

Seminario Internacional: El papel de los registros administrativos en relación con el análisis social y económico y el desarrollo del Sistema Estadístico Nacional

Comunicación

***La utilización de registros administrativos como base
para la investigación de políticas sociales***

Luis Ayala Cañón¹

Universidad Rey Juan Carlos / Instituto de Estudios Fiscales

Magdalena Rodríguez Coma²

Instituto de Estudios Fiscales

Instituto de Estudios Fiscales
Madrid 30 y 31 de enero de 2006

¹ *Instituto de Estudios Fiscales, Cardenal Herrera Oria, 378, 28035 Madrid, luis.ayala@ief.meh.es.*

² *Instituto de Estudios Fiscales, Cardenal Herrera Oria, 378, 28035 Madrid, magdalena.rodriguez@ief.meh.es.*

Contenido

- 1) Introducción
- 2) Ventajas e inconvenientes de los datos administrativos
- 3) Métodos y estrategias para la utilización de registros administrativos en investigación
 - 3.1. Evaluación de calidad y depuración de datos
 - 3.2. Creación de ficheros longitudinales
 - 3.3. Enlace de registros de datos administrativos
 - 3.4. Método probabilístico y determinístico de enlace de registros
 - 3.5. Estandarización y depuración de datos al enlazar registros
 - 3.6. Precisión del enlace de registros
 - 3.7. Algunas recomendaciones
- 4) Aplicación práctica al caso español: BADIMI, base de datos de investigación creada a partir de los registros administrativos del IMI
 - 4.1. Características de los registros administrativos del IMI
 - 4.2. Elaboración de la base de datos de investigación: Objetivos, selección de variables, control de calidad, enlace de registros, estandarización
 - 4.3. Proyectos de investigación realizados con la base de datos BADIMI
- 5) Conclusiones
- 6) Anexos
- 7) Bibliografía

1. INTRODUCCIÓN

La investigación de las políticas sociales necesita fuentes de datos complementarias a las que proporcionan las estadísticas oficiales. En este contexto, juegan un papel primordial los registros administrativos que genera el proceso de gestión, en muchos casos descentralizada, de los diferentes programas sociales. En programas con cierta trayectoria, la posibilidad de explotar estos registros permite disponer de bases de datos longitudinales, que, gracias a la existencia de un conjunto de técnicas analíticas cada vez más sofisticadas aplicables a microdatos, hacen posible abordar aspectos dinámicos en el análisis. La econometría como rama de la economía que une la teoría económica con los métodos estadísticos para interpretar los datos económicos y diseñar y evaluar políticas sociales, ha experimentado una evolución muy fuerte en los últimos treinta años, impulsada por la necesidad de analizar las nuevas fuentes de microdatos elaboradas en las últimas décadas¹. En el Estado moderno, hay una importante demanda de información para caracterizar, solucionar y averiguar las causas de los problemas sociales, y de evaluación de los objetivos de los programas dirigidos a grupos específicos. En esta tarea, los microdatos juegan un papel decisivo como base informativa y como impulso para el cambio de las metodologías econométricas tradicionales. Problemas que parecen ser poco importantes examinando medidas agregadas se convierten en centrales cuando se analizan microdatos. El nacimiento de la microeconometría moderna va ligado a la unión de la economía y la estadística con objeto de lograr resultados interpretables de los microdatos, contrastar teorías utilizando datos individuales y construir *contrafactuales* o grupos de control válidos para la evaluación de los resultados de las políticas públicas. Las bases de datos microeconómicas reflejan la diversidad y heterogeneidad del funcionamiento de la economía y permiten a los economistas analizar la eficacia de la intervención pública.

El uso de registros administrativos en la investigación aplicada en economía es muy desigual por ramas de conocimiento. En el campo de las políticas sociales, existe una limitada tradición, aunque avivada recientemente gracias a un esfuerzo notable por convertir esos registros en bases de datos para el estudio del funcionamiento y los efectos de los programas. El trabajo que aquí se presenta pretende difundir una experiencia de conversión de los

¹ Una completa síntesis de las contribuciones de la microeconometría al conocimiento económico y en particular a la evaluación científica de las políticas públicas puede verse en Heckman (2001).

registros administrativos generados por el programa IMI de la Comunidad de Madrid, iniciado a finales de 1990, en una base de microdatos longitudinal apta para investigar la dinámica de las prestaciones asistenciales y evaluar las actuaciones que combinan la ayuda a la reinserción laboral con capacitación y subsidios. Se incluye una referencia a los proyectos realizados a partir de ella, así como a las investigaciones en curso.

El trabajo se estructura como sigue. Tras esta introducción, en un segundo apartado se sintetizan las ventajas e inconvenientes de los datos administrativos. A continuación se establecen los métodos y estrategias para la utilización de datos administrativos en la investigación, contemplando la evaluación de calidad y depuración, la creación de ficheros longitudinales, el enlace de registros de datos administrativos con otras fuentes administrativas o procedentes de encuestas, sus tipos y la precisión de los mismos. El siguiente apartado se dedica a describir la aplicación práctica de la base de datos de investigación BADIMI a partir de los registros administrativos de IMI, incluye los objetivos del proyecto, la selección de variables, el control de calidad, el enlace de registros, la estandarización y el enlace con otras fuentes. A continuación se describen de forma sintética los proyectos de investigación realizados hasta el momento a partir de BADIMI y una breve referencia a los resultados obtenidos. El trabajo finaliza con una relación de las principales conclusiones.

2. VENTAJAS E INCONVENIENTES DE LOS DATOS ADMINISTRATIVOS

La utilización estadística de los registros administrativos es una práctica creciente en los países de desarrollo económico elevado. Probablemente, el país de Europa con mayor tradición de aprovechamiento de los registros administrativos como base estadística es Suecia, que ya en el siglo XVIII elaboró la estadística oficial de la población a partir del registro demográfico de cada parroquia hasta llegar al agregado nacional. Este sistema ha sido ampliamente utilizado para todo tipo de estadísticas sobre la población y características demográficas hasta la actualidad².

En el caso de España no existe esta cultura y es muy reciente y reducida, todavía, la

² Un estudio detallado sobre el control de calidad en estadísticas procedentes de registros y archivos administrativos en Suecia puede verse en Petterson (1992)

importancia de los registros administrativos como fuente de datos, aunque la progresiva disminución de los costes en el proceso automático de la información permite acceder a ellos mediante herramientas analíticas modernas y eficientes. Por una parte, la medición de cualquier macromagnitud en la sociedad actual es bastante más compleja de lo que parece y el sistema estadístico nacional debería aprovechar la gran cantidad de información disponible en la Administración Pública para mejorar la medición de las principales variables macroeconómicas. Por otra parte, las ventajas de su utilización son bastante evidentes. Si los ficheros ya existen, aunque hayan sido creados para otro propósito, como pueda ser la gestión, no parece necesario volver a obtener la misma información con fines estadísticos, de estudio o investigación. El ahorro de costes puede ser considerable, incluyendo los costes indirectos de sobrecarga de peticiones a los informantes, que precisa grandes inversiones en coordinación institucional. La gran riqueza informativa contenida en los ficheros administrativos, consecuencia de la gran cantidad de datos que los ciudadanos están obligados a facilitar a los poderes públicos a lo largo de su vida, representa un activo de conocimiento de la realidad susceptible de ser aprovechado en beneficio de la propia ciudadanía, con las garantías necesarias de protección del derecho a la intimidad y confidencialidad.

Existe una gran diferencia entre los datos de encuestas y los que proceden de registros administrativos. El diseño de un sistema de registro administrativo no tiene en cuenta los posibles usos como base de datos, sino la adecuación al proceso de control, gestión y cumplimiento de la normativa que lo define. Otra característica destacable es que los sistemas administrativos son esencialmente dinámicos y todos los cambios producidos a lo largo del tiempo en cuanto a cobertura, definiciones, contenido o precisión, deben ser evaluados con rigor si se trata de utilizarlos como fuente estadística. Entre las muchas ventajas adicionales a las ya mencionadas deben tenerse en consideración las siguientes:

- 1) Los datos administrativos suelen proporcionar medidas muy detalladas y precisas de las variables y resultados que contempla el proceso (en nuestro caso puede consultarse en el anexo I la relación de las principales variables y categorías de los ficheros del IMI)
- 2) Su gran tamaño y, muy a menudo, la cobertura de universos completos permiten tipos de análisis imposibles de realizar con otras fuentes (en nuestro caso contamos

contada la historia de la prestación desde el origen).

