

Resultados del enlace de los registros del MNP con el RPA, 12 años de seguimiento: 1998-2009

Cánovas Balboa, M^a Rosa

rosa.canovas@juntadeandalucia.es

Montañés Cobo, Víctor

victor.montanes@juntadeandalucia.es

Poza Cruz, Eva V.

evav.poza.ext@juntadeandalucia.es

Viciano Fernández, Francisco J.

franciscoj.viciano@juntadeandalucia.es

Instituto de Estadística de Andalucía

Resumen

La identificación y asignación de una clave única de persona a los apuntes registrales que nutren la Base de Datos Longitudinal de Población de Andalucía (BDLPA) es una actividad crítica en su construcción. Se han desarrollado diversas estrategias de enlace de registros adaptados a las específicas características de las fuentes nominales a cruzar. En este trabajo se evalúa la eficacia de estos procedimientos mediante un estudio de la concordancia entre las identificaciones personales en los boletines de partos y defunciones del MNP y las altas por nacimiento y bajas por defunción de los padrones. Padres y nacidos también son localizados en otros registros activos de la BDLPA.

En el periodo 1998-2009, cerca del 94% del total de defunciones y nacimientos recibidos en los Boletines de MNP de Andalucía han sido localizados en las variaciones padronales. La concordancia sube al 95% si nos limitamos a eventos de residentes. El 97% de los padres recogidos en los boletines de partos han sido localizados en algún registro activo de la BDLPA.

La identificación de las personas en MNP y Padrones permite evaluar la cobertura y calidad de ambas fuentes y, secundariamente, implementar sencillos procedimientos de recuperación de información y mejora de su calidad.

Palabras clave: enlace de registros, calidad de fuente, cobertura, itinerarios vitales, longitudinal

1. Introducción y justificación

a. *El Registro de Población de Andalucía y sus derivados*

En los últimos años el sistema de gestión de los padrones municipales de población y la legislación relacionada¹ han experimentado un cambio sustancial. El nuevo sistema de gestión de los padrones municipales, puesto en marcha en mayo de 1996 presupone una coordinación e integración de la información de los más de ocho mil padrones municipales (770 en el caso andaluz) y entre éstos y la información de otros registros administrativos², siendo primordial la incorporación de nacimientos y defunciones del Movimiento Natural de Población (MNP).

El INE es pieza fundamental en este proceso que precisa de un complejo sistema de enlace de varios de los registros administrativos de los que se derivan gran parte de las estadísticas demográficas clásicas. En 2003, merced a un acuerdo de colaboración con el INE, el Instituto de Estadística de Andalucía (IEA) se adentró en esta actividad procesando las extracciones padronales transversales que el INE enviaba anualmente. A raíz de ello, el Registro de Población se incorpora en el Plan Estadístico Andaluz 2003-2006. En esta fase, el Registro de Población de Andalucía era puramente administrativo y únicamente se procedía a la eliminación de duplicados y defunciones cuando la información personal coincidía completamente. Con el paso del tiempo se vio que el mantenimiento del sistema era poco eficiente ya que cada año, con la recepción del nuevo fichero padronal, se debía proceder a la misma depuración y actualización partiendo prácticamente de nuevo desde cero.

En el año 2005 se acordó con el INE incluir en el envío, además de los volcados de los ficheros padronales, las variaciones de los padrones municipales comunicadas mensualmente por los ayuntamientos en un fichero único acumulado semestralmente. Estos ficheros de flujos semestrales incluyen los nacimientos, defunciones, cambios de residencia e incidencias de gestión que se han incorporado a los diferentes padrones municipales andaluces en forma de altas, bajas o modificaciones en dicho periodo.

A partir de esta nueva fuente se modificó el planteamiento inicial, ya que la recepción de esta información hizo factible el manteniendo de un histórico con las variaciones de cada persona a partir del cual podría plantearse el desarrollo de un sistema longitudinal. Tras detectar una serie de lagunas en la unión de los flujos con los ficheros transversales debidas al desconocimiento que teníamos de la información previa a 2005³, se solicitó al INE el envío del histórico de variaciones padronales desde su inicio, el 1 de Mayo de 1996. Ante la ausencia de una clave de identificación única para cada persona, el IEA procedió a la creación de una clave propia.

Desde esa fecha, el Registro de Población deja de ser un producto exclusivamente administrativo y transversal para transformarse además en estadístico y longitudinal – Registro Estadístico de Población de Andalucía (REPA) -, siendo uno de los objetivos el de “integración de las estadísticas demográficas”. De hecho, en el último programa anual aparecen 3 actividades relacionadas con él:

- 01.01.001 Registro de Población de Andalucía
- 01.01.002 Estadísticas transversales del RPA

¹ Este cambio de sistema comenzó con las modificaciones: 4/1996 y 14/2003 de Ley Orgánica de Régimen Local, la Ley Orgánica de Protección de Datos de Carácter Personal (15/1999). El nuevo Reglamento General de Población (decreto 2612/1996) y diversas resoluciones e instrucciones técnicas sobre plazos, reglas para la obtención de las cifras de población oficial, como la Resolución de 28 de abril de 2005, sobre los procedimientos para acordar la caducidad de las inscripciones padronales de los extranjeros no comunitarios, que ha causado un gran impacto en las cifras de población.

² Artículo 63 del Reglamento General de Población modificado por el decreto 2612/1996: 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 mantener debidamente comprobados y actualizados dichos datos. En particular, esta remisión de datos deberá ser efectuada por las Oficinas del Registro Civil en cuanto a *nacimientos, defunciones, y cambios de nombre, de apellidos, de sexo y de nacionalidad*, con las limitaciones que imponga su legislación específica; por el Ministerio del Interior en cuanto a *expediciones de documentos nacionales de identidad y de tarjetas de residencia* y por el Ministerio de Educación y Cultura en cuanto a *titulaciones escolares y académicas* que expida o reconozca. El artículo siguiente, el 64, especifica lo que de hecho ocurre: que “Las comunicaciones de datos mencionadas en el artículo anterior podrán ser canalizadas a través del Instituto Nacional de Estadística”.

³ Actualización del Registro Estadístico de Población de Andalucía (REPA) a partir de los ficheros de variaciones del Padrón Continuo. Primer año de experiencia. Viciana, Montañés, Alba y Martínez. XV Jecas. Palma de Mallorca, 2006.

- 01.01.003 Estadística de itinerarios vitales

La integración en el fichero estadístico de los datos procedentes de otras fuentes como el censo de 2001 y la información procedente de los boletines estadísticos de parto, defunción, y en un futuro matrimonios, lleva a la constitución de la Base de Datos Longitudinal de Población de Andalucía (BDLPA).

b. El Movimiento Natural de Población (MNP)

Cuando en España se produce un parto, un matrimonio o una defunción se deben cumplimentar por ley una serie de documentos estadísticos, entre otros requisitos, en las Oficinas municipales del Registro Civil, donde se procede a la inscripción. Estos Boletines Estadísticos inscritos durante un mes en la Oficina municipal del Registro Civil, son enviados a las Delegaciones Provinciales del Instituto Nacional de Estadística (INE), donde se codifica y graba la información demográfica contenida en los respectivos boletines.

Desde que en 1996 entró en vigor el actual convenio de colaboración entre el Instituto Nacional de Estadística (INE) y el Instituto de Estadística de Andalucía (IEA) 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. Este convenio amplía las tareas y funciones en el tema de la elaboración de dichas estadísticas que comenzaron con el acuerdo de colaboración entre ambas instituciones para la elaboración de las Estadísticas de Defunciones según causa de muerte en 1992.

c. Circuitos de la información

El sistema existente en España consiste a grandes rasgos en un proceso de intercambio de información a tres bandas entre el INE, los ayuntamientos y los Registros Civiles a través de unos ficheros mensuales de comunicación, correspondiendo al INE la función de coordinación y la detección de incongruencias.

Los movimientos podrían separarse en 3 grandes bloques:

- Movimientos solicitados por el vecino: cambios de domicilio, residencia o de datos personales, nuevas altas
- Movimientos administrativos: bajas por caducidad, duplicado, cambios de seccionado...
- Movimientos generados a través de la información del Registro Civil: nacimientos y defunciones

En el primero de los casos el vecino comunica la información al ayuntamiento y éste la remite al INE por si la información enviada pudiera tener otras consecuencias (como por ejemplo la baja en el municipio de procedencia).

En el segundo caso habría que hacer una distinción entre los movimientos administrativos como el cambio de seccionado o de denominación del domicilio, realizados directamente por el ayuntamiento y los que realiza a "petición" del INE, como por ejemplo las bajas por duplicado detectadas por el INE en su proceso de coordinación.

