

TRABAJO FIN DE MÁSTER

MÁSTER DE MATEMÁTICAS

Modelización Matemática de la Estimación de Incidencia de Cáncer

Autor:

Daniel Redondo Sánchez

Tutores:

Francisco Javier Alonso Morales

DEPARTAMENTO DE ESTADÍSTICA

Miguel Rodríguez Barranco

REGISTRO DE CÁNCER DE GRANADA

Granada, junio de 2019



UNIVERSIDAD
DE GRANADA

Agradecimientos:

A mis tutores, por su siempre acertada orientación.

A todo el equipo del Registro de Cáncer de Granada, por su extraordinario y constante esfuerzo: Victoria, Maribel, Carmela, Carmen, Yoe-Ling, David, Elena, Miguel Ángel, Dafina, Miguel y María José. También a Elena Molina, por enseñarme tanto.

A mi madre, a mi hermana y a Sandra por su apoyo incondicional.

A mi hermano y a mi padre, por estar presentes cada día.

0. Índice general

Abstract	7
1. Cáncer	9
1.1. ¿Qué es el cáncer?	9
1.2. Causas del cáncer	9
1.3. Impacto del cáncer	10
1.4. Registros de Cáncer de Población	11
1.4.1. ¿Qué es un Registro de Cáncer de Población?	11
1.4.2. Situación global de los Registros	11
1.4.3. El Registro de Cáncer de Granada	13
2. Indicadores epidemiológicos del cáncer. Incidencia y mortalidad.	17
2.1. Indicadores epidemiológicos	17
2.2. Indicadores para la incidencia y mortalidad	17
2.2.1. Indicadores para la incidencia	17
2.2.2. Indicadores para la mortalidad	24
2.2.3. Razón incidencia-mortalidad de cáncer	26
3. Método Razón Incidencia Mortalidad	29
3.1. Métodos de estimación de incidencia	29
3.2. Descripción del método RIM	29
3.2.1. Fase 1. Proyección de la mortalidad	30
3.2.2. Fase 2. Modelización de la RIM	31
3.2.3. Fase 3. Estimación de la incidencia	32
4. Validación del método	33
4.1. Fuentes de información y creación de ficheros	34
4.2. Metodología de la estimación de la incidencia en Granada	37
4.2.1. Fase 1. Proyección de la mortalidad	37
4.2.2. Fase 2. Modelización de la RIM	37
4.2.3. Fase 3. Estimación de la incidencia	38

4.3. Validación de las estimaciones	38
4.4. Resultados	41
4.5. Conclusiones	47
5. Propuesta de mejora del método RIM, basada en suavizado exponencial	51
5.1. Suavizado exponencial	51
5.1.1. Recorrido aleatorio: caso $\alpha = 1$	52
5.2. Elección del mejor coeficiente de suavizado	53
5.3. Validación del método RIM	60
6. Estimaciones de la incidencia de cáncer para la provincia de Granada, año 2018	63
6.1. Fuentes de información	63
6.2. Método RIM	64
6.3. Método RIM + Suavizado exponencial	68
7. Líneas abiertas de trabajo	73
7.1. Proyectos colaborativos	73
7.2. Método de Brown	73
7.3. Actualización periódica de las estimaciones	74
Bibliografía	75
Anexo I: Código en VBA y SQL	81
Anexo II: Código en R	95

Abstract

Cancer is a worldwide public health problem. Estimates of cancer incidence are necessary to know the current burden of disease in a population. With incidence data from the Granada Cancer Registry, an estimation method based on incidence-mortality ratio modelling and mortality projection has been validated during the period 2004-2013. The best scenario for each anatomical site was chosen as the one that minimizes the mean absolute percentage error. The exponential smoothing technique for time series improves the previous method in some sites, after a process of choosing the best smoothing coefficient. Error is higher in sites with few cases (ovary, corpus uteri or lung in women), or with sudden changes in incidence or mortality (prostate). Estimates of cancer incidence in 2018 are calculated for the province of Granada, using the best method obtained for each anatomical site. 4,668 new cancer cases were estimated, with 611.6 cases per 100,000 men and 407.3 cases per 100,000 women. The most frequent cancers are prostate in men, and breast in women. These results are useful in improving health care planning and disease surveillance in Granada.

Resumen

El cáncer es un problema de salud pública a nivel mundial. Las estimaciones de la incidencia de cáncer son necesarias para conocer la carga actual de enfermedad en una población. Con datos de incidencia del Registro de Cáncer de Granada, se ha validado durante el periodo 2004-2013 un método de estimación basado en la modelización de la razón incidencia-mortalidad y en la proyección de la mortalidad. El mejor escenario para cada localización anatómica fue elegido como aquel que minimiza el error absoluto porcentual medio. La técnica del suavizado exponencial para series temporales mejora el método anterior en algunas localizaciones, tras un proceso de elección del mejor coeficiente de suavizado. El error es más alto en localizaciones con pocos casos (ovario, cuerpo uterino o pulmón en mujeres), o con repentinos cambios en incidencia o mortalidad (próstata). Se calculan las estimaciones de incidencia de cáncer en 2018 para la provincia de Granada, usando el mejor método obtenido para cada localización anatómica. Se estiman 4.668 nuevos casos de cáncer, con 611,6 casos por 100.000 hombres y 407,3 casos por 100.000 mujeres. Los cánceres más frecuentes son próstata en hombres, y mama en mujeres. Estos resultados son útiles para mejorar la planificación de la atención sanitaria y la vigilancia de la enfermedad en Granada.

1. Cáncer

1.1. ¿Qué es el cáncer?

El cáncer es un conjunto de enfermedades que tienen en común una división incontrolada de las células, que se propagan a tejidos circundantes [1]. Casi cualquier parte del cuerpo humano, que está formado por unas $3 \cdot 10^{13}$ células [2], puede desarrollar cáncer. Por ello, aunque hablamos del cáncer como una única enfermedad, en realidad hay más de 100 tipos distintos de cáncer, con diferentes características.

1.2. Causas del cáncer

Las causas del cáncer son diversas, y el proceso de creación del cáncer es complejo y multifactorial. A menudo el causante del cáncer no es un solo elemento, sino la combinación e interacción de distintos factores ambientales y genéticos. Los factores ambientales son la causa del cáncer en un 90-95 % de los casos, mientras que la genética es responsable del restante 5-10 % [3], si bien estos porcentajes varían en función del tipo de cáncer.

Los factores causantes del cáncer se pueden clasificar principalmente en dos categorías: factores no modificables (esto es, factores sobre los que no tenemos ningún control) y factores modificables (donde se centra la prevención del cáncer).

Factores no modificables:

- Edad. La edad es el principal factor de riesgo del cáncer. En general, el cáncer aumenta con la edad, aunque hay algunos cánceres que son más frecuentes en niños o adolescentes.
- Sexo. El cáncer suele ser más frecuente en hombres que en mujeres.
- Historial familiar, en algunas localizaciones anatómicas como mama, colon o próstata.

- Enfermedades genéticas, como el síndrome de Lynch y su relación con el cáncer colorrectal.

Factores modificables:

- Tabaco. Aumenta el riesgo de incidencia de algún tipo de cáncer entre 10-20 veces más.
- Alcohol. El consumo de alcohol se asocia con una mayor incidencia de cáncer de localizaciones como estómago, mama, colon, o recto.
- Dieta, sendentarismo y peso corporal. Para prevenir el cáncer es importante llevar una dieta equilibrada, realizar ejercicio físico y evitar la obesidad.
- Exposición solar. Debe controlarse para disminuir el riesgo de melanoma y otros tumores de la piel.
- Exposición a distintos carcinógenos (agentes que producen cáncer). La *International Agency for Research on Cancer* (IARC) tiene registrados 120 agentes (químicos, metales, radiaciones, productos farmacéuticos ...) que producen cáncer en humanos [4], si bien diferentes exposiciones producen distintos tipos de cáncer [5].

Otros factores no se pueden clasificar como modificables o no modificables, ya que algunos de sus aspectos no se pueden cambiar. Los principales son:

- Factores socioeconómicos. Son aquellas características del entorno del sujeto, como nivel socioeconómico, grado de cobertura sanitaria y lugar de residencia. Las desigualdades socioeconómicas se reflejan a menudo en desigualdades en cáncer.
- Factores reproductivos y hormonales, como la toma de anticonceptivos orales, la lactancia, o la terapia hormonal sustitutiva en mujeres menopáusicas.

1.3. Impacto del cáncer

El cáncer es en la actualidad uno de los mayores problemas de Salud Pública a nivel mundial. En el año 2018 se diagnosticaron en el mundo 18,1 millones de casos nuevos de cáncer, y se produjeron 9,6 millones de defunciones por cáncer [6]. El cáncer de pulmón es el más diagnosticado (12 % del total), y la principal causa

de muerte por cáncer (18%). Existen importantes variaciones geográficas en la distribución del cáncer en el mundo, debidas principalmente a las diferencias en estilos de vida y al contexto económico y sanitario de las regiones [6].

En Europa, durante el año 2018 se han diagnosticado 3,9 millones de casos de cáncer (exceptuando el cáncer de piel no melanoma) y se han producido 1,9 millones de defunciones por cáncer [7].

Los últimos datos de incidencia disponibles para España son del año 2019, donde se estima que más de 277.000 personas sean diagnosticadas de cáncer [8], mientras que hubo más de 113.000 defunciones por cáncer en el año 2017 [9].

A la vista de estos datos, podemos afirmar que el estudio del cáncer en todas sus vertientes es necesario para reducir el impacto de la enfermedad en la población, mediante acciones como la planificación de la atención sanitaria, la evaluación de la efectividad de sus tratamientos, o el desarrollo de estrategias de prevención como los programas de cribado.

1.4. Registros de Cáncer de Población

1.4.1. ¿Qué es un Registro de Cáncer de Población?

Un Registro de Cáncer de Población (RCP) es un sistema de información en el que se recogen, analizan e interpretan datos sobre todas las personas diagnosticadas de cáncer en un área geográfica bien definida.

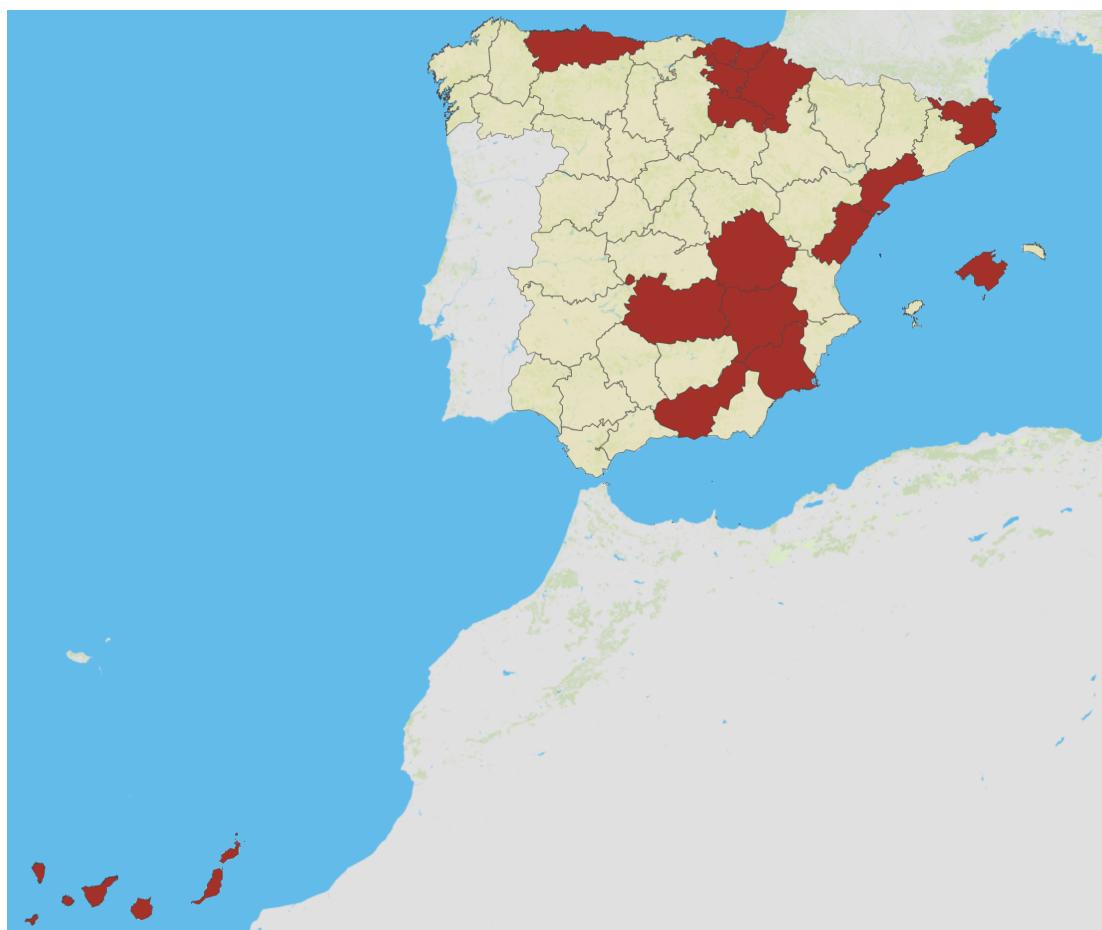
La exhaustividad debe ser un requisito indispensable de los RCP, que son necesarios para conocer la carga real de incidencia de cáncer en una población.

1.4.2. Situación global de los Registros

Cancer Incidence in Five Continents [10] es la mayor publicación sobre datos observados en RCP de todo el mundo, y recoge en su undécimo volumen datos del periodo 2008-2012 de 343 Registros en 65 países.

En algunos países como Australia, Dinamarca, Suecia o Canadá existen RCP nacionales, que cubren a la totalidad de la población del país. En otros países sólo existen RCP regionales, y ese es el caso de España, donde los RCP cubren en la actualidad aproximadamente el 27 % de la población nacional. En algunos países menos desarrollados de Asia o África (Bangladesh, Ghana, Angola...) no existe ningún Registro sobre la incidencia de cáncer.

Figura 1. Regiones de España cubiertas por los Registros de Cáncer Poblacionales que han publicado sus datos en el volumen XI de *Cancer Incidence in Five Continents*. Mapa de elaboración propia.



1.4.3. El Registro de Cáncer de Granada

El Registro de Cáncer de Granada (RCG) surge como un proyecto de la actual Consejería de Salud y Familias de la Junta de Andalucía, adscrito para su desarrollo a la Escuela Andaluza de Salud Pública.

El RCG es uno de los Registros con más antigüedad de España. Inicia su actividad en el año 1985, cubriendo desde entonces la totalidad de la provincia de Granada, que cuenta en 2018 con una población de aproximadamente 917.000 habitantes [11]. El RCG está integrado y participa desde su creación, en la *International Agency for Research on Cancer* (IARC), la *International Association of Cancer Registries* (IACR), y la *European Network of Cancer Registries* (ENCR). Además, desde el año 2010 forma parte de la Red Española de Registros de Cáncer (REDECAN).

El RCG es el único Registro de Cáncer de Andalucía acreditado por la IARC, institución perteneciente a la Organización Mundial de la Salud. En el año 2007 se crea y regula el Registro de Cáncer de Andalucía (Decreto 297/2007 del Boletín Oficial de la Junta de Andalucía), adscrito a la actual Dirección General de Salud Pública y Ordenación Farmacéutica de la Consejería de Salud y Familias de la Junta de Andalucía, integrado por registros establecidos en cada una de las provincias andaluzas. El Registro de Cáncer de Granada (RCG) se incorpora al Registro de Cáncer de Andalucía tras su creación.

Los datos de incidencia del RCG cumplen con los estándares internacionales de calidad, y por ello se han publicado en *Cancer Incidence in Five Continents*, desde el volumen VI (publicado en 1992) [12] hasta el volumen XI (publicado en 2017) [10]. Además, la incidencia registrada por el RCG ha sido utilizada para realizar las estimaciones de incidencia de cáncer a nivel nacional [8,13,14], a nivel europeo [7] y a nivel mundial [6].

Para cumplir con la exhaustividad requerida, el RCG se nutre de diversas fuentes de información para asegurar la recogida de todos los casos incidentes de la provincia. Entre sus fuentes de información, se destacan las siguientes:

- Todos los hospitales públicos y privados de la provincia:
 - Hospital Universitario Virgen de las Nieves

- Hospital Universitario San Cecilio
 - Hospital San Juan de Dios
 - Hospital de Alta Resolución de Loja
 - Hospital de Alta Resolución de Guadix
 - Hospital de San Rafael
 - Hospital de Baza
 - Hospital de Motril
 - Hospital HLA Inmaculada
 - Hospital Vithas La Salud
- Laboratorios privados de Anatomía Patológica
- Índice Nacional de Defunciones
- Registro de Mortalidad de Andalucía

En la actualidad, el RCG tiene registrados más de 125.000 casos nuevos de cáncer en la provincia. Como característica singular, se debe destacar que el RCG registra los casos de cáncer de piel no melanoma (un tipo de cáncer muy frecuente pero que cuenta con buen pronóstico), un hecho poco frecuente entre los Registros de Cáncer debido a la gran carga de trabajo que conlleva ese registro.

Figura 2. Ficha del Registro de Cáncer de Granada donde se recoge la información de cada caso incidente de cáncer.

		REGISTRO DE CÁNCER DE GRANADA				NUM. DE REGISTRO
						NUM. PREVIO:
DATOS PERSONALES	Apellidos y Nombre Sexo.....				<input type="checkbox"/>	
	NUHSA	Nº Seguridad Social	DNI	<input type="checkbox"/> NUHSA	<input type="checkbox"/> Seguridad Social	<input type="checkbox"/> DNI
	Fecha de Nacimiento	Edad				
	Lugar de Nacimiento (municipio y provincia).....				<input type="checkbox"/> <input type="checkbox"/>	
Residencia habitual: calle y número..... municipio y provincia.....				<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
DATOS DEL TUMOR	Fecha de incidencia	Fecha del D.A.P.	Localización Previo:			
	Localización.....			<input type="checkbox"/> C <input type="checkbox"/>		
	Diagnóstico anatomo-patológico.....					
	<input type="checkbox"/>					
	Base del diagnóstico.....					
	TNM	Ganglios (+ / exam.)	Factor C	cT ₁	N ₁	M ₁
	Extensión	Estadio		pT ₁	N ₁	M ₁
Breslow <input type="checkbox"/> <input type="checkbox"/>						
Otras neoplasias primarias.....						
Último contacto (Fecha - Motivo - Fuente)						
Fecha de fallecimiento		Registro Civil.....				
FUENTES INFORMANTES	Fecha de primera información					<input type="checkbox"/> <input type="checkbox"/>
	Centros Sanitarios y Servicios					<input type="checkbox"/>
Nº Historia clínica						<input type="checkbox"/>
Observaciones:						

2. Indicadores epidemiológicos del cáncer. Incidencia y mortalidad.

2.1. Indicadores epidemiológicos

Para medir en la población el impacto del cáncer se utilizan principalmente cuatro factores:

- **Incidencia** (casos nuevos). Indica el riesgo de presentar cáncer.
- **Mortalidad** (defunciones). Indica el riesgo de morir por cáncer.
- **Prevalencia** (casos nuevos y antiguos, vivos). Indica la carga asistencial de la enfermedad.
- **Supervivencia** (porcentaje de casos vivos). Indica la historia natural del cáncer y la efectividad del tratamiento.

Estos factores son clave para la vigilancia epidemiológica, las actividades de prevención y la planificación de la asistencia sanitaria.

2.2. Indicadores para la incidencia y mortalidad

2.2.1. Indicadores para la incidencia

Número de casos

La medida básica para medir la incidencia de cáncer es el número de casos nuevos. A partir de este indicador, se calculan otros más complejos que proporcionan distinta información sobre la incidencia.

Tasa bruta.

La tasa bruta (TB) por 100.000 habitantes para un periodo concreto se define como el cociente entre el número de casos nuevos y el número de personas-año a riesgo por 100.000 [15]. El número de personas-año a riesgo es la población que podría haber sido diagnosticada de cáncer, y en general se calcula como la suma

CAPÍTULO 2. INDICADORES EPIDEMIOLÓGICOS DEL CÁNCER.
18 **INCIDENCIA Y MORTALIDAD.**

de la población año a año.

$$TB = \frac{\text{Número de casos nuevos}}{\text{Personas-año a riesgo}} \cdot 100.000 = \frac{N}{P} \cdot 100.000$$

Usualmente, este cociente se multiplica por un múltiplo de 10, para facilitar la interpretabilidad del dato. Si en el caso de la natalidad, la tasa bruta se suele informar por cada 1.000 habitantes, en el caso del cáncer lo usual es representar la tasa bruta por cada 100.000 habitantes.

Las tasas brutas se usan muy a menudo debido que son fácilmente calculables e interpretables, aunque pueden enmascarar diferencias existentes entre varios grupos de edad [16]. Para analizar la incidencia en distintas edades, se utilizan las tasas específicas por edad.

Tasa específica por edad

La tasa específica por edad (TEE) se define como la tasa bruta para un grupo de edad específico [15]. Esto es, la tasa específica por edad para el grupo de 30 a 34 años se calcula dividiendo el número de casos nuevos ocurridos en personas de 30 a 34 años entre el número de personas-año en ese mismo rango de edad.

Aunque las tasas específicas se pueden calcular para cualquier rango de edad, es común que se representen por grupos de edad quinquenales. En ocasiones, puede resultar de interés analizar por separado los indicadores para niños menores de 1 año, que pueden tener características diferentes al resto de niños, y agrupar a los mayores de una determinada edad en un sólo grupo de edad. Es por ello que, en general, las tasas específicas por edad se suelen calcular en 18 grupos de edad (0-4, 5-9, ..., 80-84, 85 y más), 19 grupos de edad (0, 1-4, 5-9, ..., 80-84, 85 y más) o 21 grupos de edad (0, 1-4, 5-9, ..., 80-84, 85-59, 90-94, 95 y más).

La tasa específica por edad para el i -ésimo grupo de edad viene dada por:

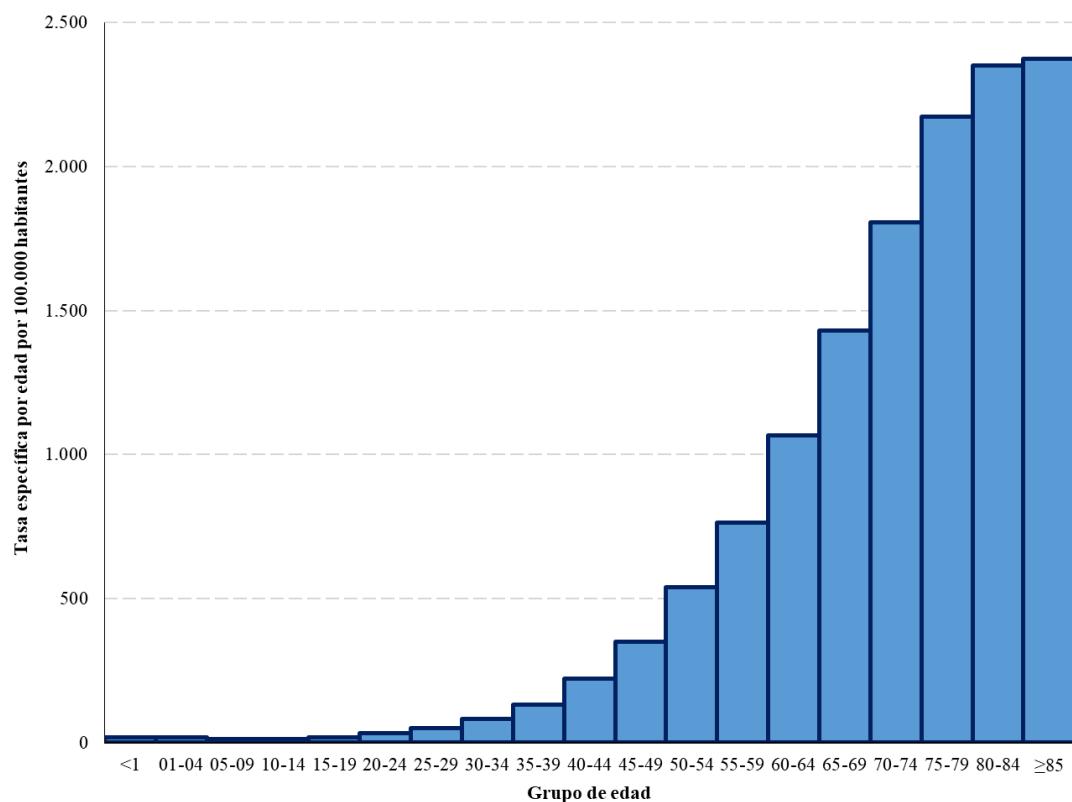
$$TEE_i = \frac{N_i}{P_i} \cdot 100.000$$

donde N_i indica el número de casos nuevos ocurridos en el i -ésimo grupo de edad,

y P_i indica las personas-año en el i -ésimo grupo de edad.

