

Nota de campo

La utilidad del uso de datos administrativos en la investigación de salud pública: la *Muestra continua de vidas laborales*



María Andrée López^{a,b,*}, Fernando G. Benavides^{a,b}, Jordi Alonso^{a,c}, Mireia Espallargues^{a,d}, Xavier Durán^{a,b} y José Miguel Martínez^{a,b,c}

^a CIBER de Epidemiología y Salud Pública (CIBERESP): Evaluación de Servicios de Salud, España

^b Centro de Investigación en Salud Laboral, Universitat Pompeu Fabra, Barcelona, España

^c IMIM-Institut Hospital del Mar d'Investigacions Mèdiques, Barcelona, España

^d Agència de Qualitat i Avaluació Sanitàries (AQuAS), Barcelona, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 20 de diciembre de 2013

Aceptado el 12 de febrero de 2014

On-line el 31 de marzo de 2014

Palabras clave:

Conexión de registros

Datos administrativos

Estudios de cohorte

Vigilancia de la salud pública

R E S U M E N

La utilización de datos administrativos es práctica habitual en la investigación en salud pública. El objetivo de este trabajo es describir la *Muestra continua de vidas laborales* (MCVL) y su utilidad en la investigación en salud. La MCVL está construida sobre registros que generan todos los contactos con la Seguridad Social (laborales, incapacidades, etc.), más datos fiscales (percepciones dinerarias, ingresos, etc.) y del padrón (nivel educativo, país de nacimiento, etc.), sin que sea posible la identificación individual. La MCVL se inició en 2004 con el 4% (1,1 millón de personas) del total de población cotizante o perceptora de la Seguridad Social. Los datos de las personas que forman parte de la MCVL son actualizados anualmente, y se sustituyen las bajas producidas. Este diseño continuo permite generar una cohorte con información de la vida laboral y económica, y valorar su asociación con la incapacidad laboral. El supuesto de conexión con la historia clínica posibilitaría el análisis de otros resultados de salud.

© 2013 SESPAS. Publicado por Elsevier España, S.L. Todos los derechos reservados.

The value of using administrative data in public health research: the Continuous Working Life Sample

A B S T R A C T

The use of administrative data is common practice in public health research. The present field note describes the Continuous Working Life Sample (CWLS) and its use in health research. The CWLS is built on records generated by all contacts with the social security system (work contracts, disability, etc.), plus tax data (monetary gains, income, etc.) and census data (level of education, country of birth, etc.), but does not allow individuals to be identified. The CWLS was started in 2004 with 4% (1.1 million persons) of the total population who were either contributors to or beneficiaries of the social security system. The information on the individuals in the CWLS is updated annually and lost individuals are replaced. This continuous design allows the construction of a cohort with information on working life and financial status and evaluation of their relationship with work disability. Future connection with clinical records would enable analysis of other health-related outcomes.

© 2013 SESPAS. Published by Elsevier España, S.L. All rights reserved.

Introducción

Desde que John Graunt (1620-1674) y William Farr (1807-1883) hicieran uso de los registros de mortalidad en Inglaterra y Gales para cuantificar los principales problemas de salud en su tiempo¹, muchos y diversos son los registros administrativos que han mostrado su utilidad para investigar y vigilar diferentes problemas de salud en la población. Un ejemplo clásico del uso de registros administrativos en salud pública lo constituyen los registros británicos². En algunos estudios, los datos administrativos han sido complementados conectándolos (*record linkage*) con datos recogidos mediante encuestas o datos clínicos procedentes de la historia

clínica². Actualmente, los registros escandinavos constituyen otro ejemplo de utilización de registros administrativos para realizar estudios epidemiológicos. Entre ellos destaca el basado en conectar el censo poblacional, el registro de calidad de agua y las altas hospitalarias en Finlandia, para determinar la relación entre la concentración de fluoruro en el agua potable y las fracturas de cadera³.

