

Estimación de los efectivos y estructuras poblacionales de Andalucía a partir de la integración de fuentes estadísticas y administrativas

Montañés Cobo, Víctor¹

victor.montanes@juntadeandalucia.es

Cánovas Balboa, M^a Rosa¹

rosa.canovas@juntadeandalucia.es

Poza Cruz, Eva¹

evav.poza.ext@juntadeandalucia.es

Viciano Fernández, Francisco¹

franciscoj.viciano@juntadeandalucia.es

1 Introducción

La información sobre los efectivos y estructura de una población con suficiente nivel de detalle en su distribución territorial es esencial para una adecuada gestión de los servicios públicos y privados. En la actualidad varias fuentes ofrecen esta información, tales como censos de población, estadísticas padronales o cifras de población y estimaciones intercensales de población, aunque con discrepancias nada despreciables en cifras, definiciones y nivel de desagregación espacial.

Las diferencias existentes en una fecha determinada entre distintas fuentes, que en Andalucía han llegado a superar las 150.000 en el año 2001, variando por nacionalidad o grupos de edad de forma más que apreciable. Más aun, no es infrecuente encontrar comportamientos anormales en la evolución de las cifras proporcionadas por alguna de estas fuentes, como ya ha sucedido en el padrón en relación a los cambios legislativos y metodológicos, tales como los que han afectado a las inscripciones y bajas de extranjeros.

Otras fuentes de cifras de población no tiene este problema, pero en contrapartida su nivel de desagregación es muy escasa al nivel municipal, y solo excepcionalmente existen desagregaciones intra-municipales.

Todo esto hace evidente la necesidad de disponer de un sistema capaz de suministrar series continuas de población, con sus correspondientes estructuras por sexo y edad y con desagregaciones territoriales inframunicipales. Además estas series poblacionales tienen que ser coherentes y similares a otras series de efectivos existentes y estar basadas en todo el conjunto de fuentes exhaustivas sobre stock y flujos poblacionales existentes.

La estimación clásica de las series de población utilizan, como fuente básica, dos tipos de agregados numéricos: por un lado recuentos poblacionales transversales, obtenidos a partir de censos periódicos o registros de población, y por otros estimaciones de los flujos o variaciones, naturales y migratorios, producidos entre los recuentos. Las cifras de estos flujos se obtienen generalmente a partir de explotaciones de sistemas de registro, tales como las clásicas estadísticas del MNP basadas en el registro civil y las variaciones residenciales (basadas en las altas y bajas padronales). Una compleja mezcla de cuestiones de cobertura, definiciones y errores registrales hace que los agregados numéricos provenientes de los recuentos y los de los flujos o variaciones poblacionales raramente sean perfectamente coherentes, por lo que es necesario realizar más o menos complejos procesos de ajuste de las cifras originales, con objeto de obtener series poblacionales coherentes y óptimamente compatible con las distintas fuentes que se hayan manejado.

Evidentemente todo este proceso de estimación de una serie poblacional se puede hacer con distintos niveles de detalle y profundidad, en base a la calidad de la información que se disponga sobre las estructuras socio-demográfica y territoriales en cada uno de los sistemas de información que se emplean. Desafortunadamente, al ser realizadas estas estimaciones clásicas mediante un procesado en serie independiente, el número, detalle y precisión de las covariables que permitirán la sub-estratificación, social y territorial de la población estimada, será el correspondiente al menos preciso y detallado de los subsistemas empleados. Por ejemplo, aunque en los recuentos censales, dispongamos del nivel educativo de la población, no será posible realizar series de población por nivel educativo, si los sistemas de información sobre defunciones o migraciones no disponen de dicha información.

Los dos problemas anteriormente enumerados, la coherencia y el nivel de detalle obtenible, provienen de una misma cuestión: los agregados numéricos en los que se basa la estimación de las series provienen de sistemas de información independientes y no integrados a nivel de microdatos. Esta es la principal causa responsable de los desajustes entre la cifras de recuentos y flujos y la consiguiente necesidad de aplicar, más o menos sofisticados, procedimientos de ajuste para que recuperen la igualdad básica de la ecuación compensadora: "la población en determinada fecha es igual a la población en una fecha anterior mas las flujos de entradas y salidas en ella producida entre las dos fechas".

La no integración de los sistemas de información de recuentos y flujos demográficos, también responsable de que se desaproveche un importante conjunto de información estadísticamente relevante, que podría ser explotada si los sistemas de información se encontraran unificados. Por ejemplo, covariables que están disponibles en subsistema de censos y padrones, no lo están en el de flujos o variaciones, igualmente hay variables muy detalladas en los subsistemas de recuentos pero que se encuentran pobremente recogidas en el de flujos, con la consecuente pérdida de precisión de estas variables. Este último es el caso de las variables territoriales, muy detalladas en censo y padrones, pero pobremente recogidas en los sistemas de registros de eventos. La integración de los sistemas a nivel de microdatos, permite que los subsistemas con menor contenido y detalle hereden la información de los sistemas más precisos y exhaustivos.

Este trabajo se propone demostrar que es posible obtener series de efectivos y estructuras poblacionales por sexo y edades simples con un muy alto nivel de detalle a nivel territorial, a partir de un sistema de información, como es la Base de

Datos Longitudinal de Población de Andalucía (BDLP) que integra a nivel de microdatos la información de los subsistemas de recuento (censos y padrones) con los de flujos (defunciones, nacimientos y variaciones residenciales).

El Instituto de Estadística y Cartografía de Andalucía está desarrollando una nueva fuente estadística: la "Base de Datos Longitudinal de Población de Andalucía" (BDLPA) a través de la integración de múltiples fuentes estadísticas y administrativas, tales como los censos de 2001 y 2011, la información de las variaciones padronales y los registros de nacimientos y defunciones. Con esta fuente se pretenden obtener unas poblaciones libres de influencias graves de cambios legislativos y con nivel de desagregación territorial suficientemente detallado.

2 Objetivos

1. El objetivo fundamental de este estudio es la obtención de unas cifras de población que no pretendan exclusivamente la obtención de cifras globales como en el caso de las estimaciones, sino que también pueda determinarse individuo por individuo quiénes son los que forman parte de ellas.
2. La geo-referenciación de las viviendas ocupadas por los individuos. Para lo cual es preciso la asignación de coordenadas geográficas a todas las viviendas que han sido ocupadas en Andalucía desde 1998 hasta 2013, fecha de fin de este ensayo. La disposición de las coordenadas permite comparaciones territoriales en cualquier momento dado, libres de la influencia de los cambios de seccionado o las segregaciones y fusiones de municipios.
3. Por último, se plantea la obtención de una serie estadística de población sobre una rejilla regular definida sobre el mapa de Andalucía.

3 Material y métodos

3.1 Fuentes utilizadas

La fuente utilizada para el desarrollo de este trabajo ha sido la Base de Datos Longitudinal de Población de Andalucía que está desarrollando el Instituto de Estadística y Cartografía de Andalucía. Esta base nutre su núcleo poblacional central con la información de los ficheros padronales de intercambio con los que el INE se comunica con los ayuntamientos y que conforman el "Registro de Población de Andalucía". Esta información es remitida al IECA semestralmente de forma que cada registro posee una causa de alta, baja o modificación y una fecha en la que se produce. El inicio de esta información se sitúa en el 1 de Mayo de 1996.

En el IECA se procede a la integración de estos ficheros con la información ya existente tras una etapa de identificación de los individuos realizado a través de un proceso de enlace de registros específico para la información padronal.

