

Nota metodológica

SART (*Statistical Analysis of Rates and Trends*): herramienta vía web para el cálculo estadístico de indicadores poblacionales

Laura Esteban^a, Ramon Clèries^{a,b}, Klaus Langohr^{c,d}, Jordi Gálvez^a,
Laura Pareja^{a,e}, Josep M. Escribà^a y Josepa Ribes^{a,b,*}

^a Pla Director d'Oncologia de Catalunya, l'Hospitalet de Llobregat, Barcelona, España

^b Departament de Ciències Clíniques, Universitat de Barcelona, l'Hospitalet de Llobregat, Barcelona, España

^c Grup de Recerca Clínica en Farmacologia Humana i Neurociències, Institut Municipal d'Investigació Mèdica, Barcelona, España

^d Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya, Barcelona, España

^e Departament d'Economia de l'Empresa, Universitat Autònoma de Barcelona, Sabadell, Barcelona, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 23 de diciembre de 2010

Aceptado el 4 de abril de 2011

On-line el 28 de junio de 2011

Palabras clave:

Aplicación web

Tasa

Porcentaje de cambio anual

Casos esperados

Predicciones

Razón de incidencia estandarizada

Razón de mortalidad estandarizada

Keywords:

Web application

Rate

Annual percent change

Expected cases

Predictions

Standardized incidence ratio

Standardized mortality ratio

R E S U M E N

Se propone una herramienta vía web (SART: <http://regstattools.net/sart.html>) que automatiza los cálculos para la obtención de distintos indicadores poblacionales importantes para el control de enfermedades o eventos de la salud. Se estructura en cuatro módulos: *a*) una descriptiva que incluye el cálculo del porcentaje, el número de casos, la tasa cruda, la tasa ajustada, la tasa truncada y la tasa acumulada; *b*) la estimación del porcentaje de cambio anual de las tasas; *c*) el cálculo de casos esperados, y *d*) la razón de incidencia o mortalidad estandarizada. La aplicación solicita unos parámetros de entrada al usuario. Una vez procesados los datos y obtenidos los resultados, éstos se envían por correo electrónico al usuario. Los resultados se obtienen para cada una de las causas de estudio (enfermedades, etnias, zonas geográficas...) y cada uno de los sexos introducidos en el fichero base.

© 2010 SESPAS. Publicado por Elsevier España, S.L. Todos los derechos reservados.

Statistical Analysis of Rates and Trends (SART): a web-based tool for statistical calculation of population indicators

A B S T R A C T

We propose a web-based tool (SART: <http://regstattools.net/sart.html>) that automates calculations to obtain various population indicators that can be used for the control of diseases or health events. SART has four modules: *a*) a descriptive module that allows calculation of the number of cases and their percentage, the crude rate, the adjusted rate, the truncated rate and the cumulative rate; *b*) the estimated annual percentage change of rates; *c*) calculation of expected cases; and *d*) the standardized incidence of mortality ratio. SART requests a base file and input parameters from the user before processing the data. The data and the results obtained are processed and then sent by email to the user. The results are provided by sex and for each of the study variables (diseases, ethnic groups, geographic areas...) introduced into the base file.

© 2010 SESPAS. Published by Elsevier España, S.L. All rights reserved.

Introducción

Los programas para la prevención y el control de enfermedades necesitan indicadores poblacionales que aseguren una adecuada medición de los eventos producidos, su monitorización en el tiempo y su comparación con otras áreas. Se dispone de pocas aplicaciones que realicen cálculos estadísticos sin requerir una instalación local y conocimientos avanzados en programación y estadística. En la *tabla 1* se presentan aplicaciones que realizan cálculos de indicadores poblacionales¹⁻⁴, y sus ventajas y desventajas respecto a SART (*Statistical Analysis of Rates and Trends*: <http://regstattools.net/sart.html>).

