



Población y Salud en Mesoamérica  
E-ISSN: 1659-0201  
[revista@ccp.ucr.ac.cr](mailto:revista@ccp.ucr.ac.cr)  
Universidad de Costa Rica  
Costa Rica

Rosero-Bixby, Luis; Antich, Daniel  
Estudio longitudinal de mortalidad de adultos costarricenses 1984-2007  
Población y Salud en Mesoamérica, vol. 7, núm. 2, enero-junio, 2010, pp. 1-24  
Universidad de Costa Rica  
San José, Costa Rica

Disponible en: <http://www.redalyc.org/articulo.oa?id=44611946004>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en [redalyc.org](http://redalyc.org)

## **Estudio longitudinal de mortalidad de adultos costarricenses 1984-2007**

### **Longitudinal mortality study of Costa Rican adults 1984-2007**

**Luis Rosero-Bixby<sup>1</sup> y Daniel Antich<sup>2</sup>**

#### **RESUMEN**

Se describen los procedimientos y se presentan algunos resultados del “estudio longitudinal de mortalidad de adultos costarricenses” (ELMAC), consistente en una muestra de cerca de 20.000 costarricenses de 30 o más años de edad del censo de población de 1984 seguidos hasta fines de 2007. El Instituto Nacional de Estadística y Censos (INEC) digitalizó los nombres de la boleta censal de los individuos en la muestra. Procesos informáticos depuraron los nombres y los enlazaron al número único de identificación –el número de cédula– del Registro Civil. El empate se efectuó en el 87% de la muestra. Los individuos identificados se siguieron en las bases de datos del Registro Civil para establecer su sobrevida. Se identificaron más de 5.000 defunciones entre el censo de 1984 y diciembre de 2007. Procesos informáticos adicionales enlazaron 92% de estas defunciones con las de las estadísticas vitales del INEC que contienen el dato de la causa de defunción. El patrón de mortalidad de la muestra, que comprende 373.000 personas-años de observación, reproduce bien las tasas de las tablas de mortalidad del país. Las gradientes socioeconómicas de la mortalidad en esta muestra no son sensibles a posibles errores en los empates o en las imputaciones efectuadas. La muestra confirma la excepcionalmente baja mortalidad de los adultos costarricenses, especialmente de los varones. Esta muestra abre la puerta para que se efectúen variedad de análisis de los determinantes socioeconómicos de la mortalidad de adultos en Costa Rica, algo rara vez intentado en un país en desarrollo.

**Palabras clave:** Mortalidad adulta; determinantes de la mortalidad; análisis longitudinal

#### **ABSTRACT**

We describe the procedures and show some results of the “longitudinal mortality study of Costa Rican adults 1984-2007” (ELMAC), which consists of a sample of about 20,000 Costa Ricans aged 30 years or more from the 1984 population census and a follow up of their survival status to December 2007. The National Institute of Censuses and Statistics (INEC) digitized the names from the census questionnaire for the individuals in the sample. Computer programs cleaned up name's errors data and linked them to the civil registration system to have the unique id-number that all Costa Ricans have since their birth—cédula number. The linkage was possible for 87% of the sample. Those individuals linked to cédula number were followed up in the computer databases of the Civil Register to establish survival or to identify deaths. More than 5,000 deaths were found between the 1984 census and December 2007. Additional computer programs linked 92% of these deaths with those in the INEC's Vital Statistics, which are the only one containing information on the cause of death. The mortality pattern in the sample, which comprises 373,000 person-years of observation, replicated nicely the rates in preexisting life tables for Costa Rica. The SES gradients in mortality in the sample are not sensitive to the imputations and assumptions made in the linkage and follow up processes. The sample confirms the exceptionally low mortality of Costa Rican adults, especially males. This sample opens up the door for further analyses of the SES determinants of adult mortality in Costa Rica, something rarely done in a developing country.

**Keywords:** Adult mortality; mortality determinants; longitudinal analysis

**Recibido: 6 set. 2009**

**Aprobado: 5 oct. 2009**

---

<sup>1</sup> Centro Centroamericano de Población, Universidad de Costa Rica. COSTA RICA lrosero@ccp.ucr.ac.cr

<sup>2</sup> Centro Centroamericano de Población, Universidad de Costa Rica. COSTA RICA dantich@ccp.ucr.ac.cr

## 1. INTRODUCCIÓN

Esta es una descripción del proceso de creación de una muestra de adultos del Censo de Población de 1984 de Costa Rica con identificaciones que permiten su seguimiento en el registro de defunciones y de estadísticas vitales. El seguimiento longitudinal de la sobrevida de esta muestra de adultos costarricenses permite el estudio de determinantes de la mortