Análisis de secuencias en el estudio del envejecimiento *

Gilbert Brenes-Camacho♦

Resumen

El análisis de secuencias busca establecer cuáles son los arreglos ordenados más frecuentes, en una cadena de eventos. El análisis de secuencias se podría utilizar para investigar eventos del curso de vida de las personas. Según la teoría del "wear-and-tear", el envejecimiento biológico se da a partir de una acumulación de problemas de salud que deterioran el estado físico y mental de los adultos mayores. Se plantea como objetivo utilizar el análisis de secuencias para investigar los procesos de morbilidad entre poblaciones de adultos mayores latinoamericanos, mediante el análisis de secuencias, con base en datos de dos encuestas de envejecimiento: CRELES (Costa Rica) y PREHCO (Puerto Rico). Las secuencias de enfermedades se determinan en la primera ronda con las preguntas retrospectivas sobre edad al diagnóstico de 7 enfermedades: hipertensión, hipercolesterolemia, diabetes, cáncer, enfermedades pulmonares, infartos y derrames y se operacionalizan como variables dicotómicas en regresiones de Cox. Se encontró que el infarto y el derrame aceleran la mortalidad independientemente de si están en una secuencia o no. El riesgo de la diabetes crece cuando la enfermedad se complica con hipertensión. Secuencias de más de 3 enfermedades son raras, y los modelos no tienen suficiente potencia para detectar sus efectos.

---

* Trabajo presentado en el IV Congreso de la Asociación Latinoamericana de Población, ALAP, realizado en La Habana –Cuba, del 17 al 19 de Octubre de 2010.

♦ Centro Centroamericano de Población, Universidad de Costa Rica

Introducción

La teoría del "wear-and-tear" es una de las teorías más utilizadas para explicar el proceso de envejecimiento biológico (Gázquez-Linares et al., 2005). Según esta, el envejecimiento biológico se da a partir de una acumulación de problemas de salud que deterioran el estado físico y mental de los adultos mayores. Si hay morbilidades que son a su vez factores de riesgo para otras enfermedades, el proceso del "wear-and-tear" se podría investigar con análisis de secuencias. El análisis de secuencias es uno de los nuevos avances de la estadística, se puede citar el análisis de secuencias. El análisis de secuencias busca establecer cuáles son los arreglos ordenados más frecuentes, en una cadena de eventos. Se ha utilizado en investigaciones sobre genomas y en investigaciones de procesos en ingeniería. El análisis de secuencias también se podría utilizar para investigar eventos en el curso de vida de las personas, por ejemplo, la serie de morbilidades que enfrenta una persona durante un período de tiempo.

Definición del problema y objetivo general

El objetivo del presente artículo es utilizar el análisis de secuencias para investigar los procesos de morbilidad y los mecanismos del "wear-and-tear" entre poblaciones de adultos mayores latinoamericanos.

Teoría del "wear-and-tear"

La teoría del "wear-and-tear", conocida en castellano como la "teoría del deterioro orgánico" ó "teoría del uso y desgaste", aduce que el envejecimiento es producido por el progresivo deterioro

de los órganos corporales al no poder responder al proceso de homeostasis, o sea, al proceso de recuperarse de situaciones perjudiciales para la salud que genera el medio ambiente (Gázquez Linares et al., 2005). En otras palabras, esta teoría implica que el organismo del ser humano, cuando joven, tiene la capacidad de recuperarse de enfermedades mediante el proceso de homeostasis, pero a medida que envejece, su organismo no puede recuperarse tan bien y empieza a acumular condiciones no adecuadas de salud. La "teoría del uso y desgaste" forma parte de un conjunto de teorías denominadas Teorías biológicas secundarias (Gázquez Linares et al., 2005). Todas ellas se refieren al daño celular provocado por la exposición a patologías. Estas otras teorías son complementarias del "wear-and-tear", en el sentido de que se centran en la acumulación de estas condiciones no saludables, que a su vez causan el envejecimiento. La teoría del "wear-and-tear" también está relacionada con el concepto de carga alostática. La carga alostática (medida a partir de biomarcadores) refleja la acumulación de respuestas del organismo a las condiciones del medio-ambiente. Las personas de mayor edad han estado expuestas en promedio a una mayor cantidad de condiciones perjudiciales para su salud. A la vez, cuando las respuestas regulatorias a la interacción individuo-ambiente exceden los rangos óptimos de operación, el proceso de envejecimiento avanza. La acumulación de las desregulaciones fisiológicas estaría entonces medida por la carga alostática (Seeman et al., 2004).

El análisis de secuencias no mide directamente todos los posibles mecanismos del envejecimiento que la teoría delinea, pero sí se podría esperar que la acumulación de enfermedades esté altamente asociada con este proceso.