El Plan Director de Oncología de Cataluña, después de su experiencia en el desarrollo de las aplicaciones estadísticas vía web (<http://regstattools.net/>) para los análisis de supervivencia relativa (WAERS)⁵, y del efecto del riesgo y de los factores demográficos en la evolución de las tasas (RiskDiff)⁶, ha diseñado la aplicación SART, que automatiza el cálculo y la representación gráfica de los indicadores poblacionales: *a*) número de casos, porcentajes y tasas; *b*) porcentaje de cambio anual estimado (EAPC); *c*) casos esperados de un evento en un periodo de tiempo o área geográfica, y *d*) razón de incidencia o mortalidad estandarizada (SIMR).

Método

La aplicación SART está compuesta por cuatro módulos: *Descriptive*, *EAPC*, *Expected cases* y *SIMR*.

* Autora para correspondencia.

Correo electrónico: j.ribes@iconcologia.net (J. Ribes).

Tabla 1
Aplicaciones para análisis estadísticos¹⁻⁴. Ventajas y desventajas o limitaciones respecto a SART

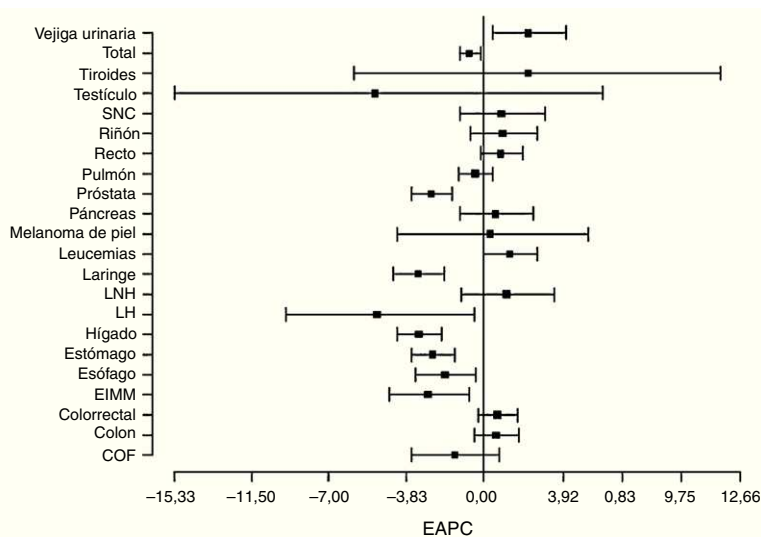
Aplicación	Ventajas	Desventajas o limitaciones
NORDCAN	Es de libre distribución Permite calcular las proyecciones de las tasas de enfermedades a corto y largo plazo	Sólo ofrece estadísticas predefinidas: sobre cáncer y en el área de los países nórdicos Dependiendo del análisis que quiera realizarse, es necesario instalarla No ofrece el cálculo de la razón de incidencia o la mortalidad estandarizada
MIAMOD AND PIAMOD software	Es de libre distribución Permite el cálculo de la estimación de la prevalencia	Sólo engloba el área del cáncer El usuario necesita un previo aprendizaje sobre la utilización de la herramienta Necesita ser instalada Requiere la estimación de la supervivencia relativa para el cálculo de proyecciones
SEER*STAT software	Es de libre distribución Realiza análisis de supervivencia y de prevalencia	No calcula proyecciones Sólo engloba el área del cáncer Necesita ser instalada
ORIUS software	Ofrece estadísticas distintas a las de SART, como años potenciales de vida perdidos, razones de tasas y figuras de incidencia comparativas	No es una herramienta para el usuario final, es un software para calcular la información para una web concreta. Ej.: Health Canada utiliza internamente Orius para producir las salidas para la página web Cancer Statistics Online (http://dsol-smed.phac-aspc.gc.ca/dsol-smed/cancer/index_e.html). No ofrece el cálculo de proyecciones

En el módulo *Descriptive*, para cada causa de estudio y sexo se calcula el número de casos, el porcentaje y las tasas cruda, ajustada según el método directo de estandarización, truncada y acumulada⁷. Para el cálculo de las tasas ajustadas y truncadas se utilizan los pesos de una población de referencia que proporciona el usuario (p. ej., la estándar mundial).

