

CENSOS DE POBLACIÓN, DATOS VINCULADOS Y EL FUTURO DE LA INVESTIGACIÓN SOCIAL

POPULATION CENSUSES, LINKED DATA AND THE FUTURE OF SOCIAL RESEARCH

MIGUEL REQUENA
UNED

mrequena@poli.uned.es

ORCID iD: 0000-0002-4490-6029

Cómo citar este artículo / Citation: Requena, M. 2021: Censos de población, datos vinculados y el futuro de la investigación social, *Revista Internacional de Sociología*, 79 (1), e181. <https://doi.org/10.3989/ris.2021.79.1.19.181c>

Copyright: © 2021 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

RESUMEN

En el diseño del censo español de 2021 se han abandonado los procedimientos tradicionales de recogida de datos mediante cuestionarios aplicados al universo poblacional y se ha adoptado una nueva estrategia basada en la vinculación de registros administrativos, el uso de *grandes datos* y el recurso a una encuesta complementaria. Más allá del diseño del nuevo censo, la vinculación de datos administrativos y la integración de otras fuentes está adquiriendo cada vez más importancia en los países con sistemas estadísticos avanzados. En este trabajo se evalúan las reacciones a estos cambios de productores y usuarios de las estadísticas oficiales en España y se defiende la idea de que, con los debidos controles metodológicos, la vinculación de datos es un ingrediente crucial para el avance de la investigación social cuantitativa.

PALABRAS CLAVE

Datos administrativos, combinación de fuentes, estadísticas oficiales, vinculación de registros.

ABSTRACT

For the design of the 2021 Spanish Census, traditional data collection methods involving sending questionnaires to the whole population have been replaced by the combination of administrative data linkage, big data and an ad-hoc survey. Beyond the design of new census, administrative data linkage and integration of other non-administrative data sources are becoming more and more relevant in countries with advanced statistical systems. Reactions to these changes among producers and users of official statistics in Spain are addressed and assessed in this paper. The idea is defended that, with due methodological controls, data linkage is a crucial element for the advancement of quantitative social research.

KEYWORDS

Administrative data, combination of sources, official statistics, record linkage

INTRODUCCIÓN

Para elaborar el censo español de 2021, el Instituto Nacional de Estadística (INE) ha abandonado los procedimientos tradicionales de recogida de información mediante cuestionarios aplicados al universo poblacional y ha adoptado una nueva estrategia basada en la vinculación de registros administrativos, el uso de *grandes datos* y el recurso a una encuesta complementaria *ad hoc*. El estudio “El nuevo censo de 2021 en España: un debate metodológico, epistemológico y político pendiente” invita a discutir la conveniencia, oportunidad y trascendencia de la transformación. El debate es importante, más allá de la producción de la información censal, debido a las implicaciones que la vinculación de bases de datos tiene para el futuro de la investigación social.

Como todos los países desarrollados, España cuenta con una plétora de datos administrativos, pero también con una práctica de vinculación relativamente reciente y más bien escasa en términos comparados. Ello nos sitúa en una situación de clara desventaja frente a otros países con sistemas estadísticos avanzados que, de persistir, lastrará con toda probabilidad el avance de la investigación científica y académica en una amplia variedad de campos. En los países avanzados, la integración de datos administrativos se está convirtiendo en un imperativo crucial para el avance de la producción estadística. En la medida en que se apoya en datos estadísticos, la investigación social no puede, ni debe, eludir ese imperativo. De hecho, tanto entre los académicos como entre los responsables públicos de la producción y distribución de datos hay un creciente consenso en torno a la idea de que uno de los más importantes retos que tienen las oficinas estadísticas nacionales (OEN) es el de vincular datos dispares que proceden de fuentes diversas. Las posibilidades de vincular datos son hoy tantas y de tanta trascendencia que, para algunos autores (Lothian, Holmberg y Seyb 2017), esta es precisamente la cuestión que va a dominar la agenda de trabajo de las OEN en los próximos años. Puede que no sea exagerado decir que, en las sociedades que disponen de sistemas estadísticos avanzados, el futuro de la investigación social va a depender, en no poca medida, de la disponibilidad de estas bases de datos vinculados.