Análisis de secuencias

El análisis exploratorio de secuencias (o de datos secuenciales) es definido por Sanderson y Fisher (1994) como una actividad empírica para analizar datos (computacionales, ambientales, o de comportamiento) en el que la integridad secuencial de los eventos o las unidades ha sido preservada (p. 255). El conjunto de técnicas dentro del análisis secuencial está encaminado a una etapa exploratoria, más que inferencial o de establecimiento de causalidad.

El análisis de secuencias ha sido ampliamente usado en investigaciones genómicas, para describir y relacionar secuencias de Ácido Dexorribonucleico (ADN). En bioestadística, este tipo de técnicas ha sido usada (a) para determinar distintos niveles de composiciones de proteínas en el DNA que determinen la organización genética de seres vivos, (b) para en el análisis de variaciones dentro y entre poblaciones, o (c) en la comparación de una secuencia con otras para determinar las similitudes que indiquen una relación evolucionaria ó funcional (Curnow and Kirkwood, 1989).

Como se puede ver, el análisis de secuencias no es ni un modelo ni una técnica como tal, sino un conjunto de herramientas matemáticas, estadísticas y computacionales para comprender mejor datos planteados en secuencias. Un análisis inicial de este tipo de datos recomendaría: (a) una descripción de las secuencias, sus longitudes y sus componentes; (b) una visualización de las secuencias para detectar similitudes; (c) comparación de secuencias vía medidas de distancias; (d) el agrupamiento de secuencias similares vía análisis de conglomerados o escalamiento multidimensional; y (e) la utilización de las secuencias como variables dependientes o independientes en otro tipo de análisis (Brzinsky-Fay et al., 2004). Análisis de secuencias más complicados visualizan las secuencias como una matriz de frecuencias de elementos, las cuales pueden ser resumidas con valores característicos (van Heel, 1991). En el análisis de este artículo

no recurriremos a métodos multivariados, sino que se limitará a la parte descriptiva, para usarla en modelos de regresión.

Fuentes de información.

En el análisis se utiliza los datos de dos encuestas de envejecimiento en América Latina: CRELES (Costa Rica: Estudio sobre Longevidad y Envejecimiento Saludable) y PREHCO (Condición de Salud de los Adultos de Edad Mayor en Puerto Rico).

La población de estudio de CRELES se refiere a los habitantes de Costa Rica de 60 años ó más en el período 2004-2006. El diseño de la muestra es probabilístico bietápico y estratificado. En la primera etapa, se seleccionaron aleatoriamente 9600 individuos de 55 años y más según el Censo de Población de Costa Rica de 2000; cada grupo quinquenal de edad constituyó un estrato, cuyas fracciones de muestreo variaban entre 1.1% y 100% entre las personas nacidas antes de 1905. Para la segunda etapa, se agruparon a los individuos seleccionados según las áreas oficiales del Ministerio de Salud, y se seleccionó una muestra de 3300 individuos. La muestra final es de 2820 personas. El trabajo de campo toma dos años porque sólo hay dos equipos de trabajo constituidos cada uno por un conductor, un técnico de laboratorio (flebotomista) y dos entrevistadores; en promedio se efectúan en total cerca de 30 entrevistas por semana. El proyecto capta información de un cuestionario estructurado, pero también se toman muestras de sangre y orina en ayunas, se le toman medidas antropométricas a los entrevistados, y se les ponen a hacer ciertas pruebas de resistencia (como pruebas de respiración y repeticiones de sentarse y ponerse de pie). El procedimiento rutinario es solicitar el consentimiento personal y efectuar la entrevista un día, y a la siguiente mañana tomar las muestras de sangre por venipuntura y, si no se les ha hecho, llevar a cabo el módulo de antropometría. El consentimiento informado, así como la entrevista y el protocolo de análisis fueron aprobados por el Comité de Ética de la Universidad de

Costa Rica. El financiamiento se obtuvo de la Fundación Wellcome Trust. La información de mortalidad se obtuvo de un pareo automatizado con el Registro de Defunciones de Costa Rica, organizado por el Registro Civil, una entidad pública perteneciente al Tribunal Supremo de Elecciones.

En cuanto a PREHCO, la encuesta portorriqueña de envejecimiento, se cuenta con la primera ronda. Esta se refiere a la población no institucionalizada de 60 años y más (y cónyuges sobrevivientes al momento de la encuesta) y que residía en Puerto Rico para Junio de 2000. La muestra es polietápica y estratificada, con sobremuestreos en áreas con alto porcentaje de afrodescendientes y en edades mayores a 80 años. Las entrevistas fueron personales ("cara a cara"), llevadas a cabo entre el 2002 y el 2003, y auxiliadas con computadoras portátiles; la recolección de datos incluyó mediciones antropométricas y de capacidad física para adultos mayores. El tamaño total de la muestra fue de 4293 entrevistados, de los cuales 12.4% requirieron un informante proxy para obtener información. La tasa de no respuesta fue de 6.1%, incluyendo 4.7% de rechazos (Palloni et al., 2005). La información sobre defunciones se extrajo de la información de la segunda ronda, conducida entre el 2006 y el 2007.