3) En algunos casos, como el que presentamos en este trabajo, la descentralización de los programas permite contar con la información completa al nivel de Comunidad Autónoma

4) Ofrecen la posibilidad de seguir a los mismos individuos a lo largo de un periodo amplio. Los datos administrativos suelen ser longitudinales y si no lo fueren es sencillo obtenerlos mediante cruces en el tiempo (en nuestro caso, en la construcción de la historia completa de la participación del hogar receptor de prestaciones del IMI ha sido posible llegar a identificar los periodos de presencia y ausencia del programa a lo largo de un período superior a 10 años)

5) Permiten tener información del mismo programa durante un intervalo temporal suficientemente prolongado para poder evaluar los cambios o reformas del mismo.

6) Existe, además, la posibilidad de enriquecer la base de micro-datos creada a partir del proceso administrativo mediante el cruce con otras fuentes procedentes de encuestas, estadísticas oficiales u otros registros administrativos (en nuestro caso, hemos añadido datos de paro dinámico a la historia de cada hogar para estudiar la influencia de la situación económica en los posibles procesos de duración en el programa. En este momento la disponibilidad de una fuente de datos adicional, una encuesta realizada a más de 2200 hogares tras su paso por el programa IMI, va a hacer posible la mezcla con la historia completa de cada hogar, permitiendo realizar evaluaciones de estrategias combinadas de intervención pública, utilizando como indicador de resultados un vector multidimensional que refleje si las posibles mejoras son compatibles o incompatibles entre sí).

3. MÉTODOS Y ESTRATEGIAS PARA LA UTILIZACIÓN DE REGISTROS ADMINISTRATIVOS EN INVESTIGACIÓN SOCIAL

Quando se hace referencia a los datos administrativos para su utilización en la investigación y evaluación de los programas sociales, fundamentalmente se está hablando de datos obtenidos de los sistemas de información para la gestión de una gran variedad de

programas de salud pública y servicios sociales³. Estos datos administrativos deben ser depurados antes de poder ser utilizados para la investigación, ya que existen numerosas fuentes potenciales de errores y los formatos no suelen ser adecuados para el análisis. Esta depuración incluye el proceso completo de transformación de los registros tal cual están en el sistema de información para la gestión en un fichero analítico. Entre las diferentes tareas necesarias para dichas transformación, el enlace de registros es una de las actividades esenciales, especialmente cuando la investigación requiera análisis longitudinal, necesidad siempre presente en la evaluación de programas. El enlace de registros es el proceso por el que se determina que dos registros pertenecen al mismo individuo o unidad informante. A continuación se presentan los métodos y estrategias necesarias para realizar con éxito todo el proceso de transformación de un sistema de información diseñado para la gestión en una base longitudinal de microdatos válida para la investigación de políticas sociales y la evaluación de programas sociales.

3.1 Evaluación de la calidad y depuración de los datos

Los ficheros de registros administrativos a los que pueden tener acceso los investigadores son, con frecuencia, una selección de campos del sistema de información para la gestión. El conjunto de datos suele referirse a un período de tiempo que se corresponde con el período administrativo natural, como puede ser el año o mes fiscal, el plazo establecido para el control de cumplimiento de los requisitos administrativos para la obtención de la prestación, etc. Esta aproximación temporal produce varios conjuntos de datos de sección cruzada, que no es la ideal para propósitos de investigación y evaluación.

Lo normal es que estos ficheros tengan una calidad muy diferente en cada uno de los campos y en función de los objetivos del estudio se precisa una intensa depuración adicional para lograr sistematizar el conjunto de cuestiones relevantes. Un fichero administrativo puede ser utilizado directamente para contestar algunas preguntas, mientras que puede ser casi imposible o necesitar una transformación profunda la contestación a otras. Debe tomarse en consideración toda la información disponible sobre el proceso administrativo. También puede ser interesante evaluar fuentes alternativas para contestar esas preguntas. En muchos casos el coste que supone la depuración está justificado si es la única posibilidad de contar con esos

³ Véase Goerge, R. M. and Joo Lee, B (2002)

datos. La depuración de los ficheros administrativos suele proporcionar un aprendizaje muy completo sobre las fuentes y el mantenimiento de una base de datos.

Para garantizar la fiabilidad de los datos administrativos el investigador necesita un conocimiento suficiente de cuestiones como la motivación para obtener los datos, las exigencias legales, el control de las prestaciones, las garantías de cumplimiento, etc., que pueden exigir una información muy precisa y con gran rigor conceptual en muchos de los campos del fichero administrativo. También es interesante saber si existe algún sistema de control de calidad de los datos, si la grabación es directa, a través de formularios en papel, con diseño automático de validación de campos individual y cruzada en entrada que impida la existencia de datos imposibles, o a través de otros sistemas que garanticen la calidad desde el momento inicial. Conviene conocer, además, los estudios y análisis realizados con anterioridad con los mismos datos en el caso de que existan, que pueden ayudar a explicar errores aparentes, situaciones atípicas y anomalías, aunque no eviten la necesidad de realizar controles de calidad. Por último, es importante averiguar si alguno de los campos de datos de interés es crítico para los objetivos del programa, si tiene un impacto grande desde el punto de vista de las consecuencias que se derivan para el informante o beneficiario y las exigencias de veracidad del dato.

3.2. Creación de ficheros longitudinales

La preferencia por ficheros longitudinales y de datos de tiempo continuo en la mayor parte de los proyectos de investigación y evaluación de programas⁹ exige transformaciones importantes en la estructura de los ficheros de datos administrativos que suelen ser conjuntos de datos de sección cruzada. En el caso del IMI, programa dirigido a lograr un nivel básico de protección económica y al desarrollo de medidas para favorecer la inserción socioeconómica de individuos y hogares en situación de exclusión social, el sistema de información administrativa para la gestión consta esencialmente de un fichero de hogar y otro de los individuos que lo forman, con controles semestrales de posibles cambios y del cumplimiento de requisitos para continuar en el programa. La construcción de una base de datos longitudinal hace necesario reconstruir la historia cronológica del hogar mientras se es beneficiario del programa, recogiendo las características dinámicas y teniendo en cuenta que los objetivos

prioritarios del estudio se basan en establecer las duraciones, conocer los factores que las explican, la posible dependencia del programa, los casos de episodios múltiples, la duración de los períodos de independencia en el caso de beneficiarios que se reincorporan, los tipos de salida y, por último, evaluar las intervenciones que favorecen una mayor independencia del programa. Para medir el tiempo ha sido necesario buscar la consistencia de dos fuentes existentes en los datos, de las cuales la más fiable era la fecha de percepción de la prestación, por los motivos apuntados en el apartado anterior, y a la que se ha dado pretensión de veracidad en caso de discrepancia. Como consecuencia de ello ha habido que realizar algunas imputaciones en los datos.

3.3. Enlace de registros administrativos

El enlace de registros de datos administrativos permite adecuar el tamaño de la muestra de una forma coste-efectiva en el caso de eventos poco frecuentes. También es muy ventajoso cuando el interés de la investigación se centra en una información concreta. En el caso de las investigaciones realizadas con los datos del IMI para estudiar los patrones de recurrencia en el programa, el enlace de registros en el tiempo ha sido un método excelente para construir los ficheros longitudinales que nos han permitido investigar sobre los episodios múltiples.

Debe tenerse en cuenta que las aplicaciones informáticas desarrolladas para el enlace pueden rentabilizarse mejor si en lugar de utilizar muestras se aplican a los datos poblacionales. El coste adicional del proceso para enlazar los ficheros completos en lugar de muestras, en la mayoría de los casos en que no existan restricciones de capacidad de proceso, está plenamente justificado por las grandes ventajas adicionales que proporciona. El enlace de registros individuales aplicado a datos poblacionales permite reconstruir las muestras representativas “a posteriori”, lo que no es factible cuando se realiza el enlace con muestras.

Existen diferentes aplicaciones de enlace de registros para la investigación:

- 1) Enlazar registros individuales en un mismo sistema de información a lo largo del tiempo. Ésta ha sido la tarea realizada con los datos del IMI para investigar los episodios múltiples de participación en el programa y es el tipo de aplicación más

común. Permite aprovechar la información histórica de los ficheros de datos administrativos para analizar distintos tipos de cuestiones de naturaleza longitudinal. Existe un identificador común que permite el enlace, que en nuestro caso fue el código de expediente.