En España, desde 2004 está disponible para los investigadores una base de datos administrativa ofrecida por la Seguridad Social, la *Muestra continua de vidas laborales* (MCVL), que incluye el 4% de las personas que tienen relación con la Seguridad Social, sea como cotizantes (p. ej., asalariados) o como beneficiarios (pensionistas). La MCVL se ha utilizado principalmente en estudios laborales y económicos, entre ellos los que evalúan las tendencias de entrada a la jubilación y la duración de las relaciones laborales⁴. Sólo recientemente se ha empezado a utilizar la MCVL en salud pública^{5,6}. El objetivo de esta nota de campo es describir la MCVL y mostrar sus posibles utilidades en la investigación en salud pública.

* Autora para correspondencia.

Correos electrónicos: andree.gomez01@estudiant.upf.edu, mariaandree.lg@gmail.com (M.A. López).

Tabla 1Variables que se encuentran en los ficheros que conforman la *Muestra continua de vidas laborales* (MCVL)

Ficheros	Variables
<i>Datos personales</i>	<ul style="list-style-type: none"> ■ Fecha de nacimiento ■ Sexo ■ Nacionalidad ■ Provincia de nacimiento^a ■ Provincia de primera afiliación ■ Domicilio ■ Fecha de fallecimiento ■ País de nacimiento^a ■ Nivel educativo^a
<i>Datos de convivientes</i>	<ul style="list-style-type: none"> ■ Fecha de nacimiento de convivientes^a ■ Sexo de convivientes^a
<i>Datos afiliados</i>	<ul style="list-style-type: none"> ■ Régimen de cotización ■ Grupo de cotización ■ Tipo de contrato ■ Coeficiente de tiempo parcial ■ Fechas reales de alta y de baja ■ Causa de la baja ■ Minusvalía según alta de afiliación ■ Actividad económica de la cuenta de cotización (CNAE 2009) ■ Número de trabajadores en la cuenta de cotización ■ Fecha de alta del primer trabajador ■ Identificación de la cuenta de cotización (secundaria) ■ Fechas reales de alta y de baja ■ Fechas de efecto del alta y de la baja ■ Grupo de cotización ■ Año del dato ■ Clase de la prestación ■ Situación del sujeto causante ■ Grado de incapacidad ■ Fecha de minusvalía ■ Norma SOVI ■ Clase de mínimo ■ Régimen de la pensión ■ Fecha de efectos económicos de la pensión ■ Base reguladora ■ Porcentaje aplicado a la base reguladora ■ Años bonificados ■ Años considerados cotizados para la jubilación ■ Importe mensual de la pensión efectiva ■ Importe mensual de revalorización ■ Importe mensual de complementos garantía mínimos ■ Importe mensual de otros complementos ■ Identificador anonimizado de la entidad pagadora ■ Provincia ■ Clave de percepción ■ Subclave ■ Percepción íntegra (dineraria) ■ Retenciones practicadas ■ Percepciones en especie (valoración) ■ Ingresos a cuenta efectuados ■ Ingresos a cuenta repercutidos ■ Ceuta o Melilla ■ Ejercicio de nacimiento ■ Tipo de relación laboral ■ Colectivo del trabajador ■ Tipo de empleador y de entidad ■ Domicilio social: provincia ■ Fecha de modificación de contrato ■ Tipo de contrato hasta la modificación ■ Coeficiente de tiempo parcial hasta la modificación ■ Fecha de modificación de grupo de cotización ■ Grupo de cotización hasta la modificación ■ Actividad económica, clasificación antigua (C93C) ■ Año de cotización ■ Base de cotización mensual cuenta ajena por contingencias comunes ■ Base de cotización cuenta propia y otros ■ Total base de cotización del año ■ Situación de la prestación (causa baja) ■ Fecha de situación de la prestación ■ Provincia de gestión de la prestación ■ Número de titulares de un mismo sujeto causante ■ Prorrata de convenio internacional ■ Prorrata de divorcio ■ Coeficiente reductor total ■ Tipo de situación de jubilación ■ Coeficiente de parcialidad (jubilación) ■ Prestación vitalicia (orfandad) ■ Concurrencia con prestación ajena ■ Importe anual pagas extra ■ Importe anual paga desviación IPC ■ Importe anual total de la prestación ■ Año de nacimiento del causante de pensión de supervivencia ■ Pensión limitada ■ Importe mensual total de la prestación ■ Situación familiar ■ Discapacidad ■ Tipo de contrato o relación ■ Prolongación actividad laboral ■ Movilidad geográfica ■ Reducciones ■ Gastos deducibles ■ Pensiones compensatorias ■ Anualidades por alimentos ■ Número de descendientes ■ Número de ascendentes
<i>Bases de cotización</i>	
<i>Datos prestaciones</i>	
<i>Datos fiscales</i>	