En la Figura 3 se representan las tasas específicas por edad (19 grupos de edad) para el total del cáncer en la provincia de Granada para el periodo 1985-2013.

Figura 3. Tasas específicas por edad del total del cáncer (incluyendo piel no melanoma) en la provincia de Granada durante el periodo 1985-2013.



Tasa estandarizada por edad

Exceptuando algunos cánceres que son típicos en personas jóvenes, en general la frecuencia del cáncer aumenta con la edad. Este hecho hace que la comparación de tasas brutas entre diferentes poblaciones (o incluso entre la misma población en distintos periodos) no sea factible, dadas las diferencias en las estructuras de edad [17]. Esto es, si una población fuese mucho más joven que otra, incluso aunque las tasas específicas por edad sean iguales en ambas poblaciones, se

CAPÍTULO 2. INDICADORES EPIDEMIOLÓGICOS DEL CÁNCER.

20 INCIDENCIA Y MORTALIDAD.

diagnosticarían más casos en la población más anciana. Para resolver este inconveniente surgen las tasas estandarizadas por edad (ASR, por sus siglas en inglés: Age-Standardised Rate).

Las tasas estandarizadas por edad son las tasas de incidencia que observaríamos en la población de estudio si esa población tuviese exactamente la misma estructura de edad que la población estándar predefinida [15]. Una población estándar consiste en la distribución de 100.000 habitantes ficticios en grupos de edad (generalmente, 18, 19 o 21 grupos de edad). Esta distribución teórica intenta reflejar la verdadera distribución de una determinada área geográfica, y permite la comparabilidad de tasas estandarizadas entre sus distintas regiones. El método más frecuente y recomendado de obtención de tasas estandarizadas por edad es el llamado “método directo” [18].

En la práctica, para el cálculo de una tasa estandarizada por edad se calculan primero las tasas específicas por edad. A continuación, la tasa de cada grupo de edad es multiplicada por un peso, que consiste en dividir entre 100.000 la población estándar de ese grupo de edad. Para N grupos de edad, la fórmula para obtener la tasa estandarizada por edad es:

$$ASR = \sum_{i=1}^N PEE_i \cdot TEE_i$$

donde TEE_i es la tasa específica por edad del i -ésimo grupo de edad y PEE_i es la población estándar para el i -ésimo grupo de edad.

Las poblaciones estándar más utilizadas son:

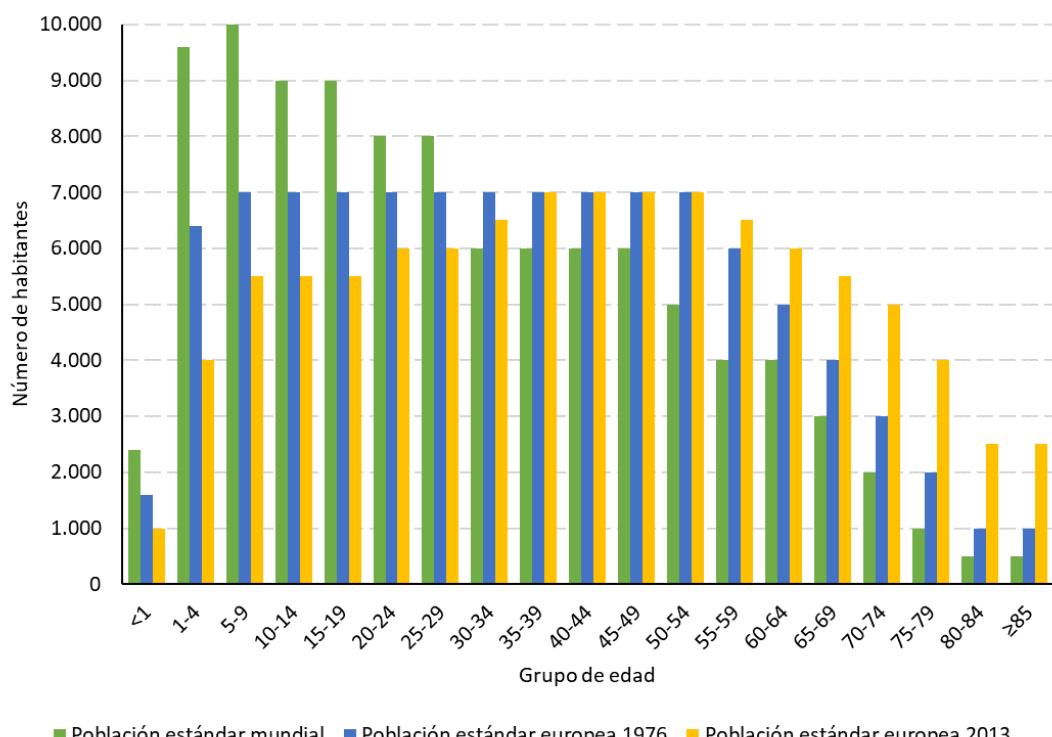
- Población mundial 1960. Propuesta por primera vez por Segi en 1960 [19], y modificada más tarde por Doll et al. en 1966 [20].
- Antigua población estándar europea (1976). Propuesta en 1976 por Waterhouse et al. [21] basándose en la estructura de edad de varias poblaciones escandinavas.
- Nueva población estándar europea (2013). Revisión de la población estándar europea de 1976 por parte de la Oficina Europea de Estadística (EUROSTAT), representa mejor el envejecimiento de la población europea [22]. El

uso de esta población estándar aún no está ampliamente extendido, aunque se espera que en unos años sustituya completamente a su predecesora. Además de redistribuir el peso de los determinados grupos de edad de la población estándar europea de 1976, permite la posibilidad de ampliar el número de grupos de edad a 21, añadiendo los grupos de edad 90-94 años y ≥ 95 años, adaptándose al aumento de habitantes muy ancianos.

Se notará ASR-W a la tasa estandarizada por la población mundial, y ASR-E a la tasa estandarizada por la población europea, especificando siempre el año de referencia (1976 o 2013).

En la Figura 4 se muestra una representación gráfica de las distintas poblaciones estándar descritas anteriormente, cuyos valores se encuentran en la Tabla 1.

Figura 4. Tasas específicas por edad del total del cáncer en la provincia de Granada durante el periodo 1985-2013.



CAPÍTULO 2. INDICADORES EPIDEMIOLÓGICOS DEL CÁNCER.
22 **INCIDENCIA Y MORTALIDAD.**

Tabla 1. Poblaciones estándar para el cálculo de tasas estandarizadas para 19 grupos de edad.

Grupo de edad	Población estándar mundial 1960	Población estándar europea 1976	Población estándar europea 2013
<1	2.400	1.600	1.000
1-4	9.600	6.400	4.000
5-9	10.000	7.000	5.500
10-14	9.000	7.000	5.500
15-19	9.000	7.000	5.500
20-24	8.000	7.000	6.000
25-29	8.000	7.000	6.000
30-34	6.000	7.000	6.500
35-39	6.000	7.000	7.000
40-44	6.000	7.000	7.000
45-49	6.000	7.000	7.000
50-54	5.000	7.000	7.000
55-59	4.000	6.000	6.500
60-64	4.000	5.000	6.000
65-69	3.000	4.000	5.500
70-74	2.000	3.000	5.000
75-79	1.000	2.000	4.000
80-84	500	1.000	2.500
≥85	500	1.000	2.500
Total	100.000	100.000	100.000

A continuación se muestra un ejemplo para el cálculo de tasas de incidencia: tasas brutas, específicas por edad y estandarizadas por edad.

Ejemplo 1. Incidencia de cáncer de próstata en hombres. Provincia de Granada, 2013.

Ejemplo de cálculo de tasas de incidencia: tasas específicas por edad (TEE), tasa bruta (TB) y tasa estandarizada por la población europea (ASR-E) tomando como referencia la población estándar europea de 1976 (PEE).

i	Grupo de edad	Población (P_i)	Número de casos (N_i)	TEE _{i}	PEE _{i}	TEE _{i} * PEE _{i}
1	<1	4.500	0	0	1.600	0
2	1-4	20.075	0	0	6.400	0
3	5-9	27.120	0	0	7.000	0
4	10-14	25.726	0	0	7.000	0
5	15-19	25.237	0	0	7.000	0
6	20-24	28.262	0	0	7.000	0
7	25-29	30.438	0	0	7.000	0
8	30-34	36.044	0	0	7.000	0
9	35-39	38.508	0	0	7.000	0
10	40-44	36.546	0	0	7.000	0
11	45-49	36.850	8	21,7	7.000	1,5
12	50-54	32.594	17	52,2	7.000	3,7
13	55-59	26.740	35	130,9	6.000	7,9
14	60-64	21.895	67	306,0	5.000	15,3
15	65-69	19.737	88	445,9	4.000	17,8
16	70-74	15.123	92	608,3	3.000	18,3
17	75-79	14.092	75	532,2	2.000	10,6
18	80-84	10.671	47	440,4	1.000	4,4
19	≥85	6.807	18	264,4	1.000	2,6
						ASR-E = 82,1

Fuentes de información:

- Registro de Cáncer de Granada. Número de casos de cáncer de próstata en la provincia de Granada, 2013.
- Instituto Nacional de Estadística. Población de hombres residentes en la

CAPÍTULO 2. INDICADORES EPIDEMIOLÓGICOS DEL CÁNCER. 24 INCIDENCIA Y MORTALIDAD.

provincia de Granada a 1 de Julio de 2013.

$$\text{TEE}_{14} = 100.000 * \frac{N_{14}}{P_{14}} = 100.000 * \frac{67}{21.895} = 306,0$$

La tasa específica por edad para los hombres de 60 a 64 años es de 306 casos por 100.000 hombres.

$$\text{TB} = 100.000 * \sum_{i=1}^{19} \frac{N_i}{P_i} = 100.000 * \frac{447}{456.965} = 97,8$$

La tasa bruta es de 97,8 casos por 100.000 hombres.

$$\text{ASR-E} = \sum_{i=1}^{19} \text{TEE}_i * \text{PEE}_i = 82,1$$

La tasa estandarizada por la población europea es de 82,1 casos por 100.000 hombres.

Otros indicadores

Existen otros indicadores epidemiológicos que miden distintos aspectos de la incidencia, como las tasas truncadas, las tasas acumulativas, o las tendencias de la incidencia. Más información sobre estos indicadores puede encontrarse en el capítulo 9 de “Registros de Cáncer: Principios y Métodos” (IARC) [15].

2.2.2. Indicadores para la mortalidad

Para la mortalidad, se pueden calcular de forma similar las mismas tasas que calculamos para la incidencia, cambiando el número de casos incidentes por el número de defunciones por cáncer.

Una defunción por cáncer se define como una muerte que es causada directamente por el cáncer, y no debe confundirse con defunciones de personas que tienen cáncer pero fallecen por otras causas (infarto, accidente de tráfico, ...).

A continuación se muestra un ejemplo para el cálculo de tasas de mortalidad: tasas brutas, específicas por edad y estandarizadas por edad.

Ejemplo 2. Mortalidad por cáncer de mama en mujeres. Provincia de Granada, 2013.

Ejemplo de cálculo de tasas de mortalidad: tasas específicas por edad (TEE), tasa bruta (TB) y tasa estandarizada por la población europea (ASR-E) tomando como referencia la población estándar europea de 1976 (PEE).

i	Grupo de edad	Población (P_i)	Número de defunciones (D_i)	TEE _{i}	PEE _{i}	TEE _{i} * PEE _{i}
1	<1	4.149	0	0	1.600	0
2	1-4	18.380	0	0	6.400	0
3	5-9	25.345	0	0	7.000	0
4	10-14	23.548	0	0	7.000	0
5	15-19	23.674	0	0	7.000	0
6	20-24	26.741	0	0	7.000	0
7	25-29	29.494	0	0	7.000	0
8	30-34	34.431	0	0	7.000	0
9	35-39	36.199	4	11,1	7.000	0,8
10	40-44	35.204	5	14,2	7.000	1,0
11	45-49	36.540	6	16,4	7.000	1,1
12	50-54	32.853	11	33,5	7.000	2,3
13	55-59	27.334	19	69,5	6.000	4,2
14	60-64	23.124	9	38,9	5.000	1,9
15	65-69	21.835	8	36,6	4.000	1,5
16	70-74	17.918	12	67,0	3.000	2,0
17	75-79	18.563	12	64,6	2.000	1,3
18	80-84	15.643	14	89,5	1.000	0,9
19	≥ 85	12.954	23	177,6	1.000	1,8

$$\text{ASR-E} = 18,8$$

Fuentes de información:

- Ministerio de Sanidad, Consumo y Bienestar Social. Número de defunciones por cáncer de mama en mujeres en la provincia de Granada, 2013.
- Instituto Nacional de Estadística. Población de mujeres residentes en la provincia de Granada a 1 de Julio de 2013.

$$\text{TEE}_{19} = 100.000 * \frac{N_{19}}{P_{19}} = 100.000 * \frac{23}{12.954} = 177,6$$

CAPÍTULO 2. INDICADORES EPIDEMIOLÓGICOS DEL CÁNCER.

26 INCIDENCIA Y MORTALIDAD.

La tasa específica por edad de mortalidad para las mujeres de 85 años o más es de 177,6 defunciones por 100.000 mujeres.

$$TB = 100.000 * \sum_{i=1}^{19} \frac{N_i}{P_i} = 100.000 * \frac{123}{463.929} = 26,5$$

La tasa bruta de mortalidad es de 26,5 defunciones por 100.000 mujeres.

$$ASR-E = \sum_{i=1}^{19} TEE_i * PEE_i = 18,8$$

La tasa de mortalidad estandarizada por la población europea es de 18,8 defunciones por 100.000 mujeres.

2.2.3. Razón incidencia-mortalidad de cáncer

La razón incidencia-mortalidad es una combinación de indicadores de incidencia y mortalidad, y se define como el cociente entre el número de casos de cáncer y el número de defunciones por cáncer. Evidentemente, esta definición se puede extender para un sexo o grupo de edad específico.

$$\text{Razón incidencia mortalidad} = \frac{\text{Casos de cáncer}}{\text{Defunciones por cáncer}}$$

La interpretación de esta razón incidencia-mortalidad presenta un importante inconveniente ya que la mortalidad se refiere a un grupo distinto de pacientes que la incidencia [23]. Por ejemplo, si analizamos la razón incidencia-mortalidad de un año en concreto, los casos incidentes son casos nuevos de cáncer detectados en ese mismo año, mientras que las defunciones pueden ser de personas diagnosticadas con anterioridad.

Por lo general, la razón incidencia-mortalidad oscilará entre 0 y 1, aunque para cánceres muy letales (pulmón, páncreas, ...) es posible que la razón incidencia-mortalidad sea mayor que uno, por el motivo visto anteriormente. Aunque en el

siglo XX se asociaba erróneamente la razón incidencia-mortalidad con la supervivencia [12], en la actualidad se consideran términos no relacionados [24] al no haber ninguna relación matemática entre ellos.

**CAPÍTULO 2. INDICADORES EPIDEMIOLÓGICOS DEL CÁNCER.
28 INCIDENCIA Y MORTALIDAD.**

3. Método Razón Incidencia Mortalidad

3.1. Métodos de estimación de incidencia

En la actualidad, existen varios métodos de estimación de incidencia. La elección de método de estimación depende de la disponibilidad de información de incidencia y mortalidad de la región en cuestión. En las estimaciones más completas publicadas hasta el momento, realizadas por el *Global Cancer Observatory (International Agency for Research on Cancer)* en 185 países para el año 2018 [6], se utilizan 7 métodos diferentes [25], mientras que en las estimaciones del mismo organismo realizadas para el año 2012 se llegaron a utilizar 9 métodos distintos [26].

El método Razón Incidencia Mortalidad (RIM) es el más adecuado a la información disponible en España [26], donde existe un Registro de Mortalidad Nacional, y se puede conocer la incidencia de algunas regiones con Registros de Cáncer de Población. Por ello, el método RIM se ha utilizado periódicamente en España para realizar estimaciones de la incidencia a nivel nacional.

Las primeras estimaciones nacionales realizadas con este método fueron publicadas en 2001 con estimaciones para el periodo 1993-1996 [27]. Distintas actualizaciones se han publicado para periodos sucesivos, como 1997-2000 [28], o 2006-2012 [29]. Recientemente, las estimaciones de incidencia de cáncer para España son realizadas por la Red Española de Registros de Cáncer (REDECAN), que ha publicado estimaciones para distintos años: 2014 [13], 2015 [14] y 2019 [8].

3.2. Descripción del método RIM

El método RIM consiste, de forma resumida, en 3 fases:

1. Proyección de la mortalidad
2. Modelización de la RIM

3. Estimación de la incidencia, multiplicando la mortalidad proyectada por la RIM modelizada

3.2.1. Fase 1. Proyección de la mortalidad

En primer lugar, se usa el método NORDPRED [30] para estimar el número de muertes que ocurren en el año objetivo para cada localización anatómica, sexo, y grupo de edad.

Formalmente, se estima la mortalidad como:

$$R_{ap} = (A_a + D \cdot p + P_p + C_c)^5$$

donde R_{ap} es la mortalidad para un grupo de edad (a) y un periodo (p) específicos, A_a es la componente de la edad para un grupo de edad concreto a , D es el parámetro de deriva que refleja la componente lineal de la tendencia, P_p es la componente no lineal de un periodo específico p , y C_c es la componente no lineal de una cohorte específica c .

Este método supone una mejora del modelo edad - periodo (fecha de diagnóstico) - cohorte (fecha de nacimiento) (APC, por sus siglas en inglés: *age-period-cohort*), ampliamente usado para estimar mortalidad [31].

Partiendo de los datos de población y mortalidad de 4 períodos conocidos (de 5 años de duración cada uno), se estima la mortalidad de un quinto periodo del que sólo se conoce la población, cuyo año central es el año objetivo. Bajo la hipótesis de que la mortalidad del año central del periodo es igual a la mortalidad media anual del periodo estimado, se obtiene así una proyección de la mortalidad para el año objetivo.

Por ejemplo, para estimar la mortalidad del año 2013, se selecciona información de mortalidad y población de 4 períodos (1991 - 1995, 1996 - 2000, 2001 - 2005, 2006 - 2010) y se proyecta la mortalidad para un quinto periodo, 2011-2015. Se obtiene así la mortalidad de 2013, suponiendo que la mortalidad de ese año es la media anual de la mortalidad del periodo 2011-2015.

A menudo, se pretende estimar la incidencia de un año reciente, y se desconoce la población del quinto periodo quinquenal. En estos casos, serán de utilidad las estadísticas de proyección de población, como las que publica periódicamente el Instituto Nacional de Estadística [32].

3.2.2. Fase 2. Modelización de la RIM

En segundo lugar, la RIM se modeliza para cada sexo y localización, mediante un modelo lineal generalizado mixto (GLMM), donde:

- El número de casos incidentes (N) es el término dependiente y sigue una distribución de Poisson
- El número de defunciones (D) entra en el modelo como un *offset*
- El efecto de la edad fue suavizado con un spline lineal con 4 nodos (percentil 10, primer tercil, segundo tercil y percentil 90) de la mortalidad de cada sexo y localización anatómica
- Los términos independientes son los grupos de edad (a_i , con $i = 1, \dots, 4$) y el año de diagnóstico (y). En el caso de que se usen datos de incidencia de varias regiones, como en las estimaciones realizadas para España, se puede incluir un término adicional para la provincia del Registro
- El efecto del año de diagnóstico se tuvo en cuenta mediante un polinomio de segundo grado

El GLMM es el siguiente:

$$\log(N) = \log(D) + \beta_0 + \beta_1 \cdot y + \beta_2 \cdot y^2 + \sum_{i=1}^4 \alpha_i \cdot a_i$$

donde α_i y β_i son los coeficientes que se obtienen del modelo.

Transformando la ecuación, tenemos que:

$$\log(N) - \log(D) = \beta_0 + \beta_1 \cdot y + \beta_2 \cdot y^2 + \sum_{i=1}^4 \alpha_i \cdot a_i$$

$$\log\left(\frac{N}{D}\right) = \beta_0 + \beta_1 \cdot y + \beta_2 \cdot y^2 + \sum_{i=1}^4 \alpha_i \cdot a_i$$

$$\log(RIM) = \beta_0 + \beta_1 \cdot y + \beta_2 \cdot y^2 + \sum_{i=1}^4 \alpha_i \cdot a_i$$

$$RIM = e^{(\beta_0 + \beta_1 \cdot y + \beta_2 \cdot y^2 + \sum_{i=1}^4 \alpha_i \cdot a_i)}$$

Una de las ventajas de usar este modelo es que expresa la RIM como una exponencial, lo que evita la estimación de RIM negativas, que no serían interpretables desde el punto de vista epidemiológico.

La estimación de los parámetros de los modelos se realiza usando estadística Bayesiana, mediante cadenas de Markov Montecarlo (MCMC).

Se utilizaron los paquetes de R BRugs v.0.9-0 [33] y R2WinBUGS v.2.1-21 [34], y el software OpenBUGS v.3.2.3. [35], la versión *open source* de WinBugs. Se realizan 3 cadenas de simulaciones con 10.000 ciclos de calentamiento y 40.000 ciclos para el análisis. Posteriormente, 1 de cada 5 ciclos fue seleccionado para obtener una matriz de 24.000×3 estimaciones de los parámetros.

3.2.3. Fase 3. Estimación de la incidencia

Finalmente, se estima la incidencia para el año objetivo, aplicando la RIM modelizada a la mortalidad estimada por NORDPRED, con varias posibles hipótesis sobre la evolución temporal de la RIM. En el apartado 4.2.2 se mencionan algunas de estas hipótesis.

$$RIM = \frac{\text{Casos incidentes}}{\text{Defunciones}} \Rightarrow \text{Casos incidentes} = RIM * \text{Defunciones}$$

Para cada combinación de sexo y localización anatómica, se obtiene así el número de casos incidentes por grupos de edad. Posteriormente, se pueden calcular indicadores como número total de casos, tasas brutas y tasas estandarizadas. Se pueden considerar los percentiles 2,5 y 97,5 de la distribución de casos incidentes estimados, para obtener un intervalo de credibilidad de los indicadores estimados al 95 %.

4. Validación del método

En este capítulo se realizará una validación del método Razón Incidencia Mortalidad (RIM) usando incidencia del Registro de Cáncer de Granada. El método usado es una adaptación del método RIM desarrollado por REDECAN. Se estimará el número de casos para seis localizaciones anatómicas y el total del cáncer (excepto piel no melanoma) por sexos, para cada uno de los años del periodo 2004-2013. Posteriormente, se usará un indicador que mide la bondad de ajuste para comparar los casos estimados con los casos observados.

La principal motivación de esta validación es la de evaluar, con medidas cuantitativas, cuál es la mejor hipótesis sobre la Razón Incidencia Mortalidad para cada localización anatómica. Hasta ahora, la elección de una hipótesis u otra se ha realizado por consenso en comités de expertos [14], teniendo en cuenta aspectos como:

- El conocimiento epidemiológico sobre cada tipo de cáncer
- La evolución temporal de la incidencia
- La selección de escenarios iguales ante el mismo patrón de evolución RIM
- La coherencia de selección de escenario entre hombres y mujeres, en función del tipo de cáncer
- Evitar situaciones poco probables como decrecimientos en la RIM

Pese a la utilidad de estos aspectos (que son subjetivos en su mayoría), sería de gran interés complementarlos con datos objetivos que ayuden en la importante toma de decisiones.

El trabajo desarrollado en este capítulo se encuentra recogido en un reciente artículo científico que se encuentra en revisión en *Population Health Metrics* [36], y cuyos resultados han sido presentados en varios congresos nacionales e internacionales:

- Comunicación oral presentada en la Reunión de la Ascensión del Grupo de Registro y Epidemiología del Cáncer de los países de Lengua Latina (GRELL). Trento (Italia), 16-18 de mayo de 2018
- Póster presentado en la XXXVI Reunión Científica de la SEE y el XIII Congresso de Associaçao Portuguesa de Epidemiologia. Lisboa (Portugal), 11-14 de septiembre de 2018
- Comunicación oral defendida en la I Jornada de la Unidad Científica de Excelencia “Salud, Cerebro y Comportamiento”. Granada, 27 de marzo de 2019
- Póster presentado en el congreso conjunto de la Asociación Internacional de Registros de Cáncer (IACR) y Asociación Norteamericana de Registros de Cáncer centrales (NAACCR). Vancouver (Canadá), 12-14 de junio de 2019
- Comunicación oral presentada en la XVII Conferencia Española y VII Encuentro Iberoamericano de Biometría. Valencia, 19-21 de junio de 2019

4.1. Fuentes de información y creación de ficheros

A continuación se describe la adaptación realizada del método usado en las estimaciones nacionales [8, 13, 14], a un método que usa datos de incidencia de la provincia de Granada y estima la incidencia futura de esa misma región.

Para cada año entre 2004 y 2013 se ha desarrollado un script en R que crea un entorno de trabajo (*workspace*) para realizar las estimaciones. El código es una adaptación y mejora del código usado para realizar las estimaciones nacionales en 2015 [14], y cuenta con más de 2.000 líneas. La cabecera del fichero, donde se definen los parámetros iniciales (año a estimar, número de ciclos seleccionados, periodos para NORDPRED, ...) se encuentra en el [Anexo II: Código en R](#) (caja de código 1).