Esta información, que actúa como eje central y que posibilita un seguimiento en el tiempo de cada uno de los habitantes de Andalucía, se completa e integra con la información de una serie de registros administrativos y estadísticos.

Si bien hasta la fecha se ha trabajado con más fuentes, las utilizadas para este trabajo han sido:

- Registros de partos y matrimonios del Movimiento Natural de Población (desde 1998).
- Registro de defunciones del Movimiento Natural de Población (desde 1996).
- Censo de Población y Viviendas de 2001.
- Fichero precensal del Censo de Población de 2011.
- Afiliados a la Seguridad Social (disponible desde el tercer trimestre de 2011).

Estas fuentes se emplearán como prueba de verificación de ausencia o presencia de los individuos sobre los que se alberguen dudas sobre su continuidad en la comunidad autónoma.

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 generará 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 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.

3.2 Construcción del fichero inicial de episodios

Como ya hemos indicado, el eje central de la BDLPA es la información padronal recogida desde mayo de 1996. El objetivo inicial del Padrón Municipal de Habitantes, al margen de su utilidad administrativa, es obtener una cifra de población a 1 de Enero de cada año. Sin embargo, durante el proceso administrativo existe un conjunto de altas y bajas que no tienen una justificación puramente demográfica, sino de gestión, y que pueden causar una serie de incongruencias. De hecho, el Consejo de Empadronamiento es el que se encarga de

decidir la actuación a llevar a cabo en muchos de estos casos. Por lo tanto, hasta que esta decisión llega, los datos relativos a un grupo de personas tienen algún tipo de problema de compatibilidad.

Específicamente, existen una serie de causas directamente vinculadas con estos procesos:

- Altas por omisión
- Bajas por duplicado
- Bajas por inclusión indebida
- Bajas por caducidad. Se realizan desde el año 2006.

Estas causas se emplean en el proceso administrativo y resultan útiles para la gestión, pero causan que la serie de movimientos de una persona a lo largo del tiempo pueda presentar falsas discontinuidades o solapamientos.

Por tanto, se hace necesario dotar de consistencia los datos estableciendo una serie de normas, que afectan fundamentalmente a fechas y causas de alta y baja. En ocasiones, estas inconsistencias no afectan al fondo de la modificación de datos del individuo, sino más bien a la forma. Por ejemplo, si una persona se da un día de baja por cambio de residencia y ese mismo día se da de alta por omisión en otro, la causa real de la alta será un cambio de residencia. Esto no afectaría a las cifras.

Sin embargo, si una baja se produce un día concreto y la alta el día siguiente, se sobreentiende que la persona no reside durante un día en un lugar desconocido y se procede a unificar las fechas. Este proceso suele ser bastante habitual entre los extranjeros, ya que en ocasiones una alta por omisión sucede a una baja por caducidad con diferencias de pocos días. La propia Estadística de Variaciones Residenciales (EVR) descarta como cambios residenciales este tipo de movimientos cuando se producen dentro del mismo municipio en función del tiempo transcurrido.

Tras estos procedimientos, se obtiene un conjunto de episodios residenciales personales con coherencia temporal y espacial. Sin embargo, se hace necesario aplicarle dos modificaciones adicionales.

1. Se procede a dar de baja a aquellos individuos cuya defunción se ha comunicado a través de los ficheros de MNP. Este proceso se realiza por mejorar la dinámica de funcionamiento, ya que mientras las defunciones de MNP llegan al IECA con periodicidad mensual, la información padronal lo hace de forma semestral, permitiéndose así una actualización más rápida.
2. Como ya era conocida la diferencia entre el número de nacidos obtenidos por MNP y el número de menores con 0 años obtenidos de la fuente padronal, también se ha utilizado la información de partos de MNP. Así pues, se opta por incluir de oficio a los nacidos a través de MNP, siempre y cuando

la madre o el padre residieran en Andalucía en el momento del parto ó se empadronaran más tarde (menos de un año) en él mismo municipio declarado en el boletín de parto. Muchos de estos partos se incorporaban de forma tardía al sistema como altas por omisión (*Gráfico 1*). Gráficamente, pueden verse los retrasos en la incorporación de la información.

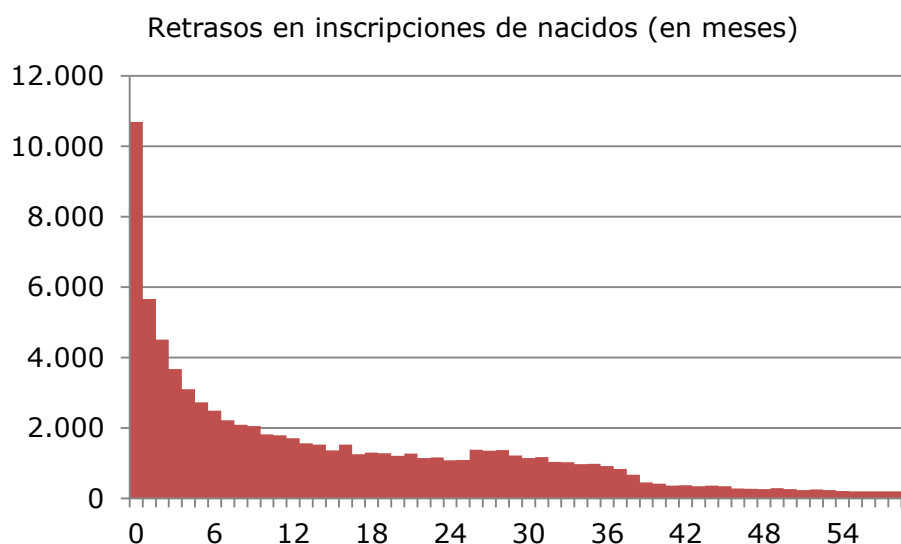


Gráfico 1. Meses de retraso en la inscripción de nacimientos

3.3 El fichero precensal de 2011 como población de referencia

La información original provoca que el total poblacional se halle muy alejado del derivado del Censo de 2011, principalmente en lo relativo a la población extranjera. La propuesta de este ensayo pasa por revisar la inscripción de cada individuo frente a su situación en el fichero precensal realizado por el INE como trabajo previo a la realización de la operación censal.

Los registros administrativos que el INE asoció a la información padronal son los ficheros de DNIs y Tarjetas de Residencia del Ministerio del Interior, los afiliados de MUFACE, afiliados a la Seguridad Social y pensionistas, Movimiento Natural de la Población y Agencia Tributaria. Adicionalmente se cruzó con información del Registro de Títulos del Ministerio de Educación.

El primer paso consiste en enlazar el fichero precensal con el fichero de episodios constituidos en la fase anterior. Esta etapa de enlace de registros fue llevada a cabo más que satisfactoriamente, con un enlace superior al 99,98%. Por tanto, el uso de este fichero parece estar suficientemente justificado. Este fichero precensal dispone de un campo denominado "factor de recuento" que para la mayoría de individuos presenta el valor 1. La distribución de este factor en la población es muy diferente cuando se analiza por nacionalidades (*Tabla 1*). Mientras para la población

española abunda el factor 1, para los extranjeros existe un gran porcentaje de valores menores.

