de inclusión de individuos en estudio prospectivo y retrospectivo



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[8] 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.

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 claras 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, empleándose esta información 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 y para establecer las relaciones de paternidad/maternidad, así como obtener información sobre la fecha del parto y el número de hijos previo.

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 del Censo de 2001 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.

5.2 Construcción de los episodios

En este estudio los episodios están formados por una combinación de dos componentes. Por un lado están marcados por los cambios de residencia y por otro por el cambio en el número de hijos.

Los episodios residenciales muestran la permanencia de un individuo en una localización geográfica determinada desde una fecha establecida como inicio del episodio hasta otra fecha

8Cánovas 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/demograficasI/resultadosdelenlacedelosregistros.pdf

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 2002-2019 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 ley que han afectado a las bajas administrativas de extranjeros, comunitarios y extracomunitarios[9], 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[10] evita un descenso súbito de la cifra de efectivos poblacionales en las fechas en que comenzó a aplicarse la ley sobre los registros administrativos.

Por otro lado, la parte relativa al número de hijos se ha realizado teniendo en cuenta la información derivada fundamentalmente de los registros de partos ocurridos en Andalucía con al menos un nacido vivo. Para determinar el número inicial de hijos se ha combinado la información de los ficheros de partos, los Censos de 1991, 2001 y 2011 y la convivencia en el hogar entre los individuos.

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.

Por lo tanto, un episodio puede iniciarse por un cambio residencial o por el nacimiento de un hijo y puede concluirse por fallecimiento, cambio residencial o nacimiento de un hijo (iniciándose automáticamente otro episodio).

5.3 Construcción del grid censal

La presente publicación continúa con la difusión de la información a través de mapas, utilizando las celdillas resultantes de la construcción de una malla regular sobre todo el territorio. Para generar la malla regular se siguen las indicaciones derivadas de la experiencia 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 Unificado de Andalucía 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 es ETRS89 / UTM 30N (EPSG: 25830). Posteriormente se completó la georreferenciación añadiendo las viviendas ocupadas entre 2002 y 2019 que no se encontraban habitadas en el Censo de 2001.

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

---

9 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.

10 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.

Inevitablemente, resta un conjunto de viviendas sin coordenadas, que actualmente es cercano al 3%. Como el fin último es conocer la celdilla y no la propia georreferencia, 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, 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 se les denominará indistintamente como celdas o celdillas.

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) con los sucesos (fallecimiento, censura, parto) y tiempos de exposición experimentados por esta población durante el periodo 2002-2019. Se han separado las tablas por sexo.

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 indicadores de fecundidad (para el primer hijo) 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 fecundidad y número de hijos de las cohortes censales

Con objeto de proporcionar a los posibles usuarios una información fácilmente accesible sobre las características de fecundidad de la población andaluza a comienzos del siglo XXI, se han definido una serie de tabulaciones que permiten estudiar la relación de la fecundidad con un amplio abanico de características personales socio-demográficas, relativas a la forma de convivencia o a la relación con la actividad que fueron recogidas 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 la cohorte. 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 el número de hijos (paridad y multiplicidad), la edad y periodo en el que se ha producido el suceso o ha existido una exposición. Los cruces de estas variables generan un hipercubo 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. La información almacenada es básicamente el número de sucesos (partos simples y múltiples y nacimientos) 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). En lo sucesivo, se hablará en este documento sobre nacimientos, si bien se hace extensible al caso de partos con al menos un nacido vivo.

Así pues, cualquiera de las tabulaciones de esta fuente estadística produce un hipercubo multidimensional de datos. Cada una de las dimensiones de este hipercubo son o bien información censal o bien una característica que cambia durante el seguimiento (grupo de edad, periodo y número de hijos). La información numérica almacenada en estos hipercubos ofrece datos bien sobre el "número de nacimientos" (o partos) en cada una de las subcategorías, o sobre el "tiempo de exposición en riesgo" en dichas subcategorías (personas-año). Además, se ofrece la tasa de fecundidad según número de hijos previos (no ajustada) expresada en tanto por mil calculada para cada cruce, definida como el cociente entre el número de nacimientos y el tiempo de exposición en riesgo de la subpoblación analizada (según paridad tanto para hombres como para mujeres).