- 2) Enlazar datos de diferentes sistemas de información administrativa entre áreas de servicios comunes. Son frecuentes las situaciones en las que los sistemas de información no tienen un identificador común y en las que debe utilizarse un tercer identificador común.
- 3) Enlazar datos de encuestas con ficheros de datos administrativos, cuando la muestra se ha extraído de los segundos. Es el caso de la fuente de información con la que se está trabajando en el proyecto de evaluación de estrategias combinadas de intervención pública con datos del IMI y los procedentes de una encuesta realizada a más de 2200 hogares tras su paso por el programa. Este tipo de enlace se realiza cuando una muestra de individuos registrados en un fichero administrativo se utiliza como estudio poblacional. Los investigadores utilizan métodos de encuesta para intentar conseguir información que no suele estar disponible en los ficheros administrativos. Este tipo de aplicación a veces no se considera como aplicación de enlace de registros aunque cumple con todas las características y exigencias del mismo.
- 4) Enlazar datos de encuestas con ficheros de datos administrativos cuando las muestras son independientes de éstos. Es el caso, en nuestras investigaciones, del enlace con la Encuesta de Población Activa (INE) para medir el paro, que se ha realizado utilizando como variable de enlace las fechas de entrada en el programa IMI y las fechas de cada semestre sucesivos en el mismo para cada hogar. El objetivo era investigar la influencia de los desequilibrios del mercado de trabajo, de forma dinámica, en el perfil de las duraciones en el programa IMI. En este tipo de enlace, los investigadores complementan la información recogida a través de encuestas con la información detallada existente en el sistema administrativo. Este enlace se realiza una vez finalizada la encuesta. Al ser una muestra independiente de los datos administrativos, no hay un identificador común. Es uno de los supuestos en los que existen riesgos de violar la confidencialidad de los datos al buscar identificadores comunes para realizar el enlace, problema que no se ha planteado en nuestro caso.

3.4. Métodos probabilístico y determinístico de enlace de registros

El método determinístico de enlace compara un identificador o un grupo de identificadores entre bases de datos. El enlace se realiza si coinciden plenamente. En el caso del IMI, por ejemplo, se ha utilizado el número de expediente del hogar como identificador de comparación. El número de expediente tiene un código bidimensional, en el que la primera parte identifica el hogar y la segunda el número de episodio que el hogar experimenta como beneficiario del IMI. Es decir, 01 la primera vez, 02 si vuelve al IMI por segunda vez después de haber salido y así sucesivamente

En la investigación suelen utilizarse combinaciones de diferentes partes de la información de identificación, con el objetivo de lograr enlaces de registros válidos. Suele hacerse en los casos en los que no se dispone de un identificador con alto grado de poder discriminante.

Lo que diferencia el método determinístico de enlace de registros es que cuando dos registros coinciden en un campo en particular, no hay información sobre si esta coincidencia aumenta o disminuye la probabilidad de que pertenezcan al mismo individuo. Que coincidan, por ejemplo, dos apellidos muy corrientes no es lo mismo que la coincidencia de apellidos muy poco frecuentes a efectos de haber enlazado verazmente, pero el método determinístico no lo tiene en cuenta.

Debido a las limitaciones del sistema de enlace determinístico se han desarrollado un conjunto de métodos probabilísticos de enlace de registros, que se basan en asumir que una simple coincidencia entre variables comunes a las fuentes de datos no identifica con certeza al mismo individuo. En lugar de ello, estos métodos calculan la probabilidad de que dos registros pertenezcan a la misma unidad de identificación, utilizando varias partes de información. Una presentación del proceso matemático en detalle del método de enlace probabilístico puede verse en Newcombe et al. (1959), Fellegi y Sunter (1969) y Winkler (1988). Los desarrollos más recientes en este terreno intentan aprovechar la velocidad y bajo coste de la automatización para asociar a un enlace determinístico una identificación con un nivel de probabilidad de ser un enlace cierto (Jaro, 1989, Winkler, 1993, 1994 y 1999).

3.5. Estandarización y depuración de datos al enlazar registros

En función del método de enlace determinístico que se utilice, los errores de grabación, tipográficos u otros errores de transmisión de datos pueden causar muchos problemas. La depuración de los datos en el proceso de enlace de registros debe incluir, a menudo, la utilización de valores consistentes en la identificación geográfica utilizada para el enlace, la división de los valores de las variables en componentes que necesiten ser comparados, el estudio de los errores tipográficos y la eliminación de los duplicados de cada fichero fuente para realizar el enlace.

En general, los datos proceden de fuentes diferentes por lo que es necesario estandarizar el formato y los valores de cada una de las variables que vayan a ser utilizadas para realizar el enlace. Esto resulta especialmente importante en el caso de las fechas y los esquemas de códigos, información geográfica y deletreo de nombres y apellidos. A partir de los años ochenta ha habido grandes avances en la investigación en el campo del enlace de registros. En concreto, en el desarrollo de algoritmos para comparar campos carácter a carácter⁴.

La eliminación de duplicados en cada fuente de datos disponible, previa al enlace de registros, es de importancia capital. Debe hacerse a través del auto-enlace dentro del propio fichero para llegar a la identificación de un único registro. Únicamente cuando esté asegurado que no existen duplicados en cada uno de los ficheros que se van a enlazar, será posible acometer el enlace.

3.6. Precisión del enlace de registros

La fase final del proceso de enlace consiste en conocer el grado de validez y precisión del enlace realizado. Sea cual sea la técnica utilizada, probabilística o determinística, hay que adoptar una hipótesis de cara a la decisión de si los registros enlazados pertenecen o no al mismo individuo o unidad de información. La decisión puede ser equivocada, ya que se trata

⁴ Puede verse, entre otros, Jaro (1985, 1989), Winkler (1990), Winkler and Thibaudeau (1991)

de un supuesto y los errores cometidos al adoptarlo se visualizarían como falsos-positivos y falsos-negativos. El error falso-positivo ocurre cuando se realiza el enlace entre dos registros que en realidad no pertenecen al mismo individuo. Este tipo de error es comparable al error de Tipo I en el contraste de hipótesis estadísticas. El error falso-negativo ocurre cuando no se hace el enlace entre dos registros y en realidad pertenecen al mismo individuo o unidad de información. Este tipo de error es comparable al error de Tipo II en el contraste de hipótesis estadísticas. De la misma forma que en los errores de Tipo I y II, la probabilidad de hacer un falso positivo es mucho más sencilla de averiguar en el proceso de enlace que la probabilidad de cometer un error falso-negativo⁵. En el caso de enlace no probabilística, un chequeo a posteriori de los registros emparejados puede dar una idea aproximada del error en falsos-positivos. La estimación del error de falsos negativos es más compleja en este caso.

Un problema adicional es que el intento de reducir un tipo de error hace que se incremente el otro tipo. Reglas muy rígidas para aceptar un tipo de error hacen muy laxas las del otro, por lo que hay que llegar a una solución de compromiso en función de objetivos concretos. En la práctica, puede ser aconsejable considerar las tasas de error de falsos-positivos y falsos-negativos como una medida para comparar diferentes métodos de enlace de registros.

3.7. Algunas recomendaciones

La utilización de registros administrativos como fuente de datos para la investigación requiere la realización de algunos esfuerzos en la depuración, que deben incluir como actividades necesarias el examen de la consistencia interna de los datos, el conocimiento de cómo han sido obtenidos, cómo han sido procesados y mantenidos antes de que le hayan sido entregados al investigador y la comparación con otras fuentes de datos procedentes de encuestas u otros registros administrativos mediante el enlace de registros. Además, se debe estudiar con detenimiento cómo funciona el programa que ha generado los datos para comprender las causas de las posibles inconsistencias y ver la forma de solucionarlas.

⁵ Puede verse un método de estimación de errores en el proceso de enlace de registros probabilístico en Berlin and Rubin (1995).

4. APLICACIÓN PRÁCTICA AL CASO ESPAÑOL: BADIMI, BASE DE DATOS DE INVESTIGACIÓN CREADA A PARTIR DE LOS REGISTROS ADMINISTRATIVOS DEL IMI

La fuente de datos utilizada en este estudio son los registros administrativos de los beneficiarios del programa de renta mínima de la Comunidad de Madrid. A finales de los años ochenta comenzó a desarrollarse en España una red de sistemas de asistencia social completamente descentralizada. La ausencia de un programa nacional de cobertura de las situaciones generales de necesidad económica llevó a algunas regiones a desarrollar sus propios sistemas de prestaciones asistenciales. La región madrileña fue una de las pioneras en el establecimiento de un programa de rentas mínimas, con la puesta en marcha del Ingreso Madrileño de Integración (IMI) en el último trimestre de 1990. La creación de este programa supuso una decisión novedosa en la región, dada la ausencia previa de planes globales en este campo de la intervención pública. Los ejes sobre los que durante estos años ha pivotado el IMI han sido, fundamentalmente, dos: el establecimiento de un nivel básico de protección económica y el desarrollo de medidas para favorecer la inserción socioeconómica de individuos y hogares en situación de exclusión social.