Factor	Europa¹	África	América	Asia	Oceanía	España	Total
0	2,5%	6,9%	10,0%	5,7%	16,9%	0,4%	0,8%
(0-0,5)	20,6%	9,3%	9,9%	4,2%	17,3%	0,1%	1,5%
[0,5-1)	0,9%	4,3%	7,4%	7,2%	0,0%	0,3%	0,6%
1	76,0%	79,5%	72,6%	82,9%	65,8%	99,3%	97,1%

Tabla 1. Distribución del factor de recuento por continente

Se observa cómo la población española tiene un 99,3% de personas con factor 1, es decir, una práctica seguridad de que la persona debe pertenecer al colectivo bajo estudio, el de personas residentes a 1 de Noviembre de 2011.

Sin embargo, la población extranjera tiene un gran porcentaje de individuos con factor 0, o al menos menor que 0,5, que van desde el 9,9% de la población asiática al 34,2% de la de Oceanía o el 23,1% de la europea.

Se deduce que la población con factor 0 es aquella de la que se tiene una certeza absoluta de que no reside en la actualidad en ese municipio. Como indicio, para más del 20% de los españoles con factor 0, se localizó en la BDLPA una baja por defunción. En otros casos, el factor 0 puede producirse cuando se detecta un duplicado, asignando factor 0 a una de las inscripciones del individuo y factor 1 al otro.

Además de estos factores, se dispone de un conjunto de personas teóricamente activas en la fecha censal (en base a las altas y bajas padronales recibidas) que no aparecen en el fichero precensal. Más de 50.000 personas se encuentran en esta situación, lo que significa que no estaban en la población de referencia. Para ellos, se asumirá que su factor de recuento es 0, tratándose a partir de este momento como si tuviera dicho valor.

Este censo se toma como primer punto de control, reajustando la población individuo a individuo en este instante temporal después de una serie de comprobaciones adicionales. En primer lugar, se admiten todos los individuos que, teniendo un itinerario coherente, poseen factor 1. Este factor 1 ya vimos que era el mayoritario.

Así pues, todos los procedimientos que se realizan a continuación tratan de estudiar cómo actuar con el 2,9% de población restante con factor inferior y otro 0,6% que

¹ Excluyendo España

ni siquiera se hallaba presente en el fichero precensal. Pese a ser aparentemente unos porcentajes pequeños, suponen tomar una decisión sobre más de 250.000 personas. Por lo tanto, se precisa establecer una serie de criterios que llevarán a la aceptación o rechazo de estas personas.

Se aceptan para el cálculo de las poblaciones a todos aquellos individuos que, no poseyendo factor de recuento (no estaban en el fichero censal) o teniendo un factor inferior a 1 hubieran experimentado algún tipo de alta por cambio de residencia, omisión o nacimiento entre la fecha censal y el 1 de Enero de 2013, así como aquellos que hubieran comunicado un cambio de domicilio o una renovación, en el caso de ciudadanos no españoles.

Se hace una revisión masiva de todos los individuos no confirmados en fases anteriores en base al enlace de registros efectuado con bases de datos complementarias que posee el IECA. Con ello, se acepta como residente a todo aquel individuo que, independientemente de su factor de recuento, aparece como dado de alta en el fichero de afiliados a la Seguridad Social de Andalucía (trabajando) en el último semestre de 2011 o el primer trimestre de 2012, fechas de realización del fichero precensal y el trabajo de campo. Por el tipo de fichero y sus implicaciones económicas en el individuo, lo consideramos una fuente con suficiente fiabilidad.

Sobre el conjunto restante se establecen relaciones con MNP, tanto de matrimonios como de partos. Si en el último municipio de residencia en que consta un individuo en el fichero de BDLPA, teóricamente activo en la fecha censal, se tiene conocimiento de un parto o matrimonio entre 2011 y 2012, se asume que si permanecía en nuestros ficheros y se casa o tiene un hijo en fechas cercanas posteriores, es porque no ha emigrado, ignorando el posible bajo valor de recuento.

Al margen de éstos, se utilizan otras reglas de inclusión o exclusión como las siguientes:

Exclusión

- Se localiza la defunción en MNP.
- No aparecen en el fichero poblacional de 1 de Enero de 2012 aunque no se haya confirmado la baja a través de las variaciones.
- Individuos con factor de recuento inferior a 0,5.
- Mayores de 100 años no presentes en el fichero de pensionistas de 2011.

Inclusión

- Se acepta a los extranjeros que hubieran renovado su inscripción en los dos años previos al censo, es decir, su inscripción sigue legalmente vigente. En el caso más desfavorable, se ampliará su estancia unos meses.
- Se aceptan a todos los españoles restantes con un factor de recuento superior a 0,5.
- Para los menores de edad si se ha aceptado la presencia de una de las personas mayores de edad con las que convive.

Finalmente, se consigue aceptar la presencia de un 55% de los individuos de presencia no segura.

A partir de este momento, la pregunta que hay que responder es: ¿desde cuándo no residen las personas excluidas de la población en Andalucía?

Para ello, recurrimos a un segundo punto de control basado en la información del Censo de Población y Viviendas de año 2001 y en los itinerarios padronales contruidos desde 1996. El objetivo que se persigue con el contraste con el Censo de 2001 es conocer si 10 años atrás ya había dudas fundadas sobre la ausencia de los individuos. En base a esta información, separaremos a este subcolectivo que hay que dar de baja antes de 1 de Noviembre de 2011 en tres grupos.

- Poseen inscripciones posteriores a 1 de Noviembre de 2001. A este criterio se acogerán todas las altas intercensales y aquellos individuos que hayan modificado su domicilio.
- Se encontraban en el censo de 2001 y sus datos no fueron imputados, por lo que existe seguridad de que contestaron el cuestionario del cuadernillo censal como residentes en Andalucía.
- No aparecen en Censo de 2001 y no hay ninguna inscripción posterior.

Esta separación va a ser útil a la hora de determinar una fecha a esa baja no recibida.

3.4 Imputación de fechas con incertidumbre

Los procesos anteriores sirven para conocer una fecha en la cual tenemos cierta seguridad de que una persona ya no reside en Andalucía. Sin embargo, la creación de un fichero longitudinal de poblaciones hace necesario que se concrete la fecha, ya que de lo contrario se producirían cambios o bajas masivos en determinadas fechas, como en este caso las fechas censales.

Al margen de las fechas censales, existen otras fechas concretas relacionadas con procesos administrativos como las bajas por caducidad, para las que se conoce en

qué fecha se efectuó la baja administrativa pero se desconoce cuándo la persona abandonó la comunidad autónoma. Estas fechas, también están sujetas a revisión.

Por lo tanto, se necesita un proceso para estimar estas fechas de baja. De antemano, se toma el criterio de asignar la misma fecha de baja a aquellas personas que llegaron a la misma vivienda en la misma fecha y se tiene conocimiento de su "no presencia" también en la misma fecha.

Básicamente, la idea conductora será la de establecer para cada individuo tres tipos de tramos temporales (*Gráfico 2*) sobre su itinerario residencial.