El plan de tabulación ha definido un conjunto de cruces complejos que han generado una serie de hipercubos, 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 hipercubos originales es enorme.

Este plan de tabulación se ha diseñado limitando 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 c*aracterísticas demográficas sexo, edad y paridad, que* se incluyen en todos los hipercubos, son combinadas con algunas variables correspondientes a otras categorías temáticas: *características socio-económicas personales, estructuras de convivencia o características relativas a la actividad económica*.

2. Criterio territorial: se han definido dos agrupaciones con diferente nivel de detalle territorial: Andalucía y sus provincias, y las unidades territoriales del Plan de Ordenación Territorial de Andalucía (POTA)[11].

3. Colectivo sub-poblacional: determina el subconjunto de la población a la que hace referencia una tabulación concreta. La información se ofrece para *población residente en viviendas familiares, permitiendo los distintos cruces establecer filtros posteriores.*

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

Con información relativa a la población residente en viviendas familiares	
Características demográficas básicas	• Sexo • Edad y generación • Nacionalidad • Número de hijos: paridad y multiplicidad • Estudios alcanzados (a fecha censal y a fecha de fin del estudio) • Estudios en curso (sólo 2001)
Características socioeconómicas personales	• Propiedad de la vivienda • Disponibilidad de 2ª vivienda • Disponibilidad de vehículo (sólo 2001)
Características del hogar	• Convivencia: con pareja, sin pareja • Tipo de hogar: hogares con núcleo de pareja, hogares monoparentales, otras tipologías o no procede (sólo 2001)
Características en relación a la actividad económica	• Relación con la actividad: trabajando, parado, pensionista, otra situación o no procede

11 www.juntade**andalucia**.es/medioambiente/pota

6.2 Ficheros de microdatos de la cohorte censal

Para el estudio *prospectivo*, se han extraído inicialmente dos muestras del fichero censal: una de personas y otra de viviendas familiares. Adicionalmente, se ha extraído una tercera muestra de microdatos relativos a viviendas a nivel de la comunidad autónoma.

Las dos primeras muestras se distribuyen previa solicitud expresa y tras la firma de un compromiso de uso y confidencialidad. La tercera muestra se distribuye sin restricciones especiales y permite a los usuarios especializados familiarizarse con la información del estudio y realizar los análisis que precisen para el conjunto de Andalucía.

En el caso del estudio prospectivo, cuyas variables de referencia son las del Censo de 2001, 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 fija del 10%. En caso de no existir información para el elemento muestreado durante el periodo de referencia, se sustituyó dicha unidad por otra, de forma que se tenga una muestra efectiva del 10%.

De cada uno de los registros seleccionados (personas y viviendas) se ha extraído la información relativa a su seguimiento durante la ventana de observación de 1-1-2002 al 31-12-2019. La información de seguimiento de cada una de las muestras incluye los eventos fallecimiento, cambio de residencia, cambio en el número de hijos y censura por fin del estudio el 31-12-2019. El nivel de detalle de la variación residencial es el de una salida o reentrada en el conjunto de la comunidad autónoma de Andalucía en la muestra de personas y las variaciones residenciales que ocurrían entre dos unidades territoriales de las definidas en el POTA en la muestra de viviendas.

Para el estudio *retrospectivo*, y debido a que el censo de 2011 es en sí mismo un proceso muestral, se distribuyen como microdatos todos los individuos muestreados que cumplen las condiciones necesarias para formar parte de este estudio de fecundidad (año de nacimiento, residencia, conocimiento de la trayectoria reproductiva…). La muestra es de alrededor del 98% de la muestra original y las variables han sido recodificadas para evitar el posible riesgo de revelación. Se distribuye mediante descarga directa la muestra sin detalle territorial y con las categorías de las variables mas agrupadas respecto a la original con objeto de evitar el posible riesgo de revelación de datos personales. Bajo petición expresa y compromiso de uso, se proporciona una segunda muestra con detalle territorial a nivel de comarca, tanto en residencia en la fecha censal como las variaciones residenciales ocurridas durante el periodo de seguimiento.