El módulo *EAPC* permite obtener el porcentaje de cambio anual para las tasas ajustadas y su respectivo intervalo de confianza (IC) mediante un modelo lineal generalizado^{7,8}.

En el módulo *Expected cases* se calculan los valores esperados para cada una de las causas de estudio, año, grupo de edad y sexo. Se ajustan los modelos log-lineales *age-drift*⁹ y *age-specific-slope*¹⁰

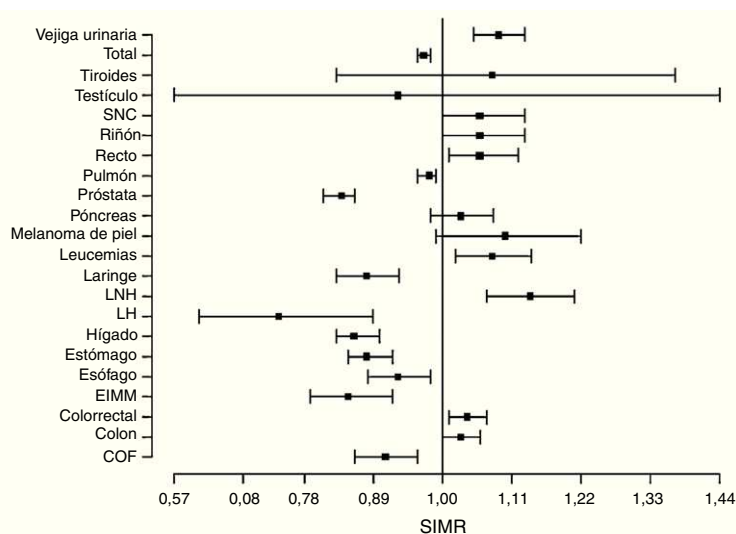
	EAPC	LCI	UCI
Vejiga urinaria	2,23	0,45	4,04
COF	-1,41	-3,54	0,76
Colon	0,63	-0,46	1,72
Colorrectal	0,69	-0,29	1,68
Esofago	-1,89	-3,36	-0,41
Estómago	-2,50	-3,58	-1,41
Hígado	-3,16	-4,24	-2,06
Laringe	-3,22	-4,46	-1,96
Leucemias	1,30	0,00	2,61
LH	-5,25	-9,82	-0,46
LNH	1,16	-1,12	3,49
EIMM	-2,72	-4,68	-0,73
Melanoma de piel	0,33	-4,28	5,17
Páncreas	0,61	-1,16	2,42
Pulmón	-0,39	-1,22	0,45
Recto	0,87	-0,14	1,90
Riñón	0,97	-0,67	2,64
SNC	0,90	-1,18	3,01
Tiroides	2,22	-6,42	11,66
Total	-0,67	-1,20	-0,14
Próstata	-2,57	-3,59	-1,55
Testículo	-5,32	-15,33	5,88



EAPC: porcentaje de cambio anual estimado; **LCI:** límite inferior del intervalo de confianza (IC); **UCI:** límite superior del IC; IC calculado al 95% de confianza; **COF:** cavidad oral y faringe; **LH:** linfoma de Hodgkin; **LNH:** linfoma no Hodgkin; **EIMM:** enfermedad inmunoproliferativa y mieloma múltiple; **SNC:** sistema nervioso central.

Figura 1. Porcentaje de cambio anual estimado e intervalo de confianza del 95% de la mortalidad global por cáncer en hombres, en Cataluña, en el periodo 1995-2004.

