



Islas Canarias
Del 15 al 19 de noviembre de 2021



Junta de Andalucía

Consejería de Transformación Económica,
Industria, Conocimiento y Universidades
Instituto de Estadística y Cartografía de Andalucía

Utilidad de la red informatizada de registros civiles de Andalucía para generar estadísticas rápidas de mortalidad en ámbitos subregionales

Francisco Viciano Fernández

Instituto de Estadística y Cartografía de Andalucía

franciscoj.viciano@juntadeandalucia.es

Soledad Márquez Calderón

Consejería de Salud y Familias de la Junta de Andalucía

soledadm.marquez@juntadeandalucia.es

16 de noviembre de 2021

Índice

1. Introducción	2
2. Objetivos	3
3. Material y Método	4
3.1. Fuentes de datos para la construcción de la serie de defunciones	4
3.2. Combinación de los ficheros de microdatos del MNP e INFOREG . . .	5
3.3. Cobertura y exhaustividad de las defunciones de INFOREG-MoMo . .	5
3.4. Relación entre el número de defunciones por lugar de inscripción y de residencia	6
3.5. Reconstrucción de series semanales de defunciones por lugar de ins- cripción	8
3.6. Exposición: población inscripción-vinculada (PIV)	8
3.7. Series semanales de poblaciones para predicción	9
3.8. Estimación de las defunciones esperadas	10
4. Resultados	11
5. Conclusiones	13
6. Referencias	14
7. Apéndice gráfico	15

1. Introducción

Desde el inicio de la pandemia de COVID-19, la fuente de información más utilizada y difundida ha estado constituida por los Sistemas de Vigilancia Epidemiológica, que han aportado datos diarios sobre incidencia, casos hospitalizados, ingresados en UCI, recuperados y fallecidos¹.

Estos datos, de gran utilidad para la monitorización de la epidemia en cada contexto local, han sido fundamentales para la toma de decisiones sobre medidas de control de la enfermedad; sin embargo, adolecen de limitaciones importantes cuando se trata de medir y comparar el impacto de la epidemia en diferentes poblaciones o entre distintos periodos en la misma población. Estos problemas provienen de que los criterios de definiciones de caso de las enfermedades emergentes no pueden estar perfectamente sincronizados en los distintos ámbitos y estructuras sanitarias, y cronológicamente van cambiando conforme mejoran las capacidades de detección de casos y la disponibilidad de test diagnósticos.

Una alternativa prontamente sugerida por los investigadores fue la de utilizar las series diarias o semanales de mortalidad a partir de las cuales sería posible hacer estimaciones del *exceso de defunciones* con respecto a las *esperadas* en las condiciones que se daban en el pasado (en línea con los nuevos paradigmas causales de los modelos contrafactuales). Los indicadores de impacto basados en el exceso del total de defunciones, en crisis emergentes como la de la COVID-19, estarían menos influenciados por cuestiones relacionadas con la organización y capacitación de los sistemas sanitarios locales, y por lo tanto permitirían derivar indicadores comparables más objetivos que los basados en la incidencia declarada.

Estas nuevas y urgentes demandas han puesto en evidencia los déficits de nuestros sistemas estadísticos para generar respuestas rápidas y a la vez fiables a cuestiones tales como las cifras diarias o semanales de mortalidad en distintos agregados geográficos.

Las estadísticas oficiales de mortalidad que provienen del Movimiento Natural de la Población (MNP) y que son la fuente principal de información sobre defunciones, están sujetas a un importante retraso de disponibilidad, por unos circuitos complejos y anticuados, pero de difícil reforma dado el actual marco jurídico. De ahí el interés por localizar otras alternativas que permitan disponer de datos de mortalidad con menos demora, pero preservando la exhaustividad y fiabilidad de la información.

Afortunadamente existían alternativas ya en funcionamiento, que si bien parten de la misma fuente de información que las estadísticas oficiales de mortalidad (certificados médicos de defunción y boletines estadísticos de defunción: CMD/BED), tienen circuitos de transmisión bastante más rápidos que los del MNP. Las oficinas informatizadas y conectadas del registro civil están desde 2008 realizando las inscripciones administrativas de las defunciones y la grabación en una base de datos centralizada por medio de la aplicación informática INFOREG (y su heredera DICIREG). Aunque no todas las oficinas del registro civil están conectadas a la base de datos central, actualmente la cobertura es bastante alta, del orden del 93 % de las defunciones en España, y por encima del 97 % en Andalucía.

Basados en los circuitos de los registros civiles, se han constituido algunos sistemas rápidos para la monitorización de la mortalidad estacional diaria o semanal.

En España destaca por su relevancia y antigüedad el sistema de Monitorización de la Mortalidad Diaria (MoMo) del Instituto de Salud Carlos III², relacionado con el proyecto EuroMOMO a nivel europeo³. MoMo utiliza los datos de INFOREG, y su objetivo es detectar y medir el exceso de muertes totales en relación con determinadas circunstancias, como son las epidemias de gripe o las olas de calor. Estos sistemas se fundamentan en la comparación entre el número de defunciones que se producen en una unidad de tiempo breve (generalmente día o semana) con una predicción basada en los datos observados en años previos. MoMo suministra series de mortalidad diaria para comunidades autónomas y grandes grupos de edad, así como las estimaciones de defunciones esperadas con sus dinteles superiores de confianza.

Recientemente el INE, ante las circunstancias de la pandemia y las demandas de los organismos estadísticos internacionales, ha empezado a publicar una estadística experimental⁴, que al igual que MoMo, utiliza los datos más recientes de INFOREG, proporcionando series semanales de defunciones, por sexo, grupos de edad y provincia de residencia, corregidas por infra-registro y actualizadas semanalmente. Estos datos están siendo utilizados para actualizar la base de datos *Short-term Mortality Fluctuations* (del proyecto *Human Mortality Database*) que contiene series semanales de mortalidad con datos de 38 países⁵.

La estadística experimental del INE aporta datos de defunciones semanales del año en curso y años previos, hasta el nivel de provincia, pero no incluye análisis del exceso de mortalidad, por lo que no puede considerarse un sistema de monitorización propiamente dicho. En el caso de MoMo, se estiman los excesos de mortalidad a nivel diario, pero no aporta datos por debajo del nivel geográfico de la comunidad autónoma. Debido a las limitaciones de estas dos actividades, tiene interés el diseño de un sistema rápido de monitorización de la mortalidad propio de Andalucía, que permita disponer de datos de defunciones observadas y esperadas en niveles geográficos subprovinciales.

2. Objetivos

El objetivo general es estudiar la viabilidad de diseñar un sistema rápido de monitorización de la mortalidad en Andalucía, que permita disponer de información hasta nivel subprovincial (tales como los Distritos Sanitarios o los municipios de más de 100.000 habitantes).

En este trabajo se expondrán los aspectos clave para la preparación de dicho sistema, dando respuesta a los siguientes objetivos específicos:

1. Analizar la calidad y cobertura de los datos provenientes de MoMo-INFOREG, en relación con los datos de defunciones del circuito estadístico (MNP).
2. Analizar los sesgos y distorsiones de la estimación rápida de la intensidad de la mortalidad al ser necesario utilizar los datos por lugar de inscripción en lugar de por lugar de residencia (esta necesidad se explica con más detalle en los siguientes apartados). Estudiar las alternativas de corrección que es posible aplicar sobre los datos para poder obtener series temporales homogéneas.

3. Evaluar distintas alternativas de estimación del *exceso de mortalidad*, en función de su capacidad para: (1) detectar episodios de mortalidad aberrante como los ocurridos durante la pandemia de COVID-19, los rebrotes epidémicos de las nuevas variantes gripales o la sobre-mortalidad por olas de calor; (2) derivar indicadores estandarizados que permitan comparar el impacto diferencial de estos episodios en los ámbitos sub-provinciales analizados.

