

# LA RED DE MÉDICOS CENTINELAS DE CASTILLA Y LEÓN: APLICACIÓN DEL ANÁLISIS DE CONGLOMERADOS PARA LA OBTENCIÓN DE UNA POBLACIÓN REPRESENTATIVA

A. T. Vega Alonso<sup>1</sup> / M. Gil Costa<sup>1</sup> / C. Ruiz Cosin<sup>1</sup> / E. Zapatero Villalonga<sup>1</sup> /  
Red de Médicos Centinelas de Castilla y León<sup>2</sup>

<sup>1</sup>Dirección General de Salud Pública y Asistencia. Junta de Castilla y León

<sup>2</sup>Secciones de Epidemiología de los Servicios Territoriales Provinciales de Bienestar Social y Médicos de Atención Primaria Participantes

## Resumen

Con el objetivo de establecer en Castilla y León una Red de Médicos Centinelas encargada de recoger de manera sistemática datos referidos a la morbilidad de la población, se procedió a una selección aleatoria de médicos de Atención Primaria cuya población atendida fuese representativa de la población total de la región. Para ello se agruparon mediante el análisis de conglomerados (*cluster analysis*) aquellas Zonas Básicas de Salud similares entre sí, según una serie de variables consideradas importantes en la distribución de las enfermedades. Fueron formados cinco *clusters* o agrupaciones en el medio urbano y semiurbano y veinte en el medio rural, que posteriormente sirvieron para una estratificación antes de efectuar el muestreo aleatorio de los facultativos participantes. Los resultados de la distribución en cada *cluster* de los médicos de Atención Primaria del conjunto de la comunidad y los médicos centinelas no muestran diferencias estadísticamente significativas. La diferencia hallada al comparar la edad media de la población de médicos de Castilla y León y los 127 médicos centinelas, debida a la voluntariedad de participación, no se encuentra al estudiar y comparar las poblaciones respectivas, lo que demuestra la eficacia de esta metodología en la selección de una población representativa a partir de un muestreo aleatorio de grupos de población.

**Palabras clave:** Muestra. Médicos centinelas. Análisis de conglomerados.

## THE SENTINEL PHYSICIAN'S NETWORK IN CASTILLA Y LEÓN: USE OF CLUSTER ANALYSIS TO OBTAIN A REPRESENTATIVE POPULATION

### Summary

In order to establish a sentinel physicians network in Castilla y León to collect systematically population-based morbidity data, a random sample of General Practitioners (GP's) whose covered population was representative of the regional population was obtained. A cluster analysis with the ZBS (Zonas Básicas de Salud) was performed according to a list of variables considered important in the diseases incidence. Five clusters were obtained in the urban areas and twenty in the rural areas, where, after stratification, the GP's random sample was selected. The outcome of the distribution, within each cluster, between all Castilla y León GP's and the sentinel GP's did not show statistically significant differences. The statistically significant difference found between the age of all GP's and the 127 sentinel GP's, due to the voluntary participation, was not found in the comparison of the two populations covered by them, confirming the efficacy of this method in the selection of a representative population from a random sampling of population groups.

**Key words:** Sample. Sentinels GP's. Cluster analysis.

## Introducción

Los sistemas de registro de enfermedades con base hospitalaria o de población (cáncer, malformaciones congénitas, infarto

agudo de miocardio, etc.), los sistemas de Enfermedades de Declaración Obligatoria, los registros en la Atención Primaria de Salud, junto a las encuestas y otros estudios puntuales constituyen el grueso de los sistemas de informa-

ción sobre morbilidad existentes en los países occidentales.

Debido a las dificultades de implantación de estos sistemas de información y al coste elevado que suponen algunos de ellos (como es el caso de los

Correspondencia: A.T. Vega Alonso. Dirección General de Salud Pública y Asistencia. Consejería de Cultura y Bienestar Social. Avda. Burgos, nº 5. 47071 VALLADOLID

Este artículo fue recibido el 16 de agosto de 1989 y fue aceptado, tras revisión, el 23 de abril de 1990.

registros), unido a las imperfecciones aún existentes en otros, se ha comenzado en las últimas décadas a desarrollar un nuevo sistema de información epidemiológica, basado en redes de médicos que son elegidos de forma aleatoria y cuya participación es voluntaria. Este sistema de recogida de información permite hacer estimaciones válidas sobre la enfermedad y su distribución en la población<sup>1</sup>.