Group	CASES	SIMR	LCI	UCI
COF	1,403	0,91	0,86	0,96
Colon	4,252	1,03	1,00	1,06
Colorrectal	5,744	1,04	1,01	1,07
EIMM	679	0,85	0,79	0,92
Esofago	1,274	0,93	0,88	0,98
Estomago	2,662	0,88	0,85	0,92
Higado	2,395	0,86	0,83	0,90
LH	120	0,74	0,61	0,89
LNH	1,141	1,14	1,07	1,21
Laringe	1,163	0,88	0,83	0,93
Leucemias	1,468	1,08	1,02	1,14
Melanoma de piel	336	1,10	0,99	1,22
Pancreas	1,849	1,03	0,98	1,08
Prostata	4,094	0,84	0,81	0,86
Pulmon	13,090	0,98	0,96	0,99
Recto	1,492	1,06	1,01	1,12
Riñon	931	1,06	1,00	1,13
SNC	1,067	1,06	1,00	1,13
Testiculo	20	0,93	0,57	1,44
Tiroides	66	1,08	0,83	1,37
Total	47,314	0,97	0,96	0,98
Vejiga urinaria	2,782	1,09	1,05	1,13



SIMR: razón de mortalidad estandarizada; **LCI:** limite inferior del Intervalo de confianza (IC); **UCI:** limite superior del IC; IC calculado al 95% de confianza; **COF:** cavidad oral y faringe; **LH:** linfoma de Hodgkin, **LNH:** linfoma No Hodgkin, **EIMM:** enfermedad inmunoproliferativa y mieloma múltiple; **SNC:** sistema nervioso central.

Figura 2. Razón de mortalidad estandarizada (SIMR) en hombres, en Cataluña. Comparación de la mortalidad del periodo 2000-2004 respecto a la del periodo 1995-1999.

para cada una de las causas de estudio y sexo, asumiendo una distribución de Poisson o binomial negativa para el número de casos observados⁹. Se utiliza el criterio de información de Akaike (AIC) y la bondad de ajuste (mediante el test de ji al cuadrado) para la elección del mejor modelo⁸.

En el módulo *SIMR* se calculan las razones de mortalidad (incidencia) estandarizada y sus respectivos IC mediante la metodología descrita por Breslow y Day¹¹, calculando la razón entre el número de casos observados en la población en estudio y el número esperado según una población de referencia.

Formato de los ficheros

El usuario envía los ficheros de datos y parámetros necesarios para los análisis a la aplicación SART y recibe los resultados por correo electrónico. La aplicación SART se ha desarrollado con los lenguajes de programación R¹², PHP: *Hypertext Preprocessor* (PHP)¹³ e *Hypertext Markup Language* (HTML).

El fichero base de datos debe contener seis variables: a) *sex* (1: hombre; 2: mujer; 0: ambos sexos); b) *age.group* (identificador numérico para cada grupo de edad); c) *year* (año/s de referencia si corresponde a un periodo); d) *group* (etiqueta descriptiva de los eventos); e) *cases* (número de casos de cada evento); y f) *population* (poblaciones a riesgo). El usuario también deberá crear ficheros específicos según el módulo utilizado:

1. *Descriptive:* a) fichero de los grupos de edad (identificador del grupo de edad del fichero base, sus límites inferior y superior); b) fichero de pesos de la población estándar (identificador del grupo de edad del fichero base y respectivo peso).

2. *EAPC:* fichero de pesos de la población (definido en el punto 1).
3. *Expected cases:* fichero con la población para la cual se estimarán los casos: *sex* (1: hombre; 2: mujer; 0: ambos sexos), *age.group* (identificador numérico del grupo de edad), *year* (año/s de referencia si corresponde a un periodo) y *population* (población a riesgo).
4. *SIMR:* fichero que define la población de referencia con la cual quiera compararse (mismo formato que el fichero base).

Parámetros de cada módulo

1. *Descriptive:* periodo de estudio, grupos de edad para truncar y acumular, eventos incluidos para el cálculo de los porcentajes e indicador oportuno para la representación gráfica.
2. *EAPC:* periodo de estudio y nivel de confianza (habitualmente 0,95).
3. *SIMR:* nivel de confianza (habitualmente 0,95).