Metodología

Las secuencias se determinan a partir de las respuestas a las preguntas sobre diagnóstico de enfermedades por parte de un médico y sobre la edad (o año) del diagnóstico. En ambos países, se utilizó información sobre hipertensión arterial, diabetes mellitus, cáncer (excepto cáncer de piel), enfermedad pulmonar, infarto, y derrame cerebral. La redacción de las preguntas en ambas encuestas es prácticamente idéntica. Adicionalmente, para Costa Rica, se incluyó también información sobre hipercolesterolemia.

Para el análisis de secuencias, se utilizaron las siguientes reglas:

a) La secuencia fue definida como un arreglo ordenado de las enfermedades que han sido diagnosticadas.

b) Se creó una variable dicotómica ("dummy") para cada secuencia, donde el 1 es la existencia de la secuencia.

c) Cada secuencia está definida por todas las enfermedades que han sido diagnosticadas a un adulto mayor. Esto quiere decir que si un adulto mayor ha tenido las enfermedades A, B y C en este orden, la secuencia queda incluida como ABC (secuencia de longitud 3), pero no queda incluida en las secuencias que representan subconjuntos de la secuencia más larga (o sea, no se incluye ni en AB, ni en AC, ni en BC).

d) Si dos enfermedades A y B fueron diagnosticadas en el mismo año o a la misma edad, el caso fue incluido como uno, tanto en la secuencia AB como en la secuencia BA. En otras palabras, las secuencias de la misma longitud no representan categorías mutuamente excluyentes.

Después de describir las secuencias, se utilizan las variables dicotómicas en un modelo de regresión de riesgos proporcionales de Cox para determinar la asociación de cada secuencia con la mortalidad, como una forma de evaluar la teoría del "wear-and-tear" del envejecimiento biológico. Una regresión de Cox se puede definir de la siguiente forma:

$$h(t|X_i) = h(0) * \exp\left[\alpha + \beta X_i + \varepsilon_i\right] \qquad (3)$$

donde:

$h(t \mid X_i)$ el riesgo en el momento t, condicional a las variables independientes.

$h(0)$ La función de riesgo o hazard en la línea de base

X_i El vector de covariables para cada individuo i.

α,β El vector of coeficientes para las variables de control.

ε_i Término de error.

Este modelo con secuencias se estima después de estimar un modelo previo en que cada enfermedad (sin clasificar dentro de las secuencias) se incorpora como una variable dicotómica. Compara este modelo con el modelo con secuencias permite analizar qué se gana si se analizan secuencias en lugar de analizar las enfermedades como efectos independientes.

En ambos modelos, se controlan por otras variables explicativas: sexo, edad, residencia metropolitana, residencia urbana, y nivel de educación. Edad se incorporó como variable continua. En CRELES, la residencia metropolitana se definió como las personas que habitan en el Gran Área Metropolitana de Costa Rica. En PREHCO, la residencia metropolitana es residir en San Juan. En CRELES, residencia urbana se refiere a vivir en la GAM o en cabeceras de provincia (Alajuela, Cartago, Heredia, Liberia, Puntarenas, o Limón) o en otras ciudades regionales (Guápiles, San Isidro del General, Ciudad Quesada). En Puerto Rico, residencia urbana se define como residir en las cabeceras de San Juan, Ponce, Mayagüez y Arecibo. En ambas encuestas, educación se refiere a haber aprobado al menos 6 años de educación.

Resultados

Para analizar los resultados, recuérdese que en PREHCO no se analizó colesterol pues no se hacía esa pregunta. Por esta razón es que el máximo de secuencias es menor en PREHCO que en CRELES: 5 vs. 6 (Cuadro 1). Esta distribución de frecuencias sugiere además que en ninguno de

los dos países se encontró la presencia de todas las enfermedades en un solo entrevistado. En CRELES, se tenían 7 enfermedades, pero la longitud máxima de secuencias fue 6; en PREHCO, eran 6 enfermedades, mientras que la longitud máxima fue 5. Además, en ambos países, menos del 4% de los entrevistados presentaban más de 3 morbilidades al mismo tiempo.

Dado lo numeroso en la cantidad de secuencias, para cada país se escogieron aquellas secuencias con frecuencias absolutas de 15 ó más (Cuadro 2): o sea, combinaciones de enfermedades que las padecieran al menos 15 personas en la muestra. En Costa Rica, de las 7 secuencias, 6 están relacionadas con el síndrome metabólico: hipertensión, hipercolesterolemia y diabetes. Adicionalmente, está la enfermedad pulmonar, como padecimiento único (4.7%), como enfermedad que sigue a la diabetes (1.7%) ó como precedente de la hipertensión y la hipercolesterolemia (1.6%). Sin embargo, y tal como lo expresa también el Cuadro 1, una tercera parte de la muestra presenta una sola enfermedad.