El tercer grupo es el que más nos interesa en este estudio. En principio el mecanismo es el que sigue. El Registro Civil envía la información de nacimientos y defunciones al INE y este procede a un enlace masivo con su base de datos padronal. Aquellas defunciones localizadas se remiten al correspondiente ayuntamiento para que proceda a aplicar la baja. En los partos es imposible la localización ya que el recién nacido no podía encontrarse empadronado (lógicamente), con lo que se busca a la madre o el padre. De encontrarse, se indica al ayuntamiento de residencia de los padres que proceda a realizar la inscripción del nacimiento en el domicilio de la madre (o del padre, si no se localiza a la madre).

Una vez recibidas, los ayuntamientos dan su conformidad (o no) y lo remiten en sus ficheros mensuales, así como otros nacimientos y defunciones detectados por el ayuntamiento y que no habían sido correctamente enlazados en los procesos anteriores.

Se produce así un intercambio de información bidireccional que causa retrasos en el "asentamiento" de la variación, ya que este intercambio no es instantáneo.

Un buen esquema del proceso completo de intercambio de información puede ser el que se muestra a continuación.

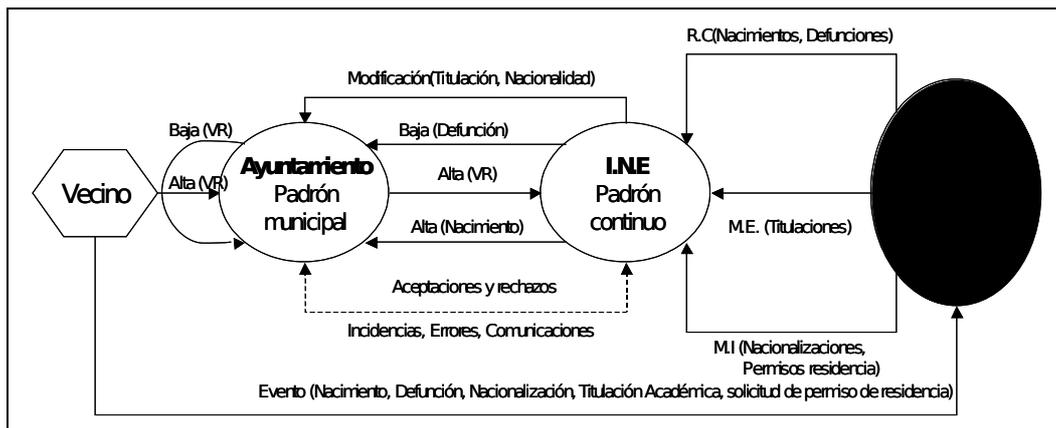


Gráfico 1. Esquema del flujo de información padronal y de los registros civiles.

En todo este proceso el IEA es un mero receptor de información de las distintas fuentes administrativas sin poder realizar ningún tipo de intervención directa en el proceso.

d. ¿Por qué enlazar la información del RPA y el MNP?

En el apartado anterior ya se ha mencionado el proceso de enlace que el INE precisa realizar entre la base padronal y los nacimientos y defunciones. Por motivos que veremos más adelante, este enlace no es (y no puede ser) perfecto. No obstante, el enlace es necesario por una serie de argumentos.

En primer lugar, el enlace de ambas fuentes permite la recuperación de información faltante en los boletines recibidos por el IEA (lugar y fecha de nacimiento, nacionalidad...), evitando llamadas y molestias innecesarias a los Registros Civiles.

En segundo lugar, permite incorporar las defunciones registradas en los Registros Civiles al Registro de Población antes de que sean recibidas en los flujos semestrales, ya que la periodicidad de la información recibida a través de los boletines es mensual. Del mismo modo, nos permite conocer qué parte de las defunciones de un fichero no se encuentran en el otro para estudiar potenciales déficit de información y detectar posibles enlaces aún no comunicados por el INE.

En cuanto a los nacimientos, el enlace permite corregir el déficit de individuos de 0 a 2 años que existe en el padrón debido a los retrasos administrativos y que causa que las cifras oficiales de población de este colectivo a 1 de Enero se encuentren por debajo de lo teóricamente esperable.

Sin embargo, existe un último motivo fundamental para realizar este tipo de enlaces. Podemos realizar un diagrama de Lexis (ver Gráfico 2) para una serie de individuos en concreto, de forma que se comprueben de una manera sencilla algunas necesidades de información para la constitución de una base de datos demográfica longitudinal.

Teóricamente, la información relativa a nacimiento, defunción y cambios de residencia podría obtenerse a partir de los flujos padronales. Sin embargo, para la incorporación de los datos de maternidad, paternidad y matrimonio, se hace necesaria la incorporación de los datos de MNP. Datos de potencial interés como el peso al nacer de los boletines de parto o el estado civil indicado en el boletín de defunción, se perderían si únicamente atendemos a la información padronal de dichos eventos.

La incorporación de los datos de los progenitores en los boletines de parto no sólo permite añadir la información sobre cuándo ocurrió el evento, sino que también permite conocer la relación de convivencia entre padres e hijos. Además, mediante una mera triangulación, pueden obtenerse otras relaciones como las de "hermano", simplemente con buscar aquellos que tengan iguales padres. Podrían estudiarse por tanto en un futuro, datos como la edad de emancipación o tipos de hogar simplemente con la combinación de esta información de MNP con el REPA.

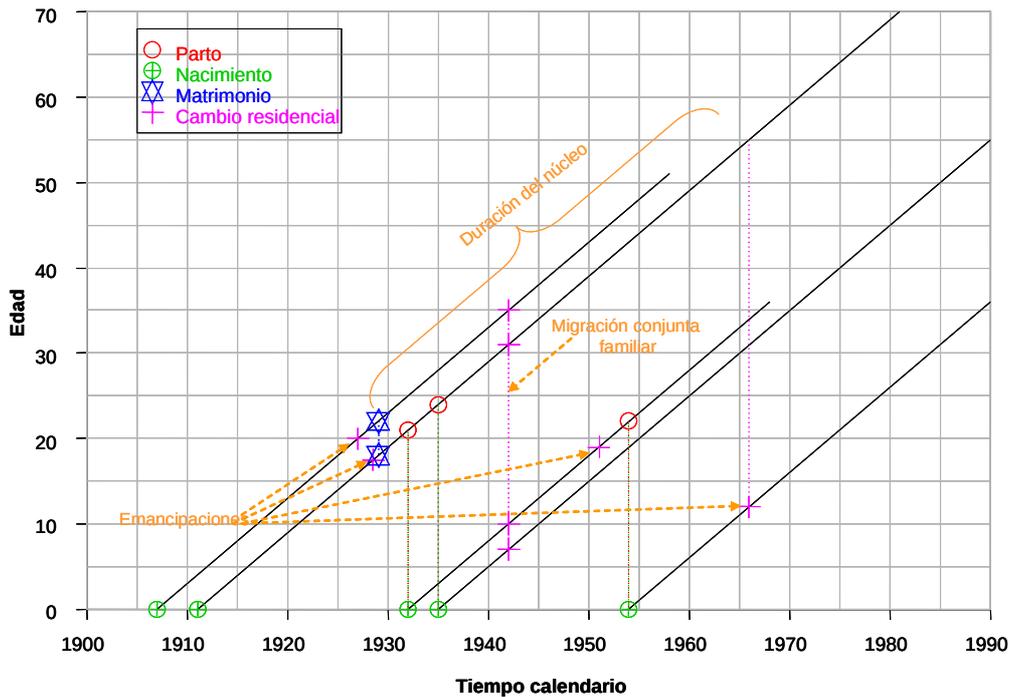


Gráfico 2. Diagrama de Lexis con eventos demográficos significativos..

e. Problemas de identificación

La decisión de qué personas son las mismas parece sencilla a priori. El problema de base en España es que se carece de un registro único para los administrados, provocando que el individuo tenga que dar sus datos cada vez que realiza un trámite y con los consiguientes errores que se puedan producir. Con el desarrollo de la administración electrónica este problema tenderá a desaparecer, pero en la actualidad es necesario tenerlo en cuenta.

Las bases de datos convencionales poseen una estructura rígida de búsqueda. Es decir, podríamos asegurar que dos individuos son la misma persona cuando tienen el mismo nombre y apellidos, sexo, número de documento, fecha y lugar de nacimiento. Sin embargo, todos ellos pueden ser cambiantes en el tiempo o susceptibles de presentar errores.

Como ejemplo, un español puede invertir el orden de sus apellidos e incluso tener un cambio de sexo y por consiguiente cambiar su nombre. Un extranjero podría llegar a España teniendo como documento un pasaporte, conseguir posteriormente la tarjeta de residencia e incluso nacionalizarse (3 números de documento diferentes).

A ésto habría que sumarle los errores de transcripción, como intercambiar el día y el mes de nacimiento, que el funcionario se equivoque al pasar los apellidos del papel al ordenador...