4.1. Características de los registros administrativos del IMI

El tiempo transcurrido desde la puesta en marcha del programa –más de tres lustros– y la disponibilidad de registros administrativos sobre el proceso de participación y las características socioeconómicas de los hogares beneficiarios, hacen del IMI una fuente especialmente útil para el estudio de la dinámica de los programas de lucha contra la pobreza. El tratamiento informático dado a estos registros por la Dirección General de Servicios Sociales de la Comunidad de Madrid (CAM) permite contar con información sobre más de 50.000 episodios de participación en el programa, con una información muy completa sobre las características particulares de cada hogar⁶. Buena parte de las variables socioeconómicas que contiene la base de datos coinciden con las señaladas por diversos estudios realizados sobre las características óptimas de las fuentes de información para el análisis de las poblaciones

⁶ En los distintos trabajos que hemos realizado con estos registros utilizamos la información disponible para el período comprendido entre 1990 y 2001. Justo después de esa fecha, el programa registró una importante reforma, con una ruptura importante en el método previo de recogida y sistematización de la información.

participantes en programas asistenciales⁷.

Los ficheros originales proporcionados por la CAM son cuatro, con una estructura relativamente parecida a los habituales en las encuestas de presupuestos familiares, que suelen dividir la información en un fichero del hogar y otro de individuos presentes en la muestra. Los dos primeros ficheros del IMI ofrecen información general sobre el hogar y cada uno de sus miembros en el momento de tramitación de la solicitud (Anexo 1).

Se trataría, por tanto, del equivalente a “una foto a la entrada”, con variables como edad, parentesco, ocupación, sexo, nivel de estudios, estado civil, recursos económicos de la familia, situación de la vivienda o las iniciativas desarrolladas por los trabajadores sociales en cada hogar para favorecer su inserción socioeconómica. Además de esas variables, se incluyen otras en una hoja de seguimiento que elaboran los profesionales de los centros de servicios sociales cada semestre. Dichas variables recogen posibles cambios en la situación de la vivienda, en la empleabilidad de los beneficiarios, en los diversos problemas sociales que afectan a estos hogares y en el nivel del trabajo de integración.

4.2. Elaboración de la base de datos de investigación: objetivos, selección de variables, control de calidad, enlace de registros y estandarización

Al iniciar los trabajos se definió como objetivo fundamental la elaboración de una base de datos de perceptores de prestaciones del IMI apta para la investigación de los siguientes temas:

- 1) Estudio de los aspectos dinámicos de los programas asistenciales. Entradas y salidas en la pobreza y en los programas de protección social. Factores que las determinan y relaciones entre ellos.
- 2) Estudio de los procesos de cronicidad y dependencia de los programas. Posibles efectos de la heterogeneidad inobservada. Tipologías de beneficiarios
- 3) Estudio de las pautas de abandono de los programas. Tipos de salida de los programas asistenciales. Modelos con salidas alternativas.

⁷ Estas características se apuntan en Manieri y Dazinguer (2001) y Goerge y Boo Lee (2001)

- 4) Estudio de las intermitencias en los programas. Episodios múltiples de participación. Características de los beneficiarios que se reincorporan. Estudio de los perfiles del tiempo de independencia del programa en los beneficiarios con mas de un episodio de participación.
- 5) Evaluación de los denominados “proyectos de inserción”, programas que se diseñan con objeto de facilitar la reinserción laboral de los beneficiarios del IMI combinada con capacitación y subsidios

Las metodologías de los modelos de duración y análisis de supervivencia en las que se apoyan estos estudios, bien se apliquen en tiempo continuo o discreto, exigen definir de forma precisa los tiempos de permanencia en el programa o lo que es lo mismo elaborar una base longitudinal de microdatos de hogares. En primer lugar se creó el fichero BADIMI1, que consta de 138.009 registros, obtenidos de la fusión de los ficheros administrativos IMI93 y SEGUI. En ellos se incluye la información administrativa correspondiente a la apertura del expediente en el momento de solicitud de la prestación (IMI93) y los datos que se rellenan semestralmente en el caso de que el expediente siga “vivo” por concesión de la prestación IMI, y que permiten seguir la evolución del mismo a lo largo de 10 años (SEGUI).

Para obtener el fichero BADIMI1, primera versión de la base de datos con fines de investigación, se realizó un estudio previo del contenido de los ficheros administrativos IMI93 y SEGUI, tratando de determinar cuáles eran las variables relevantes a efectos de investigación y valorar la calidad de las mismas⁸. A partir de aquí se seleccionaron y depuraron las variables originales, se crearon variables de control y se transformaron otras, con el fin de lograr una información mas adecuada a nuestros objetivos. También se realizó una validación de campos cruzada, que permitió detectar valores imposibles ó erróneos que no se ponen de manifiesto cuando no se cruzan los datos.

La variable fundamental, que identifica el registro es el número de expediente del hogar. BADIMI1 está ordenado por este campo. Según los registros originales, el número de expediente puede estar repetido hasta un máximo de 17 veces (en teoría podrían llegar hasta

⁸ Puede verse un diagrama con el esquema inicial de transformación de los ficheros administrativos al final del documento

veinte), en función de su permanencia en el fichero SEGUI, que no tiene porqué ser continua sino que puede haber entradas y salidas del mismo. Tras la depuración se han considerado válidos 113.770 registros del fichero SEGUI y 60.864 del IMI93. La diferencia de 24.239 registros que supone comparar con los 138.009 de BADIMI1, son los expedientes que, estando en IMI93 no están en SEGUI. Es decir, tras la apertura del expediente no han vuelto a estar en el programa. La inversa no es posible, al no poder existir ningún expediente en SEGUI que no estuviese previamente en IMI93. Éste ha sido, precisamente, uno de los criterios de depuración. Se han incluido tres variables de control con el objeto de facilitar estos cálculos. El campo COUNT indica el número de veces que el expediente aparece en BADIMI1. También se han añadido etiquetas a todas las variables con el objetivo de facilitar la interpretación de los datos.

La elaboración de un fichero idóneo para el análisis de la duración, exigió definir la entrada en el programa, la salida del mismo y los semestres de estancia en el IMI, con arreglo a la información disponible en los registros administrativos existentes. Se hizo necesario cruzar dos fuentes y llegar a la coherencia entre ellas a través de una serie de imputaciones. Después de la eliminación de inconsistencias en las fechas registradas de posible entrada y salida (valores imposibles), se consideró como fecha de entrada en el programa el primer momento de cobro de la prestación y como fecha de salida aquella en que se rellenó por última vez la hoja de seguimiento.

Definida una secuencia cronológica de 23 semestres (desde el segundo semestre de 1990, que sería el semestre 1, hasta el segundo semestre de 2001, que haría el semestre 23) y realizado un cambio de estructura en el fichero de seguimiento para lograr asociar a cada expediente el número de hojas de seguimiento cumplimentadas, se procedió al cruce de fuentes para buscar las posibles inconsistencias entre fechas de entrada y salida y el número de semestres de estancia o, lo que es lo mismo, el número de hojas de seguimiento rellenas⁹.

En los ficheros de seguimiento existe la exigencia de rellenar una nueva hoja a un hogar previamente a la tramitación del cobro del período, lo que prueba de forma clara la alta calidad

⁹ Para ayudar a detectar las inconsistencias en las fechas de entrada y salida se diseñó una tabla que puede consultarse a final del documento

de esta fuente de información para recoger correctamente la duración de la estancia en el programa¹⁰. El número de semestres de permanencia en el programa, o número total de hojas de seguimiento rellenas, debería ser el mismo que el que puede deducirse a partir de las fechas de entrada y salida del mismo. No obstante, al cruzar ambas fuentes se encontraron bastantes casos en los que no existía tal correspondencia y se realizaron imputaciones de hojas de seguimiento o de semestre de entrada/salida siguiendo los criterios que se exponen mas adelante.

A partir de BADIMI1 fue necesario adoptar diferentes decisiones para elaborar criterios de imputación para el estudio de la duración en el programa. Las tareas realizadas incluyeron:

- 1) Conversión de hojas de seguimiento en número de semestre según una cronología fija
- 2) Definición de las reglas de comprobación de errores
- 3) Definición de las reglas de imputación
- 4) Definición de los valores de la variable “calidad” en función del tipo de imputación realizada

Cada expediente de la base de datos para el análisis de la duración en el programa lleva asociada una variable denominada “calidad” que toma valores de 1 a 6 con arreglo a las imputaciones realizadas. El valor 1, por ejemplo, indica coincidencia plena entre el número de semestres de permanencia en el programa calculado como la diferencia entre las fechas de salida y entrada y el número de hojas de seguimiento obtenido del fichero administrativo de seguimiento como fuente alternativa.