IPC: índice de precios al consumo.

^a Datos que se extraen del Padrón continuo.

Muestra continua de vidas laborales

La MCVL se inició en 2004, desde la Dirección General de Ordenación de la Seguridad Social, con una muestra representativa del 4% de los individuos afiliados en alta a la Seguridad Social en ese año⁴. Desde entonces, cada año se genera una nueva muestra que, al utilizar los mismos algoritmos, vuelve a seleccionar a los que ya estaban, si siguen en contacto con la Seguridad Social, y en caso de que no lo estén (por desempleo y finalización de su prestación asociada, o por muerte del beneficiario o afiliado) son sustituidos por personas con similares características, que pasarán a formar parte de la MCVL hasta que dejen de estar en contacto con la Seguridad Social. De este modo, la muestra es similar a una muestra tipo panel, donde se recoge información sobre las mismas personas en diferentes olas, lo que permite mantener la representatividad de la muestra y acumular la información que se haya generado en ese año. Por ejemplo, cambios de asalariado a autónomo, o pasar a pensión por incapacidad permanente o por jubilación. Así, puede construirse una cohorte dinámica de individuos que son seguidos

mientras estén en contacto con la Seguridad Social. En un seguimiento desde 2004 a 2009, basado en los que son cotizantes del régimen general, el 88% de los que estaban en 2004 fueron seguidos durante todo el período⁶.

Para cada persona se dispone de datos procedentes de diversos registros administrativos, la mayoría de la Seguridad Social, pero también fiscales y demográficos. En la [tabla 1](#) se resume toda la información disponible para cada persona, como tipos de contrato, con fechas de altas y bajas, grupo de cotización y bajas por incapacidad permanente, lo que se complementa con información de ese año procedente de la Agencia Tributaria y del Padrón continuo.

Ventajas y limitaciones de la MCVL en salud pública

La riqueza de datos hace de la MCVL una herramienta útil en varios ámbitos de investigación en salud pública. Por un lado, permite reconstruir la trayectoria laboral de cada persona desde su primer contrato hasta su jubilación, identificando en estas personas la fecha de inicio y finalización de cada contrato, junto con el

Tabla 2
Características de los individuos de la *Muestra continua de vidas laborales* (MCVL) correspondiente a 2004

Características de todos los incluidos en la MCVL (n = 1.089.490)	n (%)		n (%)
Sexo		Actividad económica^e	
Mujeres	617.737 (56,7)	Agricultura, ganadería, pesca e industrias extractivas	26.856 (2,6)
Hombres	471.753 (43,3)	Industria manufacturera	95.446 (9,2)
Edad (años)^a		Producción y distribución de energía	3.483 (0,3)
<30	203.585 (18,7)	Construcción	81.723 (7,9)
31-45	363.942 (33,4)	Comercio	103.814 (10,0)
46-60	230.125 (21,1)	Hostelería, transporte y telecomunicaciones	77.809 (7,5)
>60	291.833 (26,8)	Intermediación financiera	15.526 (1,5)
Comunidad autónoma^b		Actividades inmobiliarias	87.762 (8,5)
Andalucía	173.673 (15,9)	Administración pública	200.439 (19,4)
Aragón	26.521 (2,4)	Educación, actividades sanitarias y servicios a la comunidad	85.069 (8,2)
Asturias	22.470 (2,1)	Actividades en los hogares	9.435 (0,9)
Cantabria	12.248 (1,1)	Organizaciones extraterritoriales	127 (0,012)
Castilla-La Mancha	34.571 (3,2)	Clase social ocupacional según grupo de cotización del régimen general^f	
Castilla y León	50.969 (4,7)	Cualificados no manuales	108.731 (15,5)
Cataluña	157.100 (14,4)	Cualificados manuales	234.079 (33,4)
Ceuta y Melilla	2.011 (0,2)	No cualificados no manuales	219.902 (31,4)
Comunidad Valenciana	97.480 (9,0)	No cualificados manuales	137.822 (19,7)
Extremadura	20.887 (1,9)		
Galicia	58.340 (5,4)		
Islas Baleares	19.603 (1,8)		
Islas Canarias	37.716 (3,5)		
La Rioja	6.302 (0,6)		
Madrid	127.631 (11,7)		
Murcia	26.275 (2,4)		
Navarra	12.841 (1,2)		
País Vasco	47.767 (4,4)		
Relación con la Seguridad Social			
Cotizantes^c			
Régimen general	796.924 (77,0)		
Régimen autónomos y especiales	237.451 (23,0)		
Beneficiarios^d			
Jubilación	176.186 (57,7)		
Viudedad, orfandad y otros	94.054 (26,1)		
Incapacidad	35.261 (11,5)		