Esto casos no son más que una pequeña ilustración de qué puede suceder para que la identificación no sea sencilla. Con un simple vistazo se podría determinar, probablemente, si dos conjuntos de datos pertenecen a la misma persona...pero habría que comparar anualmente más de 300.000 pares de individuos.

2. Objetivos

El objetivo de este trabajo es enlazar las series de partos y defunciones de MNP de Andalucía desde 1998 hasta 2009 con la serie histórica de variaciones del Registro de Población de Andalucía con varias metas.

Por un lado, estudiar los errores más habituales en los campos básicos de identificación y el grado de similitud de los pares identificados como de la misma persona.

En segundo lugar, ver el grado de cobertura de los enlaces para poder establecer colectivos con poco porcentaje de enlace.

La elección del periodo no es aleatoria. El inicio en 1998 se debe a que a partir de ese año se dispone de la información de nacimientos y además se produce la primera revisión padronal

desde la instauración del padrón continuo. El final en 2009 se debe a que se trata del último año con información completa tanto para flujos padronales como para eventos de MNP.

3. Metodología

a. Adaptación de la información

El primer paso, que inicialmente podría pasarse por alto para aquellos no habituados al enlace de registros, sería el de adaptar la información de ambos ficheros para que compartan el mismo formato. En REPA se ha optado por crear una base de datos básicos de personas para los datos padronales y otra para MNP. Estas bases de datos pueden verse con más detalle en otra ponencia presentada a estas JECAS⁴. Estos datos son una selección de la información más elemental y sólida de las personas en ambos ficheros, los datos de una persona que (a priori) no deberían cambiar. Por ejemplo, el lugar de residencia de una persona puede cambiar a lo largo del tiempo, pero sin embargo el lugar de nacimiento no debería hacerlo. Además de ello, se han de realizar una serie de modificaciones (estructura, normalización y estandarización) en los ficheros con objeto de mejorar la comparabilidad de las fuentes. Además de esto, se dispone de una clave de identificación de persona (IDP) para los datos padronales creada por el IEA y que puede tener varias versiones de datos relativos al mismo individuo. De forma análoga, existe una clave identificadora de la versión de persona de MNP, denominada PKVP.

- Cambios en la estructura

No en todas las bases de datos la información se estructura de igual manera. Algunos campos necesitan una adaptación que permita su homogeneización.

Los apellidos son uno de estos campos. En ambos ficheros, el primer y el segundo apellido aparecen por separado, pero la estructura en padrón es la de Partícula + Apellido, mientras que en MNP la partícula, de existir, se incluye en el mismo campo. Se optó por extraer las partículas de apellidos de MNP contenidas los mismos y crear un nuevo campo con ellas.

En cuanto al documento de identidad, hasta la aparición de los nuevos boletines de MNP en 2007 sólo había un campo para la declaración de la documentación, ya fuera DNI, tarjeta de residencia o pasaporte. El padrón, sin embargo, distinguía 2 campos. Por tanto, se hace necesario que se distingan patrones en MNP que permitan decidir si un número se corresponde con una u otra documentación. En la misma línea, en estos nuevos boletines, el DNI aparece como 8 dígitos más 1 letra, mientras que en el padrón municipal esa letra se rellena en un campo aparte.

- Normalización y estandarización

Tras los cambios de estructura se procede a normalizar y estandarizar los campos. En los procedimientos de normalización se pueden distinguir dos acciones principales. La primera de ellas es la eliminación de anomalías que no aportan información o existe evidencia de que sean errores. La segunda es decidir qué valores consideraremos nulos.

La idea de la normalización es la de adaptar los ficheros para que todo aquello que represente exactamente lo mismo se trate de igual forma. Para el caso de campos de tipo carácter, algunas de las decisiones tomadas en este proceso de normalización son las siguientes:

- Eliminación de dobles espacios en blanco
- Eliminación de espacios en blanco al inicio y al final de cada campo
- Eliminación de tildes
- Eliminación de puntos y otros signos de puntuación que no aportan información.
- M^a se sustituye por MARIA en nombres

En los campos formados teóricamente por números, se opta por sustituir todos los caracteres no numéricos (espacios en blanco, guiones, letras) por un cero y luego transformar a numérico. Si el resultado de esta transformación es cero, se le asigna el valor considerado nulo.

Con esta transformación, los códigos de provincia '4' y '04', representan lo mismo, Almería, aunque originalmente estuvieran escritos de forma diferente.

⁴ Base de Datos Longitudinal de Población de Andalucía (BDLPA): Modelo de datos y sistema de gestión. Francisco Viciano, Rosa Cánovas, Víctor Montañés, Eva V. Poza

La transformación es especialmente útil en el caso de la documentación. Tanto en padrón como en MNP se utilizan indistintamente los valores '-' y '00000000' para indicar que no hay DNI. Incluso aparecen valores del tipo 'XXXXX'... y sin embargo, intuitivamente sabemos que representan la misma idea. Con la normalización todos pasarían a presentar el mismo valor, en este caso '00000000'.

La otra fase de la normalización es la de detectar valores fuera de rango o claramente equívocos. Un código de provincia que no se corresponda con la codificación INE sería descartado (considerado nulo). En el caso de apellidos es frecuente que aparezcan las inscripciones "NO CONSTA" O "DESCONOCIDO" que deberían ser claramente transformadas a nulo.

La estandarización, por su parte, va un paso más allá y afecta fundamentalmente a los nombres y apellidos. Tras un primer proceso de normalización ampliado (necesario para la elaboración de la Estadística de Nombres y Apellidos de los andaluces) se hace un segundo acercamiento con la creación de un literal estándar. Esta estandarización se basa en una idea tan sencilla como un criterio fonético: todo lo que suena igual debe representarse igual. Con esta idea de fondo se realizan una serie de transformaciones como las que se indican

- CA, CO, CU, QU ->KA, KO, KU, KU
- B=V
- CE,CI ->ZE, ZI
- GE,GI ->JE, JI
- H desaparece
- Desaparecen las siglas sueltas ya que no constituyen un nombre completo
- Unificación de nombres que inician con J e Y
- Eliminación de partículas (DE, LOS, LA...)
- Eliminación de dobles caracteres consecutivos (excepto R y L)
- ...

Como ejemplo, "Christian", "Cristian", "Cristhian" pasarían a ser representados por "KRISTIAN". Como curiosidad o como potencial de estas transformaciones diremos que el anterior nombre poseía originalmente 79 versiones diferentes.

Estos procedimientos de normalización y estandarización reducen en gran medida la variabilidad del conjunto de nombres y apellidos. Para nombres, la normalización permite reducir en un 12% el número total de denominaciones diferentes y la estandarización otro 7% adicional. En caso de los apellidos, la normalización obtiene una reducción del 23% y la estandarización, un 12% adicional.

b. Estrategia y procedimientos de enlace

La estrategia más "elegante" sería emplear la metodología de enlace de registros por el método probabilístico de Fellegi-Sunter, empleado en otras comunidades autónomas⁵. De hecho, en el IEA se está llevando a cabo un proyecto en este sentido. No obstante, hasta el momento, los tiempos de respuesta distan aún de la eficiencia requerida, por lo que hasta el momento se ha aplazado su uso a la espera de la mejora de dicha eficiencia. Además, como simplificación para la resolución del problema, se debe suponer que las componentes de los vectores comparados (conjunto de campos empleados) sean estadísticamente independientes todas con respecto a cada una de las distribuciones condicionales. Tras varias pruebas, esta independencia podría ser puesta en tela de juicio a tenor de algunos de los resultados obtenidos. Por ejemplo, en el registro de población, la probabilidad empírica de haber nacido el 1 de Enero siendo marroquí es del 0.25, siendo español es de 0.0047 y siendo de Mali es de 0.54.

La probabilidad de tener un documento de identidad nulo es de 0.05 para los nacidos después de 1990, mientras que para el resto supera el 0.9.

La probabilidad de tener el segundo apellido nulo es prácticamente cero para personas de nacionalidad española y para otras nacionalidades europeas es prácticamente 1.

Por tanto, hay valores de ciertas variables que no parecen mostrar la deseada independencia y que utilizados como condicionamiento dan lugar a distribuciones que no tienen nada que ver la una con la otra.

⁵ Métodos automáticos de fusión de registros y su utilización en Eustat. Leire Legarreta y Laura Otero, EUSTAT 2007

Nuestra estrategia, sin embargo, lleva una serie de etapas.

1. Enlace masivo y determinístico con los datos originales.
2. Sobre el conjunto restante se realiza un nuevo enlace determinístico pero empleando literales estandarizados.
3. Sobre el conjunto restante se buscan candidatos “no exactos”, pero con suficiente grado de similaridad (enlace pseudo-probabilístico).