Al estudiar las discrepancias, el mayor volumen de casos se daba en expedientes en los que no aparecía la primera hoja de seguimiento pero sí las siguientes. En estos casos se decidió imputar al hogar correspondiente como hojas cumplimentadas por los trabajadores sociales todas las comprendidas desde el semestre de entrada hasta la primera hoja de seguimiento que aparecía en los registros. La mayoría de los casos son anteriores al segundo semestre de 1993, fecha en que los responsables de los registros administrativos del IMI introdujeron algunos

¹⁰ Véase el apartado 2.2

cambios en la recogida de información y en la codificación de las variables, lo que explica las discordancias. Los expedientes a los que hemos aplicado este criterio de imputación tienen el valor 2 en la variable de calidad. El resto de los casos posibles son bastante menos numerosos, razones de espacio impiden comentarlos todos.

Para el estudio de los modelos de supervivencia se crearon dos ficheros (MODIMI1 y HOJASIMI). Son ficheros que contienen un registro para cada expediente con la serie de 22 variables registradas cada semestre y toman valor 0 y 1 que define la ausencia / presencia del hogar en el período considerado. Se siguieron los siguientes criterios:

a) El fichero MODIMI1 se generó a partir de la base de datos BADIMI1 con arreglo a las siguientes pautas:

- 1) Eliminación de inconsistencias en fechas (conversión en valores missing)
- 2) Consideración de la fecha de entrada como primera fecha de cobro
- 3) Consideración como fecha de salida de la última fecha en que se rellena la hoja de seguimiento
- 4) Cálculo de los semestres de entrada y salida en el programa con arreglo a la siguiente cronología:

Semestre1 → de julio a diciembre de 1990

Semestre2 → de enero a junio de 1991

Semestre3 → de julio a diciembre de 1991

.....

Semestre 20 → de enero a junio de 2000

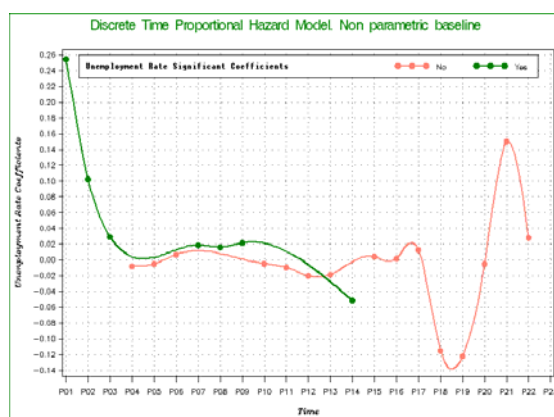
Semestre 21 → de julio a diciembre de 2000

Semestre 22 → de enero a junio de 2001

- 5) Definición de la estructura del fichero con un registro por expediente y variables dicotómicas de presencia/ausencia en cada uno de los 22 semestres considerados en el estudio.
- 6) Exclusión de los expedientes extinguidos en la versión MODIMI1. El fichero MODIMI2 incluye todos los expedientes independientemente de su situación.

b) HOJASIMI se generó a partir del fichero de datos administrativos HOJASEG , de acuerdo a las siguientes pautas:

- 1) Creación de un fichero básico que contiene las variables seleccionadas depuradas: datos de titular, estado, parentesco...
- 2) Creación de variables dicótomicas correspondientes a las 22 hojas de seguimiento, incluyendo otra variable indicativa de que se han rellenado los datos de la solicitud pero todavía no se ha rellenado la primera hoja de seguimiento
- 3) Definición de la estructura del fichero con un registro por expediente



c) Para el estudio de los episodios múltiples de participación en el programa fue necesario convertir los ficheros analíticos anteriores, una vez realizado el enlace con los datos de paro trimestral de la EPA, en un fichero denominado (IMIPAROH), con un registro para cada hogar en el que se incluyen todas las variables relevantes, además de los tiempos de presencia y ausencia en el programa en cada uno de los episodios habidos. Los estudios realizados con modelos de duración en tiempo continuo se apoyan en esta estructura del fichero, con contrapartida en un fichero de episodios (IMIPARO). Para analizar las duraciones en tiempo discreto, muy aconsejable en nuestras investigaciones debido a la forma un tanto imprecisa de medir el tiempo en el programa mediante controles semestrales, y con objeto de contrastar los resultados obtenidos con los modelos en tiempo continuo se generó otro fichero (IMIDISCRE), con tantas observaciones por hogar como semestres permanecen en el programa. Este fichero también incluye los datos de paro dinámico, para analizar en que medida los desajustes del mercado de trabajo mantienen los efectos sobre la permanencia en el

programa o se modifican con el paso del tiempo. Los efectos de intensidad decreciente y con cambio de signo pueden verse en el gráfico adjunto.

4.3. Proyectos de investigación realizados con la base de datos BADIMI

Como se ha señalado, la elaboración de los diferentes ficheros a partir de los datos administrativos del programa IMI obedece a la necesidad de contar con una base de datos adecuada para el análisis de la dinámica de participación en los programas de lucha contra la pobreza. El interés por los aspectos dinámicos de los programas de lucha contra la pobreza ha crecido considerablemente durante los últimos años. Por un lado, para un diseño más eficiente de la intervención pública parece precisa la diferenciación de las estancias de larga duración de las de carácter transitorio. Por otro, existe una creciente convicción de que los programas favorecen los comportamientos proclives a la dependencia de la intervención pública, con la reducción de la intensidad de búsqueda de empleo.

La disponibilidad de información longitudinal sobre la relación con el programa del conjunto de hogares que durante más de una década participaron en él permite dar respuesta a una variada gama de interrogantes, que han ido dando forma a distintos proyectos de investigación. Una primera fase del programa de investigación de explotación de los registros del IMI trató de dar respuesta a dos preguntas generales: Qué características socioeconómicas determinan una mayor duración en los programas? ¿Cómo cambia la probabilidad de salida al prolongarse las estancias? La contestación a la primera pregunta permitiría confirmar las diferencias en el ritmo de entrada y salida de los programas de hogares con características socioeconómicas distintas. La segunda informaría sobre la existencia o no de cadenas de dependencia motivadas por el desarrollo de la actuación pública. Mediante el desarrollo de diferentes modelos de duración, principalmente semi-paramétricos y paramétricos fue posible aislar algunas de las principales características determinantes de la mayor probabilidad de permanencia en el programa. Concretamente, los contrastes realizados coincidieron en señalar la pertenencia a minorías étnicas y la empleabilidad como los principales factores determinantes de la prolongación de las estancias en el programa.

Respecto a los posibles problemas de dependencia de la duración, el ajuste de un modelo log-normal a la dinámica seguida por el programa desde su puesta en marcha confirmó que las pautas de salida seguían un perfil muy característico: la probabilidad de abandono aumentaba durante los tres primeros años de participación en el programa, para reducirse paulatinamente a partir de dicho umbral. Tales resultados se confirmaron cuando se controla la heterogeneidad inobservable, que parecía tener poca entidad debido a la inclusión de información muy detallada de las características del hogar.

Una extensión natural de esta línea de estudio fue plantear la necesidad de diferenciar los factores que afectaban a cada tipo de salida. Cuando un hogar deja un programa de estas características puede deberse a motivos muy variados. Las estimaciones realizadas confirman la necesidad de completar los juicios que pueden extraerse de los modelos que consideran las salidas de manera homogénea, al existir factores muy diferentes en cada tipo de abandono del programa (“competing risk”). Los contrastes realizados ponen de manifiesto una notable similitud entre los perfiles de las salidas del programa consideradas como exitosas y fraudulentas por los responsables de los servicios sociales y un comportamiento claramente diferencial de aquellas causadas por motivos administrativos. Este resultado alerta contra las clasificaciones excesivamente restrictivas que consideran los dos primeros tipos como realidades distintas.

Una de las grandes ventajas de la base de datos creada es la posibilidad de observar no sólo las entradas y salidas de los programas, sino también la posibilidad de que los hogares retornen a ellos en el medio y en el largo plazo. Un límite presente en buena parte del cuerpo teórico y empírico especializado es que a menudo se obvia la existencia de una notable heterogeneidad entre los beneficiarios, determinada por la variedad de experiencias en las secuencias y en las duraciones de los episodios de participación. Buena parte de la literatura reciente volcada en la dinámica de los programas asistenciales se centra en el efecto de las reincorporaciones sobre las duraciones estimadas. La consideración de las duraciones desde un contexto de episodios múltiples resulta crucial para la medición y el diagnóstico de los programas. Si se considera cada estancia del mismo hogar o individuo como independiente de las anteriores, se tiene un retrato incompleto de su dinámica y, con ello, estimaciones de las duraciones sesgadas a la baja. La identificación, además, de la diversidad de patrones de

beneficiarios, con la posibilidad de encontrar hogares con participación recurrente, temporal o crónica, puede ayudar, además, a dar respuesta a otro de los interrogantes abiertos en la investigación sobre este tipo de políticas sociales, referido a los efectos de los cambios en las condiciones macroeconómicas sobre la dinámica de participación. Los cambios en la tasa de paro deberían, a priori, ser más significativos en el caso de los beneficiarios temporales que en los cíclicos, debido a los problemas estructurales que sufren estos últimos.