Limitaciones

SART no calcula el IC para las tasas de incidencia (mortalidad) ni para los casos esperados. Asimismo, en el módulo *Expected cases* no pueden introducirse los años del fichero de entrada por los que se quieren ajustar los datos ni el nivel de confianza para el test de ji al cuadrado.

En comparación con otras herramientas, SART no tiene implementados los años potenciales de vida perdidos, las razones de tasas, las figuras de incidencia comparativa ni la prevalencia.

a						b	
group	sex	model.chosen	pv.test	significant	method.AIC.GOF		
1	Vejiga urinaria	2	Poisson age-drift	1.00	No	Minimum AIC	sex;age.group;year;group;cases;population
2	Vejiga urinaria	1	Poisson age-drift	0.99	No	Minimum AIC	2;1;2010;Vejiga urinaria;1;226087
3	COF	2	Poisson age-drift	1.00	No	Minimum AIC	2;2;2010;Vejiga urinaria;0;201760
4	COF	1	Poisson age-drift	0.96	No	Minimum AIC	2;3;2010;Vejiga urinaria;0;169395
5	Cérvix	2	Poisson age-drift	0.99	No	Minimum AIC	...
6	Colon	2	Poisson age-drift	0.99	No	Minimum AIC	1;15;2010;Colorrectal;199;126764
7	Colon	1	Poisson age-drift	1.00	No	Minimum AIC	1;16;2010;Colorrectal;277;117000
8	Colorrectal	2	Poisson age-drift	0.99	No	Minimum AIC	1;17;2010;Colorrectal;275;80914
9	Colorrectal	1	Poisson age-drift	0.99	No	Minimum AIC	1;18;2010;Colorrectal;279;58830
10	Endometrio	2	Poisson age-drift	1.00	No	Minimum AIC	...
11	Esófago	2	Poisson age-drift	1.00	No	Minimum AIC	2;11;2015;Total;279;266867
12	Esófago	1	Poisson age-drift	0.99	No	Minimum AIC	2;12;2015;Total;360;243213
13	Estómago	2	Poisson age-drift	1.00	No	Minimum AIC	2;13;2015;Total;454;214651
14	Estómago	1	Poisson age-drift	0.93	No	Minimum AIC	2;14;2015;Total;611;202024
15	Hígado	2	Poisson age-drift	1.00	No	Minimum AIC	2;15;2015;Total;729;169875
16	Hígado	1	Poisson age-drift	0.36	No	Minimum AIC	2;16;2015;Total;856;138975
17	Laringe	2	Poisson age-drift	1.00	No	Minimum AIC	2;17;2015;Total;1149;135362
18	Laringe	1	Negative Binomial age-drift	1.00	No	Minimum AIC	2;18;2015;Total;1836;152490
19	Leucemias	2	Poisson age-drift	0.35	No	Minimum AIC	...
20	Leucemias	1	Poisson age-drift	0.73	No	Minimum AIC	1;11;2015;Total;610;272200
41	Tiroides	2	Poisson age-drift	1.00	No	Minimum AIC	1;12;2015;Total;885;236435
42	Tiroides	1	Poisson age-drift	1.00	No	Minimum AIC	1;13;2015;Total;1113;201504
43	Total	2	Poisson age-drift	0.11	No	Minimum AIC	1;14;2015;Total;1533;181839
44	Total	1	Poisson age-specific slope	0.19	No	Minimum AIC	1;15;2015;Total;1525;146223
45	Útero	2	Poisson age-drift	1.00	No	Minimum AIC	1;16;2015;Total;1658;109756
46	Próstata	1	Poisson age-drift	1.00	No	Minimum AIC	1;17;2015;Total;2165;91548
47	Testículo	1	Poisson age-drift	1.00	No	Minimum AIC	1;18;2015;Total;2567;77593
						...	1;16;2015;Testículo;0;109756
							1;17;2015;Testículo;1;91548
							1;18;2015;Testículo;1;77593