- Tiempo de estancia segura (línea negra)
- Tiempo de ausencia segura (línea roja)
- Tiempo de incertidumbre (línea naranja)

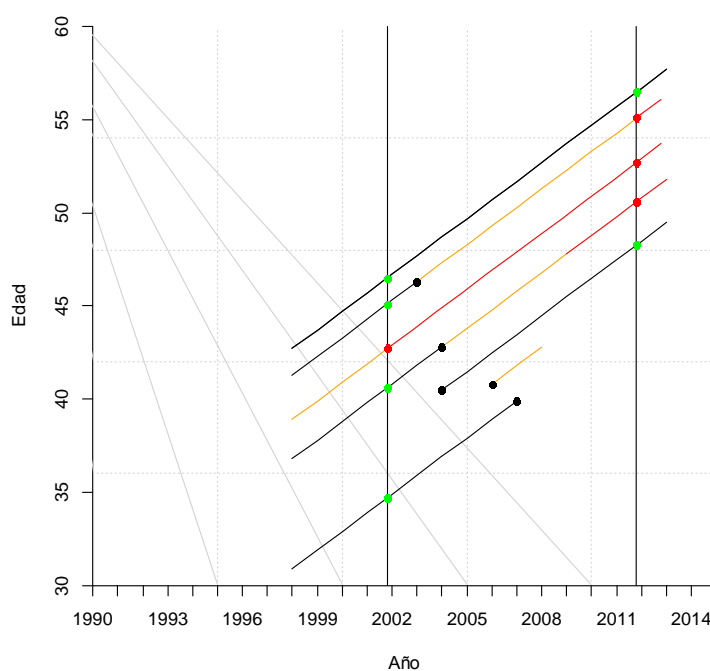


Gráfico 2. Tramos temporales de ausencia o presencia

Sobre este gráfico, realizado a modo de ejemplo, se muestran puntos de colores verdes (que indican confirmación censal), rojos (que señalan ausencia censal) y negros, que señalan comunicaciones de variaciones padronales o confirmaciones en otros registros. Estas marcas, junto con variables como nacionalidad, sirven para establecer el rango temporal de incertidumbre.

En este proceso de estimación de bajas se distingue entre ciudadanos españoles, comunitarios y extracomunitarios, debido a que reciben un trato diferente en la gestión padronal. La legislación distingue dos colectivos entre los extranjeros, los no comunitarios sin autorización de residencia permanente (ENCSARP) y el resto.

Como se desconoce qué extracomunitarios no tienen permiso de residencia permanente en Andalucía (ENCSARP), se empleará como aproximador de este colectivo los no comunitarios que no dispongan de número de identificación de extranjero a falta de mejor información.

La legislación padronal ha evolucionado en los últimos años para paliar los efectos de sobreestimación debidos a la no recogida de la salida de los extranjeros del país.

La Ley 7/1985, de 2 de abril, reguladora de las Bases del Régimen Local, en su artículo 16 señala:

“La inscripción en el Padrón Municipal sólo surtirá efecto de conformidad con lo dispuesto en el artículo 15 de esta ley por el tiempo que subsista el hecho que la motivó y, en todo caso, deberá ser objeto de renovación periódica cada dos años cuando se trate de la inscripción de extranjeros no comunitarios sin autorización de residencia permanente.

El transcurso del plazo señalado en el párrafo anterior será causa para acordar la caducidad de las inscripciones que deban ser objeto de renovación periódica, siempre que el interesado no hubiese procedido a tal renovación. En este caso, la caducidad podrá declararse sin necesidad de audiencia previa del interesado.”

Esta normativa comenzó a aplicarse a finales de 2006, pero no afecta a ciudadanos de la zona del espacio económico europeo ni a ciudadanos con permiso de residencia permanente. Para este colectivo, un acuerdo del Consejo de Empadronamiento del INE de fecha 27 de junio de 2008, indicaba la obligación de los Ayuntamientos de iniciar expediente de baja de oficio de los extranjeros comunitarios en función de la última información disponible en la base del INE, de modo que únicamente se realizarían cuando no se hubieran producido movimientos en un período de cinco años (o dos años, si no se encontraban inscritos en el Registro Central de Extranjeros). En el mismo acuerdo se incluían los extranjeros asimilados (Schengen) y extranjeros con Autorización de residencia permanente, con certificado de inscripción en el Registro Central de Extranjeros expedido hace más de cinco años y cuyo último movimiento en el Padrón tenga más de cinco años de antigüedad.

Nuevamente, como no se dispone de información del Registro Central de Extranjeros, se opta por aplicar esta categoría únicamente a los comunitarios o aquellos que dispongan de número de identificación de extranjero.

Esta legislación nos va a servir para establecer la fecha tope para ejecutar la baja, es decir, para establecer la zona de incertidumbre, que quedaría de la siguiente forma:

- No comunitarios con número de identificación de extranjero y comunitarios. Cinco años desde la última inscripción conocida.
- No comunitarios sin número de identificación de extranjero. Dos años y tres meses (estos tres meses se ha constatado que se dan en algunos ayuntamientos como periodo de gracia).
- Españoles. No existe un periodo máximo establecido.

Cuando hablemos de última inscripción lo haremos en sentido amplio. No nos referiremos exclusivamente a la última alta o modificación padronal, sino que se consultará en los ficheros auxiliares la presencia o ausencia de la persona. Estos ficheros auxiliares son fundamentalmente los matrimonios y partos del MNP, la presencia en los ficheros trimestrales de afiliados a la seguridad social y la presencia (efectiva, no imputada) en el Censo de 2001.

Inicialmente, se aplicarán estas fechas de baja sobre el conjunto de personas descartadas como presentes en la fecha censal de 2011 para las que hasta entonces no se había comunicado la baja asociada.

En primer lugar se establece una fecha máxima de permanencia como:

$$F_{\text{max}} = \text{mín}(\text{Fecha última observación/confirmación} + \text{Amplitud temporal legal}, \\ \text{Fecha detección/aplicación de ausencia})$$

Siendo la amplitud temporal legal la determinada por la legislación anteriormente indicada (infinito en el caso de españoles).

Para obtener la fecha estimada de baja se optará por un modelo del siguiente tipo:

$$\text{Fecha de baja} = \text{Fecha de última observación/confirmación} + \\ x * (F_{\text{max}} - \text{Fecha última observación/confirmación})$$

Donde x es un valor aleatorio de una función de probabilidad con valores entre 0 y 1. Esta función es una beta(4,2) en el caso de los extranjeros y una uniforme para los españoles. El uso de la beta se debe a que asigna una menor probabilidad de baja en los primeros meses posteriores al alta en el municipio, algo bastante lógico.

En el caso de los españoles el rango disponible era por lo general muy amplio, en muchos casos de 15 años, partiendo de la fecha de la última renovación padronal de 1 de Mayo de 1996 o de la fecha censal de 2001. Por lo tanto, aplicar este criterio de menor probabilidad al inicio no parece muy útil. Se pensó en una distribución normal, pero esto daría lugar teóricamente a una acumulación de bajas entre los años 2003 y 2007. Finalmente se optó por el método más sencillo, tomar valores de una distribución uniforme.

A partir de aquí se asigna una fecha a cada individuo, incidiendo una vez más en que individuos que entran en la misma vivienda observados por última vez en la misma fecha, se fuerzan a salir en la misma fecha.

Finalmente, se opta por aplicar esta metodología a todas las bajas de tipo administrativo (caducidades, inclusiones indebidas y duplicados).