Los resultados de la estimación de diferentes modelos, tanto en tiempo continuo como discreto, apuntaron a la existencia de tipologías claramente diferenciadas de acuerdo a la recurrencia y a la duración de la participación. Los contrastes realizados con el fichero elaborado a partir de los registros del programa mostraron también que el desempleo es un factor relevante, pero tiene un peso menor que las características sociodemográficas de los beneficiarios. En cualquier caso, los resultados revelaron también que la combinación de factores macroeconómicos e individuales difiere notablemente en cada uno de los grupos definidos según la participación sea crónica, recurrente o temporal. Algo más de un tercio de los beneficiarios, por ejemplo, entran de manera transitoria en el programa por la influencia de shocks macroeconómicos adversos para no volver a él, mientras que otro colectivo de beneficiarios presenta una elasticidad prácticamente nula a los cambios en las condiciones macroeconómicas, determinando sus características individuales una participación mucho más prolongada.

La importancia de las reincorporaciones al programa obligó a que una tercera línea dentro del proyecto de investigación con los registros administrativos del IMI fuera el estudio de los determinantes que elevaban la probabilidad de retornar al programa después de haber salido de él. Se trató de contestar a interrogantes del tipo de quiénes son los que salen y vuelven, si existen diferencias palpables entre los hogares que retornan y los que salen para no volver o si las características de los primeros se alejan de las que están en los programas de forma crónica. Un segundo tipo de interrogantes se refieren a cuáles son los principales factores explicativos de la duración del tiempo fuera del programa, siendo varias las posibilidades y las preguntas abiertas, como el grado en que el tiempo de la primera estancia en el programa y el tipo de salida de él podían explicar que se prolongara o acortara la duración fuera del programa, cuáles eran las características individuales asociadas a períodos más largos

sin cobrar la prestación y en qué medida variables exógenas, como las condiciones macroeconómicas, podían contribuir también al acortamiento del tiempo fuera del programa.

Para tratar de dar respuesta a estos interrogantes se utilizaron varios procedimientos econométricos. Entre los resultados de un modelo de regresión logística de la probabilidad de reentrada destacaba la necesidad de actuaciones que mejoraran los niveles de empleabilidad de los beneficiarios para afianzar una posición más estable en el mercado de trabajo, las medidas para complementar las remuneraciones de los beneficiarios que dejaban el programa y la definición de actuaciones específicas para rebajar el número y la intensidad de los problemas sociales. La maximización del tiempo fuera del programa en los hogares que tienen mayores probabilidades de reentrada debería centrarse, según nuestros resultados, en la introducción de reformas que mejoraran las salidas de los beneficiarios hacia formas más estables de participación en el mercado de trabajo y en la concentración de una importante dotación de los recursos destinados a promover la inserción en determinados colectivos. Aparte de los individuos con empleabilidad baja e intermedia, sería el caso de las personas solas con déficits relacionales y problemas de desestructuración personal, o de aquellas en las que la pertenencia a una minoría étnica supone un problema para la integración social. No obstante, la evidencia empírica encontrada revela también que existen factores exógenos, como posibles shocks macroeconómicos, que reducen notablemente las posibilidades de la intervención pública para alcanzar sus objetivos. Las simulaciones realizadas con escenarios alternativos relativizan las posibilidades de los decisores públicos para alargar el tiempo fuera del programa si existen perturbaciones macroeconómicas de carácter negativo. No obstante, de ellas se deduce también que la experiencia de la primera estancia y, en menor medida, la empleabilidad, pueden contribuir a alargar el tiempo fuera del programa, por lo que, incluso en contextos especialmente recesivos, quedarían todavía márgenes para una actuación pública más eficaz.

Una última línea de investigación ha tratado de dar respuesta a otro de los interrogantes fundamentales en la evaluación de estos programas: ¿cómo contribuyen las medidas de inserción que acompañan a las prestaciones monetarias a reducir la dependencia de la intervención pública? Fijando como objetivo de dicha intervención la eliminación de la necesidad de volver al programa se utilizaron procedimientos de evaluación no experimental para identificar el efecto de la participación en los llamados “proyectos de

inserción” sobre la probabilidad de reincorporarse al programa de los hogares que habían salido después de participar en dichas actividades. Se utilizaron, con este objetivo, métodos de emparejamiento que trataban de recrear las condiciones de los estudios experimentales. Concretamente, a través de emparejamientos basados en el “propensity score” o probabilidad de participar en esos “proyectos de inserción” se pudo crear un grupo de control con características similares a las de los participantes. Ello permitió corregir el sesgo natural en las estimaciones que obvian la distinta probabilidad de participar en esas acciones y obtener una batería de resultados con notable interés político. Entre otros, una probabilidad menor de que los hogares vuelvan a los programas si han participado en esos proyectos y un alargamiento, al menos, del tiempo pasado fuera de los programas en el caso de volver.

El efecto que tiene la intervención pública sobre la dinámica de participación en el programa y en el bienestar de los hogares constituye el núcleo de los proyectos en curso. La elaboración de una nueva base de datos a partir de una encuesta específica a realizada a más de 2200 hogares tras su paso por el programa IMI permite diferenciar el efecto de diferentes tipos de actuación sobre un vector amplio de resultados del hogar. Concretamente, el objetivo de la nueva línea de investigación es doble: en primer lugar, analizar los efectos de las reformas de los programas de lucha contra la pobreza sobre una variedad de indicadores de resultados, tratando de ver si las posibles mejoras o empeoramientos se manifiestan de forma generalizada en un conjunto amplio de dimensiones (salarios, empleo, pobreza, bienestar material y problemas sociales) o si, por el contrario, los avances en un ámbito concreto son compatibles con el empeoramiento o el estancamiento en otras dimensiones. En segundo lugar, analizar si las diferentes formas de intervención pública (formación para el empleo, seguimiento social básico, trabajo con menores, etc.) contribuyen de un modo similar a la homogeneidad o heterogeneidad de resultados.

Aparte de las líneas de investigación descritas, la base de datos creada a partir de los registros del IMI se ha utilizado para otros trabajos complementarios, no centrados en el estudio de la dinámica de participación en el programa. En los últimos años se ha desarrollado un creciente cuerpo teórico y empírico para el análisis de los factores determinantes de los cambios en las cifras de beneficiarios de las prestaciones asistenciales. Los registros del IMI

nos han permitido examinar las posibilidades de las nuevas propuestas metodológicas tomando como referencia los datos de beneficiarios del programa. Las principales novedades respecto a estudios realizados para otros países han sido la utilización de datos mensuales que comprenden todo el período de desarrollo del programa, la introducción de nuevos factores explicativos indicativos del grado de restrictividad de los programas, la diferenciación del efecto de las distintas variables por grupos demográficos y la utilización de técnicas de cointegración para el estudio de las relaciones estadísticas más significativas.

En una línea muy diferente, la disponibilidad de información sobre el proceso de gestión de las prestaciones y sobre las características de los centros de servicios sociales encargados de su tramitación los registros del IMI sirvieron como base para el estudio de cuestiones relacionadas con el análisis de la eficiencia en la gestión de los programas de rentas mínimas. Hasta la fecha, las únicas evidencias tomaban como referencia indicadores generales de gestión, careciendo de estudios que relacionen la utilización de inputs con esos indicadores. En el estudio realizado, mediante el análisis DEA se cruzaban diversos indicadores de resultados con la dotación de factores de cada centro de servicios sociales, unidad encargada de la gestión, con el propósito de generar medidas de eficiencia.

5. CONCLUSIONES

La explotación de los registros administrativos generados por el programa de rentas mínimas de la Comunidad de Madrid (IMI) iniciado a finales de 1990 y su conversión en una base longitudinal de microdatos con más de 50.000 episodios que se corresponden con algo más de 39000 hogares, puede servir para arrojar el debate sobre la reforma de los instrumentos asistenciales mediante la producción de abundante información cuantitativa y un conocimiento más preciso de la dinámica de participación. La caracterización de los procesos que determinan las duraciones dentro de estos programas contribuye a mejorar el diagnóstico previo a la introducción de cambios sustanciales en el diseño de este instrumento de la protección social. El estudio de las duraciones de las estancias en dicho programa a través de diferentes modelos no paramétricos, paramétricos y semi-paramétricos en tiempo continuo y discreto, la identificación de los principales factores determinantes de la duración en el programa, la caracterización de diferentes tipologías de beneficiarios según sus posibilidades de

inserción laboral, el estudio de los episodios múltiples de participación y sus causas, los distintos tipos de salidas del programa IMI a través de modelos “competing risk” y, por último, la evaluación, a través de la técnica “propensity-score matching”, de algunas de las actuaciones específicamente diseñadas para favorecer la inserción sociolaboral, constituyen los proyectos de investigación mas relevantes de los realizados hasta el momento con la base de microdatos longitudinales elaborada.