Los ficheros necesarios para desarrollar estimaciones de la incidencia son los siguientes:

- ***incidencia.csv***. Fichero de incidencia, con variables: provincia, año, sexo, código de la localización anatómica del tumor, y número de casos por grupos de edad (18 grupos de edad: 0-4 años, 5-9, 10-14, ..., 80-84, 85 y más).

4.1. FUENTES DE INFORMACIÓN Y CREACIÓN DE FICHEROS 35

- ***mortalidad.csv***. Fichero de mortalidad, con la misma estructura que el fichero de incidencia, con defunciones por cáncer en lugar de casos incidentes.
- ***poblacion.csv***. Fichero de población, con variables: provincia, año, sexo, y número de casos por grupos de edad (18 grupos de edad: 0-4 años, 5-9, 10-14, ..., 80-84, 85 y más).
- ***parametros.tumor.csv***. Fichero de parámetros, donde se describen los códigos empleados para identificar cada localización anatómica, con los dos terciles y los percentiles 10 y 90 de la mortalidad por cada sexo y localización anatómica.

Las fuentes de información que se han utilizado para crear estos ficheros son:

- Incidencia: Datos del Registro de Cáncer de Granada (RCG). Casos incidentes de cáncer en la provincia de Granada, periodo 1985-2013, por año, sexo, grupo de edad y localización anatómica.
- Mortalidad: Ministerio de Sanidad, Consumo y Bienestar Social. Defunciones por cáncer en la provincia de Granada, 1982-2010, por sexo, año, grupo de edad y localización anatómica [37].
- Población: Instituto Nacional de Estadística (INE). Población de la provincia de Granada a 1 de Julio, 1982-2013, por año, sexo y grupo de edad [11].

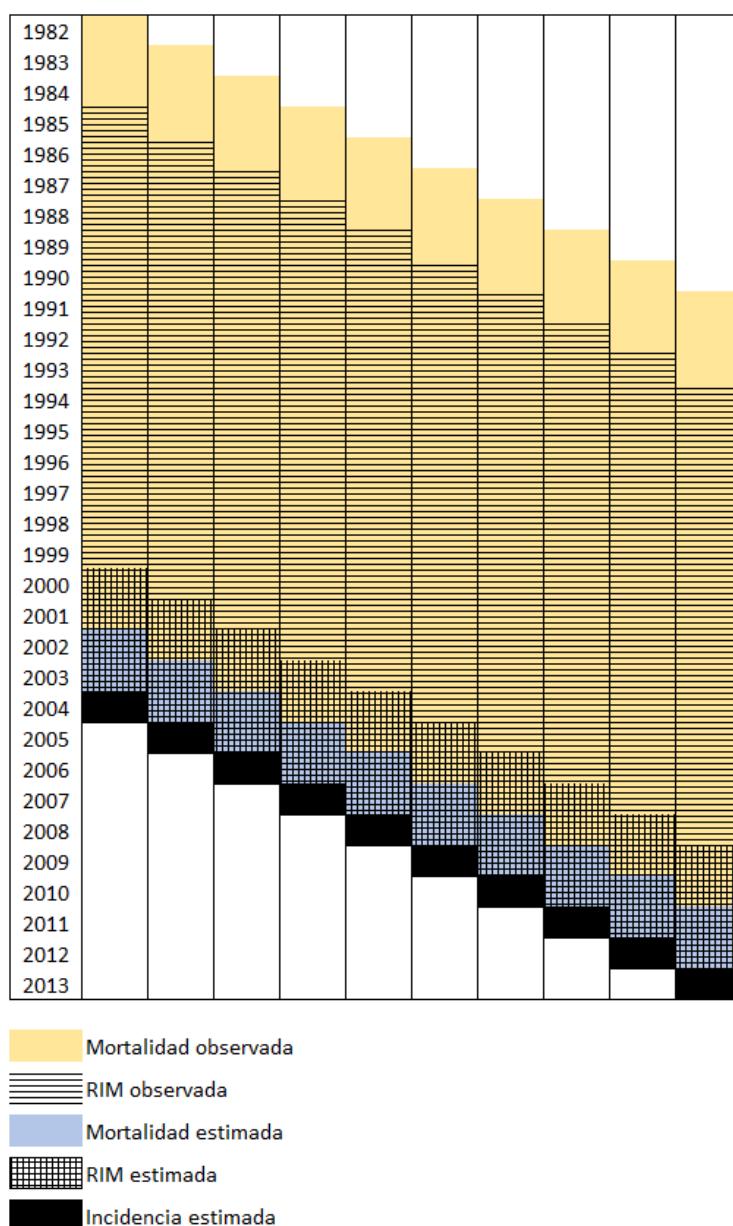
Debido a que la base de datos del Registro de Cáncer de Granada está desarrollada para un entorno de Microsoft Access, se han empleado los lenguajes de programación SQL (*Structured Query Language*) y VBA (*Visual Basic for Applications*) para extraer la información de incidencia necesaria. Para los datos de mortalidad y población, se han descargado de las fuentes de información ficheros Microsoft Excel que son importados a Microsoft Access, se procesan, y el resultado final se exporta como un fichero *.csv*. En el [Anexo I: Código en VBA y SQL](#) (cajas de código 3-5) se puede encontrar el código necesario para crear los ficheros: *incidencia.csv*, *mortalidad.csv* y *poblacion.csv*.

Para crear el fichero *parametros.tumor.csv*, se ha desarrollado un script en R que utiliza *mortalidad.csv* para calcular los percentiles necesarios para cada sexo y localización anatómica. El script puede encontrarse en el [Anexo II: Código en](#)

R (cajas de código 2 y 3).

En la Figura 5 se representan gráficamente los distintos períodos de incidencia y mortalidad utilizados en el análisis para realizar las estimaciones de la incidencia entre 2004 y 2013.

Figura 5. Resumen de información utilizada para estimar la incidencia de cáncer en la provincia de Granada durante los años 2004-2013.



4.2. Metodología de la estimación de la incidencia en Granada

En esta sección se especifica en detalle la metodología utilizada en la estimación de la incidencia.

4.2.1. Fase 1. Proyección de la mortalidad

Para usar el método de proyección de la mortalidad, se han usado las funciones `nordpred.estimate` y `nordpred.prediction` de NORDPRED implementadas en el lenguaje de programación S, compatible con R [30]. Ambas funciones se encuentran en el [Anexo II: Código en R](#) (caja de código 4).

4.2.2. Fase 2. Modelización de la RIM

Los escenarios planteados para la RIM en las estimaciones realizadas para España [8, 13, 14] son los siguientes:

- C. Se asume que la RIM sigue una tendencia constante, e igual a la RIM del último año del que se conoce la incidencia observada
- L. Se asume que la RIM sigue una tendencia lineal
- Q. Se asume que la RIM sigue una tendencia cuadrática

En este trabajo se plantean dos escenarios adicionales que complementan al escenario C. Los escenarios contemplados son entonces cinco:

- C1. Igual que el escenario C anterior, se asume que la RIM sigue una tendencia constante, e igual a la RIM del último año del que se conoce la incidencia observada
- C3. Se asume que la RIM sigue una tendencia constante, e igual a la media de las RIM de los últimos 3 años de los que se conoce la incidencia observada
- C5. Se asume que la RIM sigue una tendencia constante, e igual a la media de las RIM de los últimos 5 años de los que se conoce la incidencia observada
- L. Se asume que la RIM sigue una tendencia lineal
- Q. Se asume que la RIM sigue una tendencia cuadrática

4.2.3. Fase 3. Estimación de la incidencia

Para cada año, escenario, sexo y localización anatómica, se pueden estimar los siguientes indicadores:

- Número de casos nuevos
- Tasa bruta (TB) por 100.000 habitantes
- Tasa estandarizada por edad tomando como referencia la población estándar europea de 1976 (ASR-E) por 100.000 habitantes
- Tasa estandarizada por edad tomando como referencia la población estándar europea de 2013 (ASR-E) por 100.000 habitantes
- Tasa estandarizada por edad tomando como referencia la población mundial (ASR-W) por 100.000 habitantes

4.3. Validación de las estimaciones

Para cada año entre 2004 y 2013, se usa el método RIM para realizar estimaciones de la incidencia para la provincia de Granada de seis localizaciones anatómicas de interés por sexo.

El resto de localizaciones anatómicas de cada sexo, excluyendo la piel no melanoma (PNM), se agrupa en una categoría de “Otros” que también es estimada, para disponer información del total del cáncer excepto PNM.

Tabla 4. Localizaciones anatómicas analizadas por sexos, con sus respectivos códigos internos y códigos en la Clasificación Internacional de Enfermedades, Décima versión [38].

Sexo	Localización anatómica	Códigos CIE-10	Código interno
Hombres	Colon	C18	114
	Recto	C19, C20	115
	Pulmón	C33, C34	116
	Próstata	C61	117
	Vejiga	C67, D09, D41	118
	Estómago	C16	119
Mujeres	Otros	Total excepto C44	199
	Colon	C18	614
	Recto	C19, C20	615
	Pulmón	C33, C34	616
	Mama	C50	617
	Cuerpo uterino	C54	618
	Ovario	C56	619
	Otros	Total excepto C44	699

Debido a la inclusión de esta categoría, si tenemos dos localizaciones anatómicas: l_1 y l_2 , y denotamos:

- $l_1 \cup l_2$ a la unión de las dos localizaciones anatómicas
- I^{l_1} a un indicador I (que puede ser el número de casos, la tasa bruta, o la tasa estandarizada por edad) de una localización l_1

se tiene que:

$$N^{l_1 \cup l_2} = N^{l_1} + N^{l_2}$$

$$TB^{l_1 \cup l_2} = \frac{N^{l_1 \cup l_2}}{P} = \frac{N^{l_1} + N^{l_2}}{P} = \frac{N^{l_1}}{P} + \frac{N^{l_2}}{P} = TB^{l_1} + TB^{l_2}$$

$$\begin{aligned}\text{ASR}^{l_1 \cup l_2} &= \sum_{i=1}^{18} \text{PEE}_i \frac{N^{l_1 \cup l_2}}{P_i} = \sum_{i=1}^{18} \text{PEE}_i \frac{N_i^{l_1} + N_i^{l_2}}{P_i} = \\ &= \sum_{i=1}^{18} \text{PEE}_i \frac{N_i^{l_1}}{P_i} + \sum_{i=1}^{18} \text{PEE}_i \frac{N_i^{l_2}}{P_i} = \text{ASR}^{l_1} + \text{ASR}^{l_2}\end{aligned}$$

Por tanto, podremos calcular indicadores para el total del cáncer excepto PNM sumando los indicadores de todas las localizaciones analizadas.

Una vez realizadas las estimaciones de incidencia entre 2004 y 2013 para las localizaciones definidas para los cinco escenarios contemplados, se calcula la desviación relativa (DR) entre los casos estimados y los casos observados:

$$\text{DR} = 100 \cdot \left| \frac{E_i(s) - O_i}{O_i} \right|$$

siendo s un escenario concreto, que representa alguna hipótesis sobre la evolución de la RIM.

Como indicador de la bondad de ajuste, se define el error absoluto porcentual medio (MAPE, por sus siglas en inglés: *Mean Absolute Percentage Error*) para un escenario s y un periodo que comienza en el año y_1 y termina en el año y_2 , como:

$$\text{MAPE}(s) = \frac{100}{y_2 - y_1 + 1} \sum_{i=y_1}^{y_2} \left| \frac{E_i(s) - O_i}{O_i} \right|$$

Este indicador refleja la desviación de los valores estimados con respecto a los observados a través del periodo. Se considera por tanto que un escenario s_1 es más adecuado o “mejor” que un escenario s_2 si se cumple la condición $\text{MAPE}(s_1) < \text{MAPE}(s_2)$. El mejor escenario será entonces aquel que tiene el menor MAPE de todos.

Se han desarrollado funciones en R para medir el MAPE de cada combinación localización anatómica-sexo-escenario, y elegir el mejor escenario en cada caso.

Las funciones desarrolladas en R se pueden encontrar en el [Anexo II: Código en R](#) (cajas de código 5 y 6).

4.4. Resultados

En las Tablas 5 y 6 se describen para hombres y mujeres respectivamente, los distintos MAPE obtenidos en todos los escenarios considerados, por localización anatómica, con el número de casos observado y estimado en cada escenario.

El escenario cuadrático (Q) obtuvo altos valores de MAPE, y no fue el mejor escenario en ninguna localización anatómica de las estudiadas. De hecho, las estimaciones realizadas mediante el escenario Q tuvieron el peor MAPE en todas las localizaciones excepto en la agrupación de otros tumores en mujeres (Tablas 5 y 6, Figuras 6 y 7).

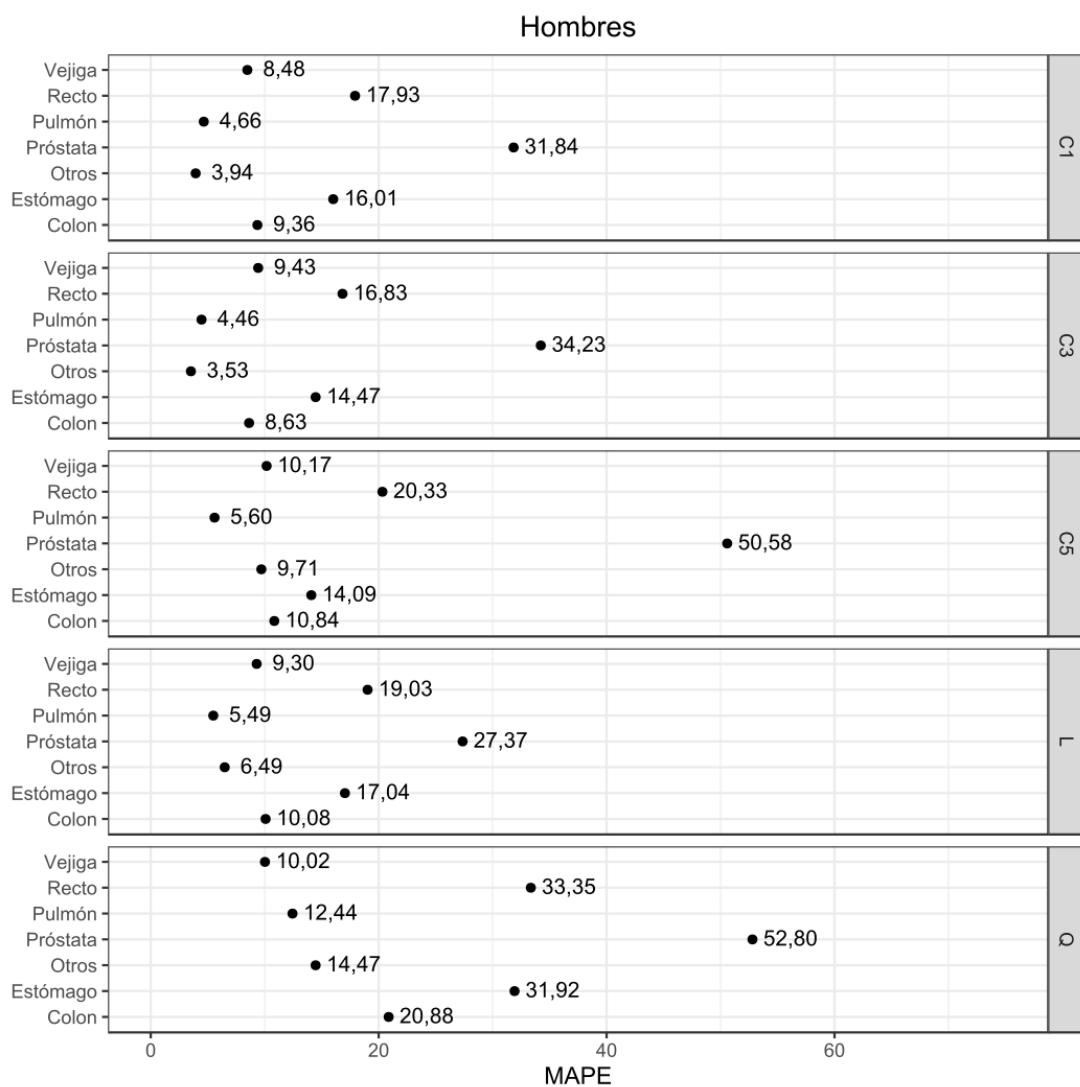
Tabla 5. Hombres. Número de casos observados (O), número de casos estimados para cada escenario (C1, C3, C5, L, Q) por año, y MAPE del periodo 2004-2013 en la provincia de Granada, por localización anatómica. Resaltado en negrita el valor del MAPE más bajo para cada localización.

Localización	Escenario	N 2004	N 2005	N 2006	N 2007	N 2008	N 2009	N 2010	N 2011	N 2012	N 2013	MAPE
Colon	O	138	203	156	180	172	227	208	214	243	222	
	C1	138	172	179	164	195	186	222	211	215	230	9,4
	C3	136	165	173	163	190	189	215	211	217	229	8,6
	C5	133	158	167	149	180	174	206	194	203	214	10,8
	L	141	182	187	178	207	199	233	227	228	245	10,1
	Q	151	230	232	161	231	158	281	207	193	229	20,9
Recto	O	90	97	93	95	115	136	112	110	115	121	
	C1	68	74	71	90	127	105	137	135	96	112	17,9
	C3	71	73	69	87	118	101	131	129	96	110	16,8
	C5	62	65	66	79	107	98	120	119	93	92	20,3
	L	76	83	75	100	143	109	151	147	99	132	19,0
	Q	47	73	89	107	216	147	182	176	95	111	33,4
Pulmón	O	307	350	332	344	377	317	348	312	353	351	
	C1	307	346	338	326	348	324	320	366	362	353	4,7
	C3	310	343	331	322	341	319	313	356	354	350	4,5
	C5	313	345	336	323	333	313	299	346	343	335	5,6
	L	303	344	336	327	358	331	336	379	375	369	5,5
	Q	281	365	394	357	399	359	363	442	417	372	12,4
Próstata	O	327	351	405	442	478	498	463	519	494	447	
	C1	224	295	269	242	266	340	330	325	486	663	31,8
	C3	216	271	251	221	247	312	309	301	436	558	34,2
	C5	160	192	181	159	173	218	218	214	306	356	50,6
	L	296	414	370	334	374	484	462	455	694	1053	27,4
	Q	250	468	388	410	383	547	468	496	932	1948	52,8
Vejiga	O	257	245	245	253	251	275	264	265	312	268	
	C1	234	226	266	305	259	259	275	243	268	259	8,5
	C3	230	223	267	301	270	263	278	242	263	254	9,4
	C5	205	202	239	273	249	248	271	239	253	237	10,2
	L	262	247	294	334	275	271	281	247	279	277	9,3
	Q	252	239	246	327	182	221	251	253	303	286	10,0
Estómago	O	87	70	89	76	94	76	83	67	86	73	
	C1	80	64	72	89	73	74	89	82	76	103	16,0
	C3	79	65	72	87	73	73	85	79	74	98	14,5
	C5	76	64	70	87	73	74	81	76	69	90	14,1
	L	84	65	74	89	73	74	95	86	82	112	17,0
	Q	85	56	68	107	75	82	122	105	89	144	31,9
Otros	O	784	891	824	872	876	939	892	950	924	947	
	C1	758	815	821	827	883	859	939	937	952	976	3,9
	C3	758	815	821	825	873	851	913	911	919	938	3,5
	C5	702	740	763	791	821	800	846	841	862	866	9,7
	L	812	888	877	860	936	911	1.015	1.016	1.020	1.063	6,5
	Q	732	777	796	826	930	900	1.119	1.117	1.204	1.259	14,5

Tabla 6. Mujeres. Número de casos observados (O), número de casos estimados para cada escenario (C1, C3, C5, L, Q) por año, y MAPE del periodo 2004-2013 en la provincia de Granada, por localización anatómica. Resaltado en negrita el valor del MAPE más bajo para cada localización.

Localización	Escenario	N 2004	N 2005	N 2006	N 2007	N 2008	N 2009	N 2010	N 2011	N 2012	N 2013	MAPE
Colon	O	117	138	135	153	144	153	152	144	168	163	
	C1	132	145	142	117	118	143	165	174	187	184	12,5
	C3	131	145	142	118	118	138	158	168	186	186	11,8
	C5	123	131	130	118	113	132	150	163	175	178	10,0
	L	140	159	153	117	122	151	175	182	198	191	16,5
	Q	140	140	136	112	116	187	228	221	185	164	20,5
Recto	O	67	61	75	60	67	70	70	77	66	56	
	C1	53	42	31	50	81	66	65	86	73	96	25,5
	C3	58	47	35	50	73	63	60	78	67	84	19,2
	C5	45	39	30	42	60	52	61	82	62	75	25,4
	L	64	48	35	57	100	80	66	85	82	111	28,6
	Q	24	19	12	46	164	88	110	184	138	253	106,8
Pulmón	O	45	52	51	53	48	62	59	71	60	56	
	C1	34	26	24	42	38	42	59	68	60	58	20,9
	C3	33	27	27	41	40	40	56	63	59	57	21,6
	C5	28	22	24	37	38	40	55	59	55	55	26,9
	L	40	31	26	46	40	44	61	74	64	61	18,3
	Q	34	15	10	41	28	58	90	111	70	62	38,3
Mama	O	338	345	389	410	415	494	419	466	450	494	
	C1	314	405	469	543	480	466	428	354	342	374	17,3
	C3	294	375	437	505	472	465	437	369	351	377	14,8
	C5	235	283	324	365	355	364	352	305	301	325	23,5
	L	391	535	629	743	624	579	516	414	389	425	34,3
	Q	459	628	694	793	477	417	334	240	263	328	46,4
Cuerpo uterino	O	114	106	99	107	103	123	127	124	107	124	
	C1	158	157	140	103	122	136	116	73	95	127	22,4
	C3	154	164	164	125	143	144	123	77	92	112	29,3
	C5	136	127	123	104	131	158	135	93	115	140	17,3
	L	178	193	171	114	125	122	106	61	78	109	34,6
	Q	175	100	36	23	36	90	78	54	135	357	60,2
Ovario	O	60	77	58	67	68	69	62	50	53	51	
	C1	85	71	73	73	71	70	92	61	64	72	22,2
	C3	83	75	80	74	71	71	88	63	67	74	23,6
	C5	83	71	78	85	76	76	93	67	70	80	30,2
	L	86	74	73	64	67	65	89	57	61	67	18,9
	Q	103	46	37	65	78	68	141	46	46	64	34,1
Otros	O	718	763	703	822	790	838	850	856	791	858	
	C1	624	655	681	635	640	700	749	767	784	849	11,3
	C3	613	641	668	624	621	673	710	725	748	815	14,3
	C5	564	578	599	574	562	590	608	634	637	704	24,1
	L	677	725	756	688	707	798	877	882	919	983	8,7
	Q	684	731	744	697	770	876	1.033	1.092	1.036	1.078	14,3

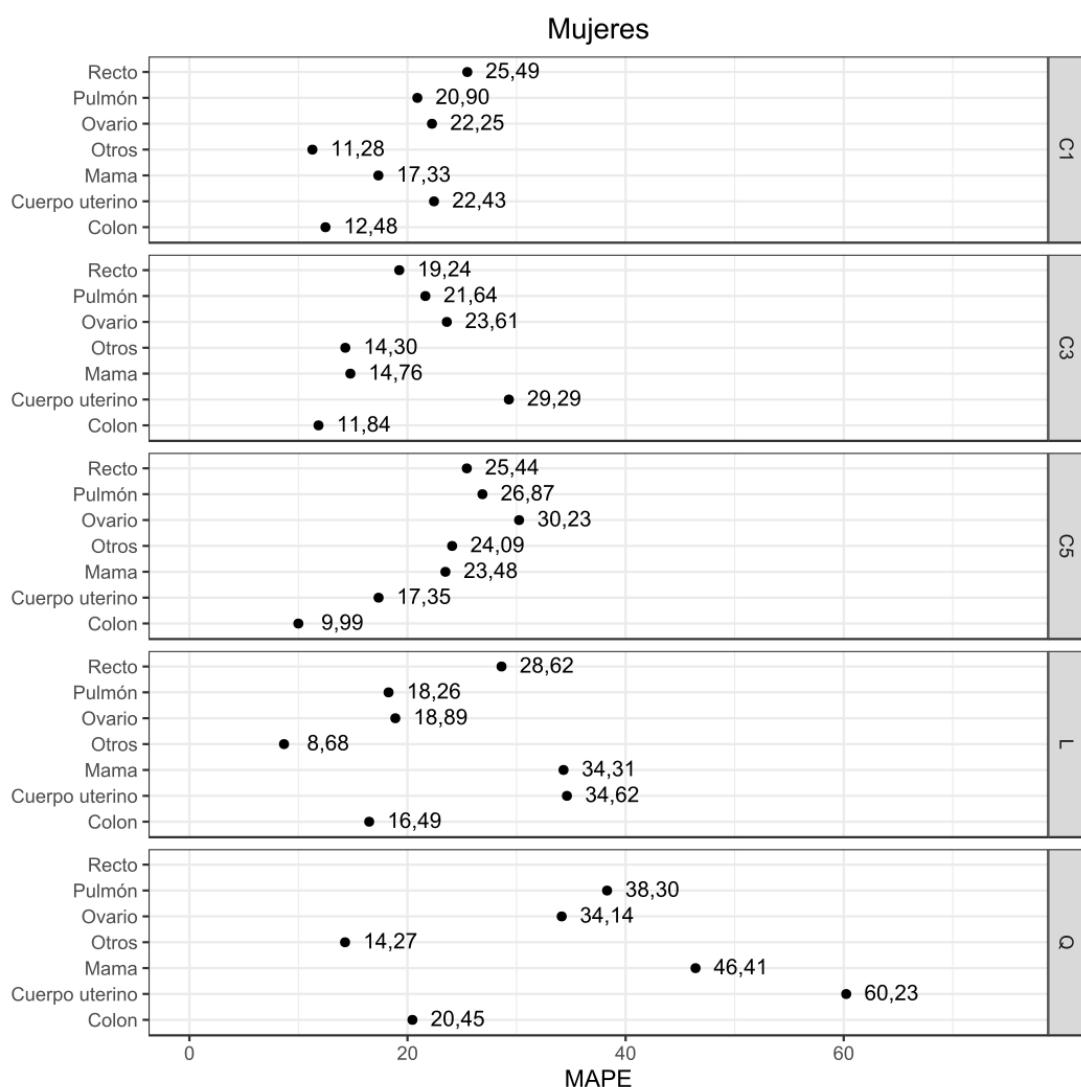
Figura 6. Error absoluto porcentual medio (MAPE) por escenario y localización anatómica para hombres.