3. Material y Método

3.1. Fuentes de datos para la construcción de la serie de defunciones

Para la construcción de una serie temporal de la mortalidad semanal que incluya datos con mínima demora y disponga de al menos una década de antigüedad, es preciso combinar dos fuentes que comparten parcialmente circuitos y documentos de declaración, pero no tienen, objetivos, definiciones y coberturas totalmente comparables: las defunciones registradas en MNP (sistema estadístico) y las defunciones grabadas en INFOREG/DICIREG.

En cuanto a MNP, contiene la información de las defunciones procedente de los CMD/BED, grabada y codificada por los organismos estadísticos. Es la fuente más fiable y exhaustiva sobre defunciones. Sus datos proporcionan las cifras de referencia para evaluar la calidad y cobertura de fuentes alternativas. Desafortunadamente, dada la complejidad de los circuitos y procesos actuales de tratamiento, es difícil disponer de datos suficientemente exhaustivos con una demora inferior a 3 o 4 meses desde la defunción.

Como se ha comentado anteriormente, INFOREG/DICIREG son las aplicaciones informáticas que recogen, en tiempo real, la inscripciones de defunciones en las oficinas municipales del Registro Civil y la almacenan en una base de datos centralizada gestionada por la Dirección General de Registros y Notarias, del Ministerio de Justicia. El Sistema de Vigilancia Epidemiológica de la Mortalidad Diaria (MoMo), gestionado por el *Centro Nacional de Epidemiología del Instituto de Salud Carlos III* (CNE-ISCIII), recibe los datos de la mortalidad diaria de INFOREG, y esta información se comparte con los sistemas de vigilancia epidemiológica de las CCAA. Para este estudio, los datos de INFOREG-Andalucía se han obtenido a través del acceso de la Consejería de Salud y Familias de Andalucía a la base de datos de MoMo (CNE-ISCIII. *Fuente de datos: División de Tecnologías y Servicios Públicos Digitales. Ministerio de Justicia*). Este acceso incluye información individualizada de las defunciones grabadas en INFOREG. Por ejemplo, los identificadores personales están censurados y la única información geográfica codificada y cumplimentada al 100 % es el municipio de inscripción. La variable 'ultimo domicilio', que en principio contendría información sobre la residencia, es un campo tipo literal, con gran heterogeneidad de contenido (mezclando nombre de entidades o instituciones, direcciones postales, etc.), lo que hace muy difícil una codificación exhaustiva. La principal ventaja de esta fuente de información es su baja demora (en la mayoría de los casos el tiempo entre la fecha de defunción y la grabación en INFOREG es menor de una semana) y su rápida disponibilidad.

3.2. Combinación de los ficheros de microdatos del MNP e INFOREG

Se construyó un 'data.set' de trabajo combinando los registros de microdatos del MNP y de MoMo-INFOREG, con el subconjunto de variables que tenían definiciones compatibles y niveles de calidad aceptable: fecha de nacimiento, fecha de defunción, sexo y municipio de inscripción de la defunción. Además se incluyó el municipio de residencia para las defunciones provenientes de MNP. La cumplimentación de estas variables es completa en la serie de MNP, sin embargo, en la serie derivada de INFOREG hay un 1,5 % de variable sexo faltante y un 0,15 % de fechas de nacimiento inconsistentes, que se han tenido que imputar antes de unir ambas fuentes.

Este 'data.set' se ha utilizado para hacer el estudio de los niveles de cobertura, el análisis de las relaciones entre las variables lugar de inscripción y lugar de residencia, y la estimación de los factores de corrección que fuera preciso aplicar sobre los datos de INFOREG.

3.3. Cobertura y exhaustividad de las defunciones de INFOREG-MoMo

Si bien las defunciones registradas en INFOREG-MoMo y en MNP provienen del mismo documento y comparten parte del circuito, no recogen la misma información con el mismo grado de detalle, ni cubren el mismo colectivo de individuos. INFOREG-MoMo en Andalucía incluye (y no con el 100 % de exhaustividad) las defunciones inscritas en la mayoría de las oficinas del registro civil localizadas en Andalucía, mientras que MNP recoge todas las defunciones inscritas en Andalucía y además, las defunciones de los residentes ocurridas en otras CCAA, tal como se recoge en el cuadro 1 (el subíndice A representa el conjunto de municipios de Andalucía, F el conjunto de localidades que no pertenecen a Andalucía, M las defunciones recogidas por MNP y R las defunciones recogidas en INFOREG).

Cuadro 1: Colectivos cubiertos por MNP e INFOREG-MoMo en Andalucía

	Residentes Inscritos Fuera	Residentes Inscritos en Andalucía	No-Residentes Inscritos en Andalucía
MNP	$M_{A,F}$	$M_{A,A}$	$M_{F,A}$
INFOREG	-	$R_{A,A}$	$R_{F,A}$

Las series de estadísticas oficiales de defunciones de Andalucía se construyen como la suma de las defunciones de los residentes, independientemente del lugar de su inscripción. Sin embargo, en la extracción de INFOREG que hemos utilizado no se dispone de las inscripciones de fuera de Andalucía, por lo que no es posible derivar directamente de INFOREG una serie compatible con las estadísticas oficiales. Otro inconveniente es que en los datos de INFOREG las variables sobre el último domicilio están pobremente cumplimentadas (más de un 9 % de registros en blanco a nivel de provincia, y en literales heterogéneos), por lo que no resultan fiables. Por ello, solo se pueden comparar las coberturas de las series de MNP e INFOREG por lugar de inscripción.

Las coberturas de INFOREG con relación a MNP son bajas en fechas anteriores a 2010 (inferior al 80 %), pero partir de 2013 superan el 98 % y su evolución es estable. Por tanto, la estimación de las defunciones por inscripción en MNP a partir de las de INFOREG resulta sencilla aplicando factores de corrección específicos para cada ámbito geográfico.

La correspondencia MNP-INFOREG a nivel de municipio de inscripción también es muy buena, a partir de 2013. Solo se han detectado 20 municipios que no han aportado datos ningún año a INFOREG. Son municipios pequeños que solo representan un 0.2 % de las defunciones inscritas en Andalucía. El resto del infra registro de INFOREG (entorno al 1.5 %) se produce entre los municipios con oficinas del registro civil ya informatizadas.

La predicción de las defunciones totales inscritas en los registros civiles de Andalucía a partir de INFOREG se obtiene multiplicando el número de defunciones reportado en INFOREG por un factor de corrección calculado como la razón entre las defunciones MNP/INFOREG en los años previos al actual. Este factor de corrección (FC^A) permite la estimación para un ámbito geográfico dado 'i' en el tiempo 't₀' en función de su historia pasada en 't-', sería:

$$FC^A_{i,t_0} = \frac{M_{A,i,t-} + M_{F,i,t-}}{R_{A,i,t-} + R_{F,i,t-}}$$

Dada la estabilidad temporal que muestra la serie del factor de corrección desde 2013 (entorno a un 2 % de infra registro), al aplicarlo sobre los datos de INFOREG obtenemos correcciones robustas de los datos que proporciona esta fuente por lugar de inscripción.

Tal y como se muestra en la figura 1 del apéndice, la corrección funciona bastante bien en los ámbitos regionales, provinciales y subregionales, ya que el factor de corrección es en la mayoría de los casos cercano a 1, por lo que solo modifica ligeramente las frecuencias originales de la fuente.

3.4. Relación entre el número de defunciones por lugar de inscripción y de residencia

Para poder derivar desde INFOREG predicciones adelantadas del número de defunciones que suministrará unos meses más tarde el circuito estadístico del MNP, hay que dar un paso más e intentar modelar las defunciones por lugar de residencia del MNP en función de las defunciones por lugar de inscripción de INFOREG.

El INE en su proyecto de estadísticas experimentales⁴, estima un solo factor de corrección basado en la razón entre las defunciones por lugar de residencia del MNP con respecto a las recopiladas tempranamente por INFOREG por inscripción. Con este factor, hace la corrección en un solo paso, de varias posibles discrepancias: el conocido infra registro de INFOREG, y las diferencias en los colectivos de referencia (residentes habituales y población flotante de un determinado ámbito).