En la muestra portorriqueña (Cuadro 3), prácticamente el 40% de los entrevistados presentaba una sola condición de morbilidad, un porcentaje mayor que en Costa Rica. Sin embargo, de las diez primeras secuencias con mayor frecuencia, sólo 4 reflejaban una sola enfermedad: hipertensión, diabetes, pulmonar y cáncer. En Puerto Rico, son altamente frecuentes las combinaciones de hipertensión con diabetes por un lado y con infarto (indiferentemente del orden). En general, entre las 15 secuencias más frecuentes en Puerto Rico, 6 de ellas se podrían entender como complicaciones de la hipertensión (hipertensión-diabetes, hipertensión-infarto, hipertensión-pulmonar, hipertensión-derrame, hipertensión-diabetes-infarto y diabetes-hipertensión-infarto).

En contraste, en ambos países, la secuencias menos frecuentes incluyen enfermedades de baja prevalencia (tanto por su baja incidencia, como por su alta letalidad): cáncer, infarto y accidentes cerebrovasculares (derrames).

Una dificultad de analizar secuencias en este tipo de encuestas es que algunas de las enfermedades pudieron ser diagnosticadas simultáneamente. Esta es la razón por la que las categorías de los cuadros 1, 2 y 3 no son mutuamente excluyentes. Este problema incide en la diferenciación de dos secuencias que tienen los mismos componentes. En el caso de Costa Rica, dos de las secuencias más comunes fueron Diabetes-Colesterol (5.8%) y Colesterol-Diabetes (3.1%). Es claro que la primera secuencia es más frecuente que la segunda, pero 61 casos (casi tres cuartas partes de la secuencia colesterol-diabetes) son casos en los que ambas enfermedades fueron diagnosticadas el mismo año (posiblemente al mismo tiempo). Algo similar ocurre en Puerto Rico, de los 330 casos de la secuencia diabetes-hipertensión (o de los 326 casos de Hipertensión-Diabetes), más de la mitad (169) son casos en los que ambas enfermedades fueron diagnosticadas simultáneamente. En forma similar, de los 72 casos de infarto-hipertensión, tres cuartas partes (55) se refieren a diagnóstico de infarto e hipertensión al mismo tiempo. El análisis tradicional de secuencias no toma en cuenta la simultaneidad, por lo que este tipo de casos genera más dificultades en el análisis.

Ahora bien, una de las formas de analizar la teoría del “wear-and-tear” del envejecimiento, es determinar el impacto diferenciado de algunas secuencias sobre la mortalidad. Antes de estimar modelos de regresión para cada país con las secuencias como variables independientes, se estimaron sendos modelos en los cuales las variables independientes no son las secuencias, sino variables dicotómicas que se refieren a la presencia o no de cada una de las enfermedades por analizar. Estos primeros dos modelos son más parsimoniosos, una característica deseable en

modelos de regresión. La comparación entre estos modelos y los modelos con las secuencias como covariables permiten ver cuánto mejora el poder explicativo del modelo al incorporar las secuencias: covariables que reflejen más detalle.

Según el Cuadro 4, el modelo de riesgos proporcionales de Cox para la mortalidad en Costa Rica muestra que el tener cualquiera de las enfermedades aumenta el riesgo de morir, aunque los riesgos relativos para cáncer y para infarto no son significativos al 10% (probablemente, por falta de potencia estadística). El hazard ratio de la hipertensión es prácticamente 1, lo cual implica nula asociación. La única excepción clara es el tener colesterol, pues disminuye significativamente los riesgos de muerte. Esta contradicción, así como el efecto de la hipertensión, se discutirá más adelante con las secuencias. En el caso de Puerto Rico, la diabetes, el cáncer, el infarto y el derrame aumentan los riesgos de morir. El riesgo relativo de la enfermedad de pulmón no es significativo pero es mayor que uno, lo cual también sugiere problemas de falta de potencia estadística. Por último, la razón de hazards de hipertensión es menor a uno, pero no es significativa. En ambos países, los coeficientes de las variables de control tienen las direcciones esperadas.