En hombres, las peores estimaciones se dieron en el cáncer de próstata. Por otra parte, el escenario Q es poco adecuado para las localizaciones anatómicas analizadas. El MAPE varió desde 3,53 (otros, escenario C3) hasta 52,80 (próstata, escenario Q).

Figura 7. Error absoluto porcentual medio (MAPE) por escenario y localización anatómica para mujeres.

Nota: No se representa el MAPE de cáncer de recto para el escenario Q (MAPE = 106,8) para mantener una escala que permita una correcta comparación visual entre los resultados.



En mujeres, las estimaciones fueron en general peores que en hombres, con un MAPE mayor de 10 en la mayoría de las localizaciones. Las peores estimaciones se encontraron en el cáncer de recto (MAPE = 19,2 en el mejor escenario) y el cáncer de ovario (MAPE = 18,9 en el mejor escenario). El MAPE varió desde

8,68 (otros, escenario L) hasta 106,8 (recto, escenario Q).

En la Tabla 7 se describe, para la provincia de Granada y el periodo 2004-2013, el número de casos observados y, para el mejor escenario, el número de casos estimados, la desviación relativa y el MAPE por sexo y localización anatómica.

Tabla 7. Número de casos observados (Obs.) y estimados (Est.) en la provincia de Granada, periodo 2004-2013, para el mejor escenario, con medidas de bondad de ajuste: desviación relativa (DR) y error absoluto porcentual medio (MAPE), por sexo y localización anatómica.

Sexo	Localización anatómica	Obs.	Est.	DR	MAPE	Mejor escenario
Hombres	Colon	1.963	1.888	3,8	8,6	C3
	Recto	1.084	985	9,1	16,8	C3
	Pulmón	3.391	3.339	1,5	4,5	C3
	Próstata	4.424	4.936	11,6	27,4	L
	Vejiga	2.635	2.594	1,6	8,5	C1
	Estómago	801	760	5,1	14,1	C5
	Otros	8.899	8.624	3,1	3,5	C3
	Total excepto PNM	23.197	23.126	0,3	6,3	
Mujeres	Colon	1.467	1.413	3,7	10,0	C5
	Recto	669	615	8,1	19,2	C3
	Pulmón	557	487	12,6	18,3	L
	Mama	4.220	4.082	3,3	14,8	C3
	Cuerpo uterino	1.134	1.262	11,3	17,3	C5
	Ovario	615	703	14,3	18,9	L
	Otros	7.989	8.012	0,3	8,7	L
	Total excepto PNM	16.651	16.574	0,5	3,8	

La hipótesis lineal para la tendencia de la RIM obtuvo el menor MAPE para cáncer de próstata en hombres, y cáncer de pulmón, ovario, y otras localizaciones en mujeres. Las hipótesis constantes (C1, C3, C5) fueron el mejor escenario para el resto de localizaciones, siendo C1 el mejor escenario exclusivamente en el cáncer de vejiga en hombres (Tabla 7).

En el total del cáncer excepto piel no melanoma, eligiendo el mejor escenario para cada localización anatómica y sexo, la desviación relativa del número de casos estimados con respecto a la incidencia observada fue de 0,3 % en hombres y 0,5 % en mujeres.

La desviación relativa varió en hombres entre 1,5 % (cáncer de pulmón) y 11,6 % (cáncer de próstata), y en mujeres entre 0,3 % (otros) y 14,3 % (cáncer de ovario) (Tabla 7). En general, la desviación relativa fue más baja en localizaciones anatómicas con mayor número de casos (Figura 8). La gran excepción se produjo en el caso del cáncer de próstata en hombres, que siendo el cáncer más frecuente en hombres obtuvo una desviación relativa cercana al 12 %.

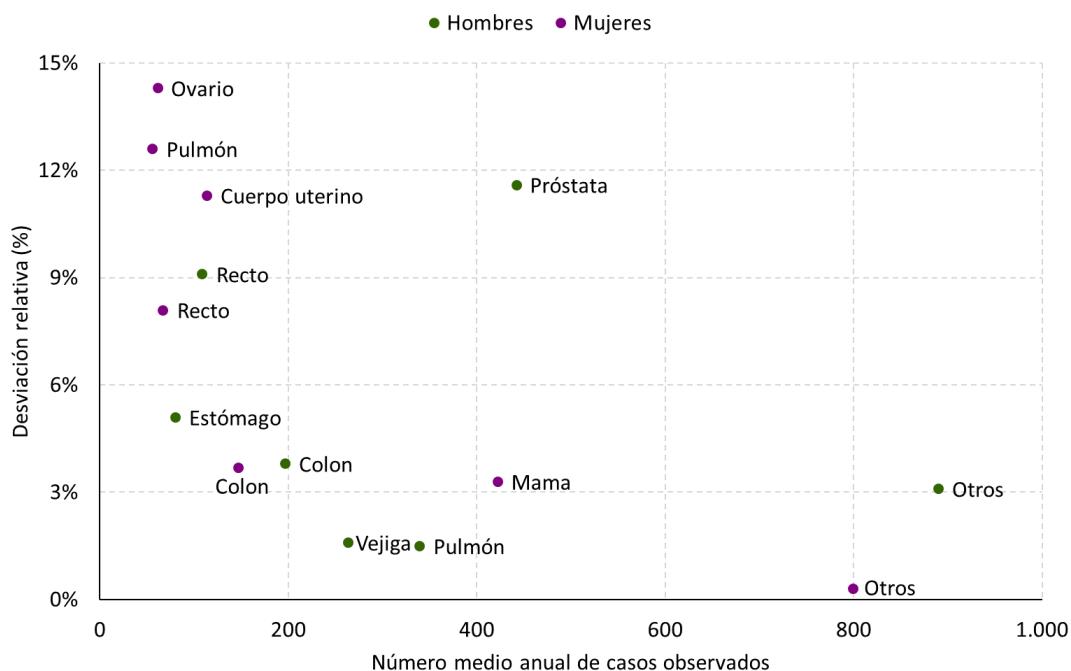
4.5. Conclusiones

La implementación del MAPE como indicador de bondad de ajuste de los distintos escenarios dota al método RIM de una validación numérica que permite discriminar cuál es la hipótesis más adecuada para cada localización anatómica.

A la vista de los resultados obtenidos, es necesario que se replantee el método actual de elección de escenario para las estimaciones nacionales, que está basado en una reunión de expertos. El ejemplo más claro se puede observar en el escenario Q, que en esta validación está considerado como el peor escenario de los planteados. Pese a ello, este escenario se está utilizando actualmente en las estimaciones de España para localizaciones anatómicas como colon, tanto en hombres como en mujeres [8]. En el análisis realizado para el cáncer de colon en la provincia de Granada, el MAPE para el escenario Q es de 20,9 % en hombres y 20,5 % en mujeres, y duplica el MAPE del mejor escenario por sexo (8,6 % para hombres y 10,0 % para mujeres).

La alta desviación relativa encontrada en localizaciones con pocos casos observados (Figura 8) indica que el método RIM es menos preciso en localizaciones con bajo número de casos incidentes o defunciones. Por otra parte, es de gran interés epidemiológico conocer la incidencia estimada en el mayor número posible de localizaciones anatómicas de interés. Es necesario, por tanto, encontrar un equilibrio que tenga en cuenta ambas circunstancias.

Figura 8. Número medio anual de casos observados y desviación relativa de la mejor estimación, por sexo y localización anatómica.



El mal funcionamiento del método RIM en el cáncer de próstata se puede explicar por la implantación del antígeno prostático específico (PSA) como método de detección este cáncer. Debido al aumento de los análisis de PSA en hombres adultos, la incidencia del cáncer de próstata ha aumentado considerablemente en los últimos años. Pese al aumento que se ha producido en la incidencia de cáncer, algunos grandes estudios indican que la detección precoz del cáncer de próstata no reduce la mortalidad [39, 40].

Este gran aumento a lo largo del tiempo en la incidencia del cáncer de próstata, sin impacto en la mortalidad, se puede constatar en la provincia de Granada. La incidencia aumentó en el periodo 1985-2013 considerablemente con un porcentaje de cambio anual de +6,3% [41], con un aumento aún mayor a partir del año 2000, mientras que la mortalidad por cáncer de próstata es estable, aunque ligeramente decreciente, con un porcentaje de cambio anual de -0,4% en el mismo periodo.

Debido al adelanto pronóstico de cáncer de próstata que se ha realizado en los últimos años, los casos observados a partir de 2011 están disminuyendo (Tabla 5), por lo que las estimaciones de los años más recientes suelen sobreestimar los casos observados, ya que el método RIM supone que el comportamiento pasado se mantendrá en el futuro.

En resumen, las estimaciones realizadas con el método RIM para localizaciones en las que hay repentinhas fluctuaciones de incidencia o mortalidad que afecten a la RIM deben ser tomadas con cautela. Algunas de las localizaciones susceptibles de tener estas características en la provincia de Granada pueden ser el cáncer de mama (debido al cribado existente en mujeres de 50 a 69 años) y el cáncer de colon-recto (debido al cribado que se está implantando actualmente). Otros métodos o escenarios son necesarios para obtener estimaciones precisas en estos cánceres.

Para localizaciones que poseen un comportamiento estable a lo largo del tiempo, el método RIM ha demostrado ser eficiente para la estimación de la incidencia de cáncer, con una desviación relativa menor al 10 % en la mayoría de las localizaciones analizadas.

5. Propuesta de mejora del método RIM, basada en suavizado exponencial

En este capítulo se realiza una aproximación al problema de las estimaciones de la incidencia de cáncer desde el punto de vista de las series temporales [42].

5.1. Suavizado exponencial

Fijada una localización anatómica, se considera la serie temporal formada por el número de casos incidentes de cada año $\{x_1, x_2, \dots, x_n\}$. El objetivo de la estimación es hallar los siguientes valores de la serie temporal x_{n+k} . En concreto, y para seguir con el criterio marcado hasta ahora, basado en la disponibilidad de los datos de los Registros de Cáncer, se considera $k = 5$.

Si consideramos que la serie $\{x_1, x_2, \dots, x_n\}$ no tiene efectos temporales y además no existe una tendencia global, podemos considerar el modelo para estimar x_t como:

$$x_t = \mu_t + \omega_t$$

donde μ_t es la media no-estacionaria del proceso en el tiempo t , y ω_t es una serie de desviaciones aleatorias independientes con media 0 y desviación estándar σ .

Una estimación de μ_t podría ser a_t , definida como una media ponderada de la observación en el tiempo t , y de la anterior estimación a_{t-1} :

$$a_t = \alpha x_t + (1 - \alpha) a_{t-1}, \text{ con } \alpha \in (0, 1)$$

El parámetro α se denomina el coeficiente de suavizado, y de su valor depende la cantidad de suavizado que se realiza. Además, si se parte de un dato inicial $a_1 = x_1$, se pueden calcular los diferentes a_t para cualquier $t = 2, \dots, n$.

Sustituyendo en la anterior ecuación a_{t-1} , posteriormente a_{t-2} y así recursivamente, se obtiene:

$$\begin{aligned}
 a_t &= \alpha x_t + (1 - \alpha) a_{t-1} = \\
 &= \alpha x_t + (1 - \alpha)(\alpha x_{t-1} + (1 - \alpha) a_{t-2}) = \\
 &= \alpha x_t + \alpha(1 - \alpha)x_{t-1} + (1 - \alpha)^2 a_{t-2} = \\
 &= \alpha x_t + \alpha(1 - \alpha)x_{t-1} + (1 - \alpha)^2(\alpha x_{t-2} + (1 - \alpha) a_{t-3}) = \\
 &= \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + (1 - \alpha)^3 a_{t-3} = \\
 &= \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \alpha(1 - \alpha)^3 x_{t-3} + \dots = \\
 &= \sum_{i=1}^t \alpha(1 - \alpha)^{t-i} x_i
 \end{aligned}$$

Como $0 < \alpha < 1$, los pesos de la serie $\alpha(1 - \alpha)^{(t - i)}$ forman una serie geométrica con suma 1.

Esta ecuación describe lo que se conoce como un “modelo de suavizado exponencial”. Como $\alpha \in (0, 1)$, en este modelo se da más peso (más importancia) a las observaciones que son más recientes.

5.1.1. Recorrido aleatorio: caso $\alpha = 1$

Un caso especial puede darse cuando en el modelo anterior, el coeficiente de suavizado es 1 ($\alpha = 1$). En ese caso, $a_t = x_t$, y la media estacionaria μ_t se considera igual a x_{t-1} , la anterior observación.

Entonces, $x_t = x_{t-1} + \omega_t$, y el único cambio que se produce entre una observación y la siguiente, se debe a una desviación aleatoria ω_t . Este modelo se denomina “recorrido aleatorio”, y su uso está extendido en diversos campos de la ciencia: biología, física, economía, Por ejemplo, se utiliza en análisis bursátil para modelizar los retornos logarítmicos.

Nótese que el otro caso especial en el que $\alpha = 0$ no debe ser considerado, ya que implica $a_t = a_{t-1}$, y por recursividad $a_t = a_{t-1} = \dots = a_1 = x_1$.

5.2. Elección del mejor coeficiente de suavizado

Se puede definir el error de predicción de cada paso t como $e_t = x_t - a_{t-1}$. Si no se especifica el coeficiente α , la función que realiza el suavizado exponencial en R (**HoltWinters**), obtiene el valor de α que minimiza la suma de estos errores al cuadrado, $\sum_{t=2}^n e_t^2$.

Este método para la obtención de α no es necesariamente la mejor forma de calcularlo [42], sino que a menudo debemos establecer otros criterios más adecuados al problema que se busca resolver.

En el caso de la estimación de la incidencia de cáncer, se dan a α los valores 0, 0,05, 0, 1, ..., 0, 95, 1. Para cada valor de α se realiza el siguiente proceso para todas las localizaciones anatómicas y para cada año objetivo entre 2004 y 2013:

1. Se modeliza la RIM desde el año 1985 (primer año del que se conoce la incidencia) hasta cinco años antes del año objetivo, utilizando la técnica de suavizado exponencial
2. Se utiliza el método NORDPRED para proyectar la mortalidad del año objetivo
3. Se obtiene el número de casos incidentes del año objetivo multiplicando la RIM modelada por la mortalidad proyectada

Las fuentes de información utilizadas son las mencionadas en el capítulo 4. Posteriormente, podemos calcular el MAPE obtenido para cada α y localización anatómica, y considerar entonces como el α más adecuado aquel que minimiza el MAPE obtenido.

El código realizado en R se basa en la función **HoltWinters** del paquete **stats**, y utiliza las funciones del método NORDPRED [30]. El código permite cambiar el número de divisiones del intervalo (0, 1] para el valor α (por defecto, 20), en caso de que se quiera una estimación más precisa de α . Tiene un coste computacional elevado, debido a las 2.800 estimaciones que se realizan (14 localizaciones \times 10 años \times 20 valores de α).

Código del suavizado exponencial.

```

1 # Lectura de ficheros
2 incidencia <- read.csv2("incidencia.csv", fileEncoding="latin1")
3 mortalidad <- read.csv2("mortalidad.csv", fileEncoding="latin1")
4 poblacion <- read.csv2("poblacion.csv", fileEncoding="latin1")
5 parametros регистра <- read.csv2("parametros регистра.csv")
6
7 # Casos observados 2004-2013 hombres ----
8 o114 <- c(138, 203, 156, 180, 172, 227, 208, 214, 243, 222)
9 o115 <- c(90, 97, 93, 95, 115, 136, 112, 110, 115, 121)
10 o116 <- c(307, 350, 332, 344, 377, 317, 348, 312, 353, 351)
11 o117 <- c(327, 351, 405, 442, 478, 498, 463, 519, 494, 447)
12 o118 <- c(257, 245, 245, 253, 251, 275, 264, 265, 312, 268)
13 o119 <- c(87, 70, 89, 76, 94, 76, 83, 67, 86, 73)
14 o199 <- c(784, 891, 824, 872, 876, 939, 892, 950, 924, 947)
15
16 # Casos observados 2004-2013 mujeres ----
17 o614 <- c(117, 138, 135, 153, 144, 153, 152, 144, 168, 163)
18 o615 <- c(67, 61, 75, 60, 67, 70, 70, 77, 66, 56)
19 o616 <- c(45, 52, 51, 53, 48, 62, 59, 71, 60, 56)
20 o617 <- c(338, 345, 389, 410, 415, 494, 419, 466, 450, 494)
21 o618 <- c(114, 106, 99, 107, 103, 123, 127, 124, 107, 124)
22 o619 <- c(60, 77, 58, 67, 68, 69, 62, 50, 53, 51)
23 o699 <- c(718, 763, 703, 822, 790, 838, 850, 856, 791, 858)
24
25 # Selección de localización anatómica y número de alphas a
   incluir en el análisis
26 loc <- 699
27 numero_alfas <- 20 # 20 -> alpha = 0.05, ..., 0.95, 1.00
28 observado <- eval(as.name(paste0("o", loc)))
29
30
31 for(i in 0:numero_alfas){
32
33   # Cálculo de alpha
34   alfa <- i / numero_alfas
35
36   if(i == 0){
37     vector <- NA # Se crea el vector que almacenará los MAPE
38     alfa <- NULL # El primer caso estima el coeficiente alpha
39   }
40
41   for(año.a.estimar in 2004:2013){

```

```
42
43 # Parámetros para estimar la mortalidad para el año a estimar
44 añoFrim <- año.a.estimar - 5
45 añoIrim <- 1985
46 crear.parametros.tumor.RCG()
47 crear.parametros.registro.RCG()
48 añoIP1 <- año.a.estimar - 22
49 añoFP1 <- año.a.estimar - 18
50 añoIP2 <- año.a.estimar - 17
51 añoFP2 <- año.a.estimar - 13
52 añoIP3 <- año.a.estimar - 12
53 añoFP3 <- año.a.estimar - 8
54 añoIP4 <- año.a.estimar - 7
55 añoFP4 <- año.a.estimar - 3
56 añoIP5 <- año.a.estimar - 2
57 añoFP5 <- año.a.estimar + 2
58
59 # Selección de incidencia y mortalidad, creación de RIM
60 incidencia.seleccion <- subset(incidencia,
61                               incidencia$AÑO %in% añoIrim:añoFrim
62                               &
63                               ! incidencia$ICDO %in% c(121, 621))
64                               [, -23]
63 mortalidad.seleccion <- subset(mortalidad,
64                               mortalidad$AÑO %in% añoIrim:añoFrim
65                               &
66                               mortalidad$PROVINCIA == 18)
66 tabla.rim <- merge(incidencia.seleccion, mortalidad.seleccion
67 )
67 tabla.rim$R <- (tabla.rim$I01 + tabla.rim$I02 +
68     tabla.rim$I03 + tabla.rim$I04 + tabla.rim$I05 +
69     tabla.rim$I06 + tabla.rim$I07 + tabla.rim$I08 +
70     tabla.rim$I09 + tabla.rim$I10 + tabla.rim$I11 +
71     tabla.rim$I12 + tabla.rim$I13 + tabla.rim$I14 +
72     tabla.rim$I15 + tabla.rim$I16 + tabla.rim$I17 +
73     tabla.rim$I18) /
74     (tabla.rim$M01 + tabla.rim$M02 + tabla.rim$M03 +
75     tabla.rim$M04 + tabla.rim$M05 + tabla.rim$M06 +
76     tabla.rim$M07 + tabla.rim$M08 + tabla.rim$M09 +
77     tabla.rim$M10 + tabla.rim$M11 + tabla.rim$M12 +
78     tabla.rim$M13 + tabla.rim$M14 + tabla.rim$M15 +
79     tabla.rim$M16 + tabla.rim$M17 + tabla.rim$M18)
80
```

CAPÍTULO 5. PROPUESTA DE MEJORA DEL MÉTODO RIM, BASADA EN SUAVIZADO EXPONENCIAL

```
81 # Selección de RIM para la localización seleccionada
82 a <- subset(tabla.rim$R, tabla.rim$ICDO == loc)
83
84 # Modelo de alisamiento exponencial
85 hw <- HoltWinters(a, alpha = alfa, gamma = FALSE, beta =
86 FALSE)
87
88 # Predicción de 5 años (para llegar hasta el año a estimar)
89 hw.p <- predict(hw, n.ahead = 5)
90
91 # NORDPRED para el año a estimar
92 prep.nordpred.lista()
93 prep.nordpred.sexo("M")
94 prep.nordpred.sexo("F")
95 escribir.prediccion.año.a.estimar()
96 prediccion.mortalidad <- read.csv2("prediccion.año.a.estimar.
97 csv")
98
99 M <- subset(prediccion.mortalidad, prediccion.mortalidad$icdo
100 == loc)
101 M <- M$M01 + M$M02 + M$M03 + M$M04 + M$M05 + M$M06 +
102 M$M07 + M$M08 + M$M09 + M$M10 + M$M11 + M$M12 +
103 M$M13 + M$M14 + M$M15 + M$M16+ M$M17 + M$M18
104
105 # Cálculo de casos incidentes estimados: mortalidad * RIM
106 total <- M * hw.p[5, ]
107
108 # Se anexan los casos estimados año a año
109 ifelse(año.a.estimar == 2004,
110        total.periodos <- total,
111        total.periodos <- rbind(total.periodos, total))
112 }
113
114 # Se imprime el MAPE, el valor de alpha, y los números de casos
115 # estimados
116 print(paste0("MAPE = ",
117             round(10 * sum(abs(total.periodos - observado) /
118                     observado), 1)))
119 print(paste0("alfa = ", alfa))
120 print("-----")
121
122 # Se guarda el valor del MAPE
123 vector <- rbind(vector, 10 * sum(abs(total.periodos - observado
```

```

) / observado))

119 if(i == numero_alfas){ # Al terminar el proceso:
120
122 # Se quita el primer elemento, que es NA
123 vector <- vector[-1, ]
124
125 # Se extrae el primer valor, resultante de la estimación
126 # automática de alpha
126 alpha_estimado <- vector[1]
127 vector <- vector[-1]
128
129 # El resto de valores se guardan en un vector. Se añaden los
130 # alphas.
130 vector <- cbind(c(1:numero_alfas) / numero_alfas, vector)
131 colnames(vector) <- c("al", "v")
132
133 # Se representará en rojo el valor más bajo de MAPE
134 colores <- rep("black", numero_alfas)
135 for(i in 1:numero_alfas){
136   if(vector[i, 2] == min(vector[, 2])) colores[i] <- "red"
137 }
138
139 # Se muestra la tabla con coeficientes de suavizado, MAPE,
140 # y color para detectar el MAPE más bajo
141 print(cbind(vector, colores))
142
143 # Mostramos el gráfico
144 plot(vector, main = paste0(loc), pch = 20, cex = 1.5,
145       col = colores, ylab = "MAPE", xlab = "alpha")
146 abline(h = alpha_estimado, lty = 3)
147
148 # Guardamos el gráfico como PDF
149 pdf(paste0(loc, "-", numero_alfas, ".pdf"))
150 plot(vector, main = paste0(loc), pch = 20, cex = 1.5,
151       col = colores, ylab = "MAPE", xlab = "alpha")
152 abline(h = alpha_estimado, lty = 3)
153 dev.off()
154 }
155 }

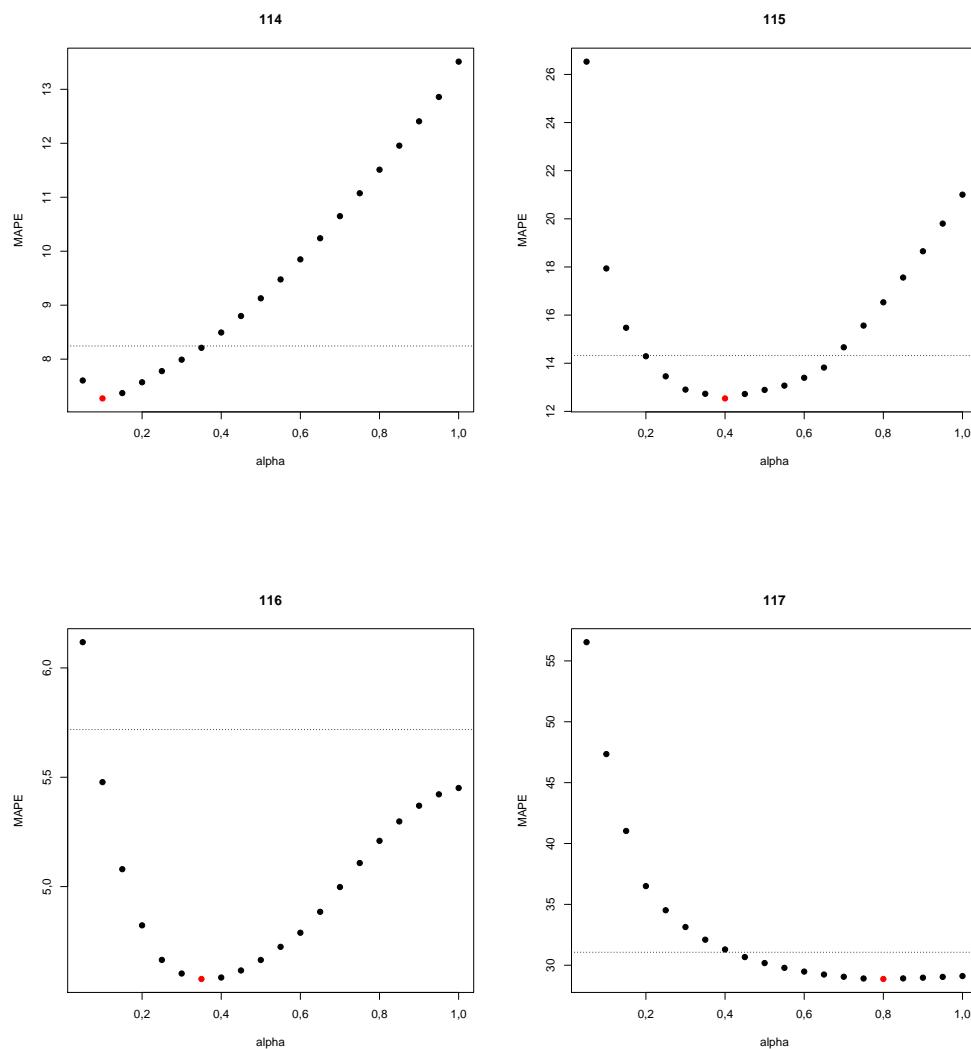
```