Las grandes posibilidades que ofrecen las bases vinculadas nacen, como es sabido, de la combinación de varios desarrollos recientes. Primero, el número de fuentes disponibles para su vinculación es potencialmente muy elevado en sociedades en las que la capacidad de recoger, almacenar y tratar grandes masas de datos ha crecido de forma exponencial en las últimas décadas. Las administraciones públicas son un claro exponente de esta multiplicada capacidad de recopilar datos administrativos con un claro potencial de contribuir al desarrollo de una

ciencia social de gran calidad y alto impacto (Connelly *et al.* 2016). Segundo, para integrar datos procedentes de bases diferentes se dan ya las condiciones técnicas necesarias soportadas por capacidades de computación muy potentes, inimaginables hasta hace poco tiempo, y por algoritmos específicos para, por ejemplo, las tareas de limpieza automática y detección y tratamiento de valores extremos. Tercero, se empiezan ya a conocer y abordar los problemas metodológicos que surgen en la integración de esos datos estadísticos (Wallgren y Wallgren 2014; Zhang 2012; Bakker y Daas 2012; Bakker 2010). Más adelante volveré a este último punto.

En el campo de las estadísticas oficiales y, por implicación, en el de la investigación social cuantitativa, la integración de bases de datos dispares se propone aprovechar al máximo la información ya existente y producir nuevos datos estadísticos a unos costes incomparablemente menores que los de crearlos *ex novo*. Estas bases de datos reducen, además, de manera significativa la carga de esfuerzo que se exige al entrevistado en unos momentos en las que las tasas de no repuesta en las encuestas tradicionales están creciendo de forma alarmante (Groves 2006). Si, como parece razonable, entre los objetivos futuros de los sistemas de estadísticas públicas está la producción de una información de alta calidad que responda a las necesidades de los usuarios de una forma más amplia, más profunda, más rápida, más barata y mejor (Holt 2000; 2007; 2008), es incuestionable que las OEN deberán incorporar de forma cada vez más intensa la vinculación de datos a sus rutinas, construir con cada vez mayor eficiencia y rigor bases de datos vinculados y distribuir con creciente agilidad al usuario potencial la información obtenida por esos medios.

CENSOS Y DATOS ADMINISTRATIVOS

Dadas las circunstancias, no es extraño que muchas OEN se hayan lanzado ya a la integración de bases de datos administrativos con información relativa a salud, ingresos, riqueza, educación, transiciones vitales y otras características de indiscutible interés para la investigación social. Un caso paradigmático de estos nuevos usos de los datos administrativos lo constituyen, sin duda, los censos de población (Poulain y Herm 2013). En relación con la ronda censal de 2011, la República Checa, Letonia y Lituania combinaron censos basados en registros y censos tradicionales; España, Alemania, Holanda, Polonia, Suiza e Israel realizaron sus censos recurriendo a operaciones basadas en registros y a encuestas por muestreo; Bélgica, Dinamarca, Finlandia, Islandia, Noruega, Eslovenia, Suecia y Turquía basaron sus censos exclusivamente en registros; y Francia recurrió a encuestas *ad hoc* con muestras rotatorias (Instituto Nacional de Estadística 2011). De los entonces

27 países de la UE, solo 11 utilizaron la metodología tradicional (United Nations Economic Commission for Europe 2011).

Es sabido que los países nórdicos se han situado desde hace tiempo en la vanguardia de esa estrategia de creciente vinculación de las bases de datos administrativos. Sus propósitos van bastante más allá de la producción censal. En esos países existe una amplia, sólida y ya larga tradición de vincular registros de población y otros datos administrativos que es única a escala europea y, por ende, mundial, y que con sobrada razón se suele considerar una auténtica mina de oro para la investigación científica. Y su voluntad de seguir en posiciones de vanguardia por el mismo camino es clara: pese a la riqueza de información vinculada existente, sin parangón en el resto del mundo, no faltan propuestas que instan a encontrar el modo de recombinar las bases de datos existentes y mejorar su vinculación para maximizar su rendimiento investigador en numerosos campos, pues, según se afirma, “no existen precondiciones equivalentes para este tipo de investigación en el resto de Europa” (Sandberg 2012: 15).