Este último punto es el que merece una explicación más detallada. En primer lugar se definen un conjunto de bloques. Estos bloques no son más que un subconjunto de pares que comparten cierta información, debiendo ser suficientemente restrictivos para poder abordar el problema. Nuestro primer bloque, por ejemplo, está formado por todos aquellos individuos que poseen DNI en el fichero de MNP para los que existe una persona con dicho DNI en RPA.

Una vez obtenidos los dos registros candidatos, se contrastan y evalúan el resto de campos. Para esta comparación se emplean un conjunto de funciones de proximidad creadas para cada uno de los campos y basadas en experiencias previas y que se relatan en el apartado siguiente. Con cada una de las puntuaciones parciales y aplicando un sistema de ponderaciones y penalizaciones, se calcula una puntuación global.

Todos aquellos pares que obtienen una puntuación global suficiente, pasan a ser considerados enlaces. Tras este proceso, se volverían a realizar nuevos bloques para el subconjunto actual de no enlaces. Nuestros criterios de bloque suelen ser la coincidencia de los dos apellidos junto con la fecha de nacimiento o la coincidencia de los dos apellidos junto con el nombre. Posteriormente podría pasarse a bloques menos restrictivos.

Para el caso de los partos se establece un bloque complementario, bastante restrictivo y efectivo. En el boletín de parto aparecen tres personas: padre, madre e hijo. Parece lógico pensar que estas tres personas podrían vivir en la misma vivienda. Si hemos conseguido enlazar al menos a uno de ellos, dispondremos de la información de las viviendas (clave de vivienda u hoja padronal) en las que ha residido desde su primera inscripción padronal posterior a 1996. Como consecuencia, propondríamos como candidatos a aquellas personas que también han habitado esas viviendas.

Así, por ejemplo, si se encuentra a la madre y no se encuentra al padre, compararíamos los datos del padre con todas aquellas personas que en algún momento han convivido con la madre, que ya es de por sí una restricción bastante grande. Esto mejora en gran medida el proceso de búsqueda, sobre todo para aquellos recién nacidos que presentan alguna errata en los apellidos en alguno de los ficheros (y que no formarían parte de los principales bloques descritos anteriormente).

Estas funciones y puntuaciones pueden ser más o menos discutibles, aunque lo que es innegable es que funcionan razonablemente bien. No obstante, se está planteando la posibilidad de utilizar técnicas de análisis discriminante y la construcción de árboles de decisión para la determinación de enlaces satisfactorios. Existe un software gratuito, WEKA, de la Universidad de Waikato (Auckland, Nueva Zelanda) que es capaz de satisfacer estas necesidades.

c. Funciones de proximidad

Para calcular la similaridad entre 2 valores pertenecientes al mismo campo se hace necesario disponer de una serie de funciones que la evalúen. Algunas de las funciones se basan en la distancia de edición de Levenshtein⁶. Esta distancia mide el número de cambios de carácter o de posición que habría que aplicar a una cadena de caracteres para transformarla en otra. Una vez obtenido el número de cambios se relativiza empleando la fórmula

$$(N^{\circ} \text{ de caracteres} - N^{\circ} \text{ de transformaciones}) * 100 / N^{\circ} \text{ de caracteres}$$

cuyo rango va desde 0 (similaridad nula) a 100 (similaridad máxima).

En el caso de nombres y apellidos se utiliza un valor de proximidad basado en dicha distancia de edición de Levenshtein, aunque ligeramente modificada para evitar ciertos efectos “nocivos”. El valor de la comparación es el de la distancia de edición, salvo en tres casos:

⁶ En ORACLE se hace a través de la función UTL_MATCH.edit_distance_similarity.

- Si los dos literales no compartan suficiente información, pero uno de ellos esté contenido en el otro, se asigna una cantidad razonable de proximidad. Se hace una corrección al alza.
- Que los dos literales no compartan suficiente información debido a un intercambio entre nombres y apellidos. Se hace una corrección al alza.
- Que los dos literales no compartan suficiente información, en un caso distinto a los anteriores. Consideramos en ese caso que la proximidad es cero.

Para la comparación de fechas de nacimiento se comparan día, mes y año en dos ocasiones, en conjunto y por separado. Cuando se realiza por separado se tienen en cuenta coincidencias exactas. Cuando se realiza en conjunto se valoran positivamente, con distinto grado, dos tipos de proximidades:

- Que la diferencia sea de tan sólo un día
- Que la diferencia sea de entre 2 y 10 días.

En sentido inverso, se penaliza cuando la fecha de nacimiento es el 1 de Enero. Este criterio no es para nada arbitrario, ya que en el caso de Andalucía, el 25% de las personas con nacionalidad marroquí declaran haber nacido en esa fecha, lo cual la hace poco fiable.

En el caso del DNI se utiliza de nuevo la distancia de edición, pero aplicando una modificación diferente a la anterior.

- Si no hay al menos una coincidencia del 75%, se considera que la proximidad es nula.
- Si únicamente difiere la última cifra se penaliza la coincidencia, ya que es habitual que suceda cuando hay gemelos, y una valoración alta de la puntuación en este caso podría llevar a un error de identificación (falso positivo)

Se comparan otra serie de campos como sexo, lugar de nacimiento o de residencia que finalmente se emplean como reajuste, siendo valores de escaso peso en la valoración final.

4. Resultados

a. Cobertura

Un dato fundamental para analizar el éxito del enlace de registros es conocer qué porcentaje de registros de un conjunto enlazan con los del otro. No obstante, y dado que ambos conjuntos no contienen a los mismo individuos, sería más correcto realizar en primer lugar un esquema para determinar las posibles limitaciones con las que podemos encontrarnos a priori.

- En el conjunto de registros de MNP podemos distinguir varios colectivos. El núcleo principal estará formado por personas que declaran residir en Andalucía. El problema es que el hecho de que declaren residir no implica necesariamente que se encuentren empadronadas en Andalucía.
- El otro núcleo estará formado por aquellas personas que declaran en el boletín que no residen en Andalucía, que se puede subdividir a su vez entre españoles y extranjeros.

Por tanto, desde este punto de definición del conjunto conocemos una limitación, ya que será bastante improbable poder realizar el enlace de los no residentes.

En cuanto al conjunto de registros de REPA, tendremos también unos grupos definidos. El fundamental serán los residentes nacidos en Andalucía, y al igual que en el anterior, los nacidos en el resto de España y el extranjero.

Para el análisis de la cobertura de MNP respecto a RPA ofreceremos varios datos, dando cada uno de ellos un matiz ligeramente distinto.

- Cobertura global: Porcentaje de registros de MNP que enlazan con RPA.
- Cobertura autóctona: Porcentaje de registros de MNP en los que se indica como residencia algún municipio andaluz que enlazan con RPA.

Las siguientes tablas muestran los resultados obtenidos durante el periodo 1998-2009.

<i>Cobertura global</i>			
	Localizado	No localizado	Tasa
Difunto	724.809	45.726	94,07%

Madre	1.018.878	26.286	97,48%
Padre	998.220	30.727	97,01%
Hijo	978.853	66.315	93,66%

Cobertura autóctona

	Localizado	No localizado	Tasa
Difunto	661.056	33.373	95,19%
Madre	930.073	20.094	97,89%
Padre	911.239	20.648	97,78%
Hijo	915.052	35.113	96,30%

Aunque en todos los casos la cobertura supera el 93%, vemos que el mejor enlace se produce en los progenitores. Entre otros motivos, esto puede deberse a que son los propios implicados los encargados de rellenar el boletín, hecho que no sucede con los difuntos. Restringiéndonos a los que declaran residencia en Andalucía, se observa que el colectivo con mayor ganancia en el enlace es el de los recién nacidos, y en esta situación el éxito del enlace se situaría por encima del 95% en todos los casos.

Estos datos muestran qué porcentaje de personas que aparecen en MNP han sido localizados en el REPA. Sin embargo, esto no implica que aparezcan los registros de alta por nacimiento o la baja por defunción en el padrón. Para estudiar dicho tipo de coincidencia, lo haremos a través de otras dos tasas equivalentes a las anteriores.

- Cobertura global del evento: Para recién nacidos y difuntos, porcentaje de registros de MNP que enlazan con la correspondiente alta por nacimiento o baja por defunción del padrón.
- Cobertura autóctona del evento: en la línea de lo anterior, porcentaje de registros de MNP que declaran residir en Andalucía que enlazan con la correspondiente alta por nacimiento o baja por defunción del padrón.

Ya de antemano, se sabe que estas tasas estarán por debajo de las calculadas previamente, ya que el numerador es un subconjunto del utilizado con anterioridad.