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 fecundidad por celda

Este producto estadístico ha sido concebido tomando como modelo la publicación previa de supervivencia y longevidad en Andalucía, con el 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 fecundidad 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 BDLPA conlleva. Con objeto de mostrar su potencial se han estimado, tanto para mujeres como para hombres, tres indicadores: número medio de hijos, edad media a la maternidad/paternidad e indicador de intensidad de la fecundidad al primer nacido.

El indicador de intensidad por el que se ha optado es análogo al usado en la en la publicación de longevidad, y se ha denominado “Razón de Fecundidad Estandarizada Suavizada” (RFES) Se ha elegido este indicador ya que su robustez y baja varianza lo hace adecuado para su uso como indicador de área pequeña. Para su cálculo se utilizan los efectivos de población sin hijos previos y los eventos primeros nacimientos.

El método indirecto de estandarización utilizado, también conocido como de las tasas tipo, precisa para su cálculo del número de casos (partos) que se producirían en una determinada zona (con estructura conocida) si ésta estuviera sometida a las intensidades recogidas por las tasas tipo (probabilidades de ampliación de familia por edad). A este resultado se le denomina casos esperados (E). En este trabajo se han usado como tasas tipo las estimadas por edad y sexo para el conjunto de Andalucía. La razón entre los primeros nacimientos observados (O) y los esperados (E) es el indicador RFE para una unidad de análisis determinada.

El indicador RFES se calcula para celdas regulares de 250 metros de lado para cada sexo. Adicionalmente, se ofrece el mismo indicador en una rejilla de 1km2 para seis periodos de observación: 2002-2004, 2005-2007, 2008-2010, 2011-2013, 2014-2016, 2017-2019. Este indicador, con menor detalle geográfico, permite estudiar cambios en el indicador en una misma zona respecto al cambio global de Andalucía en el mismo periodo, durante seis intervalos temporales.

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 RFE 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 fecundidad 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 fecundidad considerados: baja, moderadamente baja, similar a la media de Andalucía, moderadamente alta y alta.

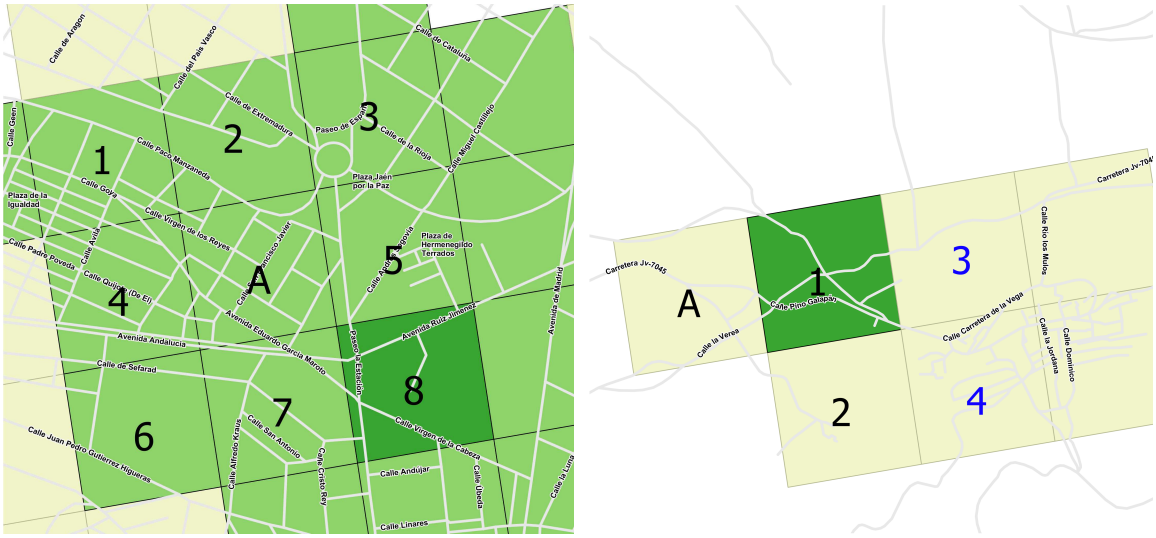


Ilustración 3. Esquema de vecindades para la estimación de la RFES de una celdilla A