Las principales ventajas de utilizar como referencia este programa son su carácter representativo de la media nacional, la riqueza de la información que ofrecen los registros, con variables hasta ahora ausentes en la mayoría de las encuestas de ingresos españolas, y la posibilidad de realizar comparaciones con los programas de otros países. Para todo ello resultó necesario un complejo proceso de evaluación de calidad, depuración, estandarización, creación de ficheros longitudinales y enlace con otras fuentes para convertir los datos administrativos en una base de datos adecuada para el estudio de de todos estos temas a partir de las duraciones en el programa. El trabajo invertido en esta tarea ha compensado ampliamente los objetivos perseguidos al iniciarla.

En la actualidad, son varias las posibilidades de análisis todavía abiertas, con un conjunto amplio de cuestiones para las que los registros administrativos ofrecen una valiosa información, a la vez que quedan otras para las que son necesarios nuevos datos, como la realización de análisis coste-efectividad de estas políticas o la estimación de sus efectos a largo plazo. Para el adecuado cumplimiento de ambas, la transformación de los registros resultará, sin duda, una herramienta imprescindible.

ANEXO I
CATEGORÍAS DE LAS DISTINTAS VARIABLES DE LOS FICHEROS DEL IMI
(CAM)

Fichero Hoja de Solicitud

PARENTESCO

1. Padre/madre
2. Esposo/a
3. Compañero/a
4. Hijo/a
5. Abuelo/a
6. Nieto/a
7. Suegro/a
8. Yerno/nuera
9. Hermano/a
10. Cuñado/a
11. Tío/a
12. Sobrino/a
13. Primo/a

SEXO

1. Hombre
2. Mujer

NIVEL DE ESTUDIOS ALCANZADO

1. No lee ni escribe
2. Sin estudios (sólo lee y escribe)
3. Estudios Primarios
4. Graduado escolar o bachiller elemental
5. BUP, COU o bachiller superior
6. Formación profesional 1.er grado
7. Formación profesional 2.o grado
8. Diplomado universitario
9. Título universitario superior
10. Sin edad escolar (entre 0 y 5 años)
11. Sin escolarizar (entre 6 y 16 años)

ESTADO CIVIL

1. Soltero/a
2. Casado/a
3. Viudo/a
4. Divorciado/a
5. Separado/a legalmente
6. Separado/a de hecho

SITUACIÓN DE LA VIVIENDA

1. Sin domicilio fijo (transeúntes, instituciones, pensión provisional, vehículo..)
2. Chabola, cueva o infravivienda.
3. Prefabricado (provisional)

4. Institución (permanente)
5. Vivienda ocupada
6. Vivienda cedida o similar
7. Vivienda alquilada
8. Pensión o similar (permanente)
9. Propia pagada
10. Propia pagando

RELACIÓN CON EL EMPLEO

1. Trabajo por cuenta propia
2. Trabajo por cuenta ajena
3. Trabajo temporal/eventual por cuenta ajena
4. Trabajo a domicilio
5. Trabajos esporádicos
6. Actividades económicas marginales
7. Empleado/a de hogar
8. Venta ambulante
9. Parado con trabajo anterior
10. Parado buscando 1.er empleo
11. Ama de casa
12. Estudiante
13. Servicio militar
14. Rentista
15. Menor de 65 años incapacitado permanente sin pensión
16. Menor de 65 años incapacitado permanente con pensión
17. Mayor de 65 años que no trabaja sin pensión
18. Mayor de 65 años que no trabaja con pensión
19. Menores de 16 años sin edad escolar (0-5 años)
20. Menores de 16 años sin escolarizar (6-16 años)
21. Estudiante

MUNICIPIO

NÚMERO DE MIEMBROS DEL HOGAR

RECURSOS ECONÓMICOS

SITUACIÓN DEL EXPEDIENTE

1. Concedido
2. Denegado
3. Extinguido
4. Modificado
5. Suspendido

CAUSAS DE EXTINCIÓN

1. Pérdida de la nacionalidad
2. Cumplir 65 años
3. Traslado fuera de la CAM
4. Recursos superiores baremo
5. No cumplir obligación firma

6. No cumplir contrato
7. Destino incorrecto prestación
8. No escolarización de menores
9. No comunicar variaciones
10. Actuación fraudulenta
11. Fallecimiento titular
12. Internamiento del titular (art. 7, Decreto 21/1992)
13. Extinción fusión expediente.
14. Cumplir objetivos acordados.
15. Rechazar medidas de inserción laboral
16. Exclusión Proyecto IMI
17. Desistimiento del titular

Fichero de Seguimiento

SITUACIÓN DE LA VIVIENDA

1. Sin domicilio fijo (transeúntes, instituciones, pensión provisional, vehículo..)
2. Chabola, cueva o infravivienda.
3. Prefabricado (provisional)
4. Institución (permanente)
5. Vivienda ocupada
6. Vivienda cedida o similar
7. Vivienda alquilada
8. Pensión o similar (permanente)
9. Propia pagada
10. Propia pagando

EMPLEABILIDAD

1. Incapacidad absoluta para el empleo normalizado
2. Necesita de un proceso de recuperación social/de salud previo
3. Parado/a que necesita formación/educación
4. Podría acceder ya a un empleo
5. Desarrolla un empleo sumergido a actividad equivalente
6. Desarrolla un empleo normal o actividad equivalente

PROBLEMÁTICA SOCIAL

1. Problemas que afectan a los menores
2. Dependencia / abuso de drogas
3. Dependencia / abuso del alcohol
4. Otros problemas de salud mental
5. Otros problemas graves de salud
6. Impagos de vivienda, desahucio
7. Acumulación de deudas, impagos
8. Mendicidad
9. Prostitución
10. Problemas de relación con la justicia penal
11. Problemas de convivencia o relación familiar
12. Aislamiento, desestructuración personal
13. Minoría étnica

NIVEL DE TRABAJO DE INTEGRACIÓN

1. Seguimiento administrativo
2. Seguimiento social básico
3. Tratamiento social intensivo

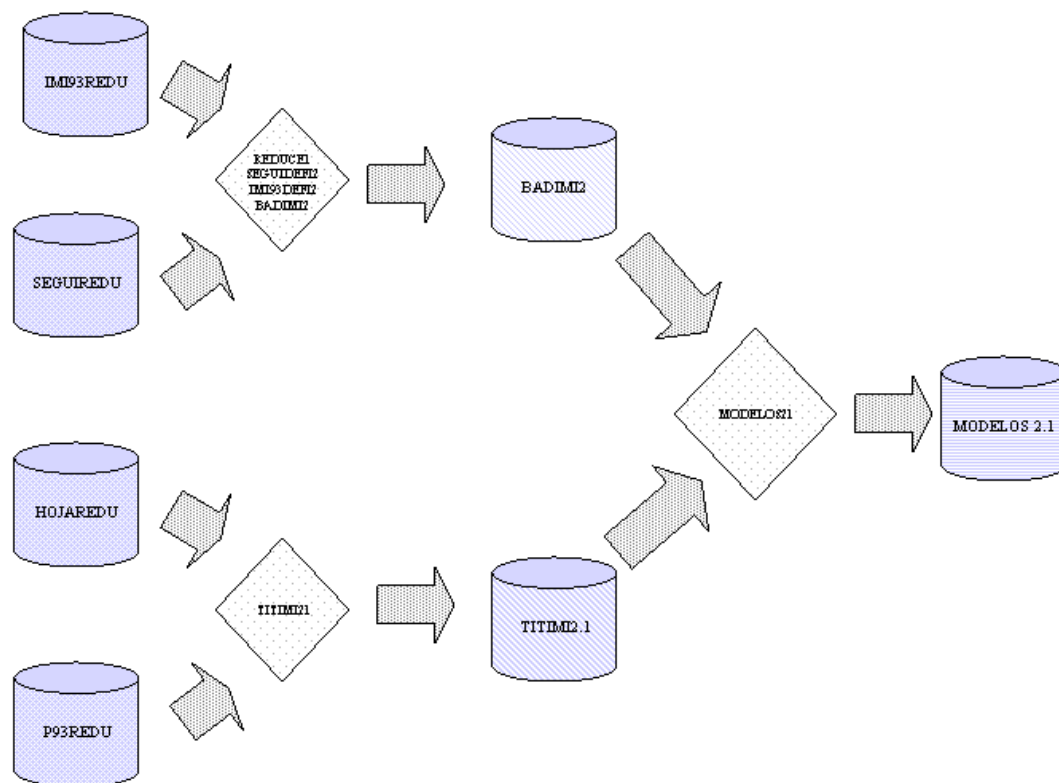
ÁREAS EN LAS QUE SE HA TRABAJADO DURANTE EL SEMESTRE