Otro caso de interés son los Países Bajos, donde la totalidad de las estadísticas demográficas se basa por completo en los registros municipales de población que se actualizan de forma casi continua cada vez que las personas registradas experimentan un evento demográfico de interés (Statistics Netherlands 2017). A partir de ahí, los progresos neerlandeses con la vinculación de datos han sido muy notables. Desde mediados de los años noventa del pasado siglo se ha venido desarrollando un sistema de bases de datos de estadísticas sociales (SSD) que, en fecha tan temprana como 2001, ya permitió realizar el censo sin recurrir al expediente tradicional de la encuesta. En la actualidad, el sistema ha crecido enormemente hasta cubrir una amplia gama de datos sociodemográficos que van desde la participación en el mercado de trabajo y la seguridad social hasta la salud, el delito, la vivienda o la migración (Bakker, van Rooijen y van Toor 2014).

Frente a los mencionados países, el acceso a datos administrativos susceptibles de ser utilizados para la investigación social es todavía limitado en otros países avanzados como EE.UU., Canadá, Australia o el Reino Unido. Sin embargo, no se debe pasar por alto el hecho de que, de una u otra manera, la experiencia de vincular datos administrativos para producir información censal ha terminado por empujar a otros países europeos a desarrollar la vinculación con propósitos que, claramente, pueden ir más allá de los censos. Tal es el caso de Austria, por ejemplo (Berka, Humer y Moser 2012; Lenk 2008). Por su parte, Bélgica cuenta con registros administrativos de población a escala municipal desde 1846 que recientemente han sido digitalizados y centrali-

zados a escala nacional y que desde los años noventa producen estadísticas demográficas nacionales (Poulain 2010). E incluso el Reino Unido, que hasta ahora ha realizado sus censos al modo tradicional, aunque parece moverse hacia una solución combinada, ha constituido una Administrative Data Research Network con el expreso propósito de “conectar el personal analítico de los departamentos de la Administración Pública con capacidades y facultades académicas para compartir conocimientos, métodos y perspectivas y transformar de forma segura información sensible en bases de datos anonimizadas que permitan el desarrollo de políticas públicas basadas en la evidencia” (<https://adrn.ac.uk>).

EL CASO ESPAÑOL

En España, el Censo de Población y Viviendas de 2011 se elaboró ya mediante la combinación de un fichero precensal realizado a partir de los registros administrativos disponibles (básicamente, el Padrón Municipal), un Censo de Edificios exhaustivo con georreferenciación de todos los inmuebles y una gran encuesta por muestreo diseñada para recoger las correspondientes características de personas y viviendas. Las cifras de población se obtuvieron mediante el recuento de los registros contenidos en el fichero precensal, ponderados con factores de recuento obtenidos de la encuesta (Instituto Nacional de Estadística 2011). Está previsto que el censo de 2021 se apoye, en la medida más completa posible, en datos procedentes de registros administrativos (Argüeso 2019). La idea es “construir más del 90 % de la información que contenía el cuestionario censal de 2011 a partir de registros administrativos y estos ofrecerán mayor calidad que los datos basados en recogida directa mediante cuestionarios” (Instituto Nacional de Estadística INE 2017a, 7).

Siguiendo el camino trazado por otras oficinas estadísticas, el esfuerzo del INE español en el ámbito de la vinculación de datos administrativos no se circunscribe a las operaciones censales, sino que comienza a extenderse a otros terrenos. Destaca, por ejemplo, el caso de la Encuesta de Población Activa (EPA) que, para capturar la información sobre el salario del empleo principal exigida por las instituciones europeas, ha optado por incluir una variable derivada que se obtiene vinculando diferentes fuentes de datos administrativos: el Impuesto de la Renta de las Personas Físicas y los registros del fichero general de afiliaciones y altas y el fichero de bases de cotización de la Tesorería General de la Seguridad Social (Instituto Nacional de Estadística INE 2017b).