Cobertura global del evento

	Localizado	Total	Tasa
Difunto	655.128	770.535	85,02%
Hijo	846.889	1.045.168	81,03%

Cobertura autóctona del evento

	Localizado	Total	Tasa
Difunto	653.926	694.429	94,17%
Hijo	845.811	950.165	89,02%

Interpretando los datos de la cobertura autóctona, podemos decir que un 1% de las defunciones de MNP no se han incorporado a la base padronal, ya sea por retrasos, falta de enlace o porque la persona había emigrado de Andalucía antes de su fallecimiento. Por su parte, un 7% de los nacimientos se inscriben con causas distintas a la de nacimiento y, como ya veremos, con retraso.

A estos cocientes habría que añadirles además otra cifra adicional para el caso de los fallecidos, que sería el número de defunciones comunicadas a través del sistema padronal y que no han sido comunicadas o enlazadas a través de MNP. Esta cifra alcanza las 16.730 defunciones en el periodo considerado.

Estos casos pueden deberse a dos motivos. El primero de ellos, como cabe pensar, es que los procesos de enlace de registros fallan. El segundo de ellos serían los individuos que oficialmente residen en Andalucía pero que han fallecido fuera de nuestra comunidad, declarándose en el boletín estadístico que ya no residía en Andalucía.

El caso de errores en la residencia puede ocurrir también a la inversa. Es decir, se ha podido declarar en el boletín que residía en Andalucía cuando en realidad la persona nunca había llegado a empadronarse en la comunidad.

En cuanto al grado de puntuación global alcanzado en los enlaces (*links*), hemos considerado para la versión de persona más similar al evento de MNP que se trataba de enlazar. En el gráfico se muestra la distribución de dichas puntuaciones.

Enlaces según puntuación y evento

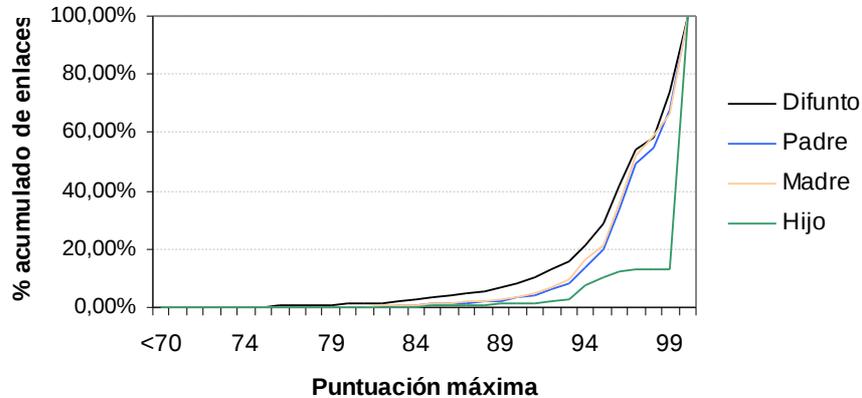


Gráfico 3. Distribución de la puntuación de aceptación según persona implicada.

Vemos que el patrón de las puntuaciones de enlace difiere según el evento que estemos tratando. La diferencia más evidente reside en puntuaciones de los recién nacidos, ya que lógicamente no nacen con un DNI y por lo tanto el campo es prácticamente ignorado. Olvidando este hecho, el otro dato a resaltar es que las defunciones se enlazan con puntuaciones ligeramente más bajas. Más adelante, en un análisis campo a campo veremos a qué se debe esta anomalía.

b. Detalles significativos de los eventos

- Defunciones

Con los datos obtenidos sobre el enlace de las defunciones podemos obtener una tabla de contingencia de resumen, al menos para la población residente, y así poder dar un dato de cobertura más amplio.

		RPA	
		Sí	No
MNP	Sí	661.056	16.730
	No	33.373	0

		RPA		
		Sí	No	Total
MNP	Sí	A	B	MS
	No	C	D	MN
	Total	RS	RN	T

El cociente A/T indica que el 92,5% de las defunciones presentan personas en común en los dos ficheros, presuponiendo que las defunciones no emparejadas no están enlazadas porque efectivamente no pertenecen a las mismas personas, es decir, ningún elemento de B está contenido en C y ningún elemento de C está contenido en B.

En otro sentido, bajo el supuesto de que las 16.730 defunciones (B) no enlazadas están incluidas en C, y el problema se debe exclusivamente a fallos en el proceso de enlace, la tasa de cobertura de RPA sobre MNP sería del 97,5% ($A/(A+B)$).

Por otro lado el 2,3% de defunciones $(C-B)/(A+C)$ que declaran su residencia en Andalucía serían de personas no empadronadas realmente.

Clasificando los enlaces según el tipo de relación que los ha causado, podemos hacernos una idea de qué porcentaje de registros poseen una identificación determinística y cuántos probabilística, y en qué grado la estandarización ayuda en el proceso de enlace.

Grado de enlace de defunciones por año del evento y forma

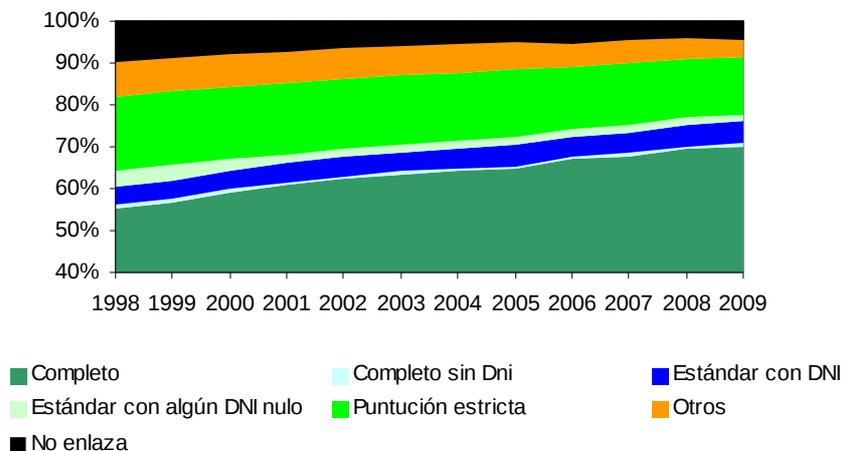


Gráfico 4. Distribución del tipo de aceptación de los pares según año del evento

El enlace determinístico pasa de un 56% a un 70%. Esta mejora con los años se debe en parte a que a medida que pasan los años se dispone de más versiones de persona para cada individuo. Por otro lado, también podemos ver que el proceso de estandarización permite una ganancia de entre el 5% y el 8%. Pero lo realmente esperanzador es el que el porcentaje de no enlazados se ha ido reduciendo progresivamente con el paso de los años, es decir, a medida que el registro de población crece se mejora el enlace.

Si hacemos el enlace en función del tamaño del municipio, los resultados nos proporcionan alguna que otra curiosidad.

Enlace por provincias y tamaño del municipio

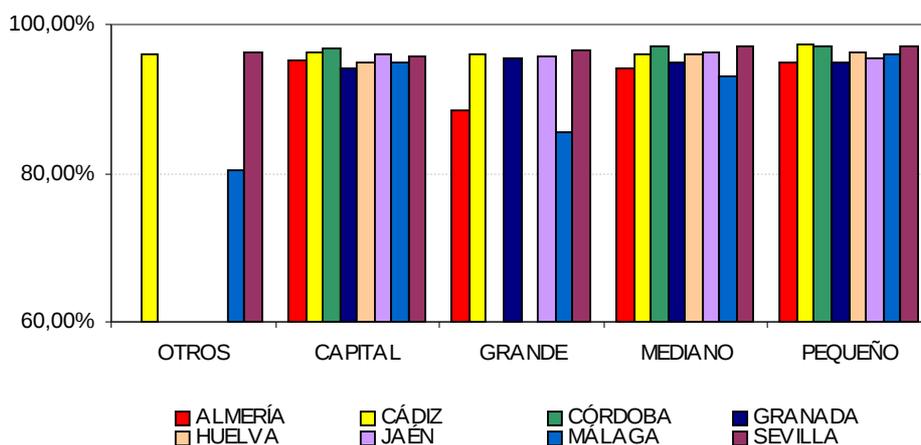


Gráfico 5. Porcentaje de enlace de defunciones según provincia y tamaño del municipio.

El grado de enlace es más o menos similar en toda Andalucía y no parece depender del tamaño del municipio. Sin embargo, se produce una anomalía en algunos municipios de Málaga y Almería que presentan el peor enlace, fundamentalmente en los municipios grandes que no son capitales, siendo el más llamativo el caso de Marbella donde se alcanza tan sólo el 80% mientras la tónica general de enlace de Andalucía supera el 95%. Este efecto de la costa puede deberse en gran parte a la población inglesa jubilada que fallece en nuestro territorio pero no se encuentra empadronada, a pesar de que en el boletín de defunción se declare que

sí reside. De hecho, únicamente el 52% de los fallecidos de nacionalidad inglesa consiguen enlazarse.