1. Dinámica familiar
2. Habilidades sociales
3. Educación adultos
4. Escolarización menores
5. Salud
6. Formación para el empleo
7. Acceso al empleo
8. Proyectos IMI
9. Vivienda
10. Ocio
11. Apoyo y acompañamiento social

TABLA 1
MATRIZ DE ENTRADA Y SALIDA EN EL PROGRAMA
(todos los casos)

So excluyen D y Z		Registro de perceptores de prestaciones C. A. de Madrid																											Total Hojas			Hoja 1 no/d	
		Semestre de salida (corregido) 99 = Expedientes en curso																											Max	Min	Mean	0	1
Año de entrada	Semestre de entrada	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	99	Total							
1990	1	0	158	180	169	130	101	121	33	37	33	20	11	10	6	5	6	5	3	2	2	5	1	1	10	1062	16	0	2	909	143		
1991	2	—	169	822	945	913	630	626	458	308	231	147	94	60	51	33	40	34	36	18	18	24	20	10	87	5774	16	0	2	5125	646		
1991	3	—	—	60	460	439	319	300	197	146	90	58	39	21	19	21	12	11	11	8	11	13	9	3	37	2287	16	0	2	1977	310		
1992	4	—	—	3	129	487	345	259	212	129	138	67	50	18	31	24	19	18	19	12	11	5	13	7	45	2033	16	0	3	1622	411		
1992	5	—	—	—	—	107	277	244	182	114	84	60	32	29	25	20	19	15	7	14	6	6	6	3	19	1269	16	0	3	980	289		
1993	6	—	—	—	—	—	111	504	411	256	198	134	122	67	49	26	32	21	27	18	18	12	10	8	45	2069	16	1	3	1051	1018		
1993	7	—	—	—	—	—	—	100	588	401	277	165	131	80	61	43	44	26	32	28	8	14	8	10	74	2090	17	0	4	6	2084		
1994	8	—	—	—	—	—	—	—	261	546	415	216	175	92	75	51	52	41	46	15	23	23	19	14	70	2137	16	0	4	133	2094		
1994	9	—	—	—	—	—	—	—	—	118	686	423	290	149	109	85	92	47	39	42	32	26	22	18	89	2267	16	0	3	802	1465		
1995	10	—	—	—	—	—	—	—	—	224	751	619	328	210	140	147	100	79	66	45	39	41	22	143	2964	14	0	4	92	2862			
1995	11	—	—	—	—	—	—	—	—	—	57	362	200	122	108	74	51	51	28	22	19	14	21	82	1211	13	0	4	12	1198			
1996	12	—	—	—	—	—	—	—	—	—	259	635	373	271	232	155	142	90	63	53	61	44	190	2571	12	0	3	415	2195				
1996	13	—	—	—	—	—	—	—	—	—	95	616	403	343	217	167	128	89	72	71	40	231	2472	11	0	4	40	2432					
1997	14	—	—	—	—	—	—	—	—	—	—	—	—	—	75	435	320	235	194	115	102	73	76	48	248	1921	10	0	4	1	1920		
1997	15	—	—	—	—	—	—	—	—	—	—	—	—	—	—	121	462	310	263	150	117	98	81	62	271	1955	5	0	4	2	1953		
1998	16	—	—	—	—	—	—	—	—	—	—	—	—	—	—	180	642	495	291	197	189	148	94	416	2652	6	0	4	12	2640			
1998	17	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	126	655	429	303	236	152	121	436	2458	7	0	3	1	2457			
1999	18	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	82	543	377	321	221	143	478	2165	6	1	3	—	2165				
1999	19	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	60	344	297	248	126	507	1584	5	1	3	—	1584	
2000	20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	50	481	367	222	718	1847	4	0	3	1	1846		
2000	21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	58	449	309	836	1652	3	1	2	—	1652		
2001	22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	52	495	1382	1829	2	1	2	—	1829		
2001	23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1315	
Total		3	327	1068	1607	2075	1781	2154	2342	2055	2376	2098	2184	1784	1625	1786	2004	2054	2348	2052	1847	2064	2091	1580	7568	46564	16	0	3	13181	56383		

Lunes, 9 de diciembre de 2002, hora: 18:53:7.1
 Fuente: Comunidad de Madrid, Consejería de Sanidad y Servicios Sociales.
 Elaboración: S. G. de Estudios Presupuestarios y de Gasto Público.
 Instituto de Estudios Fiscales.



Bibliografía

- Ayala, L., Pedraja, F., y F.J. Salinas (2004): "Assessing the Efficiency of Spanish Social Service Centers: a Data Envelopment Analysis". *Empirical Economics Letters* 3, 194–205.
- Ayala, L. y Pérez, C. (2005): "Macroeconomic Conditions, Institutional Factors and Demographic Structure: What Causes Welfare Caseloads?", *Journal of Population Economics*, 3/2005.
- Ayala, L. y M. Rodríguez (2003): *La duración de los programas de lucha contra la pobreza*. Instituto de Estudios Fiscales, Madrid.
- Ayala, L. y M. Rodríguez (2004): "Multiple Occurrence of Welfare Reciprocity: Determinants and Policy Implications". Instituto de Estudios Fiscales, Working Paper 13/2004.
- Ayala, L. y M. Rodríguez (2006): "The Latin Model of Welfare: Do 'Insertion Contracts' Reduce Long-Term Dependence?", *Labour Economics* (en prensa).
- Ayala, L. y M. Rodríguez (2006): "Exits from Social Assistance in Spain" (with Magdalena Rodríguez), *International Journal of Social Welfare* (en prensa).
- Belin, T.R. y D.B. Rubin (1995): "A method for calibrating false-match rates in record linkage". *Journal of the American Statistical Association* 90(June):694-707.
- Fellegi, I.P., and A.B. Sunter (1969): "A theory for record linkage", *Journal of the American Statistical Association* 64:1183-1210.
- Goerge, R. M., y Joo Lee, B. (2002): "Matching and Cleaning Administrative Data". En Ver Ploeg, M.; Moffitt, R. A., y Citro, C. (eds.): *Studies of Welfare Population: Data Collection and Research Issues*. Washington: National Academy Press.
- V. Joseph Hotz, Robert Goerge, Julie Balzekas, and Francis Margolin (Eds) (1998): *Administrative data for policy-relevant research: assessment of current utility and recommendations for development*, University of Chicago Joint Center for Poverty Research
- Heckman, James J. (2001). "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *Journal of Political Economy*, vol. 109(4), pages 673-748
- Jaro, M.A. (1985): "Current record linkage research". En *Proceedings of the Section on Statistical Computing*. Alexandria, VA: American Statistical Association.
- Jaro, M.A. (1989): "Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida". *Journal of the American Statistical Association* 84(June):414-420.
- Mainieri, T., y Danziger, S. (2001): "Designing Surveys of Welfare Populations". *Report from the Workshop on Designing Surveys of Welfare Recipients*, Ann Arbor, Michigan.
- Pettersson, Hans (1992). "Quality control in statistics from administrative registers and records". Instituto Vasco de Estadística . Vitoria-Gasteiz

Winkler, W. E. (1988): “Using the EM algorithm for weight computation in the Fellegi-Sunter model of record linkage”. En *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Winkler, W. E. (1990): “String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage”. En *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Winkler, W. E. (1993): “Matching and record linkage”. En *Statistical Research Division Report 93/08*. Washington, DC: U.S. Bureau of the Census.

Winkler, W. E. (1994): “Advanced methods for record linkage”. En *Statistical Research Division Report 94/05*. Washington, DC: U.S. Bureau of the Census.

Winkler, W. E. (1999): “The state of record linkage and current research problems”. En *Statistical Research Division Report 99/04*. Washington, DC: U.S. Bureau of the Census.

Winkler, W.E., y Y. Thibaudeau (1991): “An application of the Fellegi-Sunter model of record linkage to the 1990 U.S. decennial census”. En *Statistical Research Division Report 91/09*. Washington, DC: U.S. Bureau of the Census.