3.5 Reconstrucciones parciales

Una vez realizadas las estimaciones poblaciones, día a día, se detectan una serie de problemas, algunos relacionados con fechas concretas y otros relacionados con municipios con alguna particularidad. Entre estos sucesos destacan:

- En algunos municipios de Granada se asignó como fecha de alta inicial el día 1 de Mayo de 1996 a todas las personas que consideraban residentes a 1 de Enero de 1998. Los nacidos durante ese lapso temporal tenían por tanto una fecha de alta previa a su fecha de nacimiento.
- Se detectaron dos municipios que, existiendo a 1 de Mayo de 1996, carecían de población hasta el año 2001. En este caso, se tomó información de diferentes fuentes para completar estos años faltantes. En primer lugar se tomaron los nacimientos en esos años y se les asignó como fecha de alta en el municipio su fecha de nacimiento. De otro lado, se tomó el año de llegada al municipio declarado en el Censo de 2001. Dentro de ese año, se asignó una fecha aleatoria por vivienda que fue otorgada a todos sus habitantes (salvo nacidos).
- Altas por omisión el 01/11/2001 que no tengan registros previos. Para evitar un incremento súbito de la población se empleará el año de llegada a la vivienda del censo de 2001, asignando una fecha aleatoria dentro de ese año a nivel de clave de vivienda. Con los cambios de domicilio se realiza un proceso similar.

A cada uno de los residentes se le añaden una serie de campos básicos y muy estables en el tiempo: fecha de nacimiento, nacionalidad, lugar de nacimiento y sexo.

Se determina que la fecha y lugar de nacimiento deben ser únicas, y se asume que el sexo también lo es (ya que el número de cambios de sexo es muy reducido y pudiera introducir más errores). Por último, se acepta que la nacionalidad puede cambiar, pero una vez que sea española, no puede cambiar a otra. El añadido de estos campos permite la tabulación por una serie de variables demográficas básicas.

3.6 Comparación de cifras

En este apartado se comparan las cifras a 1 de Enero obtenidas con el fichero de itinerarios en bruto y las obtenidas tras los procesos de contraste y estimación de bajas. De esta forma podrán observarse las transformaciones reales que han producido en las poblaciones nuestras decisiones sobre inclusión/exclusión de individuos en cifras y la fecha en que se ha procedido a dar la baja.

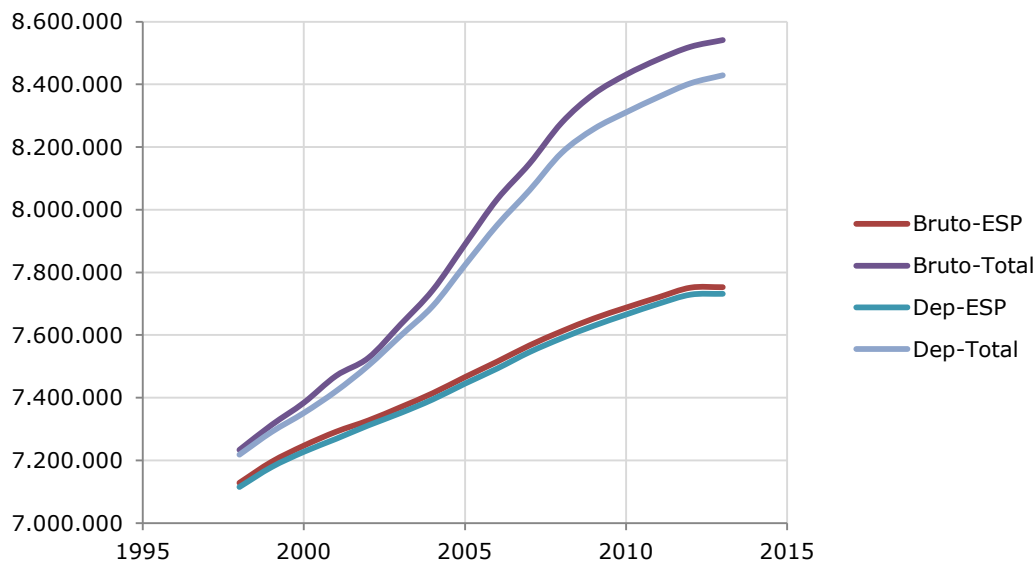


Gráfico 3. Comparación poblacional anual (española y total) entre el cálculo bruto y depurado

La cifra global comienza a separarse hacia el año 2003, años en que comenzó el proceso de entrada de inmigrantes que caracterizó la primera década de este siglo. Sin embargo, la cifra de españoles se muestra muy pareja desde el inicio, con diferencias que oscilan entre un 0,2% y un 0,4%. Por lo tanto, las verdaderas diferencias se centran en el colectivo de extranjeros, que veremos representado por separado (*Gráfico 4*).

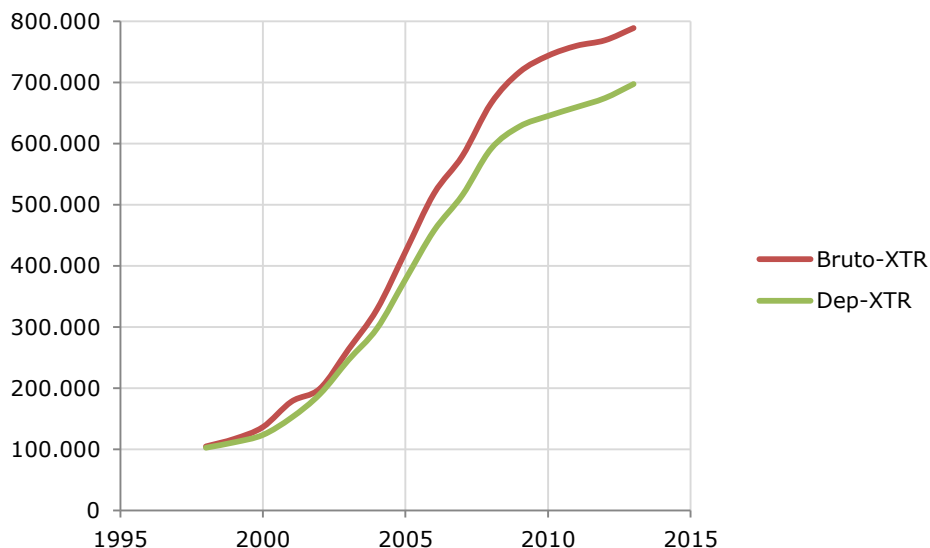


Gráfico 4. Comparación de población extranjera (bruta y depurada)

Las diferencias evolucionan desde un mínimo del 2% en 1998 hasta un 15% de máximo, muy lejos de la variación que se encontraba en la población española.

Si estudiamos las diferencias entre ambas fuentes (población bruta – población depurada) por provincias y año se obtiene una distribución claramente reveladora.

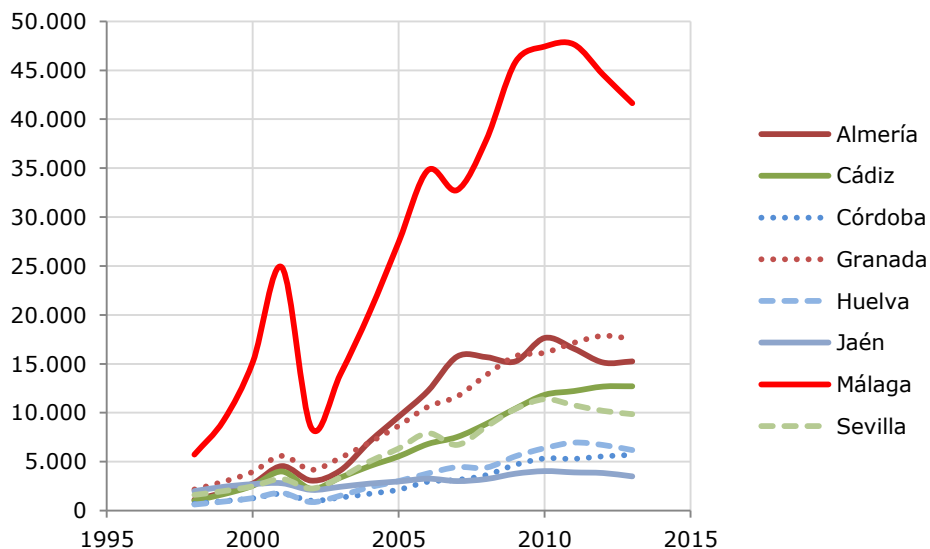


Gráfico 5. Diferencias poblacionales por provincias (bruta-depurada)

Málaga, provincia con un alto porcentaje de población extranjera (fundamentalmente comunitaria), destaca sobremanera respecto al resto de provincias (Gráfico 5). Si estudiamos estas diferencias en términos porcentuales respecto a la población, se siguen manteniendo estos órdenes. En cualquiera de los

años, la provincia de Málaga representa más de un tercio de las diferencias poblacionales.