En cuanto a la edad, podemos representar conjuntamente en un gráfico el número de defunciones y el porcentaje de enlace a dicha edad, separando entre población en general y nacidos en Andalucía.

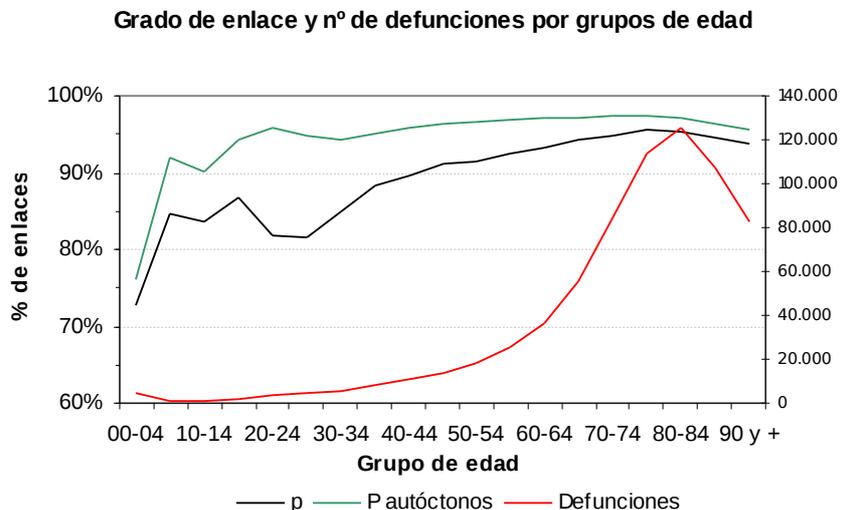


Gráfico 6. Grado global y autóctono de enlace de las defunciones según edad.

El peor enlace se produce en las edades más jóvenes, motivado fundamentalmente por la ausencia del DNI, que es el mejor decisor del que se dispone. Es en estos grupos en los que resulta más evidente la mejoría de considerar población autóctona. No obstante, el número total de defunciones a estas edades es bajo, lo que podría animar a una recuperación manual de enlaces. A partir de los 25 años mejora hasta llegar a las edades más avanzadas, donde por fortuna, el grupo con mayor número de defunciones es el que mejor porcentaje de éxito presenta.

- Recién nacidos

La clasificación por edad en los recién nacidos carece de sentido alguno. Sin embargo, la clasificación según año del evento sí proporciona una información útil que puede interpretarse junto con el Gráfico 7.

Al contrario de lo que sucedía con las defunciones, el enlace empeora conforme se acerca la fecha del parto a la fecha actual. Sin embargo, este efecto se debe al retraso en la recepción de nacimientos en los flujos padronales y no a un problema de enlace de registros. Si analizamos la población oficial de 0 años de los padrones a 1 de Enero de cada año y la comparamos con el número de nacimientos, podemos ver claramente que se debe a estos retrasos.

En el Gráfico 8 se representa la población de 0 años en Andalucía para un año X frente al número de nacimientos de madre andaluza de MNP y el número de nacidos en ese año en REPA. Además se incluye la población de 1 año empadronada en el año X+1 y la población de 2 años en el año X+2 (es decir, la evolución natural de la población de 0 años en el año X).



Gráfico 7. Grado de enlace de los recién nacidos según año de nacimiento.

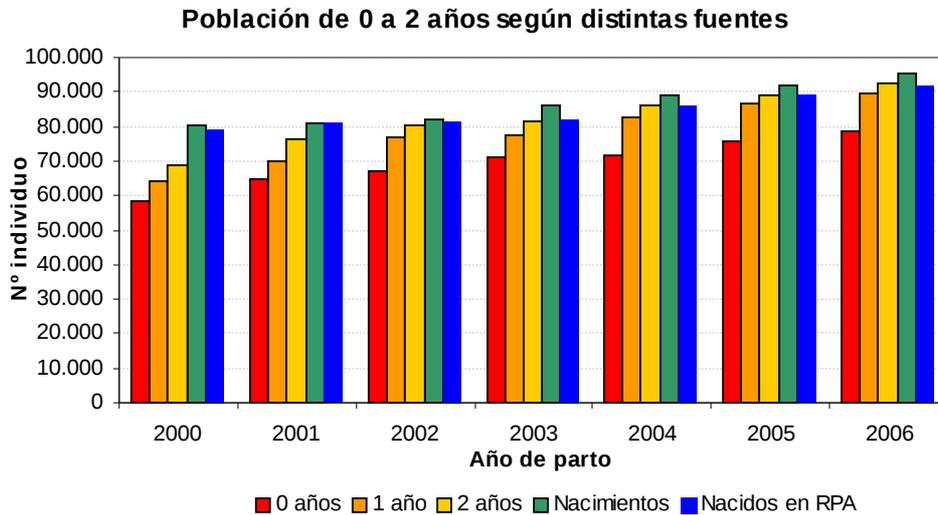


Gráfico 8. Retrasos en la inscripción padronal de nacimientos

Puede verse en el gráfico como la población nacida en un año concreto (según padrón) se va corrigiendo año tras año hasta aproximarse al número de nacidos, según MNP. Tras el segundo año, dichas cifras comienzan a emparejarse. Por este motivo, los años 2008 y 2009 presentan un porcentaje de enlace más bajo, ya que todavía falta población en la base padronal. Este desfase se corrige en unos dos años. Por tanto es de esperar que a final de 2010 se corrija en su mayor parte la cifra de 2008 y a final de 2011 la de 2009.

Es más, el enlace con REPA permite obtener la estimación del retraso de inscripción de los nacimientos en el padrón. Aquellos no inscritos como nacimientos presentan el siguiente comportamiento.

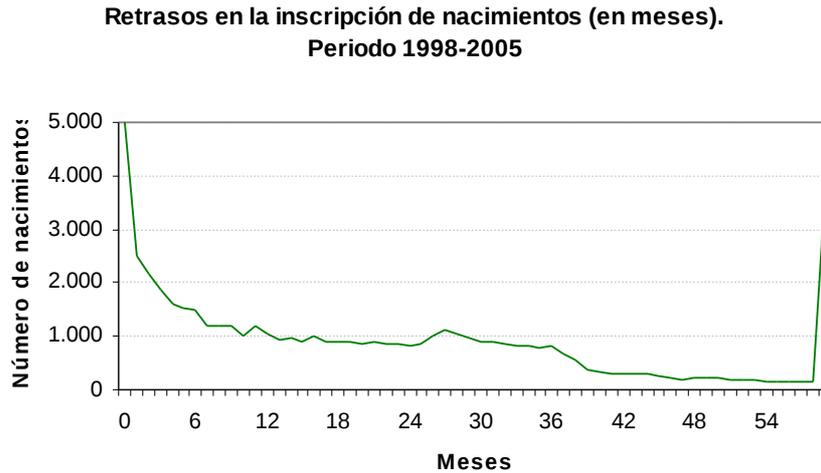


Gráfico 9. Distribución del retraso temporal de nacimientos inscritos en fechas posteriores.

Durante los 6 primeros meses de vida se produce la mayor parte de la corrección de este colectivo, pero hay un goteo incesante hasta al menos los 36 meses de edad (3 años). De hecho, en torno a un 7% de los nacimientos pasan a formar parte del padrón a través de altas por omisión, aunque es cierto que la mejora de los procedimientos de enlace del INE ha ido reduciendo esta cifra año tras año (de 11,3% en 1998 a un 4,4% en 2008).

Sobre los nacimientos también resulta de interés conocer cuántas personas se han enlazado por cada parto. Su distribución porcentual según año de parto es la que se muestra a continuación:

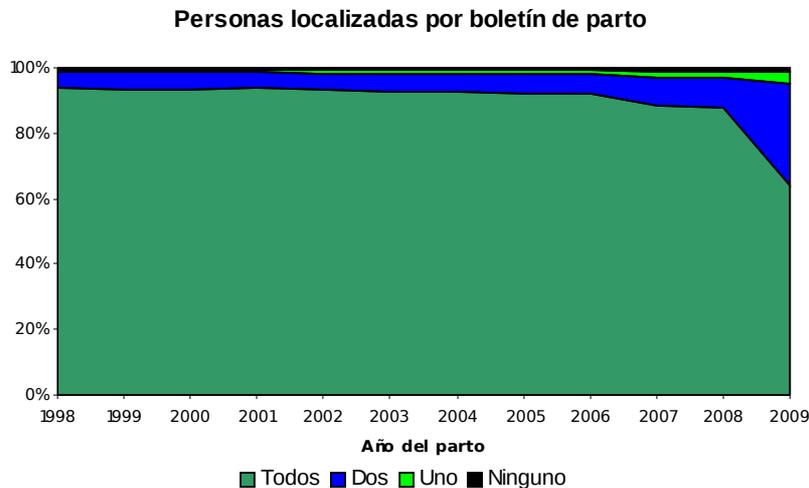


Gráfico 10. Número de personas enlazadas por boletín de parto (partos simples)

Si atendemos a la línea marcada por la categoría “Todos”, el perfil es prácticamente calcado al que obteníamos en cuanto a la proporción de enlace de los nacidos. Por tanto, lo que ya indicamos de los retrasos de los nacimientos se vuelve a poner de manifiesto. Si hay dos personas enlazadas y son los padres, es de esperar que en un breve lapso temporal estos recién nacidos aparezcan en el padrón y puedan enlazarse.

c. Comparabilidad de los campos

Otra fase interesante del estudio es saber en qué medida coinciden los campos comparados para aquellos pares que han sido identificados como enlaces. Se ha optado por incluir los principales campos empleados indicando si hay coincidencia total en los campos.