Estas redes, denominadas comúnmente redes de médicos centinelas o vigías, podrían considerarse como una combinación de elementos metodológicos del sistema tradicional de Enfermedades de Declaración Obligatoria o los sistemas de información en Atención Primaria y de las encuestas de tipo muestral, aunque en general tienen unas características y unos objetivos bastante diferentes de éstos<sup>2,3</sup>.

Una de esas características es que la participación voluntaria de un número reducido de médicos asegura una homogeneidad de criterios, una fiabilidad elevada en los resultados y permite el control permanente del sistema de registro. Por otra parte, el muestreo aleatorio, pese a la voluntariedad de participación, permite una mayor exactitud así como conocer y reducir al máximo los sesgos sistemáticos, de aparición frecuente en otros sistemas de información<sup>4</sup>.

Lobet et al. estudiaron la aplicación del análisis de conglomerados (*cluster analysis*) para la construcción de estratos homogéneos a partir de delimitaciones administrativas pequeñas, según el valor tomado por una serie de variables relacionadas con la morbilidad (aparición, frecuencia, distribución y consecuencias)<sup>5</sup>.

El objetivo de este trabajo es presentar la metodología utilizada en la implantación de la Red de Médicos Centinelas en Castilla y León, aplicando los modelos empleados con éxito en otros países, a las especiales características de nuestra organización sanitaria.

Para asegurar la validez y la representatividad de los resultados que se obtengan se ha procedido en primer lugar a un muestreo aleatorio de médicos generales que estén en situación activa en cualquiera de las nueve provin-

cias de la región y, en una segunda fase, al estudio de los resultados obtenidos en este muestreo, con la comparación, por una parte de los médicos centinelas con la población total de médicos generales, y por otra, la población cubierta por dichos médicos centinelas con la población castellano-leonesa.

## Material y métodos

Se utilizaron como base las Zonas Básicas de Salud (ZBS) a partir de las cuales se constituyeron los *clusters* o agrupaciones homogéneas. Una primera división fue la separación de ZBS rurales de las urbanas o semirurales según el nivel de jerarquía, basada en la población, de los municipios cabecera de cada Zona Básica de Salud que se recogen en el *Estudio de Comarcalización de Castilla y León*<sup>6,7</sup>.

Para cada una de las ZBS se calcularon los valores de variables relacionadas con la demografía, las condiciones socioeconómicas y las condiciones sanitarias básicas. Estas variables fueron escogidas a partir de: a) su estrecha relación con la distribución de las enfermedades (así se limita la dispersión interna en cada estrato o conglomerado obtenido de la variable estudiada, es decir, de las enfermedades en general), y b) la posibilidad de obtener datos desagregados por municipios y/o por ZBS<sup>8,9,10</sup>.

Las variables demográficas seleccionadas fueron la densidad de población por km<sup>2</sup> y los porcentajes de población menor de 16 años y mayor de 65 años. Entre las variables de tipo socioeconómico se introdujo la tasa de actividad en personas entre 16 y 65 años, la tasa de desempleo y el porcentaje de trabajadores en los sectores primario y terciario. Para las variables relacionadas con las condiciones sanitarias básicas, se utilizó la tasa de médicos por 1.000 habitantes, el porcentaje de personas que tienen una accesibilidad al Centro de Salud en menos de 5 minutos, y las que la tienen en más de 15 minutos.

Todos los valores fueron calculados por municipios con la aplicación de los datos del censo de población de 1981

a las poblaciones de habitantes de hecho reflejadas en el Padrón Municipal de Habitantes de 1986. Posteriormente se agruparon por ZBS y se calcularon los valores definitivos en tasas o porcentajes. En aquellos municipios con más de una ZBS, la información fue procesada como si se tratase de una sola ZBS.

Se procedió (separadamente para las zonas urbanas y semiurbanas por un lado, y rurales por otro) a la estandarización de los valores de cada ZBS con relación a la media de cada variable, con cálculo de su valor Z, para anular el peso específico de las distintas unidades en que están medidas las variables en la construcción de la matriz de distancias del análisis. La fórmula utilizada fue la de la distancia euclídea: distancia entre ZBS<sub>1</sub> y ZBS<sub>2</sub> =  $[\sum (ZBS_1 i - ZBS_2 i)^2]^{1/2}$ , siendo ZBS<sub>1</sub> y ZBS<sub>2</sub> dos ZBS diferentes, donde *i* representa las diferentes variables de cada zona y *r* = 2. Es decir, la distancia entre la ZBS<sub>1</sub> y la ZBS<sub>2</sub> es la raíz cuadrada de la suma de las diferencias entre los valores que cada variable toma en cada ZBS, elevadas al cuadrado<sup>11</sup>.