Para calcular la RFES en una celda A, se tienen en cuenta los valores de fecundidad 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 partos observados y esperados 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 RFE a partir de la estimación bayesiana local[12] de cada muestra.

5. Obtención del RFES de la celda, como media de sus 100 muestras bootstrap[13], 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 RFES es menos sensible a datos extremos que la RFE.

El cálculo de la RFES se realiza con los tiempos de exposición a riesgo en el periodo 2002-2019 y se ha llevado a cabo utilizando celdillas de 250 metros de lado desagregando por sexo, y por trienios (2002-2004, 2005-2007, 2008-2010, 2011-2013, 2014-2016, 2017-2019) utilizando celdillas de 1km de lado. Se ha considerado en todo momento la zona de residencia del individuo, por lo que puede darse la circunstancia de que haya zonas sin indicador en el periodo 2002-2005 pero que sí lo tengan en 2017-2019, al haberse poblado dicha zona.

En base a los niveles del indicador RFES se han representado en la herramienta MAPEA del IDE de Andalucía los mapas con los niveles de fecundidad por área pequeña para aquellas zonas pobladas, entre los años 2002 y 2019, por personas de edades comprendidas entre los 15 y los 49 años.

La RFES estimada para una determinada celda se ha representado sobre el mapa con distinto gradiente de color. Las celdas con RFES que no difieren significativamente de la fecundidad media de Andalucía se han representado con una tonalidad neutra. Los grupos con RFES 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.

Por otra parte, el indicador del número medio de hijos se calcula en base a las tasas de fecundidad por grupos quinquenales de edad y sexo para cada celdilla, de forma análoga al cálculo de un índice sintético de fecundidad. Se han considerado los grupos de edad entre 15 y 49 años para su cálculo y, una vez se han tenido todas las tasas quinquenales (siete tasas), se han acumulado para conformar este indicador.

No obstante, al igual que sucedía con el indicador RFES, la consideración de un ámbito territorial reducido hace que no siempre se tenga población suficiente. Así pues, se ha optado por un método similar al de la RFES, utilizando suavizado bayesiano y técnicas de remuestreo tipo bootstrap para cada una de las siete tasas.

Sin embargo, el cálculo de este indicador tiene la peculiaridad de que necesitamos la estimación de todas esas tasas y puede no haber población en alguno de los grupos de edad. Para solventar este problema se han utilizado técnicas de agrupación (clúster no jerárquico) aplicadas al calendario de fecundidad de cada zona para establecer tipologías de curvas de

12 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).

13 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).

fecundidad (y niveles) para poder aplicarlas sobre aquellas celdas con datos faltantes en alguna tasa, de forma que se han sustituido los valores faltantes por valores de estas curvas de referencia con una distribución similar.

El cálculo de estas tasas quinquenales también ha servido de base para el cálculo de las edades medias a la maternidad y la paternidad.

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.

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 un gran 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 el Censo de Población y Viviendas de 2001, lo que permite investigar las relaciones entre ésta y la fecundidad de los distintos subgrupos poblacionales delimitados por las covariables censales.

Las variables censales, como nivel educativo o actividad económica, tienen que ser tratadas en este estudio como **variables tiempo independientes**[14] (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 12 años 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 de 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 el intervalo intergenésico *por nivel educativo*[15] (en población mayor de 30) o por *lugar de nacimiento* o probabilidades de ampliación de familia en función de la edad y el nivel de estudios.