^a No se dispone de información sobre la edad de cinco afiliados.

^b No se dispone de información sobre la comunidad autónoma de 155.085 (14,2%) afiliados.

^c Algunos cotizantes también son beneficiarios, en total son 1.034.375 cotizantes.

^d Algunos beneficiarios también son cotizantes, en total son 305.501 beneficiarios.

^e No se dispone de información de actividad económica de 246.886 (23,9%) afiliados.

^f Corresponde a los afiliados cotizantes del régimen general, que en total son 700.534 afiliados.

tipo de éste, su base salarial y de cotización, la actividad económica de la empresa, o los períodos en que recibe la prestación por desempleo (alguno de estos datos pueden conseguirse desde 1980). Asimismo, se añade la fecha de inicio de una prestación por incapacidad permanente por enfermedad o accidente, con su grado de incapacidad⁶. Además, si el cotizante o beneficiario incluido en la MCVL está afiliado en el momento de su muerte, el registro informa también de la fecha de defunción.

La principal ventaja de la MCVL es que está basada en registros administrativos, lo que representa un reducido coste en comparación con el coste de estudios con datos primarios, en especial si son de seguimiento⁷. También, en comparación con los autorreportados, los datos de la MCVL son fiables respecto a ítems como la situación laboral, el salario, la edad de jubilación o el grado de incapacidad permanente. Además, debemos señalar como tercera ventaja su gran tamaño muestral⁷ (alrededor de un millón de personas) y su representatividad de la población de referencia, que se mantiene desde 2004 (ver [tabla 2](#)), y ya está disponible la edición de 2012⁸.

Desde el punto de vista de salud pública, la principal limitación es que la MCVL apenas incluye datos de salud, exceptuando el tipo de contingencia para la incapacidad permanente (contingencia común o profesional). Tampoco las defunciones cuentan con su causa básica. Una segunda limitación es que en la MCVL

encontramos sólo datos sobre personas que están cubiertas por la Seguridad Social, y por lo tanto quedan fuera los cotizantes a otros sistemas de previsión, como MUFACE (funcionarios), ISFA (militares) y MUGEJU (administración de justicia), si bien estos sistemas están en proceso de convergencia hacia el Sistema General de Seguridad Social⁴. Tampoco se tiene información de las personas que trabajan en la economía sumergida, y por tanto sin cobertura de la seguridad social.

Perspectivas de futuro

La MCVL representa una oportunidad singular para el futuro de la investigación en salud pública en España. Por ejemplo, desde el Programa de investigación de evaluación de programas sanitarios del CIBER de Epidemiología y Salud Pública (www.ciberesp.es) se ha iniciado un estudio de carga de la enfermedad por incapacidad laboral basado en la MCVL.