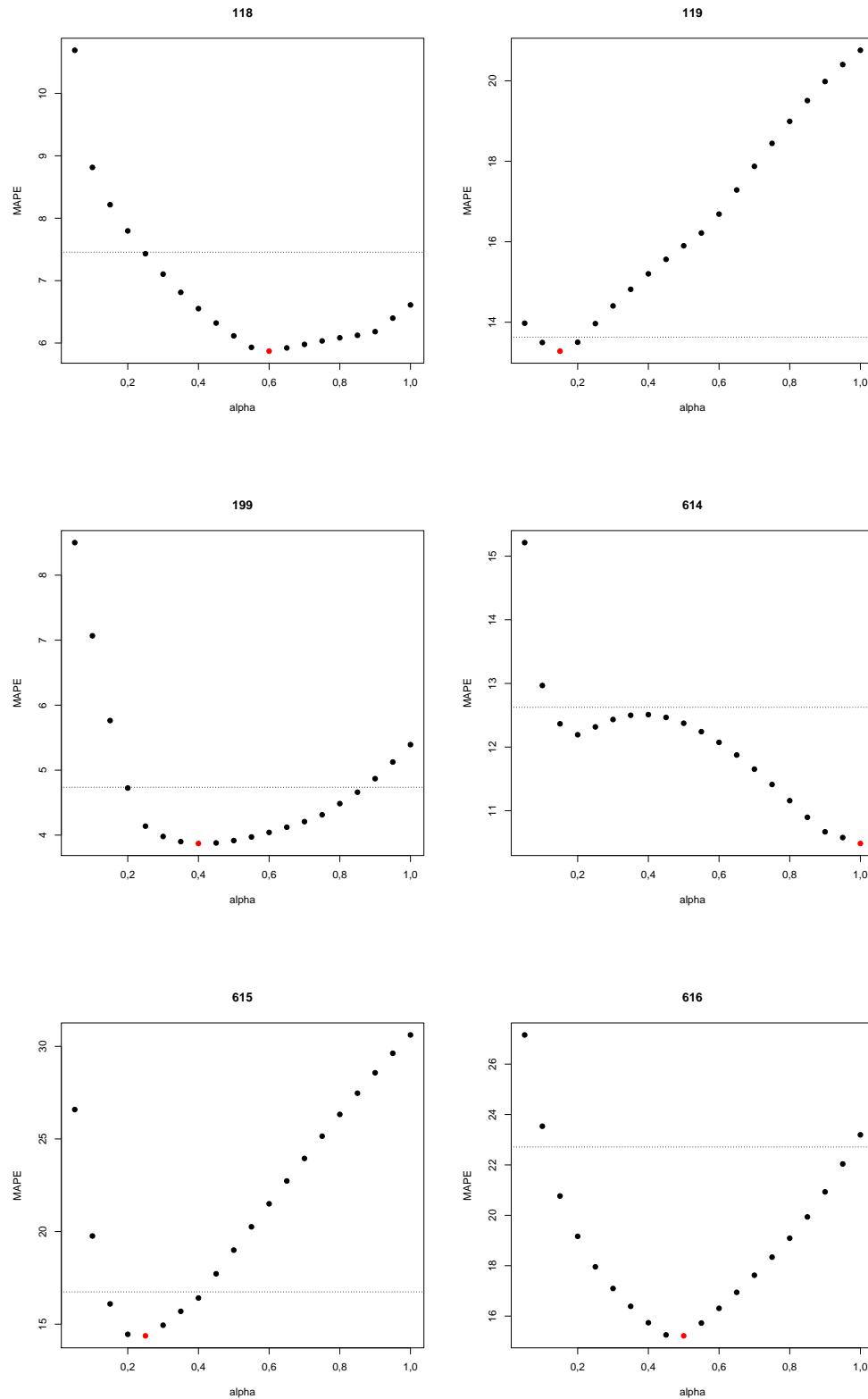
En la Figura 9 se representan, para cada localización anatómica, los valores de MAPE del periodo 2004-2013 obtenidos para los distintos valores propuestos

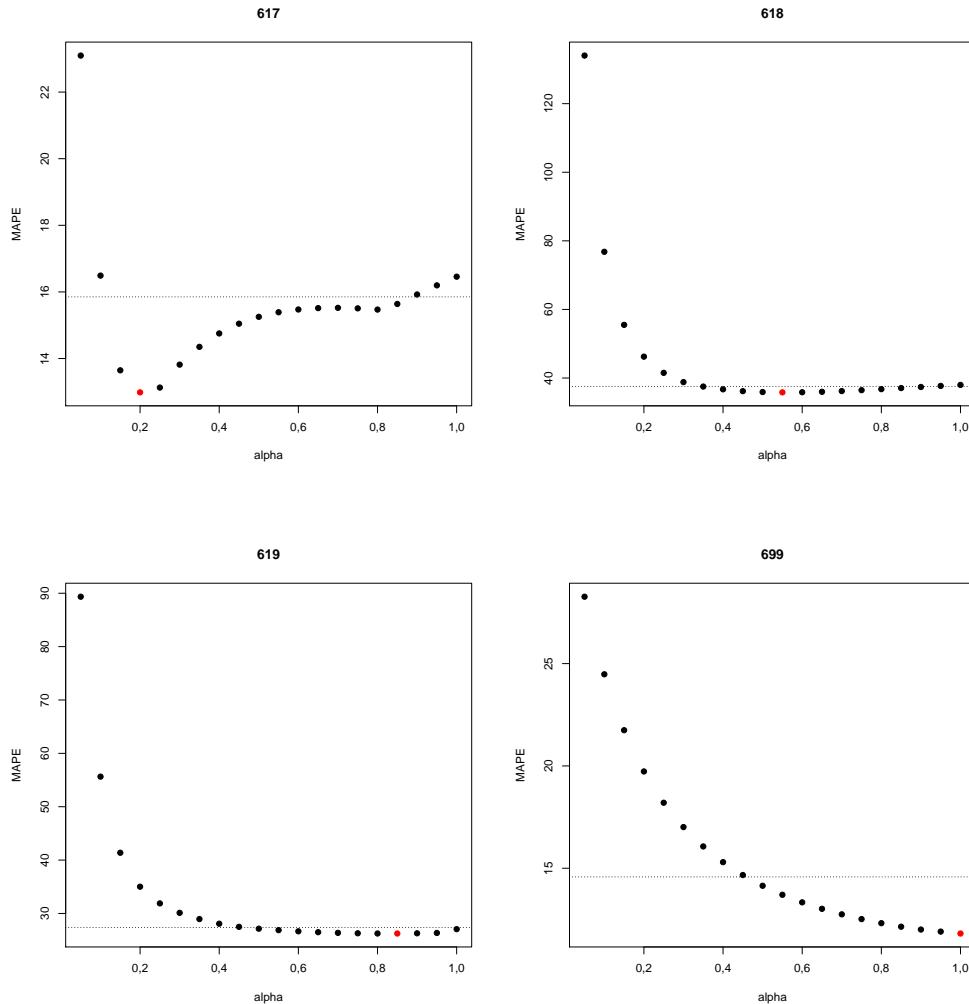
de α .

Figura 9. Error absoluto porcentual medio (MAPE) del periodo 2004-2013, por localización anatómica, para los distintos valores de α (0.05, 0.1, ..., 0.95, 1)

Se representa en línea discontinua la mejor estimación de *alpha* realizada para cada año según el comando *HoltWinters* que minimiza $\sum_{t=2}^n e_t^2$. El valor de α que minimiza el MAPE para cada localización anatómica se representa en color rojo. Los títulos de los gráficos se corresponden a los códigos internos de cada localización anatómica (Tabla 4).







En dos de los casos analizados (colon y otras localizaciones en mujeres), el mejor modelo de suavizado exponencial fue un recorrido aleatorio ($\alpha = 1$).

5.3. Validación del método RIM

En la tabla 8 se muestra, para cada localización anatómica, el mejor método de suavizado exponencial con el coeficiente α y el MAPE correspondiente. Se comparan estos errores con los mejores MAPE obtenidos por el método RIM en el capítulo 4.

Tabla 8. Error absoluto porcentual medio (MAPE) del periodo 2004-2013 mínimo para los distintos valores de α , por localización anatómica. **Comparación con los MAPE obtenidos por el método RIM.** Se señala en negrita el MAPE más bajo entre los dos métodos.

Sexo	Localización anatómica	Método RIM		Suavizado exponencial	
		MAPE	Mejor escenario	MAPE	Mejor α
Hombres	Colon	8,6	C3	7,3	0,1
	Recto	16,8	C3	12,5	0,4
	Pulmón	4,5	C3	4,6	0,35
	Próstata	27,4	L	28,9	0,8
	Vejiga	8,5	C1	5,9	0,6
	Estómago	14,1	C5	13,3	0,15
	Otros	3,5	C3	3,9	0,4
Mujeres	Colon	10,0	C5	10,5	1
	Recto	19,2	C3	14,4	0,25
	Pulmón	18,3	L	15,2	0,5
	Mama	14,8	C3	13,0	0,2
	Cuerpo uterino	17,3	C5	35,8	0,55
	Ovario	18,9	L	26,2	0,85
	Otros	8,7	L	11,8	1

El método de suavizado exponencial consigue mejorar el MAPE obtenido mediante el método RIM en 7 de las 14 localizaciones anatómicas analizadas. Las mayores mejoras se obtienen en cáncer de recto, tanto en mujeres (el MAPE se reduce en 4,8) como en hombres (-4,3), en cáncer de pulmón en mujeres (-3,1) y en cáncer de vejiga en hombres (-2,6).

En general, y como se podía esperar, el suavizado exponencial con valores de α cercanos a 1 (que indican poco suavizado) consigue valores de MAPE similares a los obtenidos mediante los escenarios constantes C1, C3 y C5 del método RIM.

6. Estimaciones de la incidencia de cáncer para la provincia de Granada, año 2018

Con el objetivo de disponer de datos recientes sobre la incidencia de cáncer en la provincia de Granada, se han realizado estimaciones para el año 2018. Estos resultados son útiles para mejorar la planificación de la atención sanitaria y la vigilancia de la enfermedad en la región de Granada.

En primer lugar, se describen las fuentes de información utilizadas. En segundo lugar, se presentan resultados usando el método RIM, con el mejor escenario hallado en el Capítulo 4 (Tabla 7). Por último, se utiliza el trabajo realizado en el Capítulo 5 para usar, en función de la localización anatómica, el mejor método de entre los dos planteados: método RIM o suavizado exponencial (Tabla 8).

6.1. Fuentes de información

Las fuentes de información son las siguientes:

- Incidencia: Datos del Registro de Cáncer de Granada (RCG). Casos incidentes de cáncer en la provincia de Granada, periodo 1999-2013 (1985-2013 para suavizado exponencial), por año, sexo, grupo de edad y localización anatómica.
- Mortalidad: Ministerio de Sanidad, Consumo y Bienestar Social. Defunciones por cáncer en la provincia de Granada, periodo 1996-2015 (1985-2015 para suavizado exponencial), por sexo, año, grupo de edad y localización anatómica [37].
- Población: Instituto Nacional de Estadística (INE). Población de la provincia de Granada a 1 de Julio, periodo 1996-2015 (1985-2015 para suavizado exponencial), por año, sexo y grupo de edad [11]. Proyección de población intercensal para 2018 (media de poblaciones a 1 de enero de 2018 y a 1 de enero de 2019) de la provincia de Granada [32].

**CAPÍTULO 6. ESTIMACIONES DE LA INCIDENCIA DE CÁNCER
64 PARA LA PROVINCIA DE GRANADA, AÑO 2018**

6.2. Método RIM

Se ha seguido el método descrito en el capítulo 4, estimando la incidencia para las mismas localizaciones: colon, recto, pulmón, próstata, vejiga, estómago y otros en hombres, y colon, recto, pulmón, mama, ovario, cuerpo uterino y otros en mujeres. Para cada combinación de sexo y localización anatómica se utilizó el mejor escenario obtenido en la validación.

Se estima que en el año 2018 se diagnosticaron 4.701 nuevos casos de cáncer en la provincia de Granada.

El 59,3 % de los casos son diagnosticados en hombres, siendo los cánceres más frecuentes los de próstata (752 casos), vejiga (356 casos) y pulmón (342 casos). En mujeres, las localizaciones más frecuentes son el cáncer de mama (565 casos), el cáncer de colon (148 casos) y el cáncer de cuerpo uterino (133 casos).

Las tasas brutas por 100.000 habitantes son 613,2 en hombres y 412,9 en mujeres. Las tasas estandarizadas por 100.000 habitantes son de 315,9 en hombres y 216,2 en mujeres por la población estándar mundial (ASR-W), de 460,7 en hombres y 296,5 en mujeres por la antigua población estándar europea (ASR-E₁₉₇₆) y de 703,2 en hombres y 391,3 en mujeres por la nueva población estándar europea (ASR-E₂₀₁₃).

En la Tabla 9 se presentan los resultados de las estimaciones de incidencia de cáncer en la provincia de Granada para el año 2018. Para cada sexo y localización anatómica, se seleccionó el mejor escenario del método RIM obtenido en la validación.

En las Figuras 10 (hombres) y 11 (mujeres) se representa, para cada localización anatómica, la evolución a lo largo del tiempo del número de casos, que es observado durante el periodo 1985-2013 y estimado para el año 2018 con el mejor escenario obtenido en la validación. Para la estimación de 2018 se representa además el intervalo de credibilidad al 95 %.

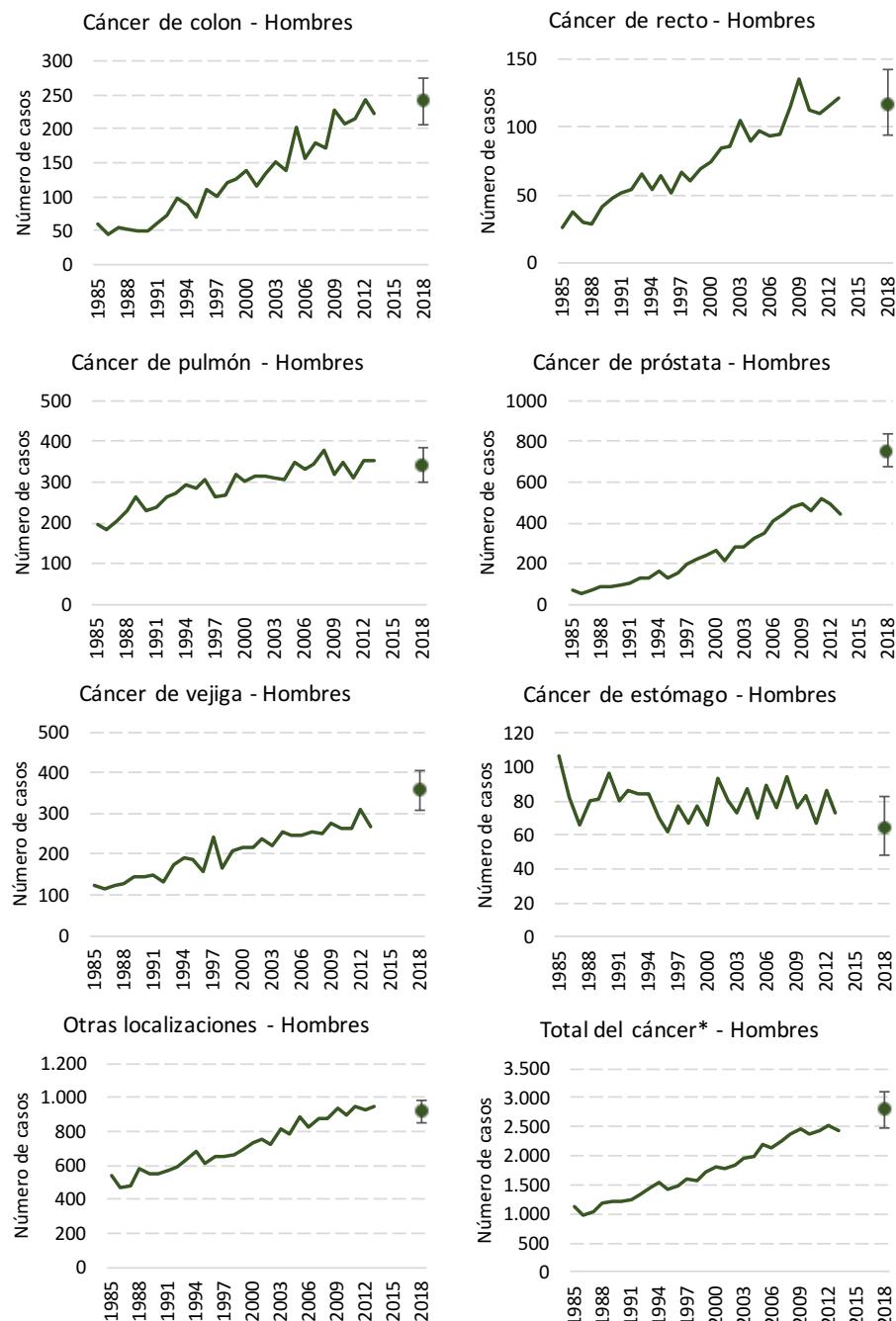
Tabla 9. Estimaciones de incidencia de cáncer en la provincia de Granada para el año 2018, para el mejor escenario obtenido en la validación del método RIM. Número de casos estimados (N), tasas bruta (TB), estandarizada por la población mundial (ASR-W), estandarizada por la población europea de 1976 (ASR-E₁₉₇₆), y estandarizada por la población europea de 2013 (ASR-E₂₀₁₃) por 100.000 habitantes. Intervalos de credibilidad al 95 %.

Sexo	Localización anatómica	N	TB	ASR-W	ASR-E ₁₉₇₆	ASR-E ₂₀₁₃	Esc
Hombres	Colon	240 (206, 275)	52,8 (45,3, 60,5)	24,9 (21,2, 28,9)	38 (32,6, 43,8)	61,8 (53, 71,1)	C3
	Recto	117 (94, 142)	25,7 (20,7, 31,2)	13,3 (10,5, 16,3)	19,4 (15,5, 23,7)	29,2 (23,4, 35,5)	C3
	Pulmón	342 (302, 384)	75,2 (66,4, 84,5)	37,8 (33,1, 42,8)	56,1 (49,4, 63,2)	85,4 (75,3, 96,1)	C3
	Próstata	752 (672, 839)	165,4 (147,8, 184,5)	82,4 (73,3, 92,2)	122,9 (109,6, 137,3)	197,5 (176,2, 220,2)	L
	Vejiga	356 (310, 405)	78,3 (68,2, 89,1)	38,2 (33,1, 43,7)	57,3 (49,9, 65,3)	91,5 (79,7, 104,1)	C1
	Estómagos	64 (48, 83)	14,1 (10,6, 18,1)	7,4 (5,4, 9,6)	10,6 (7,9, 13,7)	15,9 (11,9, 20,4)	C5
	Otros	917 (850, 985)	201,7 (187, 216,7)	111,8 (103,1, 120,9)	156,3 (144,7, 168,2)	222 (205,7, 238,7)	C3
	Total *	2.788 (2.482, 3.113)	613,2 (545,9, 684,6)	315,9 (279,8, 354,4)	460,7 (409,5, 515,1)	703,2 (625,2, 786,1)	
Mujeres	Colon	148 (123, 175)	31,9 (26,5, 37,8)	13,1 (10,6, 15,8)	19,5 (15,9, 23,4)	29,7 (24,6, 35,3)	C5
	Recto	44 (31, 60)	9,5 (6,7, 12,9)	3,9 (2,6, 5,5)	5,8 (3,9, 7,9)	9,2 (6,4, 12,4)	C3
	Pulmón	78 (55, 106)	16,8 (11,9, 22,9)	8,8 (6,1, 12,1)	12,4 (8,6, 16,9)	16,2 (11,4, 22,1)	L
	Mama	565 (513, 620)	121,9 (110,7, 133,8)	71,1 (64,2, 78,4)	95,8 (86,6, 105,4)	115,8 (105, 127,1)	C3
	Cuerpo uterino	133 (108, 159)	28,7 (23,3, 34,3)	14,4 (11,6, 17,5)	20,7 (16,7, 24,9)	27,9 (22,7, 33,4)	C5
	Ovario	47 (31, 67)	10,1 (6,7, 14,5)	5,8 (3,7, 8,4)	7,9 (5,1, 11,3)	9,8 (6,4, 13,9)	L
	Otros	898 (817, 982)	193,8 (176,3, 211,9)	99,1 (89,3, 109,4)	134,6 (122, 147,8)	182,7 (166,2, 199,9)	L
	Total *	1.913 (1.678, 2.169)	412,9 (362,1, 468,1)	216,2 (188,1, 247)	296,5 (259, 337,5)	391,3 (342,7, 444,1)	

* Total excepto piel no melanoma

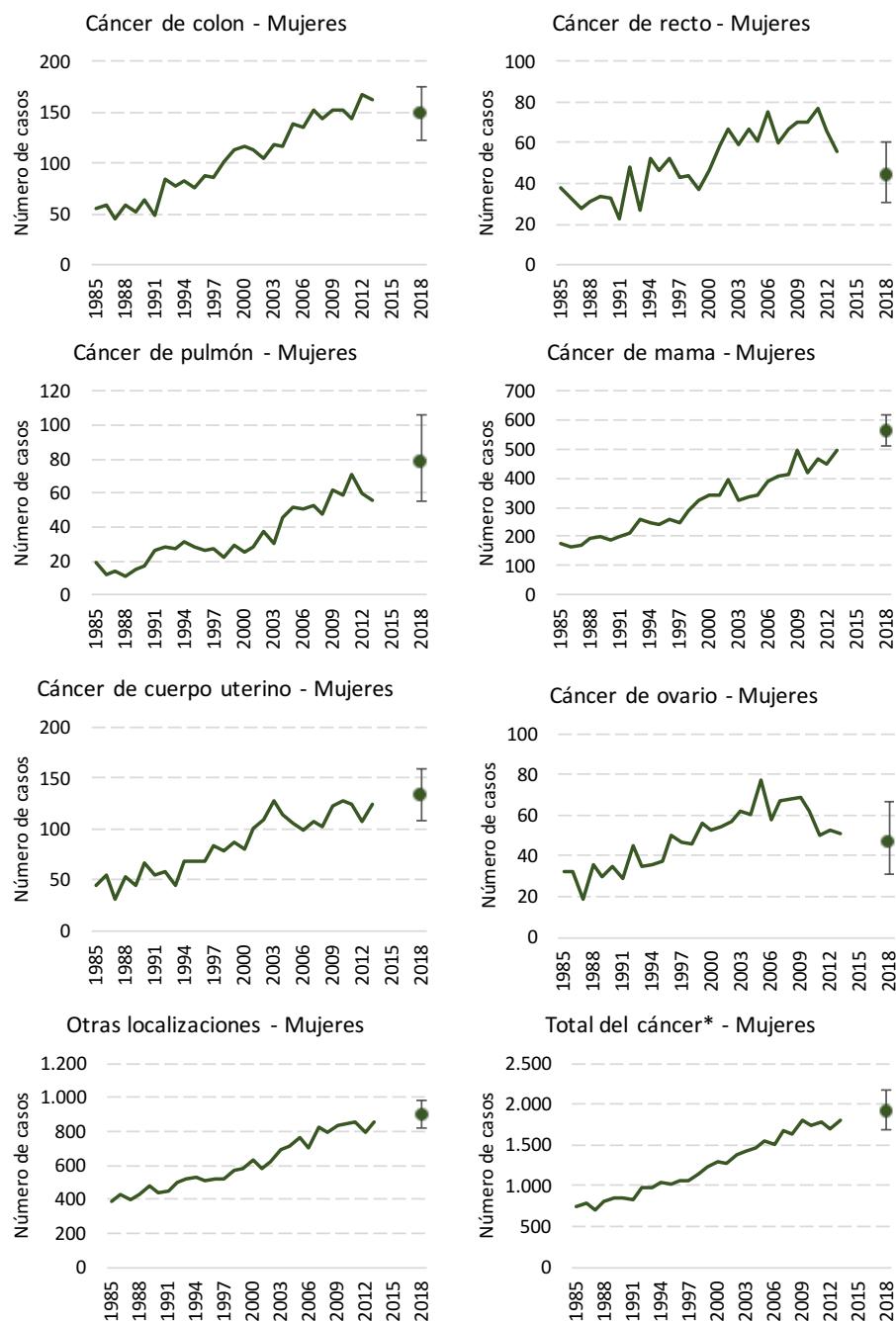
**CAPÍTULO 6. ESTIMACIONES DE LA INCIDENCIA DE CÁNCER
66 PARA LA PROVINCIA DE GRANADA, AÑO 2018**

Figura 10. Número de casos observados en la provincia de Granada 1985-2013, y número de casos estimados por el mejor escenario (método RIM) para el año 2018, con intervalo de credibilidad al 95 %. Hombres.