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 en un futuro, 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 mujeres con estudios universitarios retrasan su calendario de fecundidad en los doce años posteriores al Censo de 2001 en comparación con las mujeres con estudios inferiores. Sin embargo, con la información que suministra este diseño, no se pueden estimar cuestiones como la edad media de las "personas desempleadas" que han tenido un hijo 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 para la elaboración de los indicadores de fecundidad (se tiene la rejilla de residencia en cada momento) son las escalas de tiempo *edad y periodo* una vez discretizadas y el número de hijos.

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 la fecha censal y por lo tanto tienen que ser tratadas como *VTI* en el análisis.

---

14 No obstante, al disponer de una muestra prospectiva y otra retrospectiva, una permite conocer el estado al inicio del periodo de estudio y otra permite conocerlo casi al final del mismo.

15 En esta edición también se incluye el nivel de estudios alcanzado a fecha de fin del estudio

8.1.2 Concepto de episodio, estado, eventos y transiciones.

Para el estudio prospectivo, la población de partida fue definida como los individuos entrevistados en la operación de recuento censal de 2001, nacidos entre 1950 y 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 objetivo censada). Esta población se considera cerrada a nuevas incorporaciones, es decir, que no incorpora entradas por inmigración y va reduciéndose progresivamente con el paso de los años fundamentalmente por las emigraciones y, en menor medida, por las defunciones.

De los participantes originales que comenzaron su seguimiento el 1-1-2002, la gran mayoría mantendrá su residencia en Andalucía hasta el final del seguimiento el 31-12-2019 (algo más del 88,4%). 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, algunos de ellos fallecerán durante este periodo de 18 años de seguimiento.

Así pues, el número de participantes no se mantiene fijo a lo largo de periodo de seguimiento, ya que este efectivo poblacional está modificándose continuamente por los flujos de salida. En este estudio prospectivo se han descartado las re-entradas.

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 reproducción.

También algunos participantes cambian de residencia, lo cual puede también ser relevante para determinado tipo de investigaciones. Por ejemplo, para el cálculo de los indicadores de área pequeña, que no utilizan más información personal que la edad, la generación y el sexo, se han tenido en cuenta todos los cambios de residencia de población, incluyendo entradas a Andalucía de nueva población.

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 transcurrido hasta tener un hijo 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 compleja 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, los ficheros de microdatos de la muestra de personas que se distribuyen con esta publicación han sido diseñados 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 (tener un hijo, fallecimiento, emigración o cierre del estudio). El caso de la fecundidad se puede readaptar a este análisis clásico estudiando el evento tener un hijo si previamente se tienen x-1 hijos.

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, un parto) 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 nacido el 28-11-1981, comenzaría con 20 años el 1-1-2002 y si permaneciera residiendo en Andalucía hasta el 31-12-2019, generaría 18 episodios distintos en función de su edad, entre 20 y 38 años (de un año de duración), con una edad cumplida diferente y con distinta propensión a tener hijos en cada uno de ellos.