La fórmula de este factor de corrección, usando la nomenclatura de la tabla 1, sería la siguiente para un ámbito geográfico dato 'i' y estimación en tiempo 't₀' en función de su historia pasada en 't-'

$$FC_{i,t_0} = \frac{M_{i,A,t-} + M_{i,F,t-}}{R_{A,i,t-} + R_{F,i,t-}}$$

Este factor de corrección único producirá buenas estimaciones de las cifras oficiales adelantadas (por lugar de residencia), si se dan dos condiciones: los colectivos respectivos de fallecidos residentes en la zona, pero inscritos en otra, y los fallecidos no-residentes pero inscritos en la zona, son relativamente pequeños, su saldo neto (diferencia entre los dos) es próximo a cero o porcentualmente irrelevante. Si se dan estas circunstancias este factor no debería desviarse mucho del valor 1 y estaría fundamentalmente determinado por la parte de infra registro (que estimamos previamente bajo la denominación de 'FC^A').

Profundizando un poco más, el estudio de este factor se puede descomponer en el producto de dos factores, el 'FC^A', que definimos previamente y que está asociado al infra registro de los datos de INFOREG y otro factor (que denominaremos FC^B) asociado con la diferencia de las defunciones por lugar de inscripción y residencia, y determinado fundamentalmente por la diferencia entre los residentes fallecidos fuera y los foráneos fallecidos dentro del ámbito 'i' que estamos analizando.

$$FC_{i,t_0} = FC^A_{i,t_0} \times FC^B_{i,t_0} = FC^A_{i,t_0} \times \frac{M_{i,A,t-} + M_{i,F,t-}}{M_{A,i,t-} + M_{F,i,t-}}$$

El comportamiento del segundo factor de corrección puede estudiarse a partir de los microdatos de defunciones en MNP, donde están correctamente codificados el lugar de inscripción y el lugar de residencia. Para comprobar si este indicador sigue un comportamiento bondadoso se ha estimado su valor para distintos ámbitos geográficos, durante el periodo 1998-2019 y se ha representado en la figura 2 del apéndice. En él se muestra que el 'FC^B' tiene un efecto débil (próximo a 1) en el ámbito regional y en las provincias, por lo que en estas agregaciones su utilización no resulta estrictamente necesaria. Sin embargo, en ámbitos infra-provinciales, su comportamiento es muy irregular y sus valores están bastantes alejados del no efecto.

Se aprecia claramente que en los grandes municipios el factor de corrección es sistemáticamente menor que uno, indicando que se están inscribiendo muchas más defunciones que las que les corresponderían por residencia. En los distritos sanitarios la situación es muy variable: los distritos correspondientes con zonas rurales apartadas tienen valores superiores a 2, lo que implica que están perdiendo más de la mitad de la inscripciones de las defunciones de sus residentes. Investigando las causas de estas disparidades del factor B, se puede advertir que los municipios con centros hospitalarios ejercen una fuerte atracción para la inscripción registral de los fallecidos de otros municipios, mientras que los ámbitos sin centro hospitalario inscriben un porcentaje alto de las defunciones de sus residentes en otros municipios.

Dado que en Andalucía el mapa hospitalario esta jerarquizado a nivel provincial, las estimaciones provinciales de defunciones por residencia funcionan aceptablemente bien. Sin embargo, actualmente no es viable derivar estos eventos en ámbitos sub-provinciales.

3.5. Reconstrucción de series semanales de defunciones por lugar de inscripción

Tras todo lo comentado anteriormente, se concluyó que no era posible construir series de defunciones adelantadas por lugar de residencia a partir de la información de INFOREG-MoMo, salvo en los ámbitos regional y provincial. Por el contrario, es posible y sencillo reconstruir las series por lugar de inscripción. Esta opción, aunque no sea del todo óptima, permite derivar indicadores subprovinciales de forma rápida, lo que no será posible de otra manera, al menos mientras no se mejore la cumplimentación y codificación de las variables de residencia en INFOREG.

En consecuencia, se han reconstruido las series adelantadas por lugar de inscripción, combinando la información de los datos exhaustivos de MNP con los datos de INFOREG-MoMo una vez corregidos por subregistro.

Las series reconstruidas contienen el número de defunciones semanales, por sexo, grupos de edad y ámbito territorial (provincias, distritos sanitarios y municipios mayores de 100.000 personas). Las series se unen en un punto del tiempo, que varía conforme se procesan nuevos datos, con la condición de asegurar que los últimos datos del MNP incluidos correspondan a un mes del que se disponga de los lotes de defunciones de al menos los 6 meses posteriores, con objeto de capturar frecuentes retrasos. En ese punto del tiempo que va variando, conforme se actualizan los datos, las serie con las cifras de defunciones más recientes son las que proceden de INFOREG.

3.6. Exposición: población inscripción-vinculada (PIV)

Las series temporales de las defunciones por lugar de inscripción cubren un extenso periodo y los últimos datos disponibles tienen baja demora. Sin embargo, su utilidad resulta mermada por no disponerse de adecuadas poblaciones de referencia, ya que las poblaciones estadísticas se producen por lugar de residencia.

No disponer de poblaciones de referencia no sería un inconveniente grave si no fuera necesario trabajar con tasas para estimar las defunciones esperadas. Se puede trabajar con defunciones (en vez de tasas) si el ámbito poblacional tiene moderado o gran tamaño y se genera un número suficiente de defunciones para hacer buenas predicciones de esperadas, y secundariamente estimar indicadores estandarizados (como la razón entre defunciones observadas y esperadas). MoMo y EuroMOMO emplean este acercamiento, que funciona bien en poblaciones grandes, con estructuras de población que cambian poco de año en año y cuyos efectos sobre el número de defunciones pueden ser fácilmente capturados por el componente de tendencia del modelo. Sin embargo, en poblaciones de mediano y pequeño tamaño, con efectivos y composición por edad de evolución irregular, los altibajos en las series de población tienen importantes efectos difícilmente modelables por la componente de tendencia de un modelo basado solo en el número de defunciones. En nuestro caso, dado que se pretende realizar estimaciones de exceso de muerte en ámbitos sub-provinciales de mediano y pequeño tamaño, es un requisito prácticamente obligado, utilizar estimaciones de población de referencia (como 'offset' del modelo) para obtener predicciones de precisión suficiente en el cálculo de las defunciones esperadas.

Para ello, se ha desarrollado un constructo teórico al que se ha denominado *población inscripción-vinculada (PIV)*. La **PIV** es la población virtual, en un mundo contrafactual, que existiría si en el ámbito donde se han inscrito las defunciones de foráneos añadiéramos la población en riesgo, por sexo y edad, necesarias para producirlas, población que paralelamente deberá descontarse de los ámbitos de procedencia de los foráneos. Básicamente, se trata de emparejar las transferencias de defunciones del ámbito residencial al inscripcional, con sus correspondientes transferencias de efectivos de población en riesgo que produciría ese número de defunciones.

Las asunciones para poder estimar estas transferencias son dos:

1. Un fallecido foráneo en el ámbito donde se inscribe la defunción proviene de una subpoblación de similar fragilidad media que la de los residentes del mismo sexo y edad en dicho ámbito. Es decir, los riesgos de mortalidad de la población foránea inscrita fuera de su ámbito son similares a los de los residentes en el ámbito de procedencia.
2. Una transferencia de un fallecimiento entre un ámbito de residencia y otro de inscripción, conlleva una transferencia proporcional y calculable de población expuesta del ámbito de residencia al del inscripción.

A partir de estas asunciones y las matrices por sexo y edad, de las transferencias de defunciones entre lugar de residencia e inscripción, la población por sexo y edad por lugar de residencia y de sus respectivos riesgos de mortalidad en cada ámbito, es posible estimar las transferencias de población en riesgo, y agregándolas por lugar inscripción se obtienen las PIV.