Adicionalmente, y a una escala menor, a instancias de ciertos grupos de investigación y bajo la fórmula de las llamadas peticiones a medida, el INE también ha conseguido vincular los datos censales

y los precedentes del Movimiento Natural de la Población con propósitos puramente investigadores. Hasta la fecha, esas iniciativas han producido resultados de investigación de interés, por ejemplo, sobre diferencias educativas en mortalidad e importancia relativa de las causas de muerte (Requés *et al.* 2014), sobre el descenso de la mortalidad por grupo socioeconómico durante la crisis económica en España (Regidor *et al.* 2016) o sobre los diferenciales de mortalidad asociados a la situación de convivencia (Requena y Reher 2020).

REACCIONES AL CAMBIO

Sirvan estas apresuradas pinceladas para enmarcar las reacciones de la comunidad investigadora española en ciencias sociales —de las que tan oportunamente nos informa el estudio “El nuevo censo de 2021 en España: un debate metodológico, epistemológico y político pendiente”— a lo que con razón se ha denominado un nuevo paradigma en la producción estadística (Ramiro 2019). Geógrafos, demógrafos y sociólogos vienen a coincidir, como *usuarios*, en dos reacciones características: de un lado, una conciencia realista de la irreversibilidad del nuevo modo de producción de las estadísticas públicas basado en la vinculación de registros administrativos; de otro, una aceptación relativamente desconfiada y resignada del cambio.

Esa conciencia de que hemos emprendido un camino sin retorno hacia un nuevo modelo de producción de estadísticas oficiales se sustenta en la cabal percepción, muy notoria entre los *productores* estadísticos, de que hoy día las operaciones censales tradicionales son virtualmente inviables. Los censos tradicionales exigen procesos de gran complejidad logística que resultan muy costosos desde el punto de vista económico, chocan de manera sistemática con la fatiga de los informantes y terminan ofreciendo resultados patentemente subóptimos en lo que se refiere, por ejemplo, a la periodicidad. Como resistirse a lo inevitable produce melancolía, poco más hay que decir de esta primera reacción.

Más interés suscitan, en cambio, las actitudes de recelo ante el cambio. Al margen de algunos desacuerdos procedimentales sobre la propia gestión del cambio o sobre la idoneidad del momento elegido para llevarlo a cabo, el caldo de cultivo en el que crece esa desconfianza de los nuevos métodos es la extendida preocupación por la calidad de los datos de los cuatro pilares sobre los que se está construyendo el nuevo censo: padrón municipal, diversos registros administrativos, *big data* y encuesta *ad hoc*. Bajo el supuesto de que la calidad de los datos vinculados difícilmente podrá ser mayor que la de los datos a vincular, la suspicacia que despierta el uso de los cuatro pilares se traduce en inquietud ante la posi-

bilidad de que no se alcancen de forma satisfactoria varios de los tradicionales objetivos censales como la universalidad, la exhaustividad, el detalle territorial o la continuidad de las series históricas. Otras preocupaciones en torno a las implicaciones epistemológicas (construcción, significado y naturaleza) de los nuevos datos y a sus premisas ideológicas (“la colonización de la lógica empresarial sobre lo público” que se infiere de la negativa evaluación de los costes de los censos tradicionales) y sus repercusiones políticas (“la erosión del Estado frente a las corporaciones” que se supone implica comprar los *big data*) también promueven la difidencia de algunos.

Mientras es difícil, y posiblemente prematuro, intentar valorar las preocupaciones epistemológicas de largo alcance y las de índole político-ideológica que expresan algunos usuarios, las cautelas metodológicas sobre la calidad de los datos son, a mi juicio, pertinentes y oportunas. Hace ya más de un decenio que los expertos conocen esta premisa básica del uso enlazado de registros: “cuantos más datos administrativos se usan, más problemas metodológicos surgen” (Bakker 2009); y, aunque no se dispone todavía de una teoría completamente articulada de la vinculación de los datos administrativos, las OEN cuentan ya con inventarios de problemas y soluciones prácticas con las que remediarlos en lo posible. De hecho, los principales escollos metodológicos mencionados por los usuarios españoles que pueden deteriorar la calidad de los datos administrativos han sido detectados ya por los productores y se encuentran relativamente bien codificados (Bakker y Daas 2012).