Los modelos de regresión con las secuencias generan más información. El riesgo relativo de muerte para las personas que únicamente tienen hipercolesterolemia es significativamente menor a uno: 0.53. Sin embargo, si la hipercolesterolemia precede al accidente cerebrovascular (derrame), o es precedida por la hipertensión y seguida por la Diabetes, los riesgos de morir se multiplican por más de 4. Esto sugiere que, entre los adultos mayores costarricenses, la letalidad del colesterol alto tiene que ver más con la comorbilidad que con la presencia del colesterol por sí solo. Si un adulto mayor tiene únicamente un diagnóstico de colesterol, sus riesgos de morir disminuyen. Adicionalmente, llama la atención que la presencia del infarto por sí solo triplica el

riesgo de muerte, mientras que el infarto en secuencia no tiene algún efecto significativo. Esto no quiere decir que la comorbilidad disminuya el efecto letal del infarto. Más bien sugiere que las personas que mueren después de un infarto no han recibido diagnósticos previos, probablemente por inadecuadas prácticas preventivas. Por último, las secuencias "Cáncer-Diabetes" e "Hipertensión-Colesterol-Diabetes" también tienen mayores riesgos significativos de defunción. Si se toma en cuenta también el efecto de la secuencia Derrame-Diabetes, entonces esto sugiere que la Diabetes como complicación de una enfermedad previa tiene serias consecuencias para la mortalidad.

En Puerto Rico, varias secuencias aumentan el riesgo de muerte. Entre ellas, las secuencias "Diabetes-Cáncer", "Infarto-Diabetes", "Diabetes-Infarto-Derrame", "Diabetes-Hipertensión" y "Diabetes-Hipertensión-Derrame" sugieren que la diabetes mellitus en comorbilidad presenta una alta letalidad. Ahora bien, la hipertensión aumenta el riesgo de muerte no sólo como complicación de la diabetes, sino también del infarto. Sin embargo, a los portorriqueños que tienen diagnóstico de hipertensión únicamente, su riesgo de muerte se disminuye en un 36% (con respecto de los que no tienen ninguna de las enfermedades incluidas en el análisis). Además, las secuencias "Hipertensión-Infarto-Diabetes" e Hipertensión-Derrame-Diabetes" también disminuyen considerablemente el riesgo de morir. Estos resultados contradictorios se pueden deber a la fuerte colinealidad entre las secuencias y que fue mencionada anteriormente. Recuérdese que esta colinealidad se puede deber al diagnóstico simultáneo de dos ó más enfermedades. En otras palabras, una proporción importante de entrevistados puede estar a la vez en dos ó más de las siguientes secuencias: Hipertensión-Diabetes-Infarto (RR=2.29, no significativa), Hipertensión-Infarto-Diabetes (RR=0.19, $p<0.10$), Diabetes-Hipertensión-Infarto (RR=1.65, no significativa), Diabetes-Infarto-Hipertensión (RR=1.31, no significativa), Infarto-Hipertensión-Diabetes (RR=1.00, no significativa) e Infarto-Diabetes-Hipertensión (RR=0.68, no

significativa). Para ver el efecto conjunto de estas secuencias, se estima una prueba de hipótesis en el que la hipótesis nula es que la suma de los coeficientes de todas estas secuencias es igual a cero. El coeficiente estimado es 0.23, lo que implica un riesgo relativo de muerte de 1.26. El coeficiente de esta suma de secuencias no es significativo al 10%. Este resultado nuevamente sugiere que el que la secuencia Hipertensión-Infarto-Diabetes disminuya los riesgos de muerte se puede deber a la multicolinealidad. En general, los resultados del modelo de riesgos proporcionales de Cox para los adultos mayores portorriqueños muestra que las complicaciones de la hipertensión y, sobre todo, de la diabetes mellitus son las que producen una alta mortalidad en la isla.

Discusión.

La metodología del análisis de secuencias permite diferenciar distintos procesos de morbilidad. En Costa Rica, se observó que la hipercolesterolemia por sí sola no es tan letal, sino en comorbilidad con otras condiciones, y en Puerto Rico, se observó que las complicaciones de la diabetes mellitus y en menor medida de la hipertensión también aumentan los riesgos de muerte. En envejecimiento, la teoría del "wear-and-tear" describe cómo la acumulación de condiciones poco saludables son las que llevan al envejecimiento biológico. Esta acumulación se da secuencialmente. Ciertas enfermedades como la hipertensión o la hipercolesterolemia son enfermedades por sí mismas, pero también son factores de riego de los infartos o los accidentes cerebovasculares.

El análisis de secuencias permitió detectar estas secuencias las cuales están condicionadas por las prevalencias generales de cada enfermedad. Esto quiere decir que se encontraron más secuencias compuestas por enfermedades altamente frecuentes entre los adultos mayores, como la hipertensión y la diabetes, y pocas secuencias con los infartos, los derrames, o el cáncer.

Además, en los modelos de Cox, se encontraron varias secuencias con efectos significativos sobre la mortalidad, y estas secuencias están representadas por comorbilidades relacionadas con enfermedades cardiovasculares y, particularmente, el síndrome metabólico. Por el contrario, pocas secuencias en las que aparecía la enfermedad pulmonar tenían riesgos relativos significativamente distintos de uno. Este hallazgo era esperable pues, biológicamente, las enfermedades pulmonares no necesariamente entran en interacción con las otras enfermedades de alta prevalencia analizadas en este artículo.