Al usar las PIV para el cálculo de indicadores por lugar de inscripción, se subsana gran parte de las discrepancias observadas al comparar las tasas por residencia y la estimaciones 'naïve' por inscripción (con poblaciones según residencia de las estadísticas oficiales). A título de ejemplo, se muestra en el apéndice dos gráficos (figura 3 y 4) con las tasas específicas por edad y las estandarizadas por dos distrito sanitario según residencia e inscripción 'naïve' y ajustado por PIV.

3.7. Series semanales de poblaciones para predicción

Para construir las series semanales de población, tanto por residencia como las poblaciones inscripción vinculada o **PIV**, se partió de las poblaciones oficiales municipales a 1 de enero por sexo y edades simples para el periodo 1998-2020. A esta serie, se añadieron dos puntos más recientes, correspondientes al 1 de enero de 2021 y 2022 (aún no disponibles), para lo que se utilizó un método sencillo de proyección basado en coeficientes prospectivos generacionales derivados directamente en cada ámbito desde las propias poblaciones oficiales.

Sobre esta serie de poblaciones a principios de año, por ámbito residencial, se aplicaron las matrices de transferencia inscripción-residencia para obtener la serie de poblaciones **PIV** a 1 de enero, que servirán de población de referencia para procesar las series adelantadas por inscripción previamente construidas.

A continuación, a partir de las series de poblaciones a 1 de enero, se realizaron interpolaciones por ámbito, sexo, y edades simples en cada jueves (52 o 53 jueves en

cada año). Estos valores interpolados serán las poblaciones medias de las semanas ISO-8651, siendo la primera semana de año la que contenga el primer jueves. Estas poblaciones semanales, se convierten a personas-año, multiplicando por '7/365,2425'.

Estas estimaciones son las que se utilizan (como 'offset') para mejorar la predicción de las defunciones esperadas y estimación posterior de los indicadores estandarizados de impacto, basados en el exceso de defunciones de los episodios epidémicos.

3.8. Estimación de las defunciones esperadas

La clave para la estimación de *exceso de mortalidad* es poder disponer de un *nivel de referencia basal* que se correspondería con el número de casos esperables en un lugar y tiempo determinado en las condiciones habituales que se han observado en el pasado. Esta cuestión ha generado una abundante bibliografía en el campo de la vigilancia epidemiológica, ya que un brote epidémico se define como la aparición de un número significativamente superior de casos a los esperados en un tiempo y espacio acotado, por lo que se precisa al menos de dos 'referencias' para identificarlo: el número de casos esperados y el dintel superior partir del cual el exceso de observado sería significativo.

Aunque hay referentes previos, merece destacar que, a mediados del siglo XX, el 'Communicable Disease Center' (CDC) de Atlanta usaba varias metodologías para el cálculo de las muertes esperadas⁶. El modelo más sencillo se basaba en una regresión lineal realizada con las defunciones de las mismas semanas 'i' en los 'n' años previos.

$$Y_{t,i} = \alpha_i + \beta t + \epsilon_{t,i} | t \in \{t_0 - n, \dots, t_0 - 1\} \Rightarrow \widehat{Y}_{t_0,i} = \alpha_i + \beta t_0$$

Esta sencilla metodología, captura bastante bien el efecto tendencia y el efecto estacional. Partiendo de este esquema se han ido introduciendo sucesivas mejoras, sofisticando el análisis del componente estacional (variables 'dummy' estacionales, series de Fourier..), y mejorando el sencillo modelo lineal, con modelos lineales generalizados (Poisson, binomial negativa, etc.) o modelos aditivos generalizados (GAM).

En nuestro caso hemos optado por utilizar la estrategia propuesta por Farrington⁷, con las mejoras introducidas por Noufaily⁸. Farrington, en 1996 sintetizó muchos de los avances hasta entonces alcanzados, proponiendo un algoritmo robusto que mediante un análisis automático de las series temporales de casos fuera capaz de predecir los casos esperados y los dinteles epidémicos que permitieran identificar alarmas. La metodología propuesta permite ajustar las tendencias y la estacionalidad que solemos encontrar en las series temporales, aplicando un modelo de regresión log-lineal (Poisson con sobredispersión). El conjunto de datos que se usa para estimar los valores esperables para una fecha de referencia concreta en el año en curso, es la frecuencia de casos de los último n-años, en el mismo periodo. Por ejemplo, si se usan series semanales, se tomarían los datos de la misma semana actual, pero en los n últimos años, más los datos de frecuencia de b-semanas anteriores y posteriores de su entorno. Así, si queremos predecir los valores de la semana 35 del año actual, usando n=5 y b=2, tomaríamos los datos de la semana 35 de los 5-años anteriores y su entorno de ancho igual a 2 semanas (semanas 33 a 37) para predecir esperados y dinteles.

Esquemáticamente el algoritmo de Farrington consiste en los siguientes pasos:

1. Ajustar el modelo inicial estimando sus parámetros: media y valor de sobre-dispersión.
2. Calcular los pesos de manera automática (basado en el cálculo de los residuos de Anscombe), lo que permite corregir la serie de esperados, restando importancia en la predicción a los periodos pasados con posibles brotes epidémicos.
3. Recalcular el modelo y revisar las estimaciones de media y sobre-dispersión.
4. Comprobar que el parámetro de tendencia estimado sea significativo. Si no lo es, volver a estimar los esperados eliminando la tendencia (calculando esto como simple media de los valores de los n-últimos años).
5. Recalcular la serie completa de esperados y los valores de dintel epidémico (límite superior de confianza).
6. Marcar las alamas y posibles episodios epidémicos cuando los casos observados superen el dintel.
7. Calcular el exceso y el excedente de mortalidad, y en su caso indicadores de impacto agregado.

El método propuesto por Farrington, con diversas mejoras, es la base de la metodología utilizada para la vigilancia epidemiológica en la mayor parte de los países europeos, así como la que usa EuroMOMO.

En este trabajo se han utilizado las ampliaciones sugeridas por Noufaily⁷ al método de Farrington, que aprovechan mejor la información fuera del área del entorno de la fecha en curso, sofisticando el modelo original usando *splines*. Para las estimaciones hemos usado el paquete “Surveillance” de R⁹, que facilita la aplicación de esta metodología.

Para valorar la validez del sistema, los periodos de exceso de mortalidad detectados se confrontaron con periodos epidémicos previamente conocidos de alta mortalidad (años con epidemia de gripe importantes, ola de calor de 2015) y con las distintas olas de la pandemia de COVID-19.

4. Resultados

Los resultados sobre la exhaustividad y calidad de INFOREG como fuente de datos para las defunciones recientes en un sistema de monitorización rápida de la mortalidad en Andalucía se han adelantado en el apartado de Material y Métodos, dado que era necesario conocerlos para poder explicar la necesidad de utilizar el referente geográfico del lugar de inscripción. Igualmente se han adelantado los resultados que permiten ver la necesidad y utilidad del uso de las poblaciones inscripción-vinculadas (PIV), sobre todo en los niveles subprovinciales, y su validez como proxy de las tasas por lugar de residencia.

A continuación se exponen los resultados más relevantes sobre la **validez del sistema de monitorización de la mortalidad propuesto para la detección del exceso de mortalidad**.

En las figuras 5 y 6 se puede comprobar la validez del sistema para detectar periodos conocidos de alta mortalidad (a nivel de Andalucía, 2008-2021, por separado para hombres y mujeres): concretamente, las gripes de 2009, 2012, 2015 y 2019; la ola de calor de 2015 y las olas epidémicas de la pandemia de COVID-19 en 2020 y 2021. Al desagregar por grandes grupos de edad (figura 7) e incluso por grupos quinquenales (figura 8), esta capacidad de detección sigue siendo buena; aunque como cabe esperar en los grupos quinquenales de edad más jóvenes prácticamente no ha habido un impacto de las diferentes crisis en la mortalidad.