3.7 Geocodificación de portales ocupados

Paralelamente al cálculo poblacional, el IECA ha realizado un gran esfuerzo en la geolocalización de la población residente. Utilizando como fuentes de coordenadas el Callejero Digital de Andalucía Unificado (CDAU) y las del fichero de edificios del Censo de edificios de 2011, y empleando como apoyo la información del callejero del Censo Electoral y las modificaciones de datos territoriales derivadas de las variaciones padronales, se ha llevado a cabo una georreferenciación de todas las viviendas ocupadas en Andalucía desde el 1 de Enero de 1998 hasta 1 de Enero de 2013.

Se parte de una tabla de edificios en la que se tienen todos los descriptores postales recogidos en la BDLPA desde 1996. A partir de los campos de provincia y municipio, código de unidad poblacional, código de vía, tipo y nombre de vía y número que caracterizan cada portal, se genera una clave única que identificará los edificios.

Puesto que es habitual que un mismo nombre de vía puede aparecer escrito de diferentes formas, lo más operativo es comparar códigos, por lo que el primer paso es identificar aquellas vías donde coincidan los nombres tanto en BDLPA como en CDAU para un mismo código de vía. Este proceso hace innecesario tener en cuenta el tipo de vía, ya que el código de vía tiene en cuenta implícitamente el tipo.

En caso de no coincidir el nombre de vía, se aplica una función de limpieza de partículas y se comprueba si tras ella hay coincidencia exacta. Para los restantes, se admiten como misma vía aquellos casos en que uno de los nombres está incluido en el otro o tienen una similitud (usando la distancia Jaro-Winkler) mayor del 80%. Adicionalmente, se comprueba con el Callejero del Censo Electoral que semestralmente ofrece el INE en su web y que está disponible desde el año 2004, si la no coincidencia del nombre se debe a un cambio de denominación de la misma.

A partir de este paso, se van analizando secuencialmente los números de portal, distinguiendo si hay coincidencia exacta, si se trata de un portal cercano o un portal de la misma calle más alejado.

El proceso de enlace resultó exitoso, permitiendo asignar coordenadas a cerca de un 97% de la población. Del 3% restante, una pequeña parte es debida a restricciones propias del algoritmo de búsqueda y la mayoría se debe a deficiencias

en la cobertura de las coordenadas en el fichero censal y el CDAU, sobre todo en zonas de nueva creación (aún no incorporadas) y diseminados.

Esta asignación de coordenadas permite obtener resultados a niveles territoriales muy detallados e independientes de las zonificaciones administrativas como las secciones censales que no son estables en el tiempo.

4 Principales resultados

Como contraste para los resultados se procede a comparar los resultados con los obtenidos a través de otras fuentes como padrón o estimaciones intercensales. Las estimaciones intercensales presentan el problema de que su máximo nivel de desagregación es provincial, por lo que para ciertos estudios puede resultar insuficiente. Por otro lado, no deja de ser una población ficticia. La población padronal, por su parte, ha estado expuesta a varios cambios legislativos que han propiciado distorsiones en años concretos.

El comportamiento de la serie obtenida de la BDLPA (*Gráfico 6*) presenta en los primeros años unas cifras muy cercanas a las padronales, pero hacia el año 2008 comienza a converger hacia la cifra obtenida de las estimaciones intercensales. Este cambio en los últimos años se explica fundamentalmente con las bajas aplicadas a ciudadanos comunitarios.

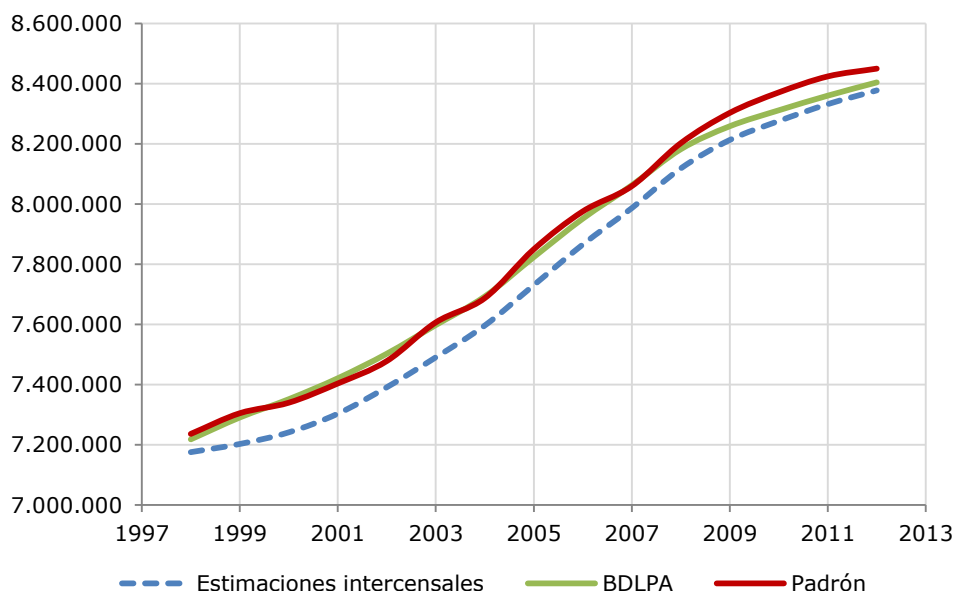


Gráfico 6. Comparación del total poblacional (intercensal, padrón y BDLPA)

El siguiente paso para estudiar las cifras es observar si causa algún tipo de efecto sobre la estructura poblacional. Para ello, se comparan las pirámides de población. La comparación se ha realizado en las fechas censales de 2001 y 2011.