*Total excepto piel no melanoma

Figura 11. Número de casos observados en la provincia de Granada 1985-2013, y número de casos estimados por el mejor escenario (método RIM) para el año 2018. Mujeres.



*Total excepto piel no melanoma

**CAPÍTULO 6. ESTIMACIONES DE LA INCIDENCIA DE CÁNCER
68 PARA LA PROVINCIA DE GRANADA, AÑO 2018**

6.3. Método RIM + Suavizado exponencial

Tras realizar la validación de las estimaciones por el método RIM y el método de suavizado exponencial, se ha encontrado el mejor escenario para cada localización anatómica. Utilizando estos escenarios, se ha calculado la estimación de la incidencia de cáncer para 2018 en la provincia de Granada.

Se estima que en el año 2018 se diagnosticaron 4.668 nuevos casos de cáncer en la provincia de Granada.

El 59,6 % de los casos son diagnosticados en hombres, siendo los cánceres más frecuentes los de próstata (752 casos), vejiga (350 casos) y pulmón (342 casos). En mujeres, las localizaciones más frecuentes son el cáncer de mama (538 casos), el cáncer de colon (148 casos) y el cáncer de cuerpo uterino (133 casos).

Las tasas brutas por 100.000 habitantes son 611,6 en hombres y 407,3 en mujeres. Las estimaciones son similares a las realizadas sólo con el método RIM para todas las localizaciones excepto para cáncer de mama, con 27 casos menos (6 casos menos por 100.000 mujeres).

Los intervalos de predicción del método de suavizado exponencial (que se calculan en R con el parámetro `prediction.interval = TRUE` en la función `predict`) son más amplios que los intervalos de credibilidad del método RIM.

Por otra parte, al realizarse el suavizado exponencial sin distinción por grupos de edad, no se pueden obtener tasas estandarizadas por la población europea o mundial.

En la Tabla 10 se presentan los resultados detallados de las estimaciones de la incidencia de cáncer en la provincia de Granada para 2018, por sexo y localización anatómica.

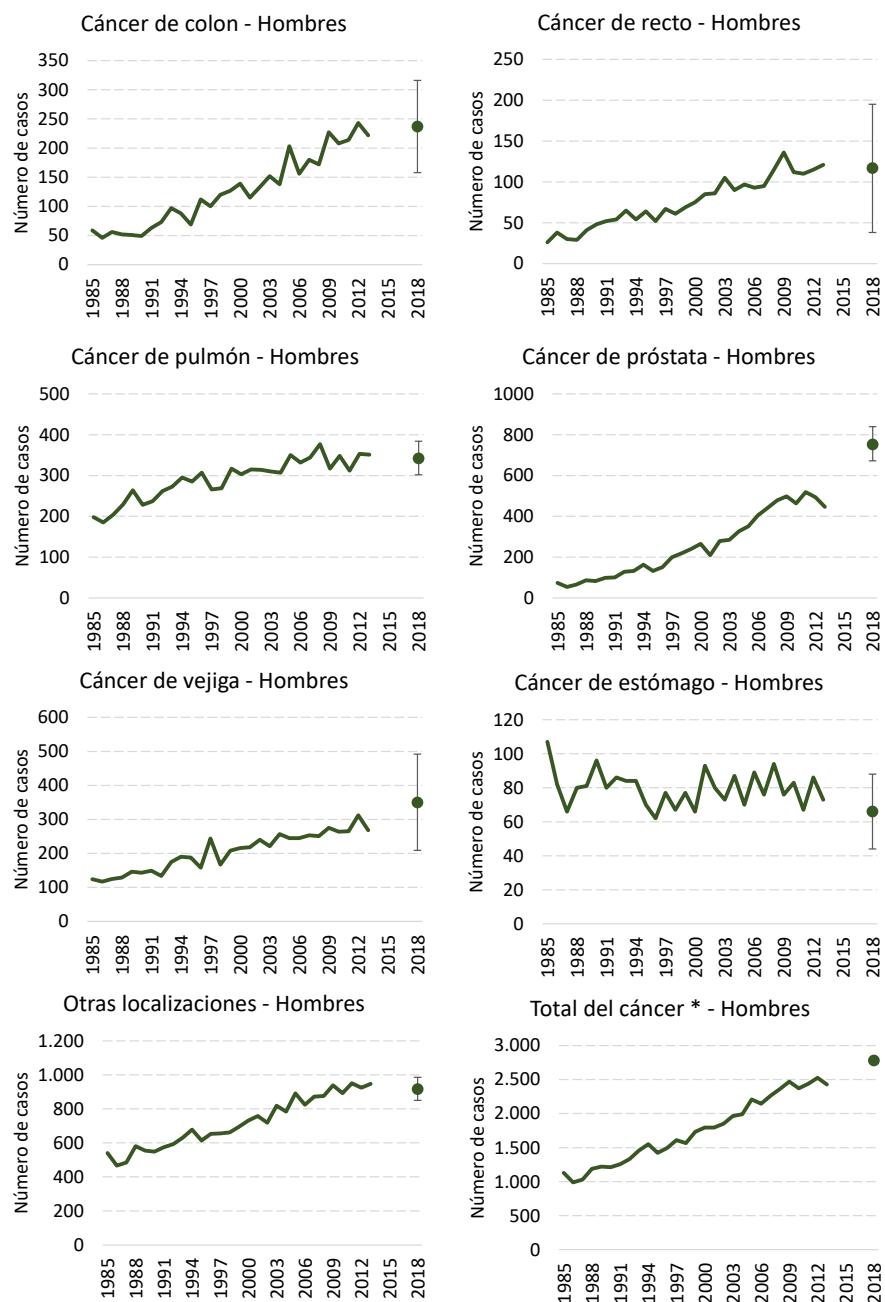
Tabla 10. Estimaciones de incidencia de cáncer en la provincia de Granada para el año 2018, para el mejor escenario obtenido en la validación del método RIM y del método de suavizado exponencial. Número de casos estimados (N) y tasas brutas (TB) por 100.000 habitantes. Intervalos de credibilidad al 95 % para el método RIM, e intervalos de predicción para el método de suavizado exponencial (SE).

Sexo	Localización anatómica	N	TB	Escenario
Hombres	Colon	237 (158, 316)	52,1 (34,8, 69,5)	$SE_{\alpha=0,1}$
	Recto	117 (38, 195)	25,6 (8,4, 42,8)	$SE_{\alpha=0,4}$
	Pulmón	342 (302, 384)	75,2 (66,4, 84,5)	C3
	Próstata	752 (672, 839)	165,4 (147,8, 184,5)	L
	Vejiga	350 (209, 492)	77 (45,9, 108,1)	$SE_{\alpha=0,6}$
	Estómago	66 (44, 88)	14,6 (9,8, 19,4)	$SE_{\alpha=0,15}$
	Otros	917 (850, 985)	201,7 (187,0, 216,7)	C3
	Total *	2.781	611,6	
Mujeres	Colon	148 (123, 175)	31,9 (26,5, 37,8)	C5
	Recto	51 (15, 88)	11,1 (3,3, 18,9)	$SE_{\alpha=0,25}$
	Pulmón	72 (16, 129)	15,6 (3,5, 27,7)	$SE_{\alpha=0,5}$
	Mama	538 (402, 674)	116,0 (86,7, 145,4)	$SE_{\alpha=0,2}$
	Cuerpo uterino	133 (108, 159)	28,7 (23,3, 34,3)	C5
	Ovario	47 (31, 67)	10,1 (6,7, 14,5)	L
	Otros	898 (817, 982)	193,8 (176,3, 211,9)	L
	Total *	1.887	407,3	

* Total excepto piel no melanoma

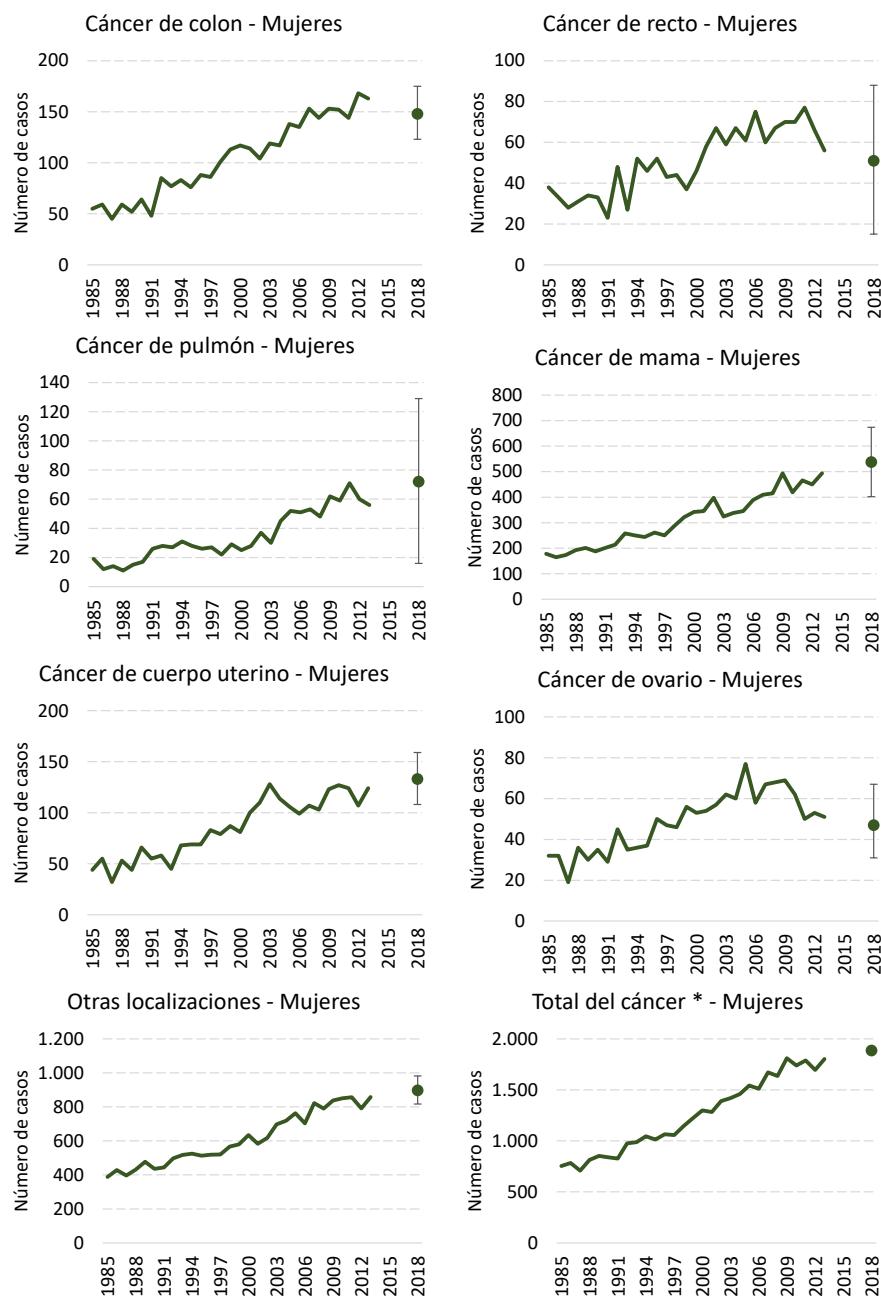
**CAPÍTULO 6. ESTIMACIONES DE LA INCIDENCIA DE CÁNCER
70 PARA LA PROVINCIA DE GRANADA, AÑO 2018**

Figura 12. Número de casos observados en la provincia de Granada 1985-2013, y número de casos estimados por el mejor escenario (método RIM + método suavizado exponencial) para el año 2018, con intervalo de credibilidad/intervalo de predicción al 95 %. Hombres.



*Total excepto piel no melanoma

Figura 13. Número de casos observados en la provincia de Granada 1985-2013, y número de casos estimados por el mejor escenario (método RIM + método suavizado exponencial) para el año 2018, con intervalo de credibilidad/intervalo de predicción al 95 %. Mujeres.



*Total excepto piel no melanoma

**CAPÍTULO 6. ESTIMACIONES DE LA INCIDENCIA DE CÁNCER
72 PARA LA PROVINCIA DE GRANADA, AÑO 2018**

7. Líneas abiertas de trabajo

7.1. Proyectos colaborativos

La participación en proyectos colaborativos con otros Registros de Cáncer de Población permitiría disponer de datos de varias regiones, ganar potencia estadística y poder estimar con confianza a regiones más grandes. En este marco, las publicaciones realizadas para España por REDECAN [8, 13, 14] han demostrado ser de gran utilidad para la vigilancia del cáncer a nivel nacional.

Con respecto a la metodología del alisamiento exponencial, un aumento del número de casos podría permitir realizar las estimaciones por grupo de edad, obteniendo así información más detallada como tasas estandarizadas por edad.

Al incluir los datos de incidencia de otras regiones, habría que tener en cuenta en los análisis las diferencias existentes en la distribución geográfica del cáncer. En España, la incidencia de cáncer sigue en general un gradiente norte-sur [10], con una incidencia más alta en regiones del norte. Por otra parte, los programas de cribado, para cánceres como el de mama o colon-recto, se han implementado en distintos años, niveles de cobertura y participación, en las diferentes regiones.

7.2. Método de Brown

En este trabajo se ha demostrado la importancia que puede tener el análisis de series temporales aplicado a la estimación de la incidencia de cáncer. Además del suavizado exponencial realizado, se pueden plantear métodos de suavizado exponencial lineal, como el método de Brown. Usando este tipo de análisis, se detecta en primer lugar una tendencia lineal que sigue la serie temporal, y posteriormente se realiza un suavizado exponencial.

7.3. Actualización periódica de las estimaciones

Debido al desfase temporal existente entre el año actual y el último año del que se dispone la incidencia en un Registro de Cáncer, es necesario realizar estimaciones de la incidencia de manera periódica, año a año, para así disponer en todo momento de datos actualizados.

Bibliografía

- [1] National Cancer Institute. What is Cancer?, 2015.
- [2] Ron Sender, Shai Fuchs, and Ron Milo. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biology*, 14(8):e1002533, aug 2016.
- [3] María José Sánchez Pérez and Juan Bayo Calero. Apuntes del curso Prevención y diagnóstico precoz del cáncer, 2016.
- [4] International Agency for Research on Cancer. Agents classified by the IARC Monographs. 1-123, 2019.
- [5] V. J. Cogliano, R. Baan, K. Straif, Y. Grosse, B. Lauby-Secretan, F. El Ghissassi, V. Bouvard, L. Benbrahim-Tallaa, N. Guha, C. Freeman, L. Galichet, and C. P. Wild. Preventable Exposures Associated With Human Cancers. *JNCI Journal of the National Cancer Institute*, 103(24):1827–1839, dec 2011.
- [6] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6):394–424, nov 2018.
- [7] IARC, JRC, ENCR, and IACR. ECIS - European Cancer Information System, 2019.
- [8] Red Española de Registros de Cáncer (REDECAN). Estimaciones de la incidencia de cáncer en España, 2019. 2019.
- [9] Instituto Nacional de Estadística (INE). Estadísticas de defunciones según la causa de muerte, 2017.
- [10] F Bray, M Colombet, L Mery, M Piñeros, A Znaor, R Zanetti, and J Ferlay. Cancer Incidence in Five Continents. Volume XI. 2017.
- [11] Instituto Nacional de Estadística (INE). Estadísticas de cifras de población.
- [12] Catherine Hill. Cancer Incidence in Five Continents. Volume VI. *Statistics in Medicine*, 14(1):104–104, jan 1992.

- [13] Jaume Galcerán, Alberto Ameijide, Marià Carulla, Antonio Mateos, José Ramón Quirós, Araceli Alemán, Dolores Rojas, Ana Torrella, Matilde Chico, Marisa Vicente, José María Díaz, Nerea Larrañaga, Rafael Marcos-Gragera, María José Sánchez, Josefina Perucha, Maria Ramos, Carmen Navarro, Eva Ardanaz, Jànnica Bigorra, Pilar Rodrigo, Rafael Peris Bonet, and REDECAN. Estimaciones de la incidencia y la supervivencia del cáncer en España y su situación en Europa. 2014.
- [14] J. Galceran, A. Ameijide, M. Carulla, A. Mateos, J. R. Quirós, D. Rojas, A. Alemán, A. Torrella, M. Chico, M. Vicente, J. M. Díaz, N. Larrañaga, R. Marcos-Gragera, M. J. Sánchez, J. Perucha, P. Franch, C. Navarro, E. Ardanaz, J. Bigorra, P. Rodrigo, and R. Peris Bonet. Cancer incidence in Spain, 2015. *Clinical and Translational Oncology*, 19(7):799–825, jul 2017.
- [15] IARC. *Registros de Cáncer: Principios y Métodos*. 1995.
- [16] World Health Organization) Silva, Isabel dos Santos (International Agency for Research on Cancer. *Cancer Epidemiology: Principles and Methods*. 1999.
- [17] Emanuele Crocetti, Tadek Dyba, Carmen Martos, Giorgia Randi, Roisin Rooney, and Manola Bettio. The need for a rapid and comprehensive adoption of the revised European standard population in cancer incidence comparisons. *European Journal of Cancer Prevention*, 26(5):447–452, sep 2017.
- [18] Hugh H. Wolfenden. On the Methods of Comparing the Moralities of Two or More Communities, and the Standardization of Death-Rates. *Journal of the Royal Statistical Society*, 86(3):399, may 1923.
- [19] Segi M. Cancer mortality for selected sites in 24 countries (1950–57). *Sendai, Japan: Department of Public Health, Tohoku University of Medicine.*, 1960.
- [20] Editors Doll R, Payne P, Waterhouse PAH. Cancer incidence in five continents, Volume I. *Geneva: Union Internationale Contre le Cancer.*, 1966.
- [21] editors. Waterhouse JAH, Muir CS, Correa P, Powell J. Cancer incidence in five continents, Volume III. *Lyon: IARC*, page 3:456, 1976.
- [22] EUROSTAT. Revision of the European standard population: Report of the Eurostat's task force. Technical report, Luxembourg: European Union., 2013.

- [23] Libby Ellis, Laura M. Woods, Jacques Estève, Sandra Eloranta, Michel P. Coleman, and Bernard Rachet. Cancer incidence, survival and mortality: Explaining the concepts. *International Journal of Cancer*, 135(8):1774–1782, oct 2014.
- [24] Libby Ellis, Aurélien Belot, Bernard Rachet, and Michel P. Coleman. The Mortality-to-Incidence Ratio Is Not a Valid Proxy for Cancer Survival. *Journal of Global Oncology*, (5):1–9, may 2019.
- [25] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D.M. Parkin, M. Piñeros, A. Znaor, and F. Bray. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International Journal of Cancer*, page ijc.31937, dec 2018.
- [26] Sébastien Antoni, Isabelle Soerjomataram, Bjørn Møller, Freddie Bray, and Jacques Ferlay. An assessment of GLOBOCAN methods for deriving national estimates of cancer incidence. *Bulletin of the World Health Organization*, 94(3):174–184, mar 2016.
- [27] V. Moreno, J.R. González, M. Soler, F.X. Bosch, M. Kogevinas, J.M. Borràs, and Grupo de Estudio de las Causas del Cáncer en España. Estimación de la incidencia de cáncer en España período 1993-1996. *Gaceta Sanitaria*, 15(5):380–388, 2001.
- [28] G López-Abente, M Pollán, N Aragonés, B Pérez-Gómez, V Barrera-Hernández, V Lope, and B Suárez. Situación del cáncer en España: incidencia. *An. Sist. Sanit. Navar. An. Sist. Sanit. Navar*, 27(272):165–173, 2004.
- [29] M. J. Sanchez, T. Payer, R. De Angelis, N. Larranaga, R. Capocaccia, and C. Martínez. Cancer incidence and mortality in Spain: estimates and projections for the period 1981-2012. *Annals of Oncology*, 21(Supplement 3):iii30–iii36, may 2010.
- [30] Bjørn Møller, Harald Fekjaer, Timo Hakulinen, Helgi Sigvaldason, Hans H. Storm, Mats Talbäck, and Tor Haldorsen. Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches. *Statistics in Medicine*, 22(17):2751–2766, sep 2003.

- [31] Theodore R. Holford. The Estimation of Age, Period and Cohort Effects for Vital Rates. *Biometrics*, 39(2):311, jun 1983.
- [32] Instituto Nacional de Estadística (INE). Estadísticas de proyecciones de población.
- [33] Andrew Thomas, Bob O'Hara, Uwe Ligges, and Sibylle Sturtz. Making BUGS Open. *R News*, 6(1):12–17, 2006.
- [34] Sibylle Sturtz, Uwe Ligges, and Andrew Gelman. R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software*, 12(3):1–16, 2005.
- [35] Andrew Thomas, David Spiegelhalter, Nicky Best, Dave Lunn, and Ken Rice. OpenBugs.
- [36] Daniel Redondo-Sánchez, Miguel Rodríguez-Barranco, Alberto Ameijide, Francisco Javier Alonso, Pablo Fernández-Navarro, José J Jiménez-Moleón, and María José Sánchez. Cancer Incidence Estimation from Mortality Data: A Validation Study within a Population-Based Cancer Registry. *Population Health Metrics (under review)*, 2019.
- [37] Ministerio de Sanidad Consumo y Bienestar Social. Estadísticas de defunciones según la causa de muerte.
- [38] World Health Organization (WHO). *ICD-10: International Statistical Classification of diseases and related health problems: 10th revision*. 1990.
- [39] Paul F. Pinsky, Philip C. Prorok, Kelly Yu, Barnett S. Kramer, Amanda Black, John K. Gohagan, E. David Crawford, Robert L. Grubb, and Gerald L. Andriole. Extended mortality results for prostate cancer screening in the PLCO trial with median follow-up of 15 years. *Cancer*, 123(4):592–599, feb 2017.
- [40] Fritz H. Schröder, Jonas Hugosson, Monique J. Roobol, Teuvo L.J. Tammela, Stefano Ciatto, Vera Nelen, Maciej Kwiatkowski, Marcos Lujan, Hans Lilja, Marco Zappa, Louis J. Denis, Franz Recker, Alvaro Páez, Liisa Määttänen, Chris H. Bangma, Gunnar Aus, Sigrid Carlsson, Arnauld Villers, Xavier Rebillard, Theodorus van der Kwast, Paula M. Kujala, Bert G. Blijenberg, Ulf-Hakan Stenman, Andreas Huber, Kimmo Taari, Matti Hakama, Sue M.