El análisis de secuencias para encuestas de envejecimiento sí presenta una limitación importante. En este tipo de estudios, generalmente se pregunta por la edad o el año del diagnóstico de una enfermedad. Este tipo de variable puede contener un fuerte sesgo de medición, debido al sesgo de recordación por parte del entrevistado. En otras palabras, es posible que un número indeterminado de entrevistados respondan una cifra aproximada en lugar de exacta. Este error puede generar que entrevistados respondan que dos ó más condiciones sean diagnosticadas simultáneamente. La misma restricción de la información hace que enfermedades que hayan sido diagnosticas en el mismo año, pero con meses de diferencia en el diagnóstico sean consideradas como de diagnóstico simultáneo.

La simultaneidad en los diagnósticos es un problema en el análisis de secuencias, pues este supone que los elementos en una secuencia están en un orden y único. Las secuencias deben ser mutuamente excluyentes, pero en este análisis, las secuencias no lo eran por la simultaneidad. Esto generó problemas de multicolinealidad cuando las secuencias fueron utilizadas como covariables en los modelos de regresión de Cox.

Un análisis de secuencias para morbilidades puede ser idóneo cuando se cuentan con bases de datos de registros médicos, en los cuales las fechas de diagnóstico están establecidas con menor

error de medición. Dichas bases de datos aún son difíciles de conseguir en los países latinoamericanos, ya sea porque no existen o porque no son suministradas para investigación por parte de las instituciones que las producen (el caso de Costa Rica con la CCSS).

En resumen, el análisis de secuencias puede ser útil para estudiar la teoría del "wear-and-tear" del envejecimiento biológico, pero se requieren instrumentos de recolección de información más detallados o bien muestras más grandes para que el problema de simultaneidad no genere multicolinealidad.

Cuadros y Gráficos

Cuadro 1. Longitud de secuencias de enfermedades en CRELES (Costa Rica) y PREHCO (Puerto Rico). (Las categorías no son mutuamente excluyentes)

Longitud de secuencias	CRELES		PREHCO	
	F. absoluta	F. relativa (%)	F. absoluta	F. relativa (%)
TOTAL	**2573**	**100.0**	**3915**	**100.0**
0	757	29.4	1186	30.3
1	911	35.4	1608	41.1
2	650	25.3	945	24.1
3	317	12.3	284	7.3
4	82	3.2	70	1.8
5	13	0.5	10	0.3
6	3	0.1	-	-

Cuadro 2. Secuencias de enfermedades detectadas con frecuencias absolutas mayores a 15 en CRELES (Costa Rica).

(Las categorías no son mutuamente excluyentes)

Secuencias de enfermedades	F. absol	F. relat (%)	Secuencias de enfermedades	F. absol	F. relat (%)
TOTAL	2827	100.0			
Hipertensión	397	14.0	Colesterol-Cáncer	26	0.9
Colesterol	192	6.8	Colesterol-Infarto	26	0.9
Diabetes-Colesterol	165	5.8	Cáncer-Diabetes	26	0.9
Pulmonar	134	4.7	Hipertensión-Colesterol-Pulmonar	26	0.9
Colesterol-Diabetes	87	3.1	Derrame	25	0.9
Diabetes	77	2.7	Colesterol-Pulmonar	23	0.8
Diabetes-Pulmonar	48	1.7	Diabetes-Derrame	23	0.8
Cáncer	47	1.7	Infarto	22	0.8
Pulmonar-Hipertensión-Colesterol	46	1.6	Diabetes-Cáncer	22	0.8
Cáncer-Colesterol	44	1.6	Cáncer-Pulmonar	21	0.7
Pulmonar-Colesterol	42	1.5	Diabetes-Hipertensión-Colesterol	21	0.7
Infarto-Colesterol	38	1.3	Pulmonar-Colesterol-Hipertensión	20	0.7
Diabetes-Infarto	31	1.1	Pulmonar-Infarto	19	0.7
Hipertensión-Diabetes-Colesterol	31	1.1	Pulmonar-Hipertensión-Diabetes	19	0.7
Colesterol-Derrame	30	1.1	Infarto-Derrame	16	0.6
Pulmonar-Cáncer	30	1.1	Derrame-Infarto	16	0.6
Derrame-Colesterol	30	1.1	Colesterol-Derrame	15	0.5
Hipertensión-Pulmonar-Colesterol	30	1.1	Pulmonar-Derrame	15	0.5
Pulmonar-Diabetes-Colesterol	29	1.0	Derrame-Diabetes	15	0.5
Pulmonar-Diabetes	28	1.0	Derrame-Pulmonar	15	0.5
			Hipertensión-Colesterol-Diabetes	15	0.5

Cuadro 3. Secuencias de enfermedades detectadas con frecuencias absolutas mayores a 15 en PREHCO (Puerto Rico) (Las categorías no son mutuamente excluyentes)