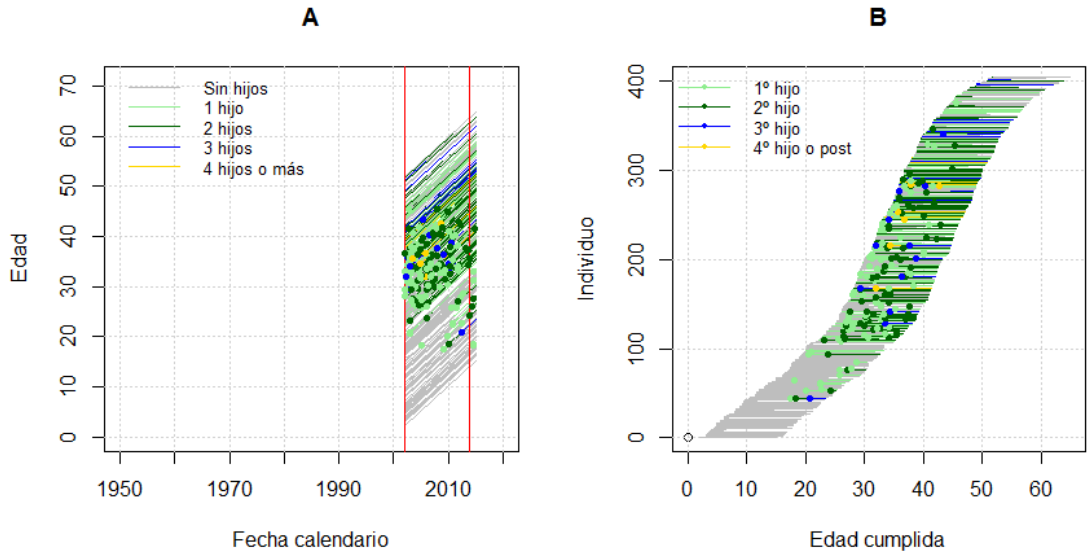
Los *episodios* se definen en función de un **espacio de estados** explícito y definido tanto en base a los objetivos de la investigación en curso como a la calidad y nivel de detalle de la información de la fuente que se utilice. En este estudio, se ha definido un sencillo *espacio de estados* en torno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

1. El número de hijos de los individuos: 0, 1, 2 y 3 o más
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 "número de hijos" 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 prospectivo. Unos pocos episodios finalizan previamente a la fecha de estudio, bien por fallecimiento, bien por emigración hacia fuera de Andalucía, sin una marca específica. Se distinguen con distinto color los tramos por número de hijos, señalando con un punto el instante en que se producen los nacimientos de sus descendientes.

Gráfico 1. Representación de una muestra de 400 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 encuentran 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.

Los puntos indican el momento en que se produce un cambio en el número de hijos.

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 suministran 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 número de hijos o residencia) permanece invariable en la dimensión por la que ha sido definido, pero no ocurre lo mismo con las escalas de tiempo durante la 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.

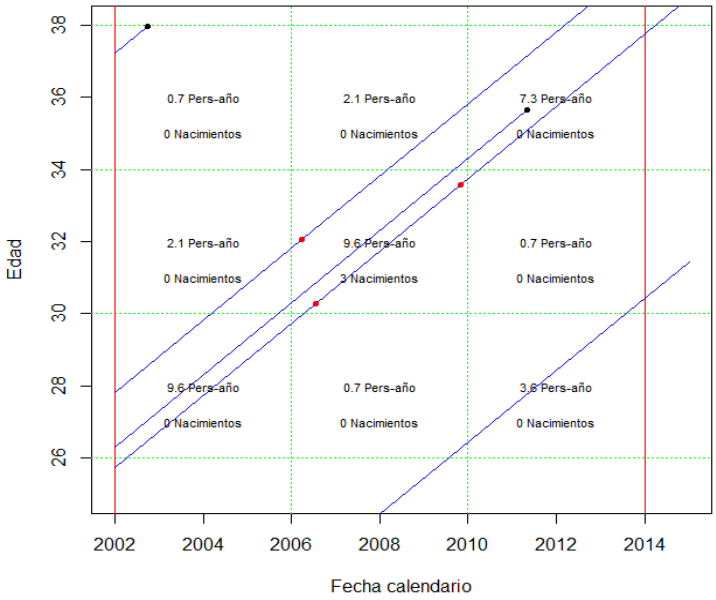
Para discretizar las escalas de tiempo se definen una serie de *fronteras temporales* en cada una de las escalas, a partir de las cuales se establece la partición discreta de éstas. A nivel práctico, este paso consiste en transformar las variables continuas (por ejemplo, la edad) en funciones escalonadas, 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 el número de hijos: una persona con 35 años de edad cumplida, se considera con igual propensión de tener un hijo el mismo 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 cinco episodios vitales de este tipo de estudio:

- Uno ha experimentado un único evento maternidad/paternidad, representado con un punto rojo.
- Otro ha experimentado dos eventos de maternidad/paternidad.
- Otros dos, que finalizan con un punto negro, han salido de Andalucía antes de concluir el periodo de observación sin tener hijos.
- El restante, permanece en Andalucía hasta el final de periodo y sin tener hijos.