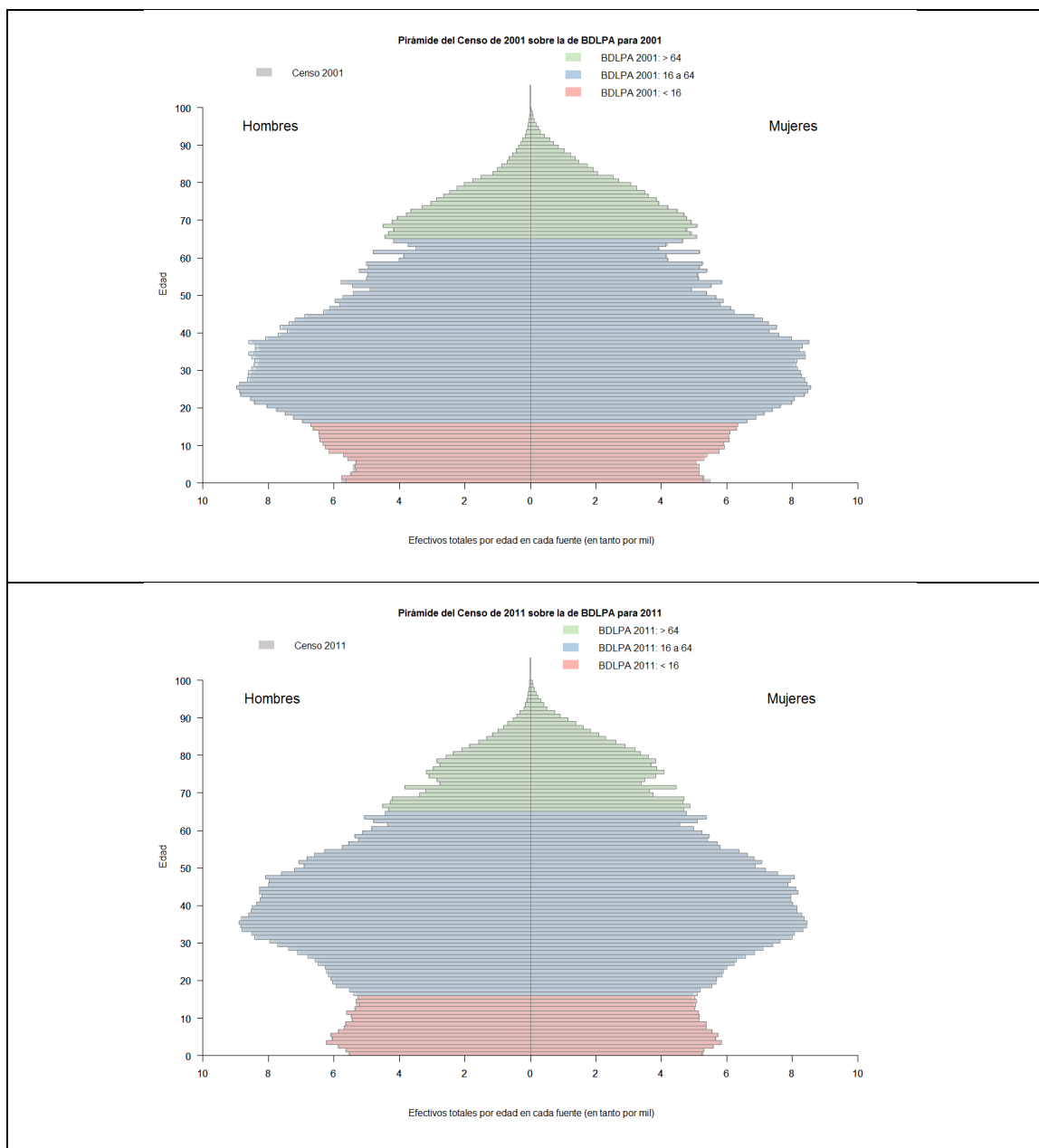


Gráfico 7. Pirámides comparativas. Andalucía. Censo vs BDLPA 2001 y 2011

En el año 2011, la estructura de las dos pirámides es prácticamente coincidente. Teniendo en cuenta que esa era la fecha tomada como punto de apoyo, es algo que no hace más que verificar que el reajuste no ha tenido ningún tipo de sesgo por edad. En el año 2001 sí que se notan una serie de diferencias y aparece más población en BDLPA en hombres de entre 25 y 45 años y menos jóvenes entre 0 y 5 años. Para mujeres únicamente haber una diferencia de infrarregistro en BDLPA entre los 65 y 70 años, pero muy leve.

Las cifras globales pueden ocultar algunos efectos sobre la población. De hecho, analizando los crecimientos poblacionales entre años consecutivos (*Gráfico 8*) se observa que las cifras resultantes de la BDLPA son muy parejas a las obtenidas de las estimaciones intercensales 2001-2011. Si bien el volumen no es el mismo, la

tendencia sí lo es. Se hallan por debajo pero discurren prácticamente en paralelo. Por el contrario, los crecimientos derivados de las estadísticas puramente padronales presentan comportamientos un tanto erráticos y de difícil justificación demográfica (que no administrativa). Por lo tanto, una primera ventaja que se obtiene es la estabilidad de la serie y la corrección (o al menos rectificación parcial) de los efectos administrativos.

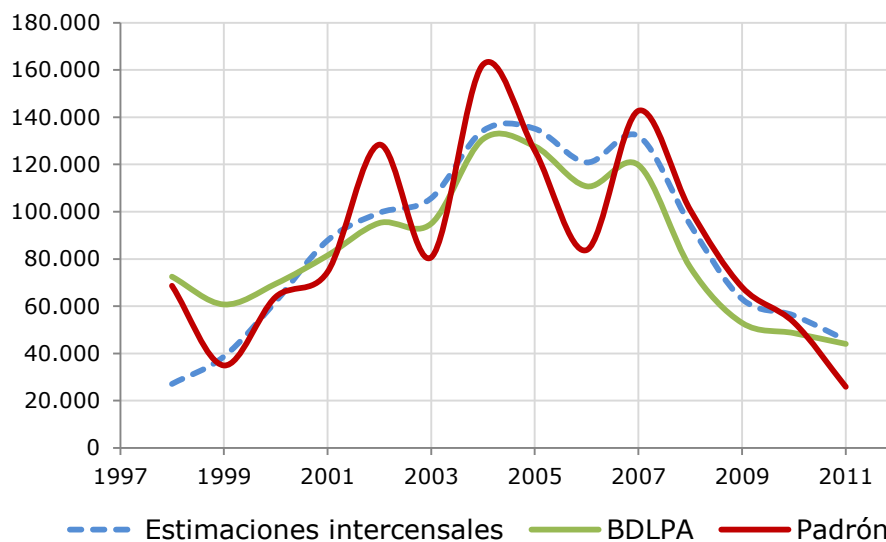


Gráfico 8. Crecimientos poblacionales anuales (intercensales, padrón, BDLPA)