Una vez calculadas las matrices de distancia se procedió a agrupar jerárquicamente las ZBS siguiendo el orden creciente de valores de la matriz. El método utilizado fue el UPGMA (*Unweighted pair-group method using arithmetic averages*, también conocido como *Average Linkage Between Groups*), que define la distancia entre dos *clusters* como la media de las distancias entre todos los pares de casos, estando un miembro del par en cada *cluster*. Así por ejemplo, si en el *cluster* A tenemos la ZBS<sub>1A</sub>, la ZBS<sub>2A</sub> y la ZBS<sub>3A</sub>, y en el *cluster* B la ZBS<sub>1B</sub> y la ZBS<sub>2B</sub>, la distancia entre los *clusters* A y el B es la media aritmética de las distancias entre los posibles pares (media aritmética de la distancia entre ZBS<sub>1A</sub> y ZBS<sub>1B</sub>, ZBS<sub>1A</sub> y ZBS<sub>2B</sub>, ..., ZBS<sub>3A</sub> y ZBS<sub>2B</sub>). En cada etapa de la agrupación se crea una nueva matriz formada por las distancias calculadas entre los *clusters* ya formados. Para estos cálculos se ha utilizado el programa SPSS/PC+<sup>12</sup>.

El número de *clusters* se seleccionó haciendo el corte en la agrupación jerárquica cuando la distancia para la

unión de los dos *clusters* siguientes superaba el valor medio del rango que entre 1 y 25 mide la diferencia entre las distancias máxima y mínima.

La muestra de médicos destinada a efectuar un primer contacto solicitando su participación voluntaria en el proyecto fue seleccionada de manera aleatoria simple en el interior de cada *cluster*, con la ayuda de una tabla de números aleatorios y en una proporción de aproximadamente 1 médico por cada 10.000 habitantes, lo que permitía obtener una población cubierta dos veces mayor que la necesaria, y compensar el 50 % previsto de médicos que no iba a estar interesado en el programa. La lista de médicos se obtuvo a través de las direcciones provinciales del Insalud y comparada con la de los colegios oficiales de médicos de cada provincia, de donde también se obtuvo la edad y el sexo de los facultativos (datos no publicados sistemáticamente).

Se realizaron tres tipos de análisis para comprobar la representatividad de la muestra: se analizó la distribución de los médicos en los *clusters* para estudiar la representatividad geográfica, comparando en cada *cluster* el número total de médicos con el número de médicos centinelas (participantes finales en el proyecto). Por otra parte, se estudió la edad y el sexo de los médicos que decidieron participar, y se comparó con la edad y el sexo de la población de médicos generales de Castilla y León. El tercer y más importante análisis que se efectuó fue la comparación entre la población cubierta por los médicos centinelas y la población de Castilla y León.

Se estimó la población cubierta por cada médico centinela a partir de los datos del Padrón Municipal de Habitantes de 1986. En aquellos municipios donde existe más de un médico (medio urbano y semiurbano) se hizo un cálculo proporcional al número de facultativos de medicina general. A estas cifras de población se le aplicaron, como en el caso de las variables utilizadas en la construcción de los conglomerados, los porcentajes por municipio de los grupos de edad, población empleada y sector de producción reflejados en el censo de 1981.

## Resultados

De las 21 ZBS urbanas o semiurbanas se formaron 5 *clusters* o agrupaciones y 20 entre las 146 consideradas como rurales.

Del total de médicos preseleccionados en todos los *clusters* (250), 127 (50,8 %) decidieron voluntariamente participar en el proyecto.

En las tablas 1 y 2 se presenta la distribución de los médicos centinelas y el total de Castilla y León en cada uno de los *clusters* formados en las ZBS rurales y urbanas. El test de  $X^2$  entre el número total de médicos y los centinelas no muestra diferencias estadísticamente significativas ( $p > 0,05$ ). Tampoco aparecen diferencias significativas entre el total de médicos y los preseleccionados, y entre los preseleccionados y los centinelas.

