

Por qué es necesario crear un estándar de datos de longevidad de la población asegurada

Juan de Ipiña // Miembro 2332 del IAE, FIA (Fellow of the Institute and Faculty of Actuaries del Reino Unido), miembro del grupo de trabajo de biometría actuarial del IAE.

La actualización de las tablas de longevidad en España ha sido un esfuerzo importante y de una importancia muy notable para la industria aseguradora en España. Creo que debería haber un **Mecanismo Institucional Continuo** en España que garantice la actualización de los datos clave de longevidad, que podría ayudar a la industria (especialmente aquellos que no tienen modelos internos parciales de longevidad o suficientes datos de longevidad) para actualizar sus hipótesis. Este mecanismo debería garantizar la calidad de los datos y **establecer una base de datos concreta**, a la que deberían contribuir con datos estándar todas las entidades españolas cada año. Este debería ser el origen de todos los análisis de longevidad en el futuro. Este artículo solo trata de ser un primer paso inicial para el análisis del establecimiento de dicho repositorio de datos estándar. He tratado de incluir los datos más básicos siendo consciente que pueden incluirse más.

En España ha pasado un período de tiempo muy largo sin que las tablas de longevidad fueran actualizadas. Por ello creo que el plazo de actualización debería estar en un intervalo más corto que el que hemos tenido entre las PER2000 y las PER2020. Creo que cada cinco años es un espacio de tiempo razonable y alineado con otros países. Por ejemplo, en Holanda la Sociedad nacional de actuarios (Actuarieel Instituut) actualiza las tablas AG cada dos años, en el Reino Unido, el Constant Mortality Investigation (CMI) propone tablas base actualizadas aproximadamente cada 6 años, aunque proporcionan datos y análisis casi todos los años para una mejor proyección de flujos de caja¹. Sabemos que este es un **reto desafiante**, pero creemos que tenemos que ser realmente ambiciosos en España.

¹ Los datos que se utilizan para completar los modelos de CMI para las proyecciones de mortalidad se publican todos los años y, por lo tanto, se puede actualizar las hipótesis cada año para las proyecciones de mortalidad.

Para poder construir una tabla de longevidad se requiere realizar un análisis de datos detallado. Esto conlleva tener **datos históricos comparables por año calendario**. Esencialmente, todo gira alrededor del número de fallecidos y los expuestos, por edad y género. Sin embargo, esto no es tan sencillo ni directo. Hace falta analizar mucho el dato y no mezclar subpoblaciones para evitar riesgo base y gestionar mejor la antiselección. El objetivo último de una tabla de longevidad es poder realizar una valoración suficiente de las obligaciones futuras que asumen las entidades aseguradoras, poder monitorizar y gestionar el riesgo biométrico y el capital económico generado por el mismo en productos que pagan un importe de dinero mientras el asegurado esté vivo. ¿Cuáles son los principales factores a tener en cuenta en su construcción? A continuación, detallo los que creo que son los principales datos a tener en cuenta.

Para empezar, hay que **seleccionar bien la muestra de datos a recoger de forma anual en el futuro**. Para ello hay que tener en cuenta los siguientes puntos:

- > Por ejemplo, para España existen datos de expuestos y fallecidos en el Human Mortality Database (HMDB)² desde el año 1908 hasta el año 2020. Sin embargo, esta información es de **población general** y no es de **población asegurada**. Esta diferenciación es importante porque en general la población asegurada es aversa al riesgo, tiene mayor poder económico y generalmente mayor nivel educativo (que conlleva mejores usos alimenticios y de estilo de vida) Es un hecho contrastado que la longevidad es mayor en la población asegurada. Por ello, es importante recabar información estándar de las entidades aseguradoras de forma periódica, para poder realizar actualizaciones en las tablas actuales de longevidad existentes cuando llegue su fecha de actualización.
- > Un problema muy común del uso de datos de población general para inferirla a la población

² [www.https://www.mortality.org/](https://www.mortality.org/)

asegurada es que tanto la tabla base como los factores de mejora de la mortalidad son distintos y puede conllevar un **riesgo base**. El riesgo base en la longevidad surge porque diferentes poblaciones, o subpoblaciones, inevitablemente experimentarán diferentes resultados de longevidad, lo que conllevaría a una estimación incorrecta en caso de asumir la base incorrecta.

- > Otro punto importante a tener en cuenta dentro de la población asegurada es diferenciar los datos de las subpoblaciones de **colectivos** e **individuales**, dado que hay mucha casuística. En carteras **individuales** es mucho más común la práctica de la **antiselección**, que puede conllevar a que en un conjunto de pólizas de rentas se integre un riesgo de longevidad superior a la técnicamente prevista a la del escenario de mejor estimación, porque son conscientes que probablemente tendrán una salud mejor que la media y serán más longevos. Puede llegar a ser un caso especialmente gravoso para carteras pequeñas. Otro punto importante en las pólizas individuales es que muchas de ellas no se tiene constancia si los asegurados fallecieron o no. Por este motivo, las entidades de seguros deben mantener una base de datos de longevidad actualizada por este motivo y verificar que los asegurados reconocidos como supervivientes en la base de datos no estén registrados en el Índice Nacional de Mortalidad (INDEF) de España, para evitar errores. En las pólizas de **colectivos** sería interesante recoger el código de industria CNAE de la compañía del trabajador asegurado. Sería ideal complementarlo con el código CNO (de ocupación). Otro punto crucial, y dado que hubo un proceso de externalización de pensiones en España hace muchos años, sería poder identificar de alguna manera el colectivo, dado que generalmente están en coaseguro en varias entidades de seguros y no queremos sobre ponderar el mismo resultado varias veces.
- > Otro factor a analizar es la **consideración del número de años a tener en cuenta** en el estudio de la experiencia de longevidad. Creo que cinco años para derivar un estudio de longevidad base es el mínimo indispensable si existe una experiencia creíble. Sin embargo, 8 años de experiencia creíble sería ideal. Por lo tanto, agregar más años ayudaría a la robustez del cálculo del vector de la **mortalidad base**. De entre todos los datos, aquellos que son más