CAMPO	COINCIDENCIA	Difunto	Madre	Padre	Hijo
APE1	NO	3,4%	2,7%	2,6%	1,7%
	SÍ	96,6%	97,3%	97,4%	98,3%
APE2	NO	4,1%	3,6%	3,7%	2,0%
	SÍ	95,9%	96,4%	96,3%	98,0%
DNI	NO	3,7%	3,2%	3,2%	0,0%
	Sin determinar	3,0%	8,5%	6,6%	6,4%
	SÍ	93,3%	88,3%	90,2%	93,6%
FNAC	NO	16,4%	5,5%	5,3%	0,8%
	SÍ	83,6%	94,5%	94,7%	99,2%
NOMB	NO	6,5%	22,4%	6,2%	2,9%
	SÍ	93,5%	77,6%	93,8%	97,1%
Todos	NO	31,7%	40,2%	25,2%	13,1%
	SÍ	68,3%	59,8%	74,8%	86,9%

Por lo general, el grado de coincidencia exacta es bastante alto campo a campo, cercano al 90%. Ahora bien, si pedimos una coincidencia absoluta de los cinco campos en conjunto, esta cifra baja de forma notable.

El dato "Sin determinar" del DNI se refiere a que en un fichero existe el DNI y en el otro es nulo. Por tanto no se puede decir que haya falta de coincidencia, sino que en todo caso existirá información faltante.

Los otros dos datos llamativos son la falta de coincidencia de la fecha de nacimiento en las defunciones y la del nombre de las madres en los partos. Para estos dos casos disponemos de una información más detallada que permite conocer un poco más el problema.

Para el caso de falta de coincidencia de la fecha de nacimiento hemos clasificado las no coincidencias en varios grupos, ya que en muchos casos coinciden parcialmente. Así, marcaremos aquellos cuya fecha de nacimiento varíe menos de 10 días, que sólo varíe el día o que fallen año, mes, mes y día... Con dicha clasificación y tabulando por edad del fallecido (edad en MNP) obtenemos un gráfico con información bastante curiosa y útil.

Coincidencia en la fecha de nacimiento de defunciones según edad

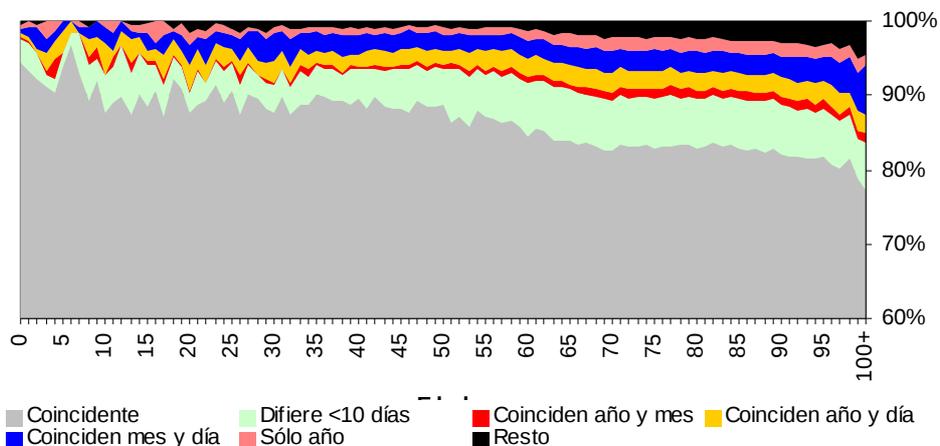


Gráfico 11. Desagregación de la coincidencia entre fechas de nacimiento según edad.

A medida que se van alcanzando edades más avanzadas, el porcentaje de errores en la edad va aumentando hasta ser fallar en uno de cada cinco casos. Es más, el porcentaje de pares para los que ambas fechas no tienen prácticamente ninguna similaridad se incrementa. Por tanto, podríamos plantear no penalizar en exceso la falta de coincidencia de la fecha de nacimiento de los ancianos en futuros enlaces.

En cuanto a la extraña no coincidencia del nombre de las madres, también podemos proceder a separar ese grupo en varios subgrupos. Por un lado tomaríamos los casos en los que el nombre estandarizado coincide. También tomaríamos los casos de inclusión de un nombre dentro de otro (María Rosa VS Rosa) y los casos en que las dos cadenas varían en un solo carácter. De acuerdo con esta clasificación y agrupando año del parto obtenemos los siguientes resultados.

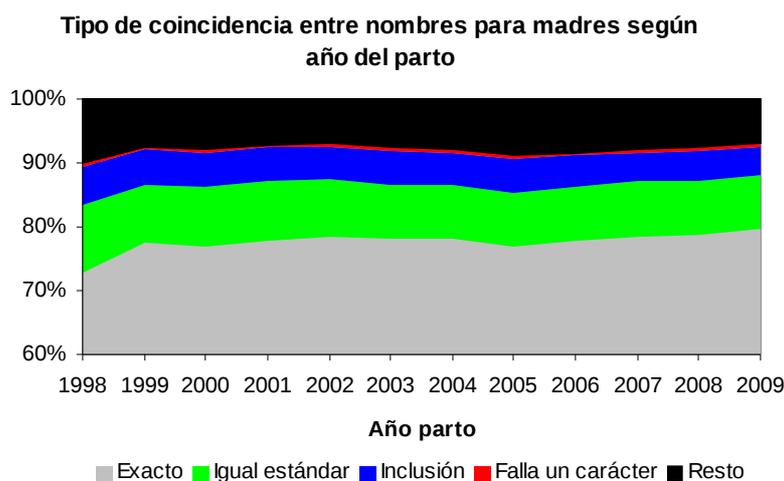


Gráfico 12. Tipo de coincidencia nominal de los nombres maternos según año del parto.

Puede comprobarse que si a la similaridad total añadimos la coincidencia estándar, se obtiene una ganancia del 10%. Además vemos que otro 5% se justifica por nombres compuestos que se han simplificado. De hecho, éste es el motivo de que en la función de proximidad de nombres se valore de forma positiva este tipo de coincidencia parcial.

5. Utilidades

Ya hemos comprobado alguna de las potenciales utilidades de este enlace entre fuentes, como por ejemplo conocer qué campos son más fiables a la hora de establecer un enlace. Otra utilidad es la de corregir la pirámide de población que se obtiene de las cifras oficiales desde el 1 de Enero de 1998. Para la BDLPA se podría corregir la fecha de incorporación de los recién nacidos al domicilio de los padres e iniciar el itinerario vital de aquellos que aún no pertenecen a la base padronal a través del sistema de flujos, pero de los que se conoce la residencia de los padres.

a. Fecundidad

En esta ocasión daremos algunas pinceladas de su aplicación a fecundidad, comprobando el patrón de fecundidad de 3 generaciones de mujeres, las nacidas en 1970, 1975 y 1980. Consideramos como colectivo inicial aquellas mujeres de dichas generaciones que se encontraban empadronadas en Andalucía el 1 de Enero de 1998. Para ellas, a través del enlace con el fichero de partos, podemos obtener la fecha en la que dieron a luz a su primer hijo (recordemos que el enlace era superior al 97%). No obstante, habría que hacer alguna consideración, ya que las mujeres de las generaciones de 1975 y 1970 tenían 28 y 23 años en la fecha de referencia considerada y es probable que ya hubieran tenido algún hijo.

Aclarado este punto, se procede al enlace entre fuentes para conocer la fecha de nacimiento del primer hijo. Para aquellos casos en que la mujer ya hubiera tenido previamente alguno, se toma la fecha del hijo anterior, que es uno de los campos que aparecen en el boletín de parto.

En los casos restantes, se procede a buscar en el censo (también enlazado con la BDLPA) la fecha de nacimiento del menor de los hijos.