Las figuras 9 y 10 ilustran la validez del sistema al utilizarlo para detectar excesos de mortalidad en poblaciones más pequeñas, concretamente en los distritos sanitarios, de forma desagregada por sexo y para los grupos de edad de menos de 65 años y de 65-74 años. La figura 9 presenta los datos para hombres en dos distritos de la provincia de Sevilla, uno con mucha población (la capital) y otro muy rural y con población pequeña (sierra norte): puede verse cómo incluso con pocos casos de defunciones observadas en la sierra norte en los hombres más jóvenes, se detectan los excesos de mortalidad asociados a la pandemia de COVID-19. La figura 10 recoge el ejemplo de dos distritos de la provincia de Granada (especialmente afectada por la pandemia), donde llama la atención la detección de un periodo de varias semanas de exceso de mortalidad durante la pandemia de COVID-19 en el Distrito Metropolitano.

Como indicador clave que puede derivarse de un sistema de monitorización de este tipo se ha analizado el impacto de los excesos de mortalidad en la expectativa de vida. En las figuras 11 y 12 se presentan los resultados para Andalucía, con datos a nivel de años epidémicos (inicio el 21 de marzo). Se observa el impacto en la expectativa de vida de las gripes epidémicas más importantes en hombres y mujeres, así como la magnitud mucho mayor del impacto de la pandemia de COVID-19, que supera el año de vida perdido en los hombres. Al realizar el análisis de la brecha en longevidad por trimestre (figura 13), se pueden observar los picos de gripe y ola de calor de 2015 diferenciados, también que el mayor impacto antes de la pandemia de COVID-19 ocurrió en la gripe de 2015 en mujeres (1.5 años de brecha de longevidad), y cómo la oleada del otoño de 2020 de la pandemia de COVID-19 es la de mayor impacto en Andalucía, superando los 2 años de brecha de longevidad en los hombres en ese trimestre. Este mismo análisis trimestral permite la comparación del impacto de los periodos epidémicos entre provincias (figura 14). Por ejemplo, puede verse el impacto en la longevidad de la ola de otoño de 2020 de la pandemia, que ha sido superior a los 2.5 años en ambos sexos en Granada, así como en los hombres en las provincias de Sevilla, Jaén y Cádiz.

5. Conclusiones

1. Un sistema de monitorización rápida de la mortalidad es factible y fiable para ámbitos subprovinciales (distritos sanitarios y municipios de más de 100.000 habitantes), a partir de la información de los certificados médicos de defunción que consta en el sistema de información de los registros civiles (INFOREG).
2. El uso de poblaciones vinculadas al lugar de inscripción resulta una estrategia útil para mejorar las estimaciones en un sistema de detección de exceso de la mortalidad.
3. Los resultados obtenidos a partir de la metodología usada en este estudio muestran su utilidad y validez para detectar periodos de exceso de mortalidad, por comparación con los periodos de sobremortalidad conocida, tanto en epidemias previas de gripe y olas de calor, como durante la pandemia de COVID-19

6. Referencias

1. Red Nacional de Vigilancia Epidemiológica (RENAVE). Instituto de Salud Carlos III. Informes COVID-19: [Hiperenlace al documento](#)
2. Instituto de Salud Carlos III. Vigilancia de la Mortalidad Diaria (MoMo): [Hiperenlace al documento](#)
3. EUROMOMO: [Hiperenlace al documento](#)
4. Instituto Nacional de Estadística (INE). Estimación del número de defunciones semanales (EDeS) durante el brote de COVID-19. Madrid: INE, 2021. [Hiperenlace al documento](#)
5. Németh L, Jdanov DA, Shkolnikov VM. An open-sourced, web-based application to analyze weekly excess mortality based on the Short-term Mortality Fluctuations data series. *PloS ONE*. 2021;16(2): e0246663. [Hiperenlace al documento](#)
6. Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Rep*. 1963;78(6):494-506. [Hiperenlace al documento](#)
7. Farrington CO, Andrews NJ, Beale AD, Catchpole MA. A statistical algorithm for the early detection of outbreaks of infectious disease. *J Royal Stat Soc*. 1996;159(3):547-63. [Hiperenlace al documento](#)
8. Noufaily A, Enki DG, Farrington P, Garthwaite P, Andrews N, Charlett A. An improved algorithm for outbreak detection in multiple surveillance systems. *Online J Public Health Inform*. 2013; 5(1): e148. [Hiperenlace al documento](#)
9. Höhle M. surveillance: An R package for the monitoring of infectious diseases. *Computational statistics*. 2007;22(4):571-82. [Hiperenlace al documento](#)

7. Apéndice gráfico

Lista de figuras

1. Razón entre las defunciones según lugar de inscripción en MNP e INFOREG (Andalucía 2013-2019)
2. Razón entre las defunciones según lugar de residencia y de inscripción en MNP (Andalucía 2013-2019)
3. Mortalidad por sexo y edad. Según residencia e inscripción (estimación ingenua y corregida) por DS. Año 2018
4. Serie de tasa estandarizadas por DS (Estimaciones ingenuas y corregidas). De 1998 a 2019
5. Defunciones observadas, esperadas y alarmas epidémicas. Hombres 2008-2021. Observado: puntos verdes y serie azul, Esperadas: serie gris, dintel epidémico: serie amarilla. Alarmas: marcas verticales
6. Defunciones observadas, esperadas y alarmas epidémicas. Mujeres 2008-2021. Observado: puntos verdes y serie azul, Esperadas: serie gris, dintel epidémico: serie amarilla. Alarmas: marcas verticales
7. Defunciones observadas, esperadas y alarmas epidémicas. Grandes grupos de edad. Total de Andalucía.
8. Defunciones observadas, esperadas y alarmas epidémicas. Grupos quinquenales. Total de Andalucía.
9. Defunciones observadas, esperadas y alarmas epidémicas. Distritos Sanitarios: Sevilla Capital y Sierra Norte
10. Defunciones observadas, esperadas y alarmas epidémicas. Distritos Sanitarios: Granada Capital y Corona Metropolitanaa Norte
11. Expectativa de vida Observada y Esperada. Andalucía. Años epidémicos con comienzo el 21-marzo
12. Brecha en longevidad: diferencias entre expectativa de vida esperada y observada. Años epidémicos con comienzo el 21-marzo
13. Brecha trimestral en longevidad: diferencias entre expectativa de vida esperada y observada. Andalucía
14. Brecha trimestral en longevidad: diferencias entre expectativa de vida esperada y observada. Provincias

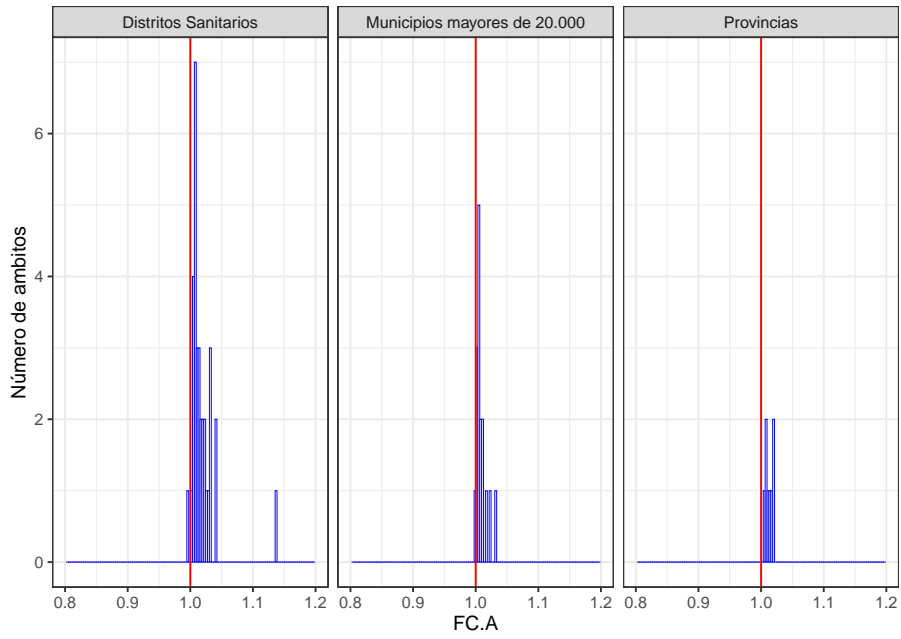


Figura 1: Razón entre las defunciones según lugar de inscripción en MNP e INFO-REG (Andalucía 2013-2019)