En la tabla 3 se observa que no existen diferencias significativas en cuanto al sexo de los médicos centinelas. Sin embargo, la edad media de los mismos es significativamente más baja que la de la población de médicos generales de la región, tanto para los hombres como para las mujeres.

Las distribuciones por sexo y por actividad laboral de la población cubierta por la red de médicos centinelas no difieren de la población general de Castilla y León. Así, los hombres representan el 49,6 % de la población cubierta y el 49,7 % en la de la comunidad, y la población activa supone el 43 % y el 43,3 %, respectivamente.

Las figuras 1 y 2 reflejan la distribución de dos de las variables analizadas en la población: la edad y los sectores de producción, no observándose diferencias importantes.

**Tabla 1. Distribución en los clusters de la población total de médicos generales y los médicos centinelas en las ZBS urbanas**

Cluster	Total médicos	Médicos centinelas
1	116	6
2	397	28
3	19	1
4	68	3
5	51	5
TOTAL	651	43
$\chi^2_4 = 1,70$ $P = 0,7910$		

**Tabla 2. Distribución en los clusters de la población total de médicos generales y los médicos centinelas en las ZBS rurales**

1	790	46
2	103	5
3	220	12
4	22	0
5	10	1
6	9	1
7	4	1
8	7	1
9	55	3
10	22	2
11	9	0
12	33	3
13	9	1
14	9	1
15	18	2
16	52	1
17	10	0
18	12	1
19	11	2
20	9	1
TOTAL	1414	84
$\chi^2_5 = 2,80^*$ $P = 0,7314$		

\* Se ha considerado una tabla de  $2 \times 6$  agrupando todos los médicos de los *clusters* 6 al 20 para evitar tener muchas celdas con  $n < 5$

**Tabla 3. Análisis comparativo de la edad y el sexo entre la población de médicos centinelas y la de médicos generales**

	N	Varones (EE) <sup>1</sup> (%)	Edad media (EE) <sup>1</sup>
Médicos generales	2.065	76,6	45,94 <sup>2</sup>
Médicos centinelas	127	78,7 (0,037)	39,02 <sup>2</sup> (1,036)

Test Z para la significación de un solo parámetro

<sup>1</sup> EE: Error estándar

<sup>2</sup>  $P < 0,001$



Figura 1. Distribución comparativa por grupos de edad.