El origen de los problemas reside en que los diseños de recogida de datos ideados para producir información con propósitos administrativos no necesariamente utilizan los mismos conceptos estadísticos, marcos muestrales, modos de administración de las operaciones ni protocolos de recogida de información que han sido de uso común en las estadísticas oficiales tradicionales. Ello puede dar lugar a posibles errores tanto de representación (infracobertura y sobrecobertura) como de medida (imprecisiones en las magnitudes que definen los atributos de los miembros de las poblaciones observadas). Además, el propio proceso de enlazamiento puede producir errores de vinculación cuando se pierden los identificadores que permiten los enlaces o no se usan identificadores inequívocos. Mención aparte merece el uso de grandes datos, de cuyos problemas de representación y más que probables sesgos hace tiempo que también se viene advirtiendo (véase, por ejemplo, Goldthorpe 2017).

Todos estos problemas son bien conocidos, como también lo son las medidas precautorias que tomar para minimizarlos (Daas y Ossen 2011), las soluciones razonables a las que recurrir una vez detectados

errores y sesgos (Bakker 2010), la teoría que aplicar a las estadísticas basadas en registros y datos integrados a partir de fuentes diferentes (Zhang 2012) y los procedimientos con que evaluar la calidad de este tipo de datos (Holmberg 2012). Considerando que los datos estadísticos libres de errores son una quimera, la cuestión pertinente es naturalmente la de la calidad comparada de unas y otras fuentes. Cabe señalar que también los censos tradicionales estaban, por unas u otras razones, lejos del ideal de la perfección estadística, de manera que el juicio último sobre el nuevo censo de 2021 —y, para el caso, de todas las operaciones estadísticas basadas en la integración de datos administrativos— dependerá de la evaluación de su calidad en comparación con las fuentes alternativas equivalentes. No está de más recordar que las previsiones de los productores son muy claras en este punto: “El censo de 2021 se realizará, por tanto, a partir de datos administrativos y su calidad será superior a la de los censos anteriores” (Argüeso 2019: 11).

El tono optimista de las reflexiones previas no pretende tanto eludir el debate que exigen algunos usuarios como llevarlo al terreno de las fuentes específicas y a la evaluación de su uso para ciertos propósitos.¹ En lo que se refiere al nuevo censo, suspendamos la desconfianza y otorguemos un voto de confianza al INE, al menos hasta contar con una evaluación de la calidad de la nueva operación y poder emitir un juicio basado en la evidencia. Y, con carácter general, reclamemos como usuarios a los productores de estadísticas oficiales que, en la medida de lo posible, hagan en todas las operaciones de integración un triple esfuerzo de control y supervisión de la calidad de los datos administrativos, de estimación de la precisión de las medidas y la magnitud de los errores producidos a resultas de los enlaces, y de transparencia respecto a los procedimientos de vinculación.

DATOS VINCULADOS E INVESTIGACIÓN SOCIAL

Más allá de la elaboración del próximo censo, la capacidad de integrar datos que proceden de fuentes diversas ofrece un potencial para la investigación social en nuestro país de inmenso alcance. La vinculación de datos puede ofrecer, a costes relativa-

mente bajos, información estadística solvente de una calidad, una riqueza y un nivel de detalle no disponibles en otras fuentes. No aprovechar esas nuevas capacidades de nuestro sistema estadístico para aplicarlas a la investigación social sería poco razonable. Por otra parte, puesto que el futuro de muchas estadísticas oficiales va a depender en muy buena medida de la combinación de datos administrativos, es urgente que la comunidad investigadora se ponga al día y se familiarice con las tareas de depuración, explotación y análisis de los datos vinculados para producir mejores resultados de investigación. Es más que probable que, también en España, una importante porción de la futura investigación social de calidad se base en el uso intensivo de este tipo de datos. Además, disponer de tales fuentes vinculadas nos situaría a los científicos sociales españoles en pie de igualdad con otras comunidades nacionales que disponen desde hace tiempo de este tipo de datos. Por último, la realización de análisis comparados con otros países que disponen de datos estadísticos sociodemográficos mucho más completos y detallados que los nuestros requiere que también nosotros los produzcamos y seamos capaces de utilizarlos de forma apropiada.