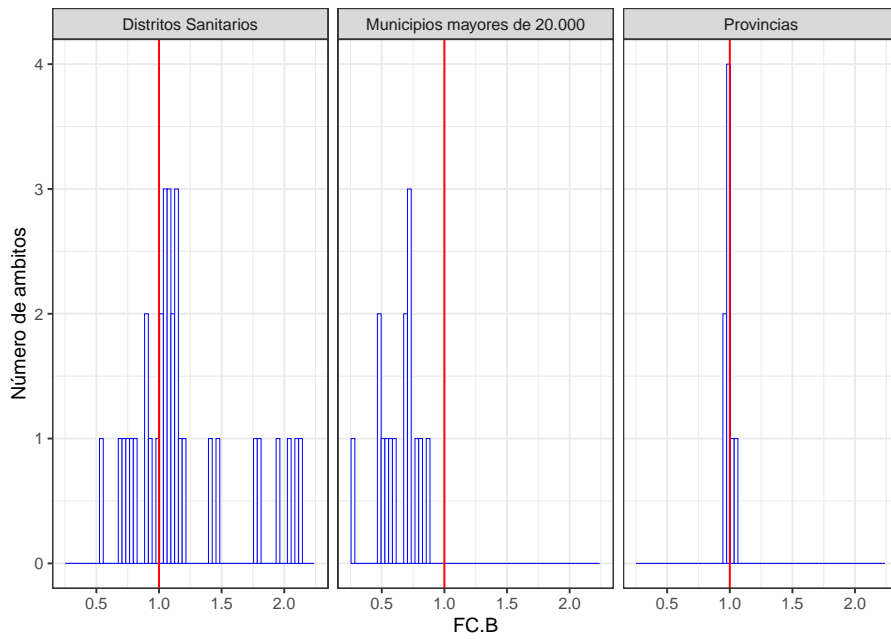


Figura 2: Razón entre las defunciones según lugar de residencia y de inscripción en MNP (Andalucía 2013-2019)

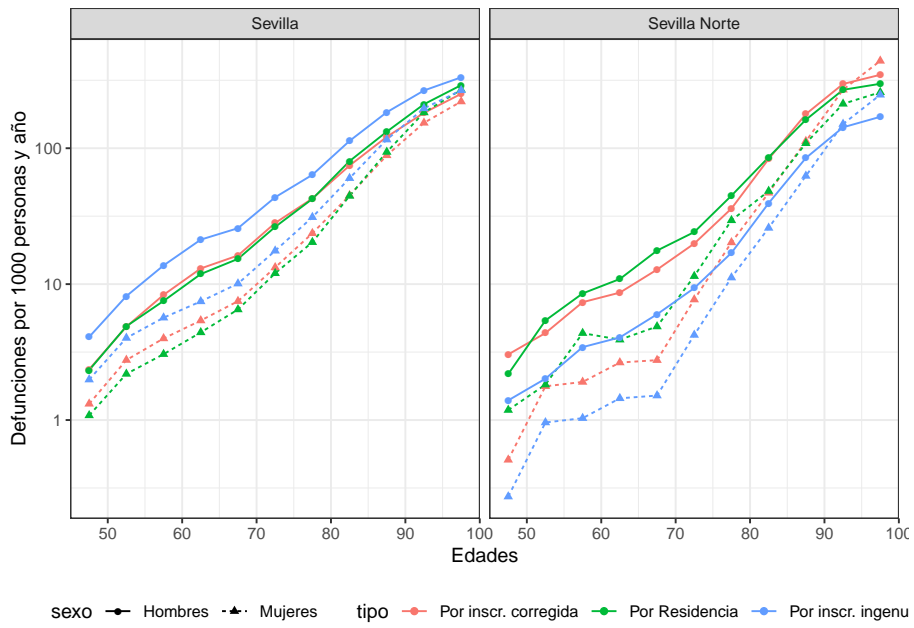


Figura 3: Mortalidad por sexo y edad. Segun residencia e inscripción (estimación ingenua y corregida) por DS. Año 2018

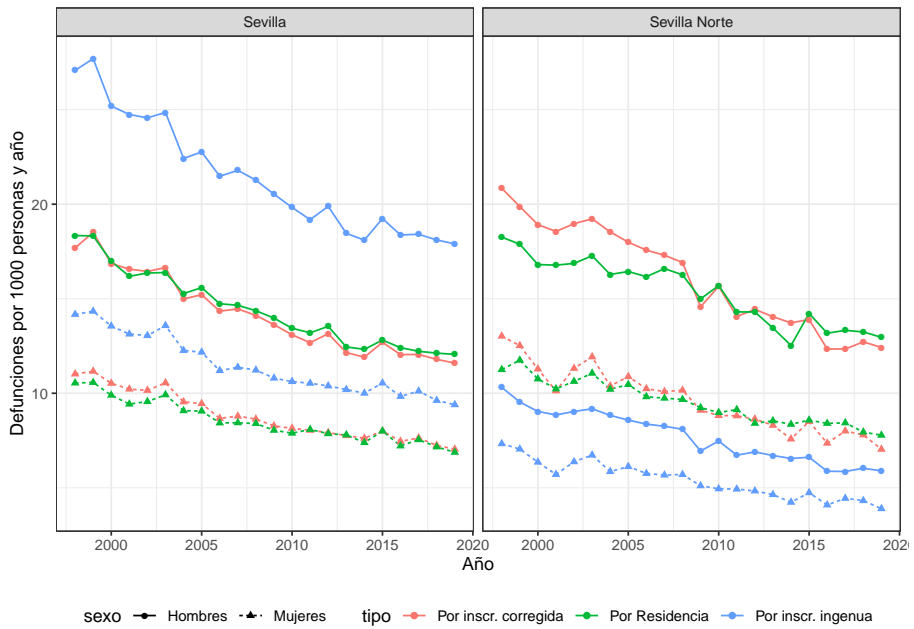


Figura 4: Serie de tasa estandarizas por DS (Estimaciones ingenuas y corregidas). De 1998 a 2019

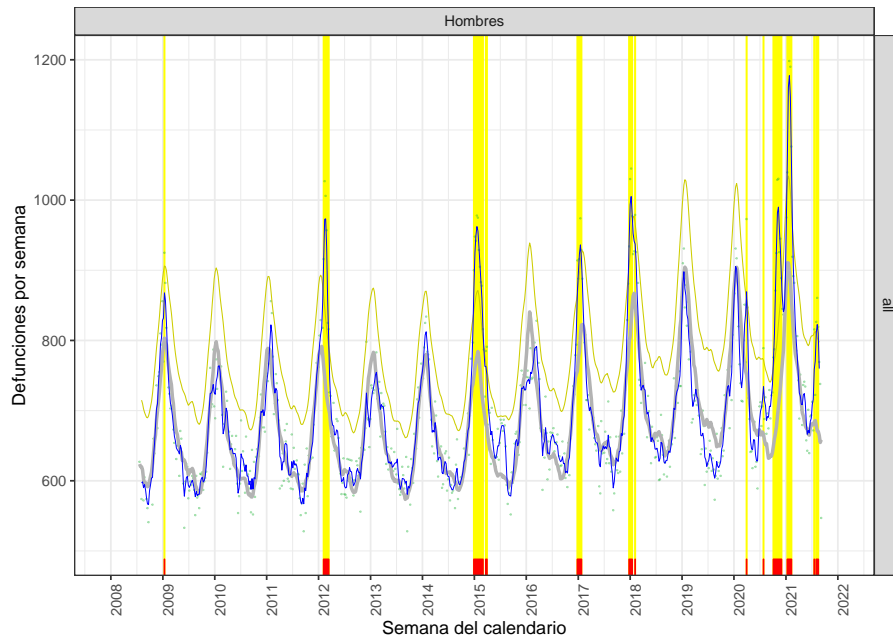


Figura 5: Defunciones observadas, esperadas y alarmas epidemicas. Hombres 2008-2021.