- Moss, Harry J. de Koning, and Anssi Auvinen. Prostate-Cancer Mortality at 11 Years of Follow-up. *New England Journal of Medicine*, 366(11):981–990, mar 2012.
- [41] Elena Molina Portillo, Daniel Redondo-Sánchez, Miguel Rodríguez-Barranco, and María José Sánchez Pérez. Informe Web: Tendencias temporales de la incidencia de Cáncer en la provincia de Granada, 1985-2013. 2017.
- [42] Paul S.P. Cowpertwait and Andrew V. Metcalfe. *Introductory Time Series with R*. Springer New York, New York, NY, 2009.

Anexo I: Código en VBA y SQL

Anexo I: Código en VBA y SQL

Cabecera

La siguiente cabecera es común para todo el código, y define las variables que se usan posteriormente.

Caja 1. Cabecera del código.

```

1 Option Compare Database
2 Dim consulta(1 To 10) As QueryDef, i As Integer, j As Integer,
   ruta As String

```

Funciones

Las siguientes funciones auxiliares se utilizan en el código para realizar conversiones.

Caja 2. Código de funciones auxiliares.

```

1 ' Conversión de edad simple a grupo de edad
2 Function edad18c(edadx As Integer) As String
3   Select Case edadx
4     Case 0 To 4:           edad18c = "1"
5     Case 5 To 9:          edad18c = "2"
6     Case 10 To 14:         edad18c = "3"
7     Case 15 To 19:         edad18c = "4"
8     Case 20 To 24:         edad18c = "5"
9     Case 25 To 29:         edad18c = "6"
10    Case 30 To 34:        edad18c = "7"
11    Case 35 To 39:        edad18c = "8"
12    Case 40 To 44:        edad18c = "9"
13    Case 45 To 49:        edad18c = "10"
14    Case 50 To 54:        edad18c = "11"
15    Case 55 To 59:        edad18c = "12"
16    Case 60 To 64:        edad18c = "13"
17    Case 65 To 69:        edad18c = "14"
18    Case 70 To 74:        edad18c = "15"
19    Case 75 To 79:        edad18c = "16"
20    Case 80 To 84:        edad18c = "17"
21    Case Else:            edad18c = "18"
22 End Select

```

```
23 End Function
24
25 ' Conversión de localización a código interno
26 Function transformacioncodigo(sexo As String, loc As String) As
    Integer
27 Select Case sexo:
28     Case 1: Select Case loc:
29         ' Colon
30         Case 13: transformacioncodigo = 114
31         ' Recto
32         Case 14: transformacioncodigo = 115
33         ' Pulmón
34         Case 21: transformacioncodigo = 116
35         ' Próstata
36         Case 39: transformacioncodigo = 117
37         ' Vejiga
38         Case 45: transformacioncodigo = 118
39         ' Estómago
40         Case 11: transformacioncodigo = 119
41         ' Colon-recto
42         Case 1314: transformacioncodigo = 121
43         ' Otros
44     Case Else: transformacioncodigo = 199
45 End Select
46 Case 2: Select Case loc:
47     ' Colon
48     Case 13: transformacioncodigo = 614
49     ' Recto
50     Case 14: transformacioncodigo = 615
51     ' Pulmón
52     Case 21: transformacioncodigo = 616
53     ' Mama
54     Case 29: transformacioncodigo = 617
55     ' Cuerpo uterino
56     Case 33: transformacioncodigo = 618
57     ' Ovario
58     Case 35: transformacioncodigo = 619
59     ' Colon-recto
60     Case 1314: transformacioncodigo = 621
61     ' Otros
62     Case Else: transformacioncodigo = 699
63 End Select
64 End Select
```

```
65 End Function
66
67 ' Conversión de sexo (de numérico a texto)
68 Function transformacionsex0(sex As Integer) As String
69 Select Case sex:
70     Case 1:           transformacionsex0 = "M"
71     Case 2:           transformacionsex0 = "F"
72 End Select
73 End Function
74
75 ' Agrupación de localizaciones CIE-10
76 Function cie10_1(cie10 As String) As String
77 Select Case Left(cie10, 3)
78     Case "C00":          cie10_1 = "01"
79     Case "C01", "C02":    cie10_1 = "02"
80     Case "C03", "C04", "C05", "C06":  cie10_1 = "03"
81     Case "C07", "C08":    cie10_1 = "04"
82     Case "C09":          cie10_1 = "05"
83     Case "C10":          cie10_1 = "06"
84     Case "C11":          cie10_1 = "07"
85     Case "C12", "C13":    cie10_1 = "08"
86     Case "C14":          cie10_1 = "09"
87     Case "C15":          cie10_1 = "10"
88     Case "C16":          cie10_1 = "11"
89     Case "C17":          cie10_1 = "12"
90     Case "C18":          cie10_1 = "13"
91     Case "C19", "C20":    cie10_1 = "14"
92     Case "C21":          cie10_1 = "15"
93     Case "C22":          cie10_1 = "16"
94     Case "C23", "C24":    cie10_1 = "17"
95     Case "C25":          cie10_1 = "18"
96     Case "C30", "C31":    cie10_1 = "19"
97     Case "C32":          cie10_1 = "20"
98     Case "C33", "C34":    cie10_1 = "21"
99     Case "C37", "C38":    cie10_1 = "22"
100    Case "C40", "C41":   cie10_1 = "23"
101    Case "C43":          cie10_1 = "24"
102    Case "C44":          cie10_1 = "25"
103    Case "C45":          cie10_1 = "26"
104    Case "C46":          cie10_1 = "27"
105    Case "C47", "C49":   cie10_1 = "28"
106    Case "C50":          cie10_1 = "29"
107    Case "C51":          cie10_1 = "30"
```

```

108     Case "C52": cie10_1 = "31"
109     Case "C53": cie10_1 = "32"
110     Case "C54": cie10_1 = "33"
111     Case "C55": cie10_1 = "34"
112     Case "C56": cie10_1 = "35"
113     Case "C57": cie10_1 = "36"
114     Case "C58": cie10_1 = "37"
115     Case "C60": cie10_1 = "38"
116     Case "C61": cie10_1 = "39"
117     Case "C62": cie10_1 = "40"
118     Case "C63": cie10_1 = "41"
119     Case "C64": cie10_1 = "42"
120     Case "C65": cie10_1 = "43"
121     Case "C66": cie10_1 = "44"
122     Case "C67", "D09", "D41": cie10_1 = "45"
123     Case "C68": cie10_1 = "46"
124     Case "C69": cie10_1 = "47"
125     Case "C70", "C71", "C72", "D18", "D21", "D32", "D33", "D42",
126         "D43": cie10_1 = "48"
127     Case "C73": cie10_1 = "49"
128     Case "C74": cie10_1 = "50"
129     Case "C75", "D35", "D44": cie10_1 = "51"
130     Case "C81": cie10_1 = "52"
131     Case "C82", "C83", "C84", "C85", "C96": cie10_1 = "
132         53"
133     Case "C88": cie10_1 = "54"
134     Case "C90": cie10_1 = "55"
135     Case "C91": cie10_1 = "56"
136     Case "C92", "C93", "C94": cie10_1 = "57"
137     Case "C95": cie10_1 = "58"
138     Case "D45", "D47", "D76": cie10_1 = "59"
139     Case "D46": cie10_1 = "60"
140     Case "": cie10_1 = ""
141     Case Else: cie10_1 = "69"
142 End Select
143 End Function
144
145 ' Conversión de provincias, de texto a código de provincia
146 Function transformacionprov(i As String) As String
147     Select Case i
148         Case "Almeria": transformacionprov = "04"
149         Case "Cadiz": transformacionprov = "11"
150         Case "Cordoba": transformacionprov = "14"

```

```

149 Case "Granada": transformacionprov = "18"
150 Case "Huelva": transformacionprov = "21"
151 Case "Jaen": transformacionprov = "23"
152 Case "Malaga": transformacionprov = "29"
153 Case "Sevilla": transformacionprov = "41"
154 End Select
155 End Function
156
157 ' Función que añade un cero a números de un dígito para conseguir
   el formato necesario
158 Function conv(j As Integer) As String
159 Select Case j
160 Case 0: conv = "00"
161 Case 1: conv = "01"
162 Case 2: conv = "02"
163 Case 3: conv = "03"
164 Case 4: conv = "04"
165 Case 5: conv = "05"
166 Case 6: conv = "06"
167 Case 7: conv = "07"
168 Case 8: conv = "08"
169 Case 9: conv = "09"
170 Case Else: conv = "" & j & ""
171 End Select
172 End Function

```

Creación del fichero de incidencia

Caja 3. Código para la creación del fichero de incidencia.

```

1 Private Sub CASOS()
2
3 ' Se parte de la tabla "Casos_RCG" que contiene todos los casos
   incidentes del RCG por año, localización, sexo y edad
4
5 ' Se piden los años para que se introduzcan por teclado (por
   defecto, 1985-2013)
6 año1 = InputBox("Define el primer año del periodo" , , 1985)
7 año2 = InputBox("Define el segundo año del periodo" , , 2013)
8
9 ' Se crea CASOS_RCG_Seleccion. Quitamos 25 = piel no melanoma, y
   se selecciona un periodo

```

```

10 DoCmd.RunSQL "SELECT NENTRADA , FECHA_DIAG , SEXO as sex , EDAD AS
   edadx , Left([cie10_1],2) AS LOC INTO CASOS_RCG_Seleccion " -
11   & " FROM CASOS_RCG WHERE ((Left(cie10_1, 2) Not Like '25') AND
   (Year([FECHA_DIAG]) between " & año1 & " and " & año2 & "))
   ORDER BY FECHA_DIAG ;"
12
13 DoCmd.RunSQL "ALTER TABLE CASOS_RCG_Seleccion ADD COLUMN
   PROVINCIA integer , AÑO integer , SEXO text (20) , EDAD integer ,
   cie_10 integer "
14
15 ' Clonamos los registros con colon-recto , con 1314 como LOC
16 DoCmd.RunSQL "INSERT INTO CASOS_RCG_Seleccion ( NENTRADA , FECHA_
   DIAG , sex , edadx , LOC ) SELECT CASOS_RCG.NENTRADA , CASOS_RCG.
   FECHA_DIAG , CASOS_RCG.SEXO AS sex , CASOS_RCG.EDAD AS edadx ,
   '1314' AS Expr1 FROM CASOS_RCG WHERE (((Year([FECHA_DIAG])))
   Between " & año1 & " And " & año2 & ") AND ((CASOS_RCG.cie10_
   1)= '13' Or (CASOS_RCG.cie10_1)= '14')) ORDER BY CASOS_RCG.FECHA
   _DIAG ;"
17
18 DoCmd.RunSQL "UPDATE CASOS_RCG_Seleccion SET PROVINCIA = 18 , EDAD
   = edad18c(edadx) , AÑO=Year(FECHA_DIAG) , cie_10 =
   transformacioncodigo(sex , LOC) , SEXO = transformacionsex0(SEX)
   "
19
20 ' Creación de casos
21 DoCmd.RunSQL "SELECT PROVINCIA , AÑO , SEXO , EDAD , cie_10 as ICDO
   INTO casos FROM CASOS_RCG_Seleccion "
22
23 ' Creación de descriptivo para ver el número de casos por cada
   localización
24 Set consulta(1) = CurrentDb.CreateQueryDef("descriptivo" , "SELECT
   casos.ICDO , Count(casos.ICDO) AS CuentaDeICDO FROM casos
   GROUP BY casos.ICDO")
25
26 DoCmd.OpenQuery "descriptivo"
27
28 End Sub
29
30
31
32
33 Private Sub INCIDENCIA()
34

```

```

35 ' Se asigna un caso a cada registro
36 DoCmd.RunSQL "SELECT casos.* , 1 as CASOS INTO Incidencia_1 FROM
37   casos"
38
39 ' Se suman los casos
40 DoCmd.RunSQL "SELECT I.PROVINCIA , I.AÑO , I.SEXO , I.EDAD , I.ICDO ,
41   Count(I.CASOS) AS CASOS INTO Incidencia_2 FROM INCIDENCIA_1 I
42   "
43 & " GROUP BY I.PROVINCIA , I.AÑO , I.SEXO , I.EDAD , I.ICDO ; "
44
45 ' Se añaden los totales: 100 para hombres y 600 para mujeres
46 MsgBox "Va a añadir el total para todas las localizaciones (100
47   para hombres y 600 para mujeres)"
48 DoCmd.RunSQL "INSERT INTO Incidencia_2 (PROVINCIA , SEXO , EDAD , AÑ
49   O , ICDO , CASOS ) "
50 & "SELECT '18' AS bbb , Z.SEXO , Z.edad , Z.año , '00' AS aaa , Sum(
51   Z.casos) AS CASOS "
52 & "FROM Incidencia_2 Z "
53 & "GROUP BY '18' , Z.SEXO , Z.edad , Z.año , '00' "
54 & "ORDER BY Z.SEXO , Z.edad , Z.año ; "
55 CurrentDb.TableDefs.Refresh
56 DoCmd.RunSQL "UPDATE Incidencia_2 SET Incidencia_2.ICDO = 100
57   WHERE (Incidencia_2.SEXO='M' AND (Incidencia_2.ICDO)=0); "
58 DoCmd.RunSQL "UPDATE Incidencia_2 SET Incidencia_2.ICDO = 600
59   WHERE (Incidencia_2.SEXO='F' AND (Incidencia_2.ICDO)=0); "
60
61 ' Se hace una tabla de referencias cruzadas pivotando por edad
62 Set consulta(2) = CurrentDb.CreateQueryDef("Refcruzadas" , "
63   TRANSFORM Sum(I.CASOS) AS SumaDeCASOS SELECT I.PROVINCIA , I.AÑ
64   O , I.SEXO , I.ICDO FROM Incidencia_2 I GROUP BY I.PROVINCIA , I.
65   AÑO , I.SEXO , I.ICDO PIVOT I.EDAD")
66 DoCmd.RunSQL "SELECT Refcruzadas.* INTO Incidencia FROM
67   Refcruzadas ORDER BY PROVINCIA , AÑO , SEXO , ICDO ; "
68
69 ' Se desactivan temporalmente las notificaciones de Access
70 DoCmd.SetWarnings (False)
71
72 ' Se renombran los campos, porque hay que ponerles una I delante
73   a todos los grupos de edad
74 ' Se actualiza a 0 casos todos los campos nulos nulos
75 For i = 1 To 9
76   CurrentDb.TableDefs("Incidencia").Fields("'" & i & "'").Name = "
77     I0" & i & ""

```

```

64 DoCmd.RunSQL "UPDATE Incidencia I SET I.I0" & i & " = 0 WHERE I
   .I0" & i & " Is Null;"
65 Next i
66
67 For i = 10 To 18
68 CurrentDb.TableDefs("Incidencia").Fields("") & i & "").Name = "I
   " & i & ""
69 DoCmd.RunSQL "UPDATE Incidencia I SET I.I" & i & " = 0 WHERE I.
   I" & i & " Is Null;"
70 Next i
71
72 ' Se vuelve a activar las notificaciones de Access
73 DoCmd.SetWarnings (True)
74
75 ' Se añaden los casos sin edad (que no hay, pero hace falta la
   columna)
76 DoCmd.RunSQL "ALTER TABLE Incidencia ADD COLUMN SE integer"
77 DoCmd.RunSQL "UPDATE Incidencia SET Incidencia.SE = 0"
78
79 ' Se exportan los ficheros
80 DoCmd.TransferText acExportDelim, "Incidencia - Especificación de
   exportación", "Incidencia", CurrentProject.Path & "\"
   incidencia.csv", True
81
82 ' Se pregunta si quieres abrir la ruta para acceder al fichero
   directamente
83 a = MsgBox("Se ha exportado 'incidencia.csv'. ¿Quieres abrir la
   ruta?", vbYesNo + vbQuestion)
84 If (a = 6) = True Then Call Shell("explorer.exe" & " " & "F:\"
   INVESTIGACION", vbNormalFocus)
85
86 ' Se limpia la base de datos
87 CurrentDb.TableDefs.Delete "Incidencia_1"
88 CurrentDb.TableDefs.Delete "Incidencia_2"
89 CurrentDb.QueryDefs.Delete "Refcruzadas"
90 End Sub

```

Creación fichero de mortalidad

Caja 4. Código para la creación del fichero de mortalidad.

```

1 Private Sub Mortalidad()
2
3 ' Se parte de la tabla "CASOS_RCG_MORTALIDAD" que contiene la
4 ' mortalidad de Andalucía por causa de muerte, provincia, año,
5 ' sexo y edad
6 ' Se quita 25 = piel no melanoma, igual que hicimos en la
7 ' incidencia
8 Set consulta(1) = CurrentDb.CreateQueryDef("Mortalidad_1", "
9     SELECT transformacionprov(C.prov) as PROVINCIA, C.año as AÑO,
10    transformacionsexo(C.sexo) as SEXO, " _
11    & "edad18c(cint(Left(C.edad18c,2))) as EDAD18, C.def, C.cie10_"
12    & "1, transformacioncodigo(C.sexo,C.cie10_1) as ICDO " _
13    & "FROM Casos_RCG_Mortalidad C WHERE C.cie10_1 Not like '25' ")
14
15 ' Se suman los casos
16 DoCmd.RunSQL "SELECT M.PROVINCIA, M.AÑO, M.SEXO, M.ICDO, M.EDAD18
17   , Sum(M.def) AS def INTO Mortalidad_2 " _
18   & "FROM Mortalidad_1 M GROUP BY M.PROVINCIA, M.AÑO, M.SEXO, M.
19   ICDO, M.EDAD18;" 
20
21 ' Se anexan los totales para hombres
22 DoCmd.RunSQL "INSERT INTO Mortalidad_2 ( PROVINCIA, AÑO, SEXO,
23   ICDO, EDAD18, def ) " _
24   & "SELECT M.PROVINCIA, M.AÑO, M.SEXO, '100' AS Expr1, M.EDAD18,
25   Sum(M.def) AS def " _
26   & "FROM Mortalidad_2 M GROUP BY M.PROVINCIA, M.AÑO, M.SEXO,
27   '100', M.EDAD18 HAVING (((M.SEXO)='M'));" 
28
29 ' Se anexan los totales para mujeres
30 DoCmd.RunSQL "INSERT INTO Mortalidad_2 ( PROVINCIA, AÑO, SEXO,
31   ICDO, EDAD18, def ) " _
32   & "SELECT M.PROVINCIA, M.AÑO, M.SEXO, '600' AS Expr1, M.EDAD18,
33   Sum(M.def) AS def " _
34   & "FROM Mortalidad_2 M GROUP BY M.PROVINCIA, M.AÑO, M.SEXO,
35   '600', M.EDAD18 HAVING (((M.SEXO)='F'));" 
36
37 ' Se hace una tabla de referencias cruzadas pivotando por edad
38 Set consulta(2) = CurrentDb.CreateQueryDef("Mortalidad_3", "
39     TRANSFORM Sum(M.[def]) AS SumaDedef SELECT M.[PROVINCIA], M.[A
40     ÑO], M.[SEXO], M.ICDO FROM Mortalidad_2 M " _
41     & "GROUP BY M.[PROVINCIA], M.[AÑO], M.[SEXO], M.ICDO PIVOT M.[
42     EDAD18];")
```

```

27 ' Se crea la tabla por grupos de edad
28 DoCmd.RunSQL "SELECT cint(M.PROVINCIA) as PROVINCIA, M.AÑO, M.
    SEXO, M.ICDO, M.[1], M.[2], M.[3], M.[4], M.[5], M.[6], M.[7],
    M.[8], M.[9], M.[10], " _
29   & "M.[11], M.[12], M.[13], M.[14], M.[15], M.[16], M.[17], M
    .[18] INTO Mortalidad FROM Mortalidad_3 M"
30
31 ' Actualizamos a 0 cuando no hay defunciones en ese grupo de edad
32 DoCmd.SetWarnings (False)
33 For i = 1 To 18
34   DoCmd.RunSQL "UPDATE Mortalidad M SET M." & i & "=0 WHERE M.[
    " & i & "] is null"
35 Next i
36 DoCmd.SetWarnings (True)
37
38 ' Se selecciona Granada
39 DoCmd.RunSQL "INSERT INTO Mortalidad ( AÑO, SEXO, ICDO, 1, 2, 3,
    4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,
    PROVINCIA ) " _
40   & "SELECT M.AÑO, M.SEXO, M.ICDO, M.[1], M.[2], M.[3], M.[4], M
    .[5], M.[6], M.[7], M.[8], M.[9], M.[10], M.[11], M.[12], M
    .[13], " _
41   & "M.[14], M.[15], M.[16], M.[17], M.[18], 101 AS Expr1 FROM
    Mortalidad M WHERE (M.PROVINCIA=18) ORDER BY AÑO, PROVINCIA,
    SEXO;"
42
43 ' Se limpia la base de datos
44 CurrentDb.QueryDefs.Delete "Mortalidad_1"
45 CurrentDb.TableDefs.Delete "Mortalidad_2"
46 CurrentDb.QueryDefs.Delete "Mortalidad_3"
47
48 ' Se renombran los campos
49 For i = 1 To 18
50   CurrentDb.TableDefs("Mortalidad").Fields("'" & i & "'").Name = "M
    " & conv(i)
51 Next i
52
53 ' Se exporta el fichero
54 DoCmd.TransferText acExportDelim, "mortalidad - Especificación de
    exportación", "Mortalidad", CurrentProject.Path & "\"
    mortalidad.csv", True
55
56 ' Se pregunta si quieres abrir la ruta para acceder al fichero

```

```

    directamente
57 a = MsgBox("Se ha exportado ''mortalidad.csv''. ¿Quieres abrir la
      ruta?", vbYesNo + vbQuestion)
58 If (a = 6) = True Then Call Shell("explorer.exe" & " " & "F:\"
      INVESTIGACION\Explotación\Estimaciones\Creación de ficheros",
      vbNormalFocus)
59
60 End Sub

```

Creación fichero de población

Caja 5. Código para la creación del fichero de población.

```

1 Private Sub poblacion()
2
3 ' Se importa el Excel
4 ruta = CurrentProject.Path
5 DoCmd.TransferSpreadsheet acImport, acSpreadsheetTypeExcel12, "
      Poblacion", ruta & "\Población INE 1971-2017 Granada>Editada.
      xlsx", True, "A7:T196"
6
7 ' Se borran las filas que sobran
8 DoCmd.RunSQL "Delete P.sex FROM Poblacion P WHERE (P.sex Is
      null OR P.año like '')";
9
10 ' Se borran los datos del 1 de Enero
11 DoCmd.RunSQL "Delete P.año FROM Poblacion P WHERE (P.año Is null
      OR P.año like '*Enero*')";
12 DoCmd.RunSQL "UPDATE Poblacion P SET P.año=Right(P.año,4)"
13
14 ' Se añade un campo para la provincia, que es siempre 18
15 DoCmd.RunSQL "ALTER TABLE Poblacion ADD COLUMN PROVINCIA integer"
16 DoCmd.RunSQL "UPDATE Poblacion P SET P.PROVINCIA=18"
17
18 ' Se recodifica el sexo
19 DoCmd.RunSQL "UPDATE Poblacion P SET P.SEXO='M' WHERE P.SEXO='
      Hombres'"
20 DoCmd.RunSQL "UPDATE Poblacion P SET P.SEXO='F' WHERE P.SEXO = '
      Mujeres'"
21
22 ' Se suman los casos

```

```

23 DoCmd.RunSQL "INSERT INTO Poblacion ( PROVINCIA , SEXO , AÑO , P01 ,
24     P02 , P03 , P04 , P05 , P06 , P07 , P08 , P09 , P10 , P11 , P12 , P13 ,
25     P14 , P15 , P16 , P17 , P18 )" _
26 & "SELECT 100 AS Expr1 , P.SEXO , P.AÑO , Sum(P.P01) AS P01 , Sum(P.
27     .P02) AS P02 , Sum(P.P03) AS P03 , " _
28 & "Sum(P.P04) AS P04 , Sum(P.P05) AS P05 , Sum(P.P06) AS P06 , Sum
29     (P.P07) AS P07 , " _
30 & "Sum(P.P08) AS P08 , Sum(P.P09) AS P09 , Sum(P.P10) AS P10 , Sum
31     (P.P11) AS P11 , " _
32 & "Sum(P.P12) AS P12 , Sum(P.P13) AS P13 , Sum(P.P14) AS P14 , Sum
33     (P.P15) AS P15 , " _
34 & "Sum(P.P16) AS P16 , Sum(P.P17) AS P17 , Sum(P.P18) AS P18 " _
35 & "FROM POBLACION P GROUP BY 100 , P.SEXO , P.AÑO ;"
36
37 ' Se seleccionan los años 1982-2013 y se ordena
38 DoCmd.RunSQL "SELECT P.PROVINCIA , P.SEXO , P.año , P.P01 , P.P02 , P.
39     P03 , P.P04 , P.P05 , P.P06 , P.P07 , P.P08 , P.P09 , P.P10 , P.P11 , P
40     .P12 , P.P13 , P.P14 , P.P15 , P.P16 , P.P17 , P.P18 INTO Poblacion_
41     Final FROM Poblacion P " _
42 & "WHERE ((cint(P.AÑO) Between 1982 And 2013)) ORDER BY CInt([
43         PROVINCIA]) , cint(P.AÑO) , P.SEXO DESC ;"
44
45 ' Se exporta el fichero
46 DoCmd.TransferText acExportDelim , "poblacion - Especificación de
47     exportación" , "Poblacion_Final" , ruta & "/poblacion.csv" , True
48 MsgBox "Se ha exportado con éxito el archivo poblacion.csv"
49
50 ' Se limpia la base de datos
51 CurrentDb.TableDefs.Delete "Poblacion"
52 CurrentDb.TableDefs.Delete "Poblacion2"
53
54 End Sub