Dado que el INE ya está en disposición de producir datos vinculados, es el momento de que la comunidad investigadora española haga la parte del trabajo que le corresponde y que básicamente debiese consistir, en opinión de quien esto escribe, en: (1) intervenir en la producción de este tipo de datos vinculados, diseñando la estructura y contenido de las bases de datos requeridas; (2) supervisar, en su caso, el proceso de elaboración de los datos y controlar la calidad de los obtenidos; (3) acometer su explotación estadística sistemática; (4) analizarlos y convertirlos en productos de investigación, y (5) difundirlos a la comunidad científica y a los eventuales decisores —tanto públicos como privados— que pudieran tener interés en ellos. Tan pronto se alcancen, consoliden y difundan resultados de investigación sobre la realidad española basados en el análisis de datos vinculados, habrá que buscar socios científico-académicos en los otros países que ya disponen de este tipo de información para proceder a los necesarios estudios comparativos. A todas esas labores ayudaría, por cierto, una mayor rapidez y agilidad del INE en la producción y distribución de estos datos a los potenciales usuarios.

NOTAS

- [1] Por ejemplo, Polavieja (2020) examina el uso de datos administrativos de la Agencia Estatal de la Administración Tributaria para analizar la movilidad de ingresos en España y evalúa certeramente los sesgos de representación en que se incurre al hacerlo.