group: grupo de interés; sex: 1 hombres; 2 mujeres; model.chosen: el modelo de ajuste elegido para predecir; pv.test: p-valor del test de bondad de ajuste (alfa = 0,05); significant: es significativo el p-valor anterior?; method.AIC.GOF: el método utilizado para la selección del modelo (AIC: criterio de información de Akaike; GOF: test de bondad de ajuste ji al cuadrado); COF: cavidad oral y faringe; LH: linfoma de Hodgkin; LNH: linfoma no Hodgkin, EIMM: enfermedad inmunoproliferativa y mieloma múltiple; SNC: sistema nervioso central.

(a) Criterio de selección de modelos para cada localización y sexo.

(b) Fichero con el cálculo de casos esperados según el mejor modelo en cada localización y sexo. Este fichero incluye la columna "age.group" que refiere a los grupos de edad (en este caso quinquenales 1: 0-4, ..., 17: 80-84, 18: ≥ 85).

Figura 3. Selección del modelo estadístico y cálculo de defunciones esperadas por cáncer en los años 2010 y 2015, para ambos sexos, en Cataluña.

Aplicación práctica

SART incluye un tutorial y un conjunto de datos utilizado en el ejemplo de aplicación de este artículo. Se presenta la tendencia, en los hombres, de la mortalidad por cáncer en Cataluña en el periodo 1995-2004 (fig. 1), la comparación de la mortalidad por cáncer entre los periodos 2000-2004 y 1995-1999 (fig. 2) y la estimación de las defunciones esperadas por cáncer para ambos sexos en el periodo 2010-2015 (fig. 3). Los IC se calculan al 95% y en el análisis se han definido 18 grupos de edad quinquenales (1: 0-4 años; 2: 5-9 años [...]; 18: >85 años).

La figura 1 muestra los EAPC de la mortalidad por cáncer en Cataluña según la localización tumoral. En la mortalidad global por cáncer en los hombres en el periodo 1995-2004 (Total) se constata una disminución anual del 0,67%, estadísticamente significativa (EAPC: -0,67; IC: -1,20 a -0,14).

La figura 2 muestra los resultados de la comparación de la mortalidad por cáncer en los hombres en el periodo 2000-2004 respecto a 1995-1999 (SIMR). La mortalidad observada por cáncer en el periodo 2000-2004 en los hombres fue un 3% menor respecto a la esperada en 1995-1997 (SIMR: 0,97; IC: 0,96 a 0,98).

La figura 3 incluye una tabla y un fichero con los resultados correspondientes a las defunciones esperadas por cáncer en Cataluña en el periodo 2010-2015. En la tabla se detallan los modelos seleccionados para el cálculo del número de casos esperados según la localización tumoral y el sexo, siendo mayoritariamente el modelo seleccionado el *age-drift* con la distribución de Poisson. Todos los modelos se seleccionaron mediante el criterio del mínimo

valor del AIC. El fichero muestra las defunciones esperadas (*cases*) por cáncer en Cataluña según el sexo, el grupo de edad, el año de estimación y la localización tumoral. En Cataluña, para el año 2015 se estiman 1113 defunciones por cáncer en los hombres entre 60 y 64 años de edad (grupo 13).

Conclusiones

SART es una herramienta estadística destinada al cálculo automatizado de indicadores poblacionales necesarios para cualquier profesional de las ciencias de la salud o sociales. A diferencia de otras herramientas estadísticas¹⁻⁴, SART engloba el cálculo de todos los indicadores en una misma aplicación, no requiere instalación local y es de uso sencillo, rápido y de fácil interpretación.

Una nueva versión incorporará los IC para las tasas crudas y ajustadas del módulo *Descriptive*. En el módulo *Expected cases* quiere ofrecerse la posibilidad de calcular su IC y seleccionar el periodo en que se desean ajustar los datos y el nivel de confianza para el test de ji al cuadrado.