recientes son más relevantes y deberían utilizarse siempre que sea posible. De hecho, el año base de la tabla de longevidad, resulta de los años utilizados para el estudio. Para la derivación del **factor de mejora de la mortalidad**, la profundidad histórica de datos debiera ser mayor, idealmente sobre los 30 años. Una cifra menor de años seguiría siendo factible, pero no sería totalmente suficiente para desarrollar una hipótesis desde cero. Como esta profundidad histórica de población asegurada no suele existir, generalmente la práctica habitual es derivar el factor de mejora de la mortalidad con datos de población asegurada (utilizando datos del HMDB que para España se basa en datos del Instituto Nacional de Estadística) y realizar ajustes al mismo para adaptarlos a población asegurada basados en juicio experto, tanto médicos como actuariales.

Un problema muy común del uso de datos de población general para inferirla a la población asegurada es que tanto la tabla base como los factores de mejora de la mortalidad son distintos y puede conllevar un riesgo base

- > En muchas tablas de longevidad se ponderan los datos de fallecidos por la **suma asegurada**, dada que esta podría considerarse como un factor de valoración de la situación socioeconómica del asegurado, y la conclusión sería que una clase socioeconómica superior supone menores tasas de mortalidad. Este dato es generalmente difícil de obtener de forma homogénea entre las entidades aseguradoras y generalmente se suele ponderar por número de pólizas.
- > Para **edades bajas y altas**, salvo que sea una cartera con muchas rentas diferidas (no es el caso en España), no es preocupante que no haya muchos datos por debajo de la edad de 60 años porque no suele haber muchas rentas pagaderas a esa edad. De forma adicional, los tantos de mortalidad que puedan obtenerse de estas edades pueden ser anómalas o inclu-

so volátiles. De tal forma que lo mejor es que se extrapolen. Lo mismo aplicaría para edades superiores a los 95 años, donde puede que no haya ni datos para alguna edad y se deba extrapolar e inferir resultados.

Esta es una propuesta base e inicial de los datos que debieran ser utilizados para la construcción de un repositorio de datos estándar para la creación de una tabla base de longevidad, lo cual no implica que puedan y deban requerirse más datos

- > En España no se recogen datos de mortalidad por **código postal**, sin embargo, en el Reino Unido y en Holanda se hacen estudios de mortalidad por código postal. La experiencia de mortalidad puede diferir materialmente por zona y este campo se está mostrando muy útil en estos países. Este dato recoge de forma indirecta, el nivel socioeconómico del rentista como también cualquier causa adicional de mortalidad que pueda existir, como la proximidad a zonas industriales o de zonas intensas de tráfico o de alta contaminación.

Hay información que se utiliza en la construcción de esta tabla de longevidad que no entraría dentro de la recopilación de los datos estándar y que me limito simplemente a mencionar, dado que creo que es importante tenerla a mano para la derivación y/o **ajuste de los factores de mejora de la mortalidad:** muertes por accidente de tráfico, SIDA, Covid, prevalencia del tabaquismo por edad y género, datos generales de nivel socio económico. Todos estos factores tienen un impacto en la mortalidad en el medio y largo plazo y han de ser tenidos en cuenta para el ajuste de los factores de mejora de la mortalidad.

Por último, quiero destacar que esta es una **propuesta base e inicial** de los datos que debieran ser utilizados para la construcción de un repositorio de datos estándar para la creación de una tabla base de longevidad, lo cual no implica que puedan y deban requerirse más datos. Pienso que esta propuesta de creación de un **repositorio de datos estándar**, con datos de cada entidad aseguradora para compartir, proporcionaría muchos datos comparables a las autoridades u organismos que quieren realizar una actualización de las tablas de longevidad de forma periódica. También se conseguiría una actualización más sencilla y rápida de las tablas. De forma adicional, solucionarían gran parte de los problemas de calidad del dato y de trazabilidad del mismo. Los datos deben tener siempre los mismos campos y formato, o al menos mantener los campos originales (para comparar los datos en los años siguientes y ser consistentes a lo largo del tiempo). ●

ANEXO – Datos inicialmente propuestos para recabar datos

Un listado muy preliminar de petición de datos podría ser el siguiente:

- Id interno (no el de póliza)
- Género: mujer/varón.
- Fecha de nacimiento.
- Asegurado principal o segunda cabeza.
- % pago a la segunda cabeza como % de la renta de la primera cabeza.
- Cantidad de renta pagadera (anualmente): especificar si son 12 o 14 pagas y las fechas de los pagos extraordinarios.
- Fecha comienzo pago renta.
- Fecha final pago renta.
- Causa final pago renta: fallecimiento/vencimiento de renta.
- Tipología incremento renta: Constante/geométrica/aritmética.
- Incremento renta: 0/% geom/ cantidad aritmética.
- Fecha de extracción de la información; importante para saber el año calendario de la información.
- Fecha de fallecimiento del rentista.
- Póliza individual/Colectivo.
- Si es Colectivo indicar código de industria CNAE de la compañía del trabajador.
- Si es Colectivo indicar código CNO.
- Renta en período de diferimiento o en período de pago: diferimiento/pago.
- Código postal del rentista.