REFERENCIAS

- Argüeso, A. 2019. "Los censos de población y viviendas de 2021 en España se basarán en registros administrativos". *Índice* 74: 10-12.
- Bakker, B.F.M. (2009). *Trek alle registers open!* Amsterdam: Vrije Universiteit.
- Bakker, B.F.M. 2010. *Micro-integration. State of the Art*. Nueva York: United Nations.
- Bakker, B.F.M. y P.J.H. Daas. 2012. "Methodological challenges of register-based research". *Statistica Neerlandica* 66(1): 2-7.
- Bakker, B.F.M., van Rooijen, J. y L. van Toor. 2014. "The System of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics". *Statistical Journal of the IAOS* 30:411-424.
- Berka, Ch., Humer, S. y M. Moser. 2012. "Combination of evidence from multiple administrative data sources: quality assessment of the Austrian register-based Census 2011". *Statistica Neerlandica* 66(1): 18-33.
- Connelly, R., Playford, Ch.J., Gayle, V. y Ch. Dibben. 2016. "The role of administrative data in the big data revolution in social science research". *Social Science Research* 59:1-2.
- Daas, P.J.H. y S.J.L. Ossen. 2011. "Metadata quality evaluation of secondary data sources". *International Journal for Quality Research* 5: 57-66.
- European Commission. 2010. *Riding the wave. How Europe can gain from the rising tide of scientific data*. Final report of the High Level Expert Group on Scientific Data. (<http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>).
- Goldthorpe, J.H. 2017. *La Sociología como ciencia de la población*. Madrid: Alianza.
- Holmberg, A. 2012. "Discussion on assessing quality of administrative data". *Statistica Neerlandica* 66(1): 34-40.
- Holt. D.T. 2007. "The Official Statistics Olympic Challenge: Wider, Deeper, Quicker, Better, Cheaper. (With discussion)". *The American Statistician* 61:1-15.
- Holt. D.T. 2008. "Official statistics, public policy and public trust". *Journal of the Royal Statistical Society Series A*, 171: 1-20.
- Holt. D.T. 2000. "The Future for Official Statistics". *Journal of the Operational Research Society* 51(9): 1010-1019.
- Instituto Nacional de Estadística. 2011. *Proyecto de los Censos Demográficos 2011*. Madrid: Instituto Nacional de Estadística (http://www.ine.es/censos2011/censos2011_proyecto.pdf).
- Instituto Nacional de Estadística. 2017a. *Anteproyecto de los Censos de Población y viviendas 2021*. Madrid: Instituto Nacional de Estadística. Madrid: Subdirección General de Estadísticas Sociodemográficas, Instituto Nacional de Estadística (http://www.ine.es/censos2021/censos2021_anteproyecto.pdf#page=7).
- Instituto Nacional de Estadística. 2017b. *El Salario del Empleo Principal en la Encuesta de Población Activa. Explotación estadística de los Registros de la Seguridad Social y de las Agencias Tributarias*. Madrid: Instituto Nacional de Estadística (http://www.ine.es/metodologia/t22/nota_epa_decil.pdf).
- Lenk, M. 2008. *Methods of register-based census in Austria*. Viena: Statistics Austria.
- Lothian, J., Holmberg, A. y A. Seyb. 2017. "Linking Administrative Data: An Evolutionary Schema". *Administrative Data Paper ArXiv e-prints* (<https://arxiv.org/ftp/arxiv/papers/1712/1712.08522.pdf>).
- Polavieja, J.G. 2020. "Grandes Datos, Grandes Sesgos, Grandes Errores: Sobre el Atlas de Oportunidades". *Revista Internacional de Sociología* 78(3), en prensa.
- Poulain, M. 2010. "Le registre de population belge". Pp. 83-117 en *Histoire de la population de la Belgique et de ses territoires. Actes de la Chaire Quetelet 2005*, compilado por T. Eggerickx y J.-P. Sanderson. Lovaina: Presses Universitaires de Louvain.
- Poulain, M. y A. Herm. 2013. "Central population registers as a source of demographic statistics in Europe". *Population* 68(2): 183-212.
- Ramiro, D. 2019. "Hacia un nuevo paradigma en la producción estadística: Las infraestructuras de datos y los registros longitudinales de población". *Índice* 74: 4-5.
- Requena, M. y D. Reher. 2020. "Residential status and health in middle and late life: a population-based study with new data from Spain". *British Medical Journal Open* 10:e033330. doi:10.1136/bmjopen-2019-033330.
- Statistics Netherlands. 2017. *Population register data, basis for the Netherlands Population Statistics*. La Haya: Statistics Netherlands.
- United Nations Economic Commission for Europe (UNECE). 2010. *Main results of the UNECE-UNSD survey on the 2010-2011 round of censuses in the UNECE region*. Informe preparado para Eurostat Working Group on Demography and Censuses (Luxembourg, 19-20 April 2010) (<http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2010/mtg1/sp.1.e.pdf>).
- Wallgren, A. y B. Wallgren. 2014. *Register-based Statistics: statistical methods for administrative*. Chichester: Wiley.
- Zhang, L.-C. 2012. "Topics of statistical theory for register-based statistics and data integration". *Statistica Neerlandica* 66(1): 41-63.

MIGUEL REQUENA es Catedrático de Sociología en el Departamento de Sociología II (Estructura Social) en la UNED y miembro del Grupo de Estudios 'Población y Sociedad'. Licenciado y Doctor en Ciencias Políticas y Sociología por la Universidad Complutense de Madrid. Ha sido profesor en la Universidad Complutense de Madrid (1983-1988), miembro del Gabinete Técnico del Centro de Investigaciones Sociológicas (1984-1987), investigador en la Fundación Investigaciones Económicas y Sociales Aplicadas (1989-1992), Doctor Vinculado al Instituto de Estudios Sociales Avanzados del Consejo Superior de Investigaciones Científicas (1993-1994), Senior Associate Member en St. Antony College, Oxford University (1994-95), Profesor invitado en Hamilton College (1992-2001) y Visiting Professor en Princeton University (2002). Su investigación se ha centrado en sociología y demografía de la familia y sociología de la estructura y la estratificación social.