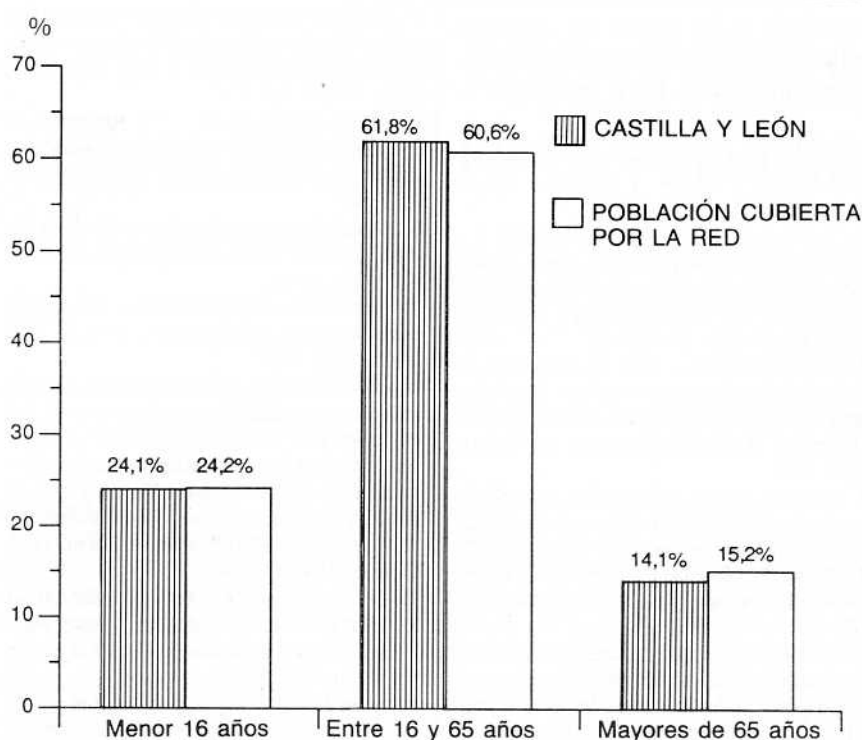
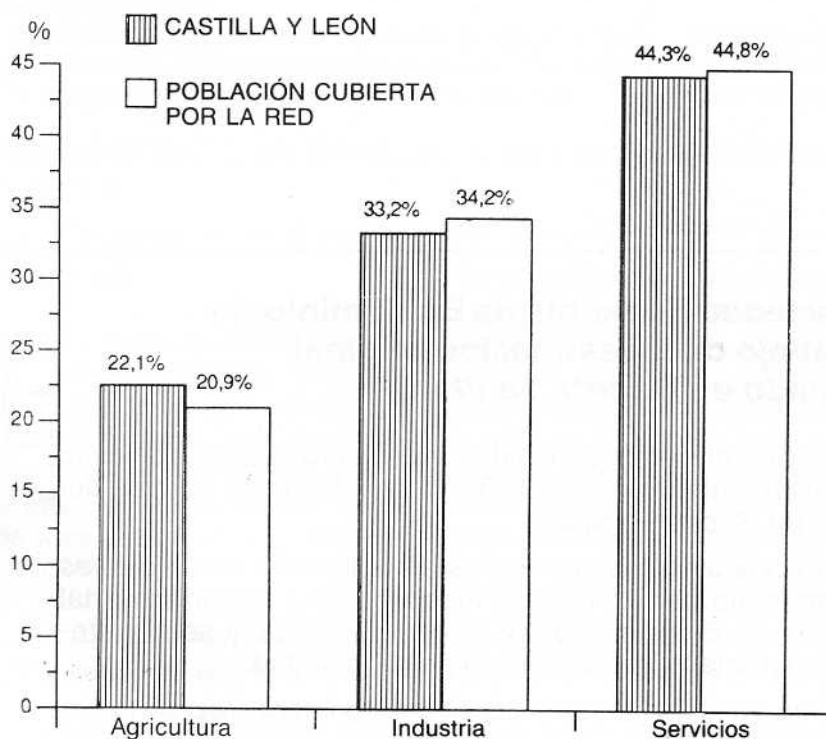


Figura 2. Distribución comparativa por sectores económicos de producción.



## Discusión

La primera cualidad de una muestra es la de ser representativa, es decir, la de proporcionar estimadores no sesgados. Para ello es necesario que todos los individuos de la población tengan una probabilidad conocida y diferente de cero de figurar en la muestra<sup>13</sup>.

El mayor problema que se planteó en nuestro muestreo es el hecho de que vamos a utilizar como base muestral la población de médicos generales de la comunidad autónoma y no la población general, que va a ser en definitiva de donde se obtengan los datos y a la cual se extrapolen los resultados. De esta manera, pese a obtener con un muestreo aleatorio simple unos estimadores no sesgados, la precisión de dichos estimadores será pequeña.

Si consideramos que la precisión de un estimador es inversamente proporcional a la varianza de ese estimador, cuya fórmula es  $\sigma^2/n$ , podemos aumentar dicha precisión de dos maneras, aumentando  $n$ , lo que nos obligaría a una cobertura mayor de la población con la consiguiente pérdida de eficacia y de control del sistema de registro, o bien reduciendo la libertad que le dejamos al azar para la elección de cada unidad de muestreo, es decir, estratificando.

La estratificación de la población en estratos o subpoblaciones más homogéneas donde se hace el sorteo aleatorio supone que las variaciones dentro de cada estrato son más pequeñas que la media total de la población y, por tanto, los estimadores dentro de esos estratos son mucho más precisos que un estimador para toda la población.

El segundo problema a solucionar era: ¿qué criterios elegir para la estratificación de la población? Los criterios únicos, como las delimitaciones geográficas o administrativas (provinciales, rurales, urbanas), parecen insuficientes, y los criterios más específicos, como edad, sexo, profesión, etc., son inviables a la hora de seleccionar a médicos que tienen un cupo fijo de personas, además de reducir considerablemente las posibilidades de elección si los estratos resultaran excesivamente pequeños.