Secuencias de enfermedades	F. absol.	F. relat (%)	Secuencias de enfermedades	F. absol.	F. relat (%)
TOTAL	4228	100.0			
Hipertensión	1132	26.8	Derrame	27	0.6
Diabetes-Hipertensión	330	7.8	Diabetes-Infarto-Derrame	26	0.6
Hipertensión-Diabetes	326	7.7	Cáncer-Infarto	23	0.5
Diabetes	258	6.1	Infarto-Pulmonar	23	0.5
Hipertensión-Infarto	126	3.0	Hipertensión-Diabetes-Pulmonar	23	0.5
Pulmonar-Hipertensión	93	2.2	Diabetes-Derrame-Infarto	23	0.5
Hipertensión-Cáncer	86	2.0	Pulmonar-Derrame	22	0.5
Pulmonar	76	1.8	Hipertensión-Pulmonar-Infarto	20	0.5
Infarto-Hipertensión	72	1.7	Infarto-Diabetes-Hipertensión	20	0.5
Cáncer	71	1.7	Cáncer-Diabetes	19	0.4
Hipertensión-Pulmonar	66	1.6	Hipertensión-Derrame-Diabetes	19	0.4
Hipertensión-Derrame	62	1.5	Infarto-Hipertensión-Diabetes	19	0.4
Hipertensión-Diabetes-Infarto	55	1.3	Diabetes-Derrame	18	0.4
Diabetes-Hipertensión-Infarto	55	1.3	Cáncer-Pulmonar	18	0.4
Infarto	53	1.3	Infarto-Derrame	18	0.4
Cáncer-Hipertensión	48	1.1	Cáncer-Hipertensión-Diabetes	18	0.4
Hipertensión-Infarto-Derrame	45	1.1	Diabetes-Infarto	17	0.4
Hipertensión-Infarto-Diabetes	43	1.0	Diabetes-Hipertensión-Cáncer	17	0.4
Derrame-Hipertensión	41	1.0	Cáncer-Derrame	16	0.4
Diabetes-Hipertensión-Derrame	39	0.9	Pulmonar-Infarto	16	0.4
Pulmonar-Diabetes	36	0.9	Infarto-Diabetes	16	0.4
Hipertensión-Derrame-Hipertensión	36	0.9	Diabetes-Hipertensión-Pulmonar	16	0.4
Diabetes-Infarto-Hipertensión	36	0.9	Derrame-Infarto-Hipertensión	16	0.4
Diabetes-Cáncer	33	0.8			
Hipertensión-Diabetes-Derrame	33	0.8			

Notas: *: $p<0.10$, **: $p<0/05$

Cuadro 4. Riesgos relativos de mortalidad en regresiones de Cox para enfermedades (sin determinar secuencias) y otras covariables en CRELES (Costa Rica) y ENASEM (México)

Covariables	CRELES		MHAS	
	Riesgo relativo	p-value	Riesgo relativo	p-value
Edad	1.08	**	1.09	**
Urbano	0.97		1.17	
Metrop.	0.96		0.74	**
Mujer	0.84		0.78	*
Educación>=6 años	1.15		0.94	
Hipertensión	0.98		0.86	
Colesterol	0.76	**	-	-
Diabetes	1.60	**	1.58	**
Cáncer	1.85	**	1.85	**
Pulmón	1.26		1.40	
Infarto	1.44		1.55	**
Derrame	2.10	**	1.69	**

Notas: *: p<0.10, **: p<0/05

Cuadro 5. Riesgos relativos de mortalidad en regresiones de Cox para secuencias de enfermedades detectadas con frecuencias absolutas mayores a 15 y otras covariables en CRELES (Costa Rica).

Covariables	RR	p-value	Covariables	RR	p-value
Edad	1.09	**	Pulmonar-Colesterol	1.16	
Urbano	0.95		Pulmonar-Diabetes	1.25	
Metrop.	0.99		Pulmonar-Cáncer	1.73	
Mujer	0.81	*	Pulmonar-Infarto	0.47	
Educación>=6 años	1.09		Pulmonar-Derrame	1.95	
Hipertensión	0.87		Infarto-Colesterol	1.71	
Colesterol	0.53	**	Infarto-Derrame	2.38	
Diabetes	1.68		Derrame-Colesterol	1.09	
Cáncer	1.39		Derrame-Diabetes	3.49	**
Pulmón	1.09		Derrame-Pulmonar	1.54	
Infarto	3.14	**	Derrame-Infarto	0.55	
Derrame	1.93				
Colesterol-Derrame	4.89	*			
Colesterol-Diabetes	0.78				
Colesterol-Cáncer	1.26		Hipertensión-Colesterol-Diabetes	4.56	*
Colesterol-Pulmonar	0.30		Hipertensión-Colesterol-Pulmonar	0.79	
Colesterol-Infarto	0.33		Hipertensión-Diabetes-Colesterol	1.16	
Colesterol-Derrame	0.70		Hipertensión-Pulmonar-Colesterol	1.41	
Diabetes-Colesterol	1.16		Diabetes-Hipertensión-Colesterol	0.87	
Diabetes-Cáncer	0.91				
Diabetes-Pulmonar	1.39				
Diabetes-Infarto	1.35		Pulmonar-Hipertensión-Colesterol	0.72	
Diabetes-Derrame	1.29		Pulmonar-Hipertensión-Diabetes	0.52	
			Pulmonar-Colesterol-Hipertensión	1.54	
Cáncer-Colesterol	1.47		Pulmonar-Diabetes-Colesterol	1.03	
Cáncer-Diabetes	2.94	**			
Cáncer-Pulmonar	0.92				