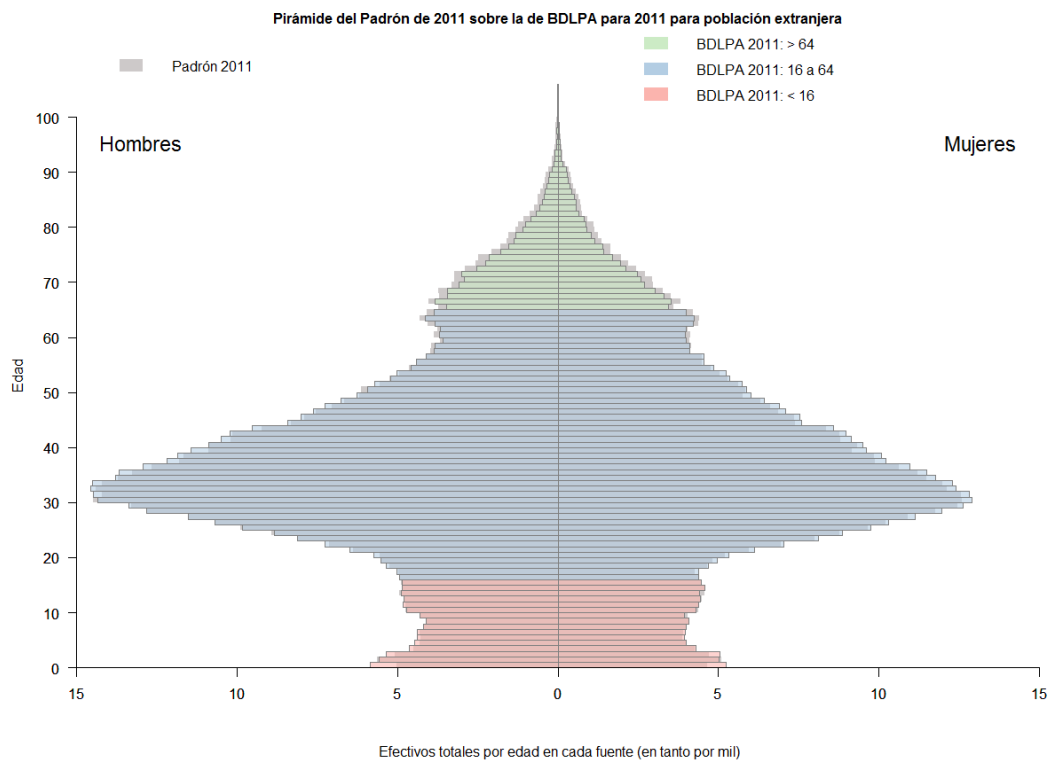


Gráfico 9. Pirámide de población comparativa. Año 2011. Extranjeros Padrón vs BDLPA

La comparativa de la pirámide de población extranjera para el año 2011 (*Gráfico 9*) da como resultado una estructura es bastante similar, como cabía esperar, hay que resaltar una serie de aspectos. La BDLPA muestra una base más ancha, es decir, un mayor número de menores de 5 años. En el lado opuesto, se halla un recorte en las edades mayores a 65 años (zona verde) en las que puede observarse que los efectivos derivados de BDLPA son sensiblemente menores a los de padrón. De hecho, la mayor parte de la población europea comunitaria (no española) residente en Andalucía, pertenece a esos tramos de edades.

Un aspecto reseñable de esta nueva forma de plantear el estudio de las poblaciones es la posibilidad de llegar a un nivel de desagregación geográfico del que no se dispone hasta el momento. Únicamente el Censo de 2011 lo permite, pero es una operación muestral en lugar de poblacional.

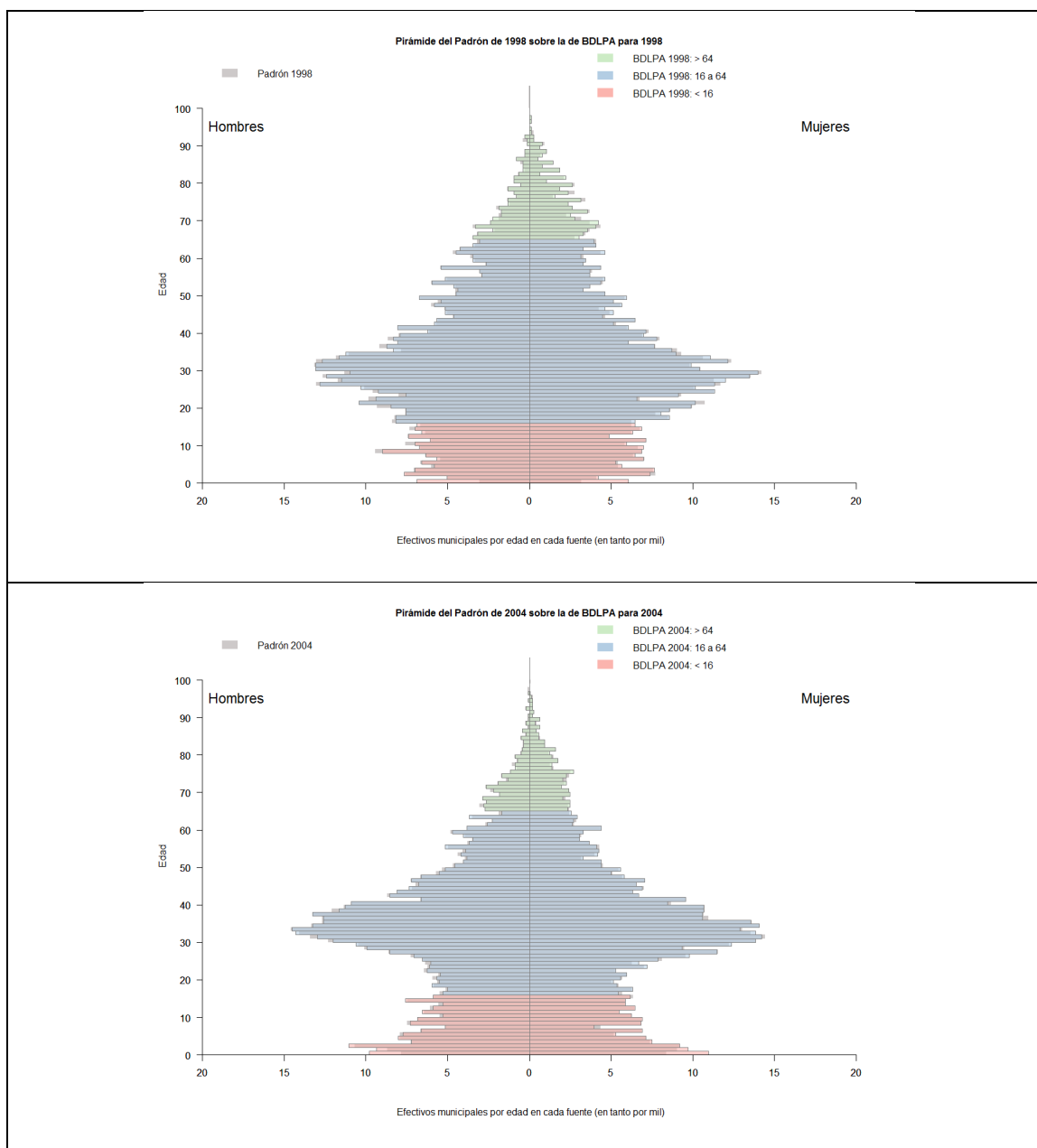


Gráfico 11. Pirámides de población inicial y final para el municipio del Gráfico 10

Este mismo proceso podría realizarse por nacionalidades o acompañarse de informaciones procedentes de otras fuentes, permitiendo por ejemplo conocer la distribución de afiliados a la seguridad social por zonas y su evolución.

Otra ventaja que posee la fuente, debido a la forma en que se ha construido, es que se pueden obtener estadísticas de nacimientos, migraciones, cambios de domicilio intramunicipales y defunciones al mismo nivel de detalle a nivel transversal (como hemos realizado con las pirámides) o longitudinal (se tienen 15 años de seguimiento continuo).

5 Conclusiones

La metodología empleada en la estimación de estas poblaciones proporciona una serie de ventajas de que no disponen las estimaciones clásicas. Al realizarse individualmente, permite utilizar la información personal para obtener un mayor nivel de detalle en cuanto a la clasificación por nacionalidad, lugar de nacimiento o residencia que ofrecen otras estimaciones. Adicionalmente, si se dispone de información complementaria de otras fuentes como el número de hijos o el nivel de estudios, se podría clasificar a la población según estos criterios sin emplear criterios de reparto.

La separación individual y la disponibilidad de coordenadas permiten una desagregación territorial de los resultados con tanto detalle como permita la salvaguarda del secreto estadístico.

La incorporación implícita de nacimientos, defunciones y cambios de domicilio que lleva asociada la creación del fichero longitudinal, permiten obtener un gran grado de detalle en la distribución espacial de estos eventos y en su incidencia sobre la población.