Contribuciones de autoría

J. Ribes concibió el estudio y junto con R. Clèries, K. Langohr y L. Esteban supervisaron todos los aspectos de su realización. L. Esteban programó la aplicación y realizó los análisis. J. Gálvez desarrolló la interfaz web. L. Esteban contribuyó a la redacción del primer borrador del manuscrito. L. Pareja trató los datos

de la aplicación práctica. J.M. Escribà contribuyó a la revisión de la aplicación y a la redacción del artículo. Todos los autores aportaron ideas, interpretaron los hallazgos de la aplicación práctica y revisaron los borradores del manuscrito. Todos los autores aprobaron la versión final. L. Esteban es la responsable del artículo.

Financiación

Ninguna.

Conflictos de intereses

Ninguno.

Agradecimientos

Este trabajo se enmarca dentro de la Red de Investigación Cooperativa de Cáncer (RTICC) del Instituto de Salud Carlos III y de la SGR 1541 Registre de Càncer de Catalunya- Grup de recerca emergent. Queremos expresar también nuestro agradecimiento a Josep Maria Borràs, responsable del Pla Director d'Oncología de Catalunya, y a Rosa Gispert, responsable del Servei d'Informació i Estudis de Catalunya. Los autores quieren agradecer los comentarios y sugerencias realizados por tres revisores anónimos, que han mejorado significativamente el artículo.

Bibliografía

1. Engholm G, Ferlay J, Christensen N, et al. NORDCAN: Cancer Incidence, Mortality, Prevalence and Prediction in the Nordic Countries, Version 2.4. [Actualizado 26/10/2010.] Disponible en: <http://www.ancr.nu/nordcan.asp>.
2. Instituto Superiore di Sanità. Centro Nazionale di Epidemiologia. Sorveglianza e Promozione della Salute. Roma. MIAMOD. Mortality, incidence analysis model. [Actualizado 2011]. Disponible en: <http://www.eurocare.it/MiamodPiamod/tabid/60/Default.aspx#software>.
3. Surveillance Epidemiology and End Results. National Cancer Institute. SEER*Stat Software. Version 6.6.2. [Actualizado 13/4/2010.] Disponible en: <http://www.seer.cancer.gov/seerstat/>.
4. On L, Semenciw RM, Mao Y. Orius software: calculation of rates and epidemiologic indicators, and preparation of graphical output. *Chronic Dis Can.* 2000;21:134-6.
5. Cleries R, Ribes J, Galvez J, et al. Cálculo automatizado de la supervivencia relativa vía web. El proyecto WAERS del Instituto Catalán de Oncología. *Gac Sanit.* 2005;19:71-5.
6. Valls J, Cleries R, Gálvez J, et al. RiskDiff: a web tool for the analysis of the difference due to risk and demographic factors for incidence or mortality data. *BMC Public Health.* 2009;9:473.
7. Esteve J, Benhamou E, Raymond L. *Statistical methods in cancer research. Vol. IV. Descriptive epidemiology.* Lyon: IARC Scientific Publications; 1994. p. 302.
8. McCullagh P, Nelder J. *Generalized linear models.* Chapman & Hall/CRC; 1989. p. 511.
9. Dyba T, Hakulinen T, Paivarinta L. A simple non-linear model in incidence prediction. *Stat Med.* 1997;16:2297-309.
10. Dyba T, Hakulinen T. Comparison of different approaches to incidence prediction based on simple interpolation techniques. *Stat Med.* 2000;19:1741-52.
11. Breslow NE, Day NE. *Statistical methods in cancer research. Vol. II. The design and analysis of cohort studies.* IARC Sci Publ. 1987;82:1-406.
12. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2010. [Consultado 18/2/2011.] Disponible en: www.r-project.org.
13. Davis ME, Phillips JA. *Learning PHP and MySQL.* O'Reilly; 2007. p. 411.