En nuestro caso, el muestreo entre los médicos de Atención Primaria, destinado a obtener una población representativa para un programa de recogida de información epidemiológica, demuestra la eficacia del sistema, pese a las diferencias encontradas entre los médicos centinelas y el total de los médicos generales de Castilla y León (que fue la base muestral empleada) y cuya causa más probable sería la voluntariedad de participación. Es poco probable que dichas diferencias interfieran en la extrapolación de los resultados obtenidos por la red, puesto que éstos corresponderán a una población representativa de la total regional. El hecho de que 50,8 % de los médicos preseleccionados decidieran voluntariamente participar supone un porcentaje superior al obtenido en otros proyectos europeos.

Como conclusión podemos decir que el análisis de conglomerados efectuado en nuestra comunidad autónoma, utilizando variables relacionadas con la distribución de las enfermedades y calculadas para cada una de las Zonas Bási-

cas de Salud de la región, permite el muestreo estratificado de unidades estadísticas o de grupos de unidades estadísticas, y reduce considerablemente la variación en dicho muestreo.

#### Bibliografía

1. Valleron AJ, Bouvet E, Garnerin P et al. A computer Network for the Surveillance of Communicable Diseases: The French Experiment. *American J Public Health* 1986; 76: 1289-92.
2. Nederlands Instituut Voor Onderzoek van de Eerstelijnsgezondheidszorg. NIVEL. *Continuous Morbidity Registration Sentinel Stations*. Utrecht: Ministry of Welfare, Public Health and Culture, 1986: 1-19.
3. Paccand F, Billo N, Somaini B. La place des registres dans la surveillance des maladies transmissibles. *Rev Epidém et Santé Publ* 1986; 36: 369-75.
4. Thacker SB, Keewhan C, Brachman P. The Surveillance of Infectious Diseases. *JAMA* 1983; 249: 1181-5.
5. Lobet MP. Application du cluster analysis à la Construction d'une typologie des Arrondissements Administratifs en Épidémiologie. Bruxelles: Institut d'Hygiène et d'Épidémiologie. Ministère de la Santé Publique et de la Famille, 1985.
6. Consejería de Presidencia y Administración Territorial. *Estudio y propuesta de comarcalización*

de Castilla y León. Valladolid: Junta de Castilla y León, 1985.

7. Consejería de Cultura y Bienestar Social. *Mapa de Atención Primaria de Salud de Castilla y León*. Valladolid: Junta de Castilla y León, 1987; vol I: 73-4.

8. Instituto Nacional de Estadística. *Poblaciones de derecho y de hecho de los municipios españoles. Padrón Municipal de Habitantes de 1986*. Madrid: INE, 1987.

9. Instituto Nacional de Estadística. *Censo de Población de 1981. Resultados Municipales*. Madrid: INE, 1985; vol IV.

10. Instituto Nacional de Estadística. *Encuesta de población activa. Enero, febrero y marzo de 1986*. Madrid: INE, 1986.

11. Sánchez Carrión JJ. Aspectos teóricos del análisis cluster y aplicación a la caracterización del electorado potencial de un partido. En: *Introducción a las técnicas del análisis multivariable aplicadas a las ciencias sociales*. Madrid: Centro de Investigaciones Sociológicas, 1984: 165-205.

12. Norusis MJ. *SPSS/PC+ for the IBM PC/XT/AT*. Chicago: SPSS Inc, 1986.

13. Rumeau-Rouquette C, Breart G, Padieu R. La Stratification. En: *Méthodes en Épidémiologie, Échantillonnage, Investigation et Analyse*. 3 ed. París: Flammarion, 1985.

14. Lobet MP, Stroobant A, Mertens R et al. Tool for Validation of the Network of Sentinel General Practitioners in the Belgian Health Care System. *International Journal of Epidemiology* 1987; 16: 612-8.

### Premio de la Sociedad Española de Epidemiología al mejor trabajo de investigación original publicado en *Gaceta Sanitaria*

Con el fin de estimular la producción científica de calidad en nuestro país, la *Sociedad Española de Epidemiología* convoca un premio de 100.000 pesetas a los autores del mejor artículo original publicado en *Gaceta Sanitaria*.

Serán candidatos al mismo todos los artículos aparecidos en la sección de *Originales* de la revista. El premio será otorgado por un tribunal independiente designado a tal efecto por la *Sociedad Española de Epidemiología y Gaceta Sanitaria*, y será entregado al primer autor del artículo durante la Reunión anual de la Sociedad.

La convocatoria tendrá carácter anual, y se concederá por primera vez al mejor artículo publicado durante el año 1990.