Notas: *: p<0.10, **: p<0/05

Cuadro 6. Riesgos relativos de mortalidad en regresiones de Cox para secuencias de enfermedades detectadas con frecuencias absolutas mayores a 15 y otras covariables en PREHCO (Puerto Rico).

Covariables	RR	p-value	Covariables	RR	p-value
Edad	1.09	**	Pulmonar-Diabetes	2.61	
Urbano	1.13		Pulmonar-Infarto	1.64	
Metrop.	0.73	**	Pulmonar-Derrame	2.27	
Mujer	0.77	*			
Educación>=6 años	0.91		Infarto-Hipertensión	1.95	**
			Infarto-Diabetes	3.73	*
Hipertensión	0.64	**	Infarto-Pulmonar	1.77	
Diabetes	1.14		Infarto-Derrame	1.84	
Cáncer	2.11		Derrame-Hipertensión	1.27	
Pulmón	0.83		Derrame-Diabetes	0.85	
Infarto	0.91		Hipertensión-Diabetes-Pulmonar	1.04	
Derrame	2.27		Hipertensión-Diabetes-Infarto	2.29	
			Hipertensión-Diabetes-Derrame	1.14	
Hipertensión-Diabetes	0.63		Hipertensión-Pulmonar-Infarto	0.27	
Hipertensión-Cáncer	1.03		Hipertensión-Infarto-Diabetes	0.19	*
Hipertensión-Pulmonar	0.96		Hipertensión-Infarto-Derrame	2.20	*
Hipertensión-Infarto	1.20		Hipertensión-Derrame-Diabetes	0.18	**
Hipertensión-Derrame	1.24		Hipertensión-Derrame-Infarto	1.14	
Diabetes-Hipertensión	1.79	**	Diabetes-Hipertensión-Cáncer	0.34	
Diabetes-Cáncer	3.97	**	Diabetes-Hipertensión-Pulmonar	0.79	
Diabetes-Infarto	1.66		Diabetes-Hipertensión-Infarto	1.65	
Diabetes-Derrame	1.78		Diabetes-Hipertensión-Derrame	5.48	**
			Diabetes-Infarto-Hipertensión	1.31	
Cáncer-Hipertensión	1.47		Diabetes-Infarto-Derrame	2.44	*
Cáncer-Diabetes	1.89		Diabetes-Derrame-Infarto	0.19	**
Cáncer-Pulmonar	1.74				
Cáncer-Infarto	0.42		Cáncer-Hipertensión-Diabetes	0.89	
Cáncer-Derrame	0.29	*	Infarto-Hipertensión-Diabetes	1.00	
			Infarto-Diabetes-Hipertensión	0.68	
Pulmonar-Hipertensión	0.83		Derrame-Infarto-Hipertensión	1.80	

Notas: *: p<0.10, **: p<0/05

Bibliografía

Brzinsky-Fay C, Kohler U & Luniak M (2006). "Sequence analysis with STATA". The STATA Journal 6:435-460.

Curnow RN and Krikwood TBL (1989). "Statistical analysis of Deoxyribonucleic acid sequence data. A review". Journal of the Royal Statistical Society. Series A (Statistics in Society) 152:199-220.

Gázquez Linares JJ, Yuste Rossell N, Pérez Fuentes MdC (2005). "Review of biological hypotheses explaining aging". Anales de Psicología. 21:323-327.

Palloni A, McEniry M, Dávila AL, García Gurucharri A (2005). "The influence of early conditions on health status among elderly Puerto Ricans". Social Biology, Social Science Module 52:132-163.

Sanderson PM & Fisher C (1994). "Exploratory Sequential Data Analysis: Foundations". Human-Computer Interaction 9:251-317.

Seeman TE, Crimmins E, Huanga MH, Singer B, Bucur A, Gruenewaldd T, Berkmane LF, and Rueben DB (2004). "Cumulative biological risk and socio-economic differences in mortality: MacArthur Studies of Successful Aging". Social Science & Medicine 58: 1985–1997.

van Heel Marin (1991). "A new family of powerful multivariate statistical sequence analysis techniques". Journal of Molecular Biology 220:877-887.

.