Observado: puntos verdes y serie azul, Esperadas: serie gris, dintel epidémico: serie amarilla. Alarmas: marcas verticales

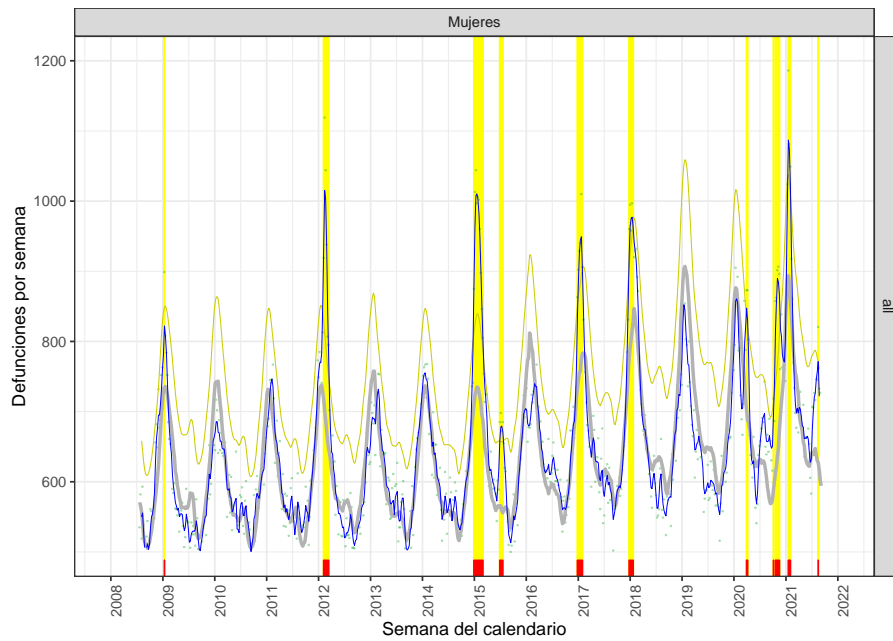


Figura 6: Defunciones observadas, esperadas y alarmas epidemicas. Mujeres 2008-2021.

Observado: puntos verdes y serie azul, Esperadas: serie gris, dintel epidémico: serie amarilla. Alarmas: marcas verticales

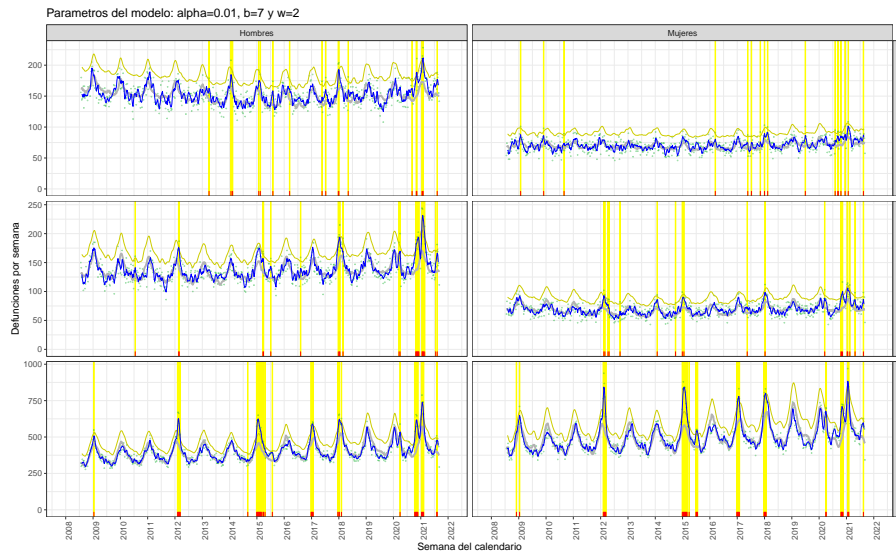


Figura 7: Defunciones observadas, esperadas y alarmas epidemicas. Grandes grupos de edad. Total de Andalucia.

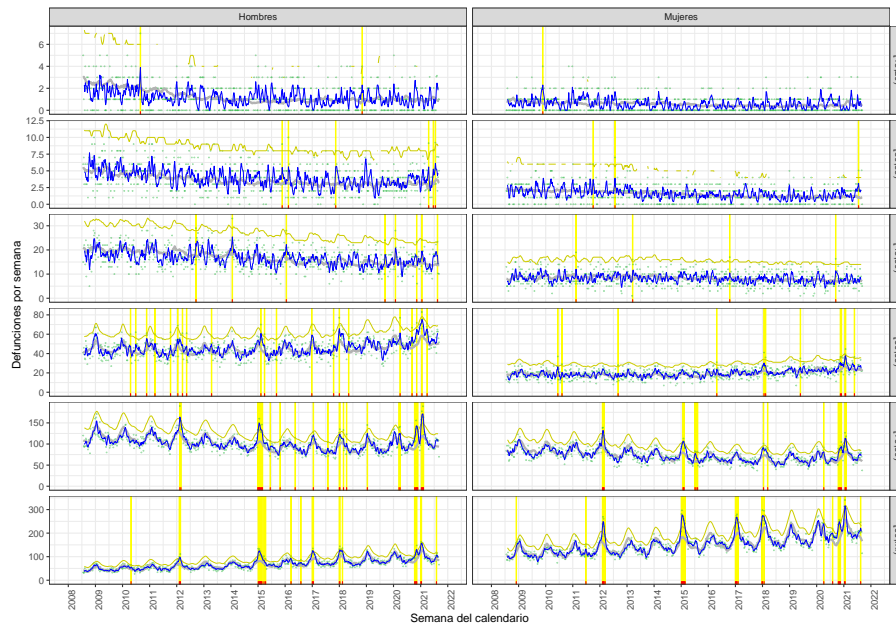


Figura 8: Defunciones observadas, esperadas y alarmas epidemicas. Grupos quinquenales. Total de Andalucia.



Figura 9: Defunciones observadas, esperadas y alarmas epidemicas. Distritos Sanitarios: Sevilla Capital y Sierra Norte

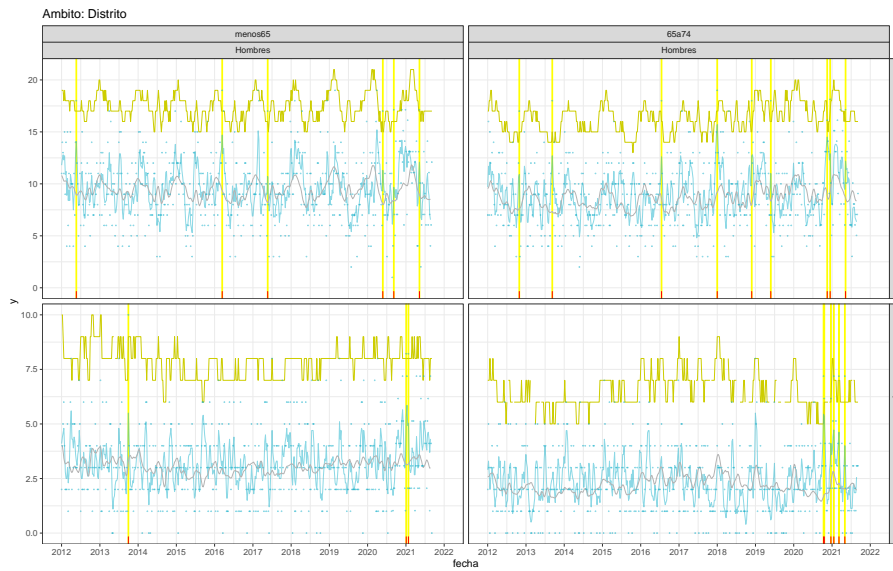


Figura 10: Defunciones observadas, esperadas y alarmas epidemicas. Distritos Sanitarios: Granada Capital y Corona Metropolitanaa Norte