```


Anexo II: Código en R

Anexo II: Código en R

Cabecera de fichero

Caja 1. Cabecera del fichero.

```

1 # 240 para 500 simulaciones, 24000 para 50000 simulaciones
2 ciclos.seleccionados <- 24000
3
4 ruta <- "RUTA"
5
6 año.a.estimar <- 2013
7
8 # Años para evolución RIM
9 total.años.evolucion.rim <- 15
10 añoFrim <- año.a.estimar - 5
11 añoIrim <- añoFrim - total.años.evolucion.rim + 1
12 # Por ejemplo:
13 # Si año.a.estimar es 2017, y se consideran 15 años de evolución
   que acaban en 2012, el año inicial será 1998
14
15 # Años NORDPRED (I = Año inicial, F = Año final, P = Periodo)
16 añoIP1 <- año.a.estimar - 22
17 añoFP1 <- año.a.estimar - 18
18 añoIP2 <- año.a.estimar - 17
19 añoFP2 <- año.a.estimar - 13
20 añoIP3 <- año.a.estimar - 12
21 añoFP3 <- año.a.estimar - 8
22 añoIP4 <- año.a.estimar - 7
23 añoFP4 <- año.a.estimar - 3
24
25 # P5 es el periodo del que obtenemos la mortalidad tras aplicar
   NORDPRED a los periodos del 1 al 4
26 añoIP5 <- año.a.estimar - 2
27 añoFP5 <- año.a.estimar + 2

```

Creación fichero de parametros.tumor

Caja 2. Definición de cuartiles en R.

```

1 cuartiles <- function(icdo)

```

```

2 {
3   # Se selecciona Granada (18 = 101) y los datos de los cuatro
      primeros periodos de NORDPRED (años añoIP1 hasta año añoFP4)
4   datos <- mortalidad[mortalidad$PROVINCIA==101 & mortalidad$ICDO
      ==icdo & mortalidad$AÑO>=añoIP1 & mortalidad$AÑO<=añoFP4 ,]
5   datos <- aggregate(datos[,4:22],list(ICDO=datos$ICDO),sum)
6   datos <- datos[,3:20]
7   suma <- sum(datos)
8   frecuencia <- (datos/suma)*100
9   delta1 <- 100
10  delta2 <- 100
11  delta3 <- 100
12  delta4 <- 100
13
14  # Se calculan P10, T1, T2 y P90.
15  # corteedad es el número del grupo de edad (de 1 a 18) que
     corresponde al P10.
16  for (i in 1:18) {
17    suma.sup <- sum(frecuencia[,1:i])
18    if (abs(10-suma.sup) < delta1) {
19      delta1 <- abs(10-suma.sup)
20      corteedad <- i
21      p10 <- i*5
22    }
23    if (abs(33.3333-suma.sup) < delta2) {
24      delta2 <- abs(33.3333-suma.sup)
25      t2 <- i*5
26    }
27    if (abs(66.6666-suma.sup) < delta3) {
28      delta3 <- abs(66.6666-suma.sup)
29      t3 <- i*5
30    }
31    if (abs(90-suma.sup) < delta4) {
32      delta4 <- abs(90-suma.sup)
33      p90 <- i*5
34    }
35  }
36  aux <- c(icdo, corteedad,p10,t2,t3,p90)
37  aux
38 }

```

Caja 3. Creación de `parametros.tumor` en R.

```

1 crear.parametros.tumor.RCG<-function(){
2   salida <- cuartiles(114)
3   salida <- rbind(salida, cuartiles(115))
4   salida <- rbind(salida, cuartiles(116))
5   salida <- rbind(salida, cuartiles(117))
6   salida <- rbind(salida, cuartiles(118))
7   salida <- rbind(salida, cuartiles(119))
8   salida <- rbind(salida, cuartiles(199))
9   salida <- rbind(salida, cuartiles(614))
10  salida <- rbind(salida, cuartiles(615))
11  salida <- rbind(salida, cuartiles(616))
12  salida <- rbind(salida, cuartiles(617))
13  salida <- rbind(salida, cuartiles(618))
14  salida <- rbind(salida, cuartiles(619))
15  salida <- rbind(salida, cuartiles(699))
16  # para respetar la estructura de parametros.tumor
17  subsalida1 <- salida[c(1:6), c(2:6)]
18  subsalida2 <- salida[7, c(2:6)]
19  subsalida3 <- salida[c(8:13), c(2:6)]
20  subsalida4 <- salida[14, c(2:6)]
21  cero5 <- rep(0, 5)
22  salida <- rbind(subsalida1, cero5, subsalida2, cero5,
23    subsalida3, cero5, subsalida4, cero5)
24  icdo <- c(114, 115, 116, 117, 118, 119, 121, 199, 100, 614,
25    615, 616, 617, 618, 619, 621, 699, 600)
26  sexo <- c("M", "M", "M", "M", "M", "M", "M", "M", "F", "F",
27    , "F", "F", "F", "F", "F", "F")
28  loc <- c("Colon", "Recto", "Pulmón", "Próstata", "Vejiga", "Est
29    ómago", "Colon-recto", "Otros", "Total excepto piel no
30    melanoma", "Colon", "Recto", "Pulmón", "Mama", "Cuerpo
31    uterino", "Ovario", "Colon-recto", "Otros", "Total excepto
    piel no melanoma")
32  tipo <- c("SP", "SP", "SP", "SP", "SP", "SP", "O", "SP", "O", "
33    SP", "SP", "SP", "SP", "SP", "SP", "O", "SP", "O")
34  escenario <- c("A", "A", "A", "A", "A", "A", "Especial", "A", "
35    Especial", "A", "A", "A", "A", "A", "A", "Especial", "A", "
36    Especial")
37  salida <- cbind(icdo, sexo, loc, salida, tipo, escenario)
38  cabecera <- c("CODIGO", "SEXO", "VALOR", "CORTEEDAD", "P10", "
39    T2", "T3", "P90", "TIPO", "ESCENARIO")
40  colnames(salida) <- cabecera
41  setwd(paste0(ruta, "\\datos"))

```

```

32   write.csv2(salida, "parametros.tumor.csv", row.names = FALSE,
33             quote = FALSE)
34   print(paste("Se ha creado parametros.tumor.csv"), quote = FALSE)
35 }

```

Funciones para NORDPRED

Caja 4. Definición de las funciones `nordpred.estimate` y `nordpred.prediction`.

```

1 nordpred.estimate <- function(cases,pyr,noperiod,starteststage,
2   linkfunc="power5") {
3   Rplatform <- exists("R.Version")
4   if ( dim(cases)[1] !=18 || dim(pyr)[1] !=18 ) {
5     stop("\ncases\" and \"pyr\" must have data for 18 age groups"
6         )
7   }
8   if ( dim(cases)[2]>dim(pyr)[2]) {
9     stop("\npyr\" must include information about all periods in \
10      \"cases\"")
11  }
12  if ((dim(pyr)[2]==dim(cases)[2])) {
13    stop("\npyr\" must include information on future rates (E.g.
14      must be greater than \ncases\"))
15  }
16  if ((dim(cases)[2]-noperiod)>5) {
17    stop("Package can not predict more than 5 periods (given by
18      sizes of \npyr\" and \ncases\"))
19  }
20  if ((dim(cases)[2]-noperiod)<0) {
21    stop("More periods specified in \nnoperiod\" than columns in
22      \ncases\"")
23  }
24  if (noperiod<3) {
25    stop("\nnoperiod\" must be 3 or larger to get enough for
      estimating")
26  }
27  dnoperiods <- dim(cases)[2]
28  dnoagegr   <- 18
29  ageno      <- rep(1:dnoagegr,dnoperiods)
30  periodno  <- sort(rep(1:dnoperiods,dnoagegr))
31  cohort     <- max(ageno)-ageno+periodno

```

```

26   y      <- c(as.matrix(pyr[,1:dnoperiods]))
27   apcdata <- data.frame(Cases= c(as.matrix(cases)), Age=ageno,
28     Cohort=cohort, Period=periodno, y=y)
29   apcdata <- apcdata[apcdata$Age>=startestage,]
30   apcdata <- apcdata[apcdata$Period>(dnoperiods-noperiod),]
31   if (Rplatform) {
32     y <- apcdata$y
33     power5link <- poisson()
34     power5link$link <- "0.2 root link Poisson family"
35     power5link$linkfun <- function(mu) { (mu/y)^0.2 }
36     power5link$linkinv <- function(eta) { pmax(.Machine$double.
37       eps, y*eta^5) }
38     power5link$mu.eta <- function(eta) { pmax(.Machine$double.
39       eps, 5*y*eta^4) }
40   } else {
41     y <- apcdata$y
42     power5link <- poisson()
43     power5link$link <- function(mu) { (mu/y)^0.2 }
44     power5link$inverse <- function(eta) { y*eta^(1/0.2) }
45     power5link$deriv <- function(mu) { 0.2*(1/y)*(mu/y)^(0.2 -
46       1) }
47   }
48   options(contrasts=c("contr.treatment","contr.poly"))
49   if (linkfunc=="power5") {
50     res.glm <- glm(Cases~as.factor(Age)+Period+as.factor(Period)+
51       as.factor(Cohort) -1,family=power5link,data=apcdata)
52   } else if (linkfunc=="poisson") {
53     res.glm <- glm(Cases~as.factor(Age)+Period+as.factor(Period)+
54       as.factor(Cohort)+ offset(log(y)) -1,family=poisson(),data
55       =apcdata)
56   } else {
57     stop("Unknown \\\"linkfunc\\\"")
58   }
59   if (Rplatform) {
60     pvalue <- 1-pchisq(res.glm$deviance,res.glm$df.residual)
61   } else {
62     pvalue <- 1-pchisq(res.glm$deviance,res.glm$df)
63   }
64   mod1 <- glm(Cases~as.factor(Age)+Period +as.factor(Cohort) +
65     offset(log(y)) -1,family=poisson,data=apcdata)
66   mod2 <- glm(Cases~as.factor(Age)+Period +I(Period^2)+as.factor(
67     Cohort) + offset(log(y)) -1,family=poisson,data=apcdata)
68   if (Rplatform) {
69

```

```

60 pdiff<-anova(mod1,mod2,test="Chisq")$"P(>|Chi|)"[2]
61 # Correction added 2015-09-19 for compability with newer R
62   versions:
63   if (is.null(pdiff))
64     pdiff <- anova(mod1,mod2,test="Chisq")$"Pr(>Chi)"[2]
65 } else {
66   pdiff<-anova(mod1,mod2,test="Chisq")[2,7]
67 }
68 if (pdiff <0.05) {
69   suggestionrecent <-T
70 } else {
71   suggestionrecent <-F
72 }
73 res <- list(glm=res.glm,cases=cases,pyr=pyr,noperiod=noperiod,
74   gofpvalue=pvalue,startestage=startestage,
75   suggestionrecent=suggestionrecent,pvaluerecent=
76   pdiff,linkfunc=linkfunc)
77 class(res) <- "nordpred.estimate"
78 attr(res,"Call") <- sys.call()
79 return(res)
80 }

81 nordpred.prediction <- function(nordpred.estimate.object,
82   startuseage,recent,cuttrend=c(0,.25,.5,.75,.75)) {
83   if (class(nordpred.estimate.object)!="nordpred.estimate") {
84     stop("Variable \"nordpred.estimate.object\" must be of type \
85       \"nordpred.estimate\"")
86   }
87   if (nordpred.estimate.object$startestage>startuseage) {
88     stop("\\"startuseage\\" is set to high compared to \
89       startestage\" in \"nordpred.estimate.object\")")
90   }
91   if (length(cuttrend)<(dim(nordpred.estimate.object$pyr)[2]-dim(
92     nordpred.estimate.object$cases)[2])) {
93     err <- paste("\\"cuttrend\\" must always be at least the same
94       length as")
95     err <- paste(err,"the number of periods with population
96       forecasts")
97     stop(err)
98   }
99   cases      <- nordpred.estimate.object$cases
100  pyr        <- nordpred.estimate.object$pyr
101  noperiod  <- nordpred.estimate.object$noperiod

```

```

94  nototper  <- dim(pyr)[2]
95  noobsper  <- dim(cases)[2]
96  nonewpred <- nototper-noobsper
97  if (length(cuttrend)>nonewpred) {
98    cuttrend <- cuttrend[1:nonewpred]
99  }
100 if (is.data.frame(pyr)) {
101   years <- names(pyr)
102 } else {
103   if (is.null(dimnames(pyr))) {
104     years <- paste("Periode",1:nototper)
105   } else {
106     years <- dimnames(pyr)[[2]]
107   }
108 }
109 datatable <- matrix(NA,18,nototper)
110 datatable[,1:(nototper-nonewpred)] <- as.matrix(cases)
111 datatable <- data.frame(datatable)
112 row.names(datatable) <- c("0-4","5-9","10-14","15-19","20-24",
113                           "25-29","30-34","35-39","40-44","45-49","50-54","55-59",
114                           "60-64","65-69","70-74","75-79","80-84","85+")
115 names(datatable)      <- years
116 for (age in 1:(startuseage-1)) {
117   obsinc <- cases[age,(noobsper-1):noobsper]/pyr[age,(noobsper
118               -1):noobsper]
119   if (sum(is.na(obsinc))) {
120     obsinc[is.na(obsinc)] <- 0
121   }
122   datatable[age,(noobsper+1):nototper] <- ((obsinc[,1]+obsinc
123               [,2])/2)*pyr[age,(noobsper+1):nototper]
124 }
125 for (age in startuseage:18) {
126   startestage  <- nordpred.estimate.object$startestage
127   coefficients <- nordpred.estimate.object$glm$coefficients
128   coh          <- (18-startestage) - (age-startestage) + (noperiod
129               +1:nonewpred)
130   noages      <- 18-startestage+1
131   drifttmp    <- cumsum(1-cuttrend)
132   cohfnd     <- noages + (noperiod-1) + 1 + (coh-1)
133   maxcoh     <- 18 - startuseage + noperiod
134   agepar      <- as.numeric(coefficients[age-startestage+1])
135   driftfind   <- pmatch("Period",attributes(coefficients)$names)
136   driftpar    <- as.numeric(coefficients[driftfind])

```

```

132 cohpar <- rep(NA,length(coh))
133 for (i in 1:length(coh)) {
134   if (coh[i] < maxcoh) {
135     cohpar[i] <- as.numeric(coefficients[cohfind[i]])
136   } else {
137     # For young cohorts not estimated:
138     cohpar[i] <- as.numeric(coefficients[length(coefficients)
139                               -(startuseage-startestage)])
140     cohpar[i][is.na(cohpar[i])] <- 0
141   }
142   if (recent) {
143     lpfnd <- driftfind + noperiod-2
144     lppar <-as.numeric(coefficients[lpfnd])
145     driftrecent <- driftpar - lppar
146   }
147   if (nordpred.estimate.object$linkfunc=="power5") {
148     if (recent) {
149       rate <- (agepar+driftpar*noobsper+driftrecent*driftmp+
150                 cohpar)^5
151     } else {
152       rate <- (agepar+driftpar*(noobsper+driftmp)+cohpar)^5
153     }
154   } else { # Possion link:
155     if (recent) {
156       rate <- exp(agepar+driftpar*noobsper+driftrecent*driftmp+
157                     cohpar)
158     } else {
159       rate <- exp(agepar+driftpar*(noobsper+driftmp)+cohpar)
160     }
161   }
162   datatable[age ,(noobsper+1):nototper] <- rate*pyr[age ,
163             noobsper+1):nototper]
164   res <- list(predictions=datatable,pyr=pyr,nopred=nonewpred,
165               noperiod=nordpred.estimate.object$noperiod,
166               gofpvalue=nordpred.estimate.object$gofpvalue,recent
167               =recent ,pvaluerecent=nordpred.estimate.object$ 
168               pvaluerecent ,
169               cuttrend=cuttrend ,startuseage=startuseage ,
170               startestage=nordpred.estimate.object$startestage
171               ,
172               glm=nordpred.estimate.object$glm)

```

```

166   class(res) <- "nordpred"
167   attr(res,"Call") <- sys.call()
168   return(res)
169 }
```

Funciones para la validación de estimaciones

Caja 5. Definición de `funcion.ajuste`, que mide el MAPE en R.

```

1 funcion.ajuste<-function(localizacion, escenario, sims = 100,
2                             nivelconfianza, imprimir = "FALSE"){
3
4   if(escenario=="TODOS"){
5     salida <- funcion.ajuste(localizacion, "E3", sims,
6                               nivelconfianza)
7     salida <- rbind(salida, funcion.ajuste(localizacion, "E1",
8                               sims, nivelconfianza))
9     salida <- rbind(salida, funcion.ajuste(localizacion, "E5",
10                           sims, nivelconfianza))
11    salida <- rbind(salida, funcion.ajuste(localizacion, "L",
12                           sims, nivelconfianza))
13    salida <- rbind(salida, funcion.ajuste(localizacion, "C",
14                           sims, nivelconfianza))
15
16    salida
17  }
18
19  else{
20    # 121 = Colon-recto en hombres
21    # 621 = Colon-recto en mujeres
22    if(localizacion != 121 && localizacion != 621){
23      observado <- estimado.observado(localizacion, escenario,
24                                         sims, nivelconfianza)$observado
25      estimado <- estimado.observado(localizacion, escenario,
26                                         sims, nivelconfianza)$estimado
27    }
28    else{
29      if(localizacion==121){
30        observado.colon <- estimado.observado(114, escenario,
31                                              sims, nivelconfianza)$observado
32        observado.recto <- estimado.observado(115, escenario,
33                                              sims, nivelconfianza)$observado
34      }
35    }
36  }
37}
```

```

24     observado <- observado.colon+observado.recto
25     estimado.colon <- estimado.observado(114, escenario, sims
26         , nivelconfianza)$estimado
27     estimado.recto <- estimado.observado(115, escenario, sims
28         , nivelconfianza)$estimado
29     estimado <- estimado.colon+estimado.recto
30   }
31   else{
32     observado.colon <- estimado.observado(614, escenario,
33         sims, nivelconfianza)$observado
34     observado.recto <- estimado.observado(615, escenario,
35         sims, nivelconfianza)$observado
36     observado <- observado.colon+observado.recto
37     estimado.colon <- estimado.observado(614, escenario, sims
38         , nivelconfianza)$estimado
39     estimado.recto <- estimado.observado(615, escenario, sims
40         , nivelconfianza)$estimado
41     estimado <- estimado.colon+estimado.recto
42   }
43
44   if(imprimir == "TRUE"){
45     print("Observado:", quote = "FALSE")
46     print(observado)
47     print("Estimado:", quote = "FALSE")
48     print(estimado)
49   }
50
51   # MAPE
52   f <- 100 * (1 / 10) * sum(abs((estimado - observado) /
53       observado))
54   icdo <- parametros.tumor[parametros.tumor$CODIGO ==
55     localizacion, "VALOR"]
56   data.frame(localizacion, icdo, escenario, f)
57 }
58 }
59 }
```

Caja 6. Definición de las funciones `mejor.escenario.ajuste`, `funcion.ajuste.total` y `tabla.mejores.escenarios`.

```

1 mejor.escenario.ajuste <- function(localizacion, sims = 100,
2   nivelconfianza, imprimir = "FALSE"){
3   esc.e1 <- funcion.ajuste(localizacion, "E1", sims,
4     nivelconfianza, "FALSE")[[4]]
5   esc.e3 <- funcion.ajuste(localizacion, "E3", sims,
6     nivelconfianza, "FALSE")[[4]]
7   esc.e5 <- funcion.ajuste(localizacion, "E5", sims,
8     nivelconfianza, "FALSE")[[4]]
9   esc.l <- funcion.ajuste(localizacion, "L", sims,
10    nivelconfianza, "FALSE")[[4]]
11  esc.c <- funcion.ajuste(localizacion, "C", sims,
12    nivelconfianza, "FALSE")[[4]]
13
14  #Vemos cuál es el mejor escenario
15  if(esc.e1<esc.e3 && esc.e1<esc.e5 && esc.e1<esc.l && esc.e1<esc.
16    .c) esc.mejor="E1" else{
17    if(esc.e3<esc.e1 && esc.e3<esc.e5 && esc.e3<esc.l && esc.e3<
18      esc.c) esc.mejor="E3" else{
19      if(esc.e5<esc.e1 && esc.e5<esc.e3 && esc.e5<esc.l && esc.e5<
20        esc.c) esc.mejor="E5" else{
21          if(esc.l<esc.e1 && esc.l<esc.e5 && esc.l<esc.e3 && esc.l<
22            esc.c) esc.mejor="L" else{
23              esc.mejor="C"
24            }
25          }
26        }
27      }
28
29  ifelse(imprimir == "FALSE",
30    salida <- paste0(funcion.ajuste(localizacion, esc.mejor,
31      sims, nivelconfianza, "FALSE")[[3]]),
32    salida <- funcion.ajuste(localizacion, esc.mejor, sims,
33      nivelconfianza, "FALSE"))
34
35  salida
36
37  funcion.ajuste.total <- function(sims = 100, nivelconfianza){
38
39  # Hombres
40  salida <- funcion.ajuste(114, "TODOS", sims, nivelconfianza)
41  salida <- rbind(salida, funcion.ajuste(115, "TODOS", sims,
42    nivelconfianza))
43  salida <- rbind(salida, funcion.ajuste(116, "TODOS", sims,
44    nivelconfianza))
45
46  
```

```

    nivelconfianza))
31 salida <- rbind(salida, funcion.ajuste(117, "TODOS", sims,
    nivelconfianza))
32 salida <- rbind(salida, funcion.ajuste(118, "TODOS", sims,
    nivelconfianza))
33 salida <- rbind(salida, funcion.ajuste(119, "TODOS", sims,
    nivelconfianza))
34 salida <- rbind(salida, funcion.ajuste(199, "TODOS", sims,
    nivelconfianza))

35
36 # Mujeres
37 salida <- rbind(salida, funcion.ajuste(614, "TODOS", sims,
    nivelconfianza))
38 salida <- rbind(salida, funcion.ajuste(615, "TODOS", sims,
    nivelconfianza))
39 salida <- rbind(salida, funcion.ajuste(616, "TODOS", sims,
    nivelconfianza))
40 salida <- rbind(salida, funcion.ajuste(617, "TODOS", sims,
    nivelconfianza))
41 salida <- rbind(salida, funcion.ajuste(618, "TODOS", sims,
    nivelconfianza))
42 salida <- rbind(salida, funcion.ajuste(619, "TODOS", sims,
    nivelconfianza))
43 salida <- rbind(salida, funcion.ajuste(699, "TODOS", sims,
    nivelconfianza))

44
45 salida
46
47 }
48
49 tabla.mejores.escenarios<-function(sims=100,nivelconfianza){
50   v<-funcion.ajuste.total(sims,nivelconfianza)
51   v<-v[order(v$f),]
52
53 # Hombres
54   salida <- t(subset(v, v$localizacion == "114"))
55   salida <- rbind(salida, t(subset(v, v$localizacion == "115")))
56   salida <- rbind(salida, t(subset(v, v$localizacion == "116")))
57   salida <- rbind(salida, t(subset(v, v$localizacion == "117")))
58   salida <- rbind(salida, t(subset(v, v$localizacion == "118")))
59   salida <- rbind(salida, t(subset(v, v$localizacion == "119")))
60   salida <- rbind(salida, t(subset(v, v$localizacion == "199")))
61

```

```
62 # Mujeres
63 salida <- rbind(salida, t(subset(v, v$localizacion == "614")))
64 salida <- rbind(salida, t(subset(v, v$localizacion == "615")))
65 salida <- rbind(salida, t(subset(v, v$localizacion == "616")))
66 salida <- rbind(salida, t(subset(v, v$localizacion == "617")))
67 salida <- rbind(salida, t(subset(v, v$localizacion == "618")))
68 salida <- rbind(salida, t(subset(v, v$localizacion == "619")))
69 salida <- rbind(salida, t(subset(v, v$localizacion == "699")))
70
71 print(salida, quote = "FALSE")
72 }
```