Una vez realizados todos estos procedimientos, podemos conocer que proporción de mujeres ha dado a luz a su primer hijo y a qué edad exacta los han ido teniendo.

Proporción de mujeres con algún hijo según edad (en meses)

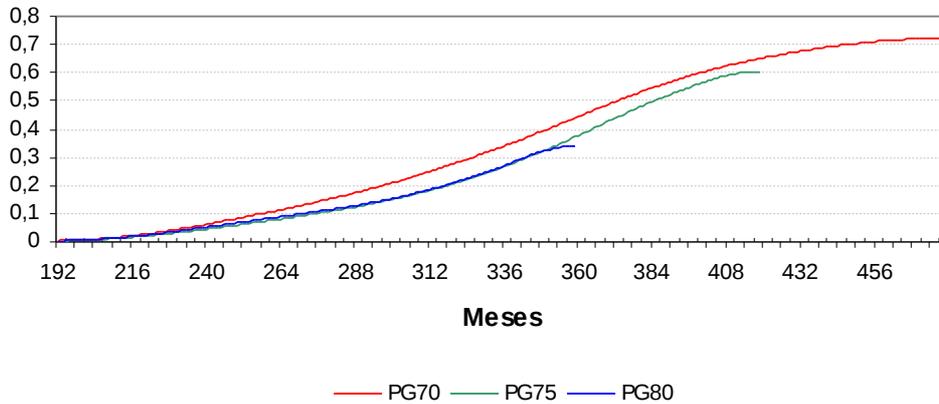


Gráfico 13. Proporción de mujeres con hijos para las generaciones de 1970, 1975 y 1980.

En el gráfico se muestra la evolución desde los 18 a los 40 años (edad máxima alcanzada por la generación de 1970). Se observa un patrón claramente diferenciado entre las generación de 1970 y las otras dos. No podemos concluir que la proporción de mujeres con algún hijo vaya a descender, aunque lo que sí es cierto es que han visto retrasado su calendario biológico. Quizás lo más adecuado sea observar la diferencia según calendario de estas dos series frente a la serie de las nacidas en 1970.

Diferencia en la proporción de mujeres con algún hijo según edad

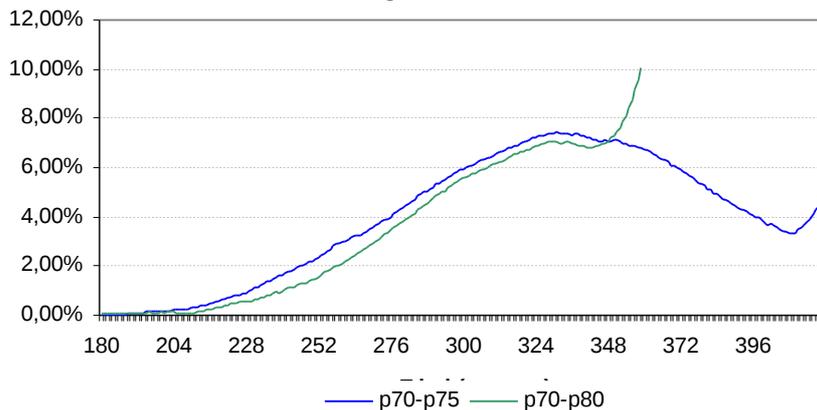


Gráfico 14. Diferencia en la proporción de madres con hijos de las generaciones de 1975 y 1980 frente a la generación de 1970.

Se observa una vez más el comportamiento tan parejo que poseen las series de 1975 y 1980. Al final de cada una se produce una desviación anómala que no es más que un mero artificio causado por la fuente, ya que los datos de los meses finales corresponden a 2010, año que aún no ha concluido.

La diferencia toca techo en torno a los 27.5 años, situándose para la generación de 1975 un 7.4% por debajo de la proporción existente en la generación de 1970. La generación de 1980, a pesar de su enorme parecido con la de 1975, presenta una reducción en cuanto a la diferencia

de proporciones. Estas diferencias se comienzan a reducir considerablemente en torno a los 30 años hasta situarse cerca del 3%. En cinco años, tiempo que tardará la generación de 1975 en alcanzar los 40 años, conoceremos si la proporción de mujeres con hijos se mantiene o si ha habido un verdadero cambio en el número de mujeres que son madres.

Siguiendo en esta línea se podría extender el estudio anotando la fecha de nacimiento de los sucesivos hijos. De esta forma podríamos tener la proporción de mujeres con x hijos a las distintas edades. También, aprovechando dicha información, podríamos estudiar la probabilidad de ampliación de familia y su calendario en función de la edad cumplida o de la distancia temporal con el parto previo.

Otros datos que se pueden obtener con la unión de los ficheros son los relativos a la movilidad residencial tras el parto. Sobre la generación de 1980 se ha utilizado el cruce de ficheros para estudiar el tiempo transcurrido desde el parto hasta la inscripción un cambio de domicilio. Un 54% de las mujeres de esta generación que han tenido un niño han formalizado un cambio de residencia tras el parto. La distribución de los tiempos se observa en el siguiente gráfico.

Cambios de domicilio tras el parto

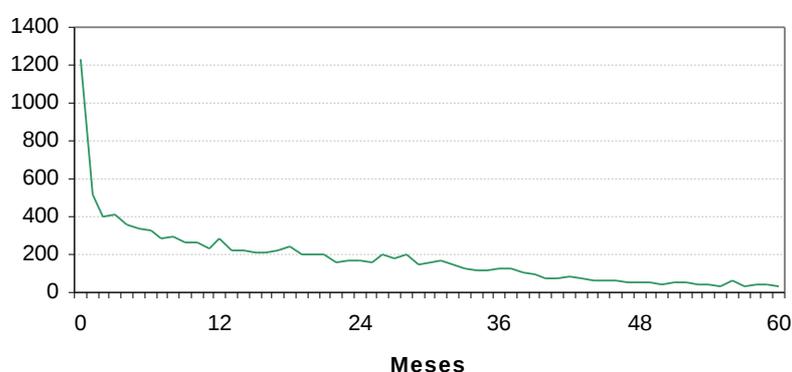


Gráfico 15. Meses transcurridos desde el parto hasta el cambio de residencia para mujeres de 1980.

Se ve que la mayoría de estos cambios se producen a los pocos meses, probablemente por necesidades legales.

b. Mortalidad infantil

En colaboración con el Instituto de Economía, Geografía y Demografía del CSIC, se ha llevado a cabo una experiencia sobre mortalidad infantil con los datos de la BDLPA. Al establecer el enlace de registros centralizado en el padrón, podemos determinar una relación entre el boletín de parto y el boletín de defunción. Esta relación es especialmente útil en la mortalidad asociada a los primeros meses de vida, ya que dicha mortalidad está muy relacionada con parámetros anómalos en el parto. El enlace permite estudiar la interrelación entre los datos del parto y de la defunción.

c. Otros estudios de mortalidad

En otra ponencia presentada a este mismo Congreso⁷, se realiza un ejemplo de explotación de la base de datos de Andalucía. Dicha explotación versa sobre el estudio de la mortalidad empleando información de RPA, MNP y Censo de 2001, permitiendo obtener datos tan novedosos como las tasas de mortalidad según nivel de estudios.

⁷ Estimación longitudinal de la mortalidad de Andalucía. Un ejemplo de explotación de la Base de Datos Longitudinal de Población de Andalucía (BDLPA). F. Viciano, V. Montañés, R. Cánovas y E. Poza

6. Conclusiones

En España existen dos principales sistemas de información demográfica que ofrecen resultados diferentes e independientes: el padrón y el MNP. Del primero se obtienen las cifras de población y las variaciones residenciales y, a pesar de que al padrón llegan las bajas por defunción y las altas por nacimiento, no se ofrecen datos sobre ese extremo.

El MNP por su parte, proporciona otra visión parcial ofreciendo las cifras de defunciones y nacimientos, e incluso información sobre los progenitores.

Sin embargo, no existe un nexo entre ambos ficheros, al menos como fuente estadística. El enlace de ambos registros pone en consonancia la información relativa a los dos ficheros, permitiendo asegurar que los eventos de MNP les suceden a personas del REPA.

Tal y como se expresan las tasas en demografía, nos permite calcular tasas donde el numerador y el denominador pertenezcan a ciencia cierta a la misma población. Esta consonancia permite combinar la información de manera que se puedan obtener productos y resultados estadísticos que, o bien son novedosos, o bien sólo podrían obtenerse con la realización de un censo.

Hasta ahora el aprovechamiento de esta información demográfica combinada no se ha utilizado en España. Al menos en Andalucía podrá hacerse en breve a través de la Base de Datos Longitudinal de Población de Andalucía (BDLPA), la cual abre una nueva vía de estudios que pueden ser de gran interés.