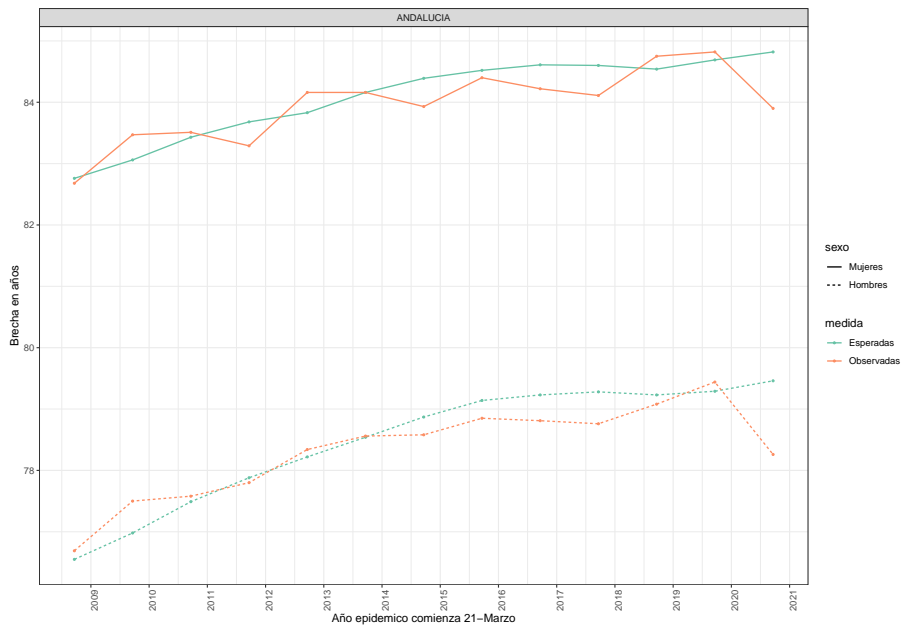


Figura 11: Expectativa de vida Observada y Esperada. Andalucía Años epidémicos con comienzo el 21-marzo

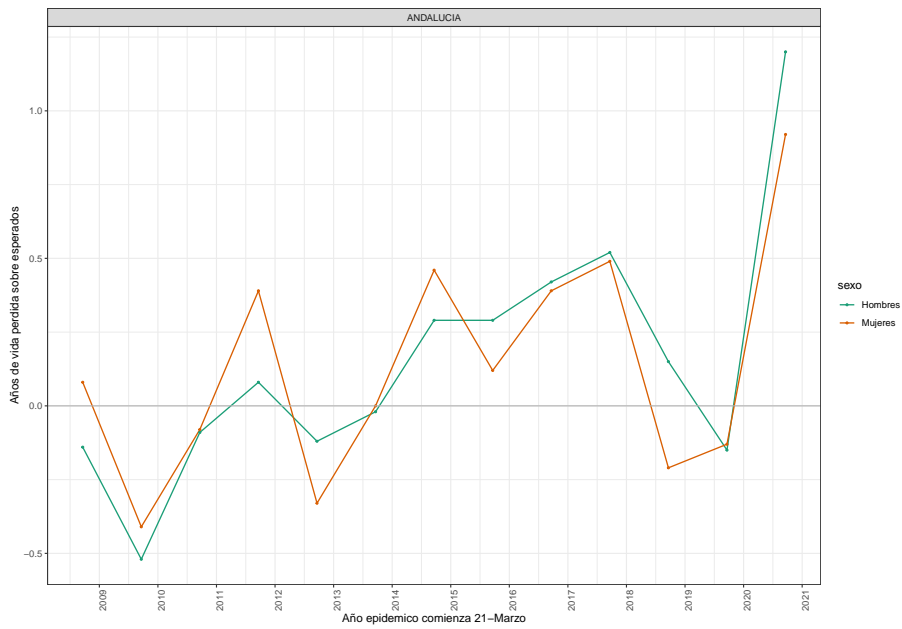


Figura 12: Brecha en longevidad: diferencias entre expectativa de vida esperada y obserbada. Años epidémicos con comienzo el 21-marzo

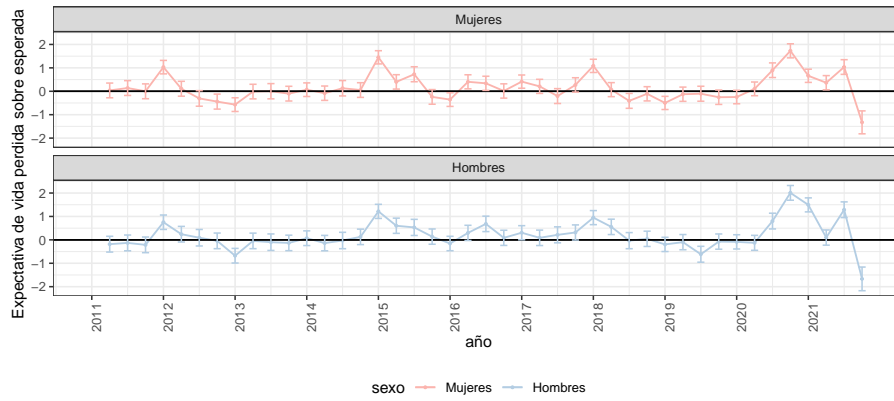


Figura 13: Brecha trimestral en longevidad: diferencias entre expectativa de vida esperada y obserbada. Andalucía

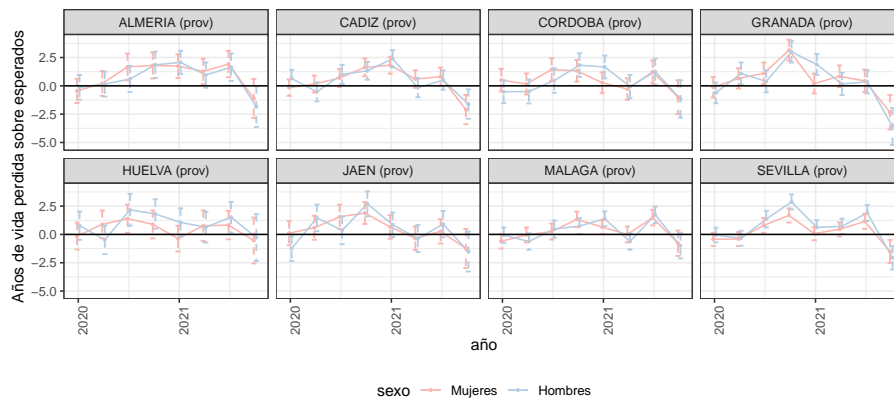


Figura 14: Brecha trimestral en longevidad: diferencias entre expectativa de vida esperada y obserbada. Provincias

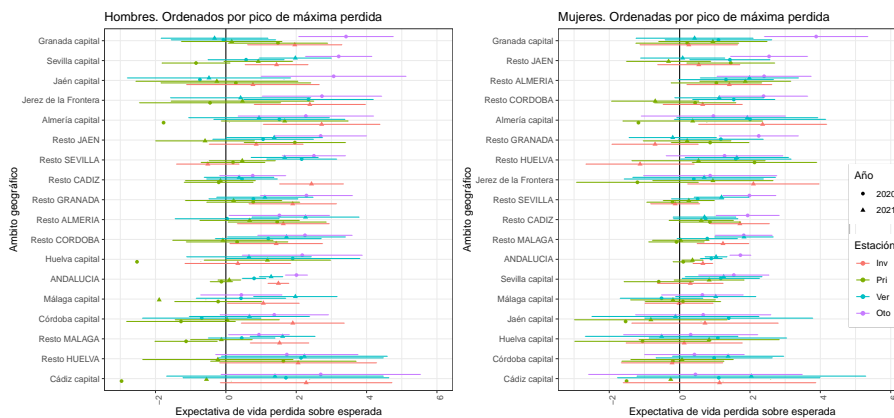


Figura 15: Perdida/ganacia trimestral en Esperanza de vida. Capitales y resto provincial