Algunas experiencias previas han avanzado en esta dirección. Por ejemplo, recientemente se han incorporado a la MCVL de 2009 los datos de incapacidad temporal con el diagnóstico médico al alta, y su fecha de inicio y finalización, que permiten identificar si esos episodios se han producido en la misma persona⁹. Estos datos están facilitando analizar la incidencia de la incapacidad temporal, infrecuente en la literatura y muy útil, ya que existe la posibilidad de

estimar con precisión los tiempos que cada persona ha estado en riesgo de poder padecer una incapacidad temporal, permitiendo descontar los períodos en que no está activo en una empresa y con el régimen que le cubre esa contingencia. De esta forma, la incidencia de la incapacidad temporal puede estimarse mediante una tasa, a diferencia de estudios que únicamente han podido utilizar la incidencia acumulada¹⁰. La repetición de esta conexión entre la MCVL y los datos de incapacidad temporal permitiría analizar trayectorias de incapacidad temporal y su asociación con la incapacidad permanente.

Finalmente, otra potencialidad es la de conectar la MCVL con información de algunos registros sanitarios, manteniendo la protección de la identidad de las personas que forman parte de la muestra, como ocurre ahora, y tras la supervisión de los respectivos comités de ética.

Editor responsable del artículo

Alberto Ruano-Ravina.

Contribuciones de autoría

F.G. Benavides concibió la idea principal de dar a conocer a la comunidad científica el uso de la *Muestra continua de vidas laborales* en estudios de salud pública. La propuesta del manuscrito original la realizaron M.A. López y F.G. Benavides. Las tablas descriptivas de la muestra las elaboró M.A. López con la contribución de F.G. Benavides, J. Alonso, X. Durán y J.M. Martínez. Todos los autores hasta ahora mencionados, con la contribución de M. Espallargues, realizaron una revisión crítica del manuscrito y aportaron contribuciones intelectuales al contenido de la nota. Todos los autores aprobaron la versión final para su publicación.

Financiación

Ninguna.

Conflictos de intereses

Ninguno.

Bibliografía

1. McMichael AJ. Prisoners of the proximate: loosening the constraints on epidemiology in an age of change. *Am J Epidemiol.* 1999;149:887–97.
2. Acheson ED. Record linkage techniques in studies of the aetiology of cancer. *Proc R Soc Med.* 1968;61:726–30.
3. Kurttio P, Gustavsson N, Vartiainen T, Pekkanen J. Exposure to natural fluoride in well water and hip fracture: a cohort analysis in Finland. *Am J Epidemiol.* 1999;150:817–24.
4. Durán A. La Muestra continua de vidas laborales de la Seguridad Social. *Rev del Minist Trab y Asuntos Soc.* 2007;231–40 (Consultado el 27/09/2013.) Disponible en: <http://www.empleo.gob.es/es/publica/pub.electronicas/destacadas/revista/numeros/ExtraSS07/Est09.pdf>
5. Solé M, Díaz-Serrano L, Rodríguez M. Disparities in work, risk and health between immigrants and native-born Spaniards. *Soc Sci Med.* 2013;76:179–87.
6. Durán X, Martínez JM, Benavides FG. Occupational factors associated with the potential years of working life lost due to a non-work related permanent disability. *Work.* 2013;45:305–9.
7. Virnig BA, McBean M. Administrative data for public health surveillance and planning. *Annu Rev Public Health.* 2001;22:213–30.
8. Seguridad Social: estadísticas, presupuestos y estudios. (Consultado el 25/10/2013.) Disponible en: http://www.seg-social.es/Internet_1/Estadistica/Est/Muestra.Continua.de.Vidas.Laborales/index.htm
9. Malo MA, Cueto B, García C, et al. La medición del absentismo: estimaciones desde la perspectiva de las empresas y de las vidas laborales. 2012. (Consultado el 02/12/2013.) Disponible en: <http://www.seg-social.es/prdi00/groups/public/documents/binario/174212.pdf>
10. Ármannsdóttir B, Mårdby A-C, Haukenes I, et al. Cumulative incidence of sickness absence and disease burden among the newly sick-listed, a cross-sectional population-based study. *BMC Public Health.* 2013;13:329.