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 rojas marcan el comienzo y el fin del periodo de seguimiento en este ejemplo, mientras que la rejilla verde representa los *tiempos frontera* que se han usado en cada *escala de tiempo* para discretizarla. Las *líneas de vida, de color azul, intersecan con estas fronteras temporales* que marcan las categorías temporales discretas (grupos de edad y de periodo). En este sentido las "intersecciones" 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 de interés (nacimientos) 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, en el caso del estudio prospectivo, ha sido diseñada para permitir la investigación de las relaciones causales entre las característica personales y del entorno, recogidas en el Censo de 2001[16], y la frecuencia o cronología con la que, en años posteriores, ocurren nacimientos[17] en dicha población.

La relativa sencillez para estimar indicadores de fecundidad 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

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

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

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 compleja 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.

Por ejemplo, puede darse la circunstancia de que individuos de 20 años que no viven en pareja en el Censo de 2001 tengan un mayor riesgo de tener hijos en 2019 que los que sí vivían en pareja. Esto no quiere decir que los que viven en pareja tengan realmente menos hijos, sino que la variable utilizada va perdiendo su significado con el paso de los 18 años transcurridos desde su medición. Es decir, puede ser que 18 años después (a los 38 años de edad) sí vivan en pareja.

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 fecundidad derivados de esta fuente, pueden no coincidir con los indicadores anuales estimados a partir de las fuentes transversales habitualmente publicadas. Varias razones justifican estas diferencias. Por un lado, la población de referencia, 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 (salidas de Andalucía). La población residente en Andalucía está continuamente incorporando población por nacimientos o inmigraciones, sin embargo ambas entradas no son incluidas en el estudio prospectivo, para el que se tiene una población cerrada a entradas. La razón para no incorporar nuevas entradas en el estudio prospectivo es la de no disponer, para los nuevos individuos, del conjunto de variables de exposición recogidas durante la operación censal de 2001.

Los eventos recopilados para la cohorte censal (nacimientos y emigraciones exteriores), tampoco coinciden con los eventos producidos en el conjunto de la población andaluza en los años posteriores al Censo de 2001 dado que los nacimientos anuales y emigraciones de la población de Andalucía incluyen los eventos ocurridos en las nuevas entradas por migración y nacimiento después de 2001, 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. No obstante, se ha estimado que esta fuente de subregistro se sitúa por debajo del 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 fecundidad y el análisis de supervivencia. Aun así, la evolución y correlación de los indicadores derivados de ambas fuentes es notable.

En el caso del estudio retrospectivo, tomando las variables tiempo independientes correspondientes a la operación censal de 2011, sí se tiene en cuenta la población que entró en el proceso migratorio de primeros de siglo, siempre y cuando estas personas permanecieran en Andalucía en la fecha censal. A la hora de interpretar los resultados, estamos viendo en este caso las características casi al final del estudio de las personas que han tenido un determinado comportamiento en relación a la fecundidad. Por ejemplo, puede darse la circunstancia de que aparezcan individuos de 14 años en el año 2003 cuyo nivel de estudios alcanzado sea el de diplomado universitario. Evidentemente, es imposible con dicha edad. Este valor habría que interpretarlo como el nivel de estudios que tenía a finales de 2011 (con 22 años) una persona que tenía 13 años en 2003.

8.2 Anexo: Fuentes primarias

El estudio sobre fecundidad en Andalucía se apoya en las tres fuentes clásicas de información demográfica de España: los Censos, 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[18] (ENCSARP).
- La inscripción contendrá obligatoriamente: nombre, apellidos, sexo, domicilio, nacionalidad, documento nacional de identidad o similar, titulación académica y lugar y fecha de nacimiento.
- 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.

18 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 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.

Incidiendo en esta interrelación, se optó por hacer llegar a cada vivienda, junto con el cuestionario censal, los datos padronales[19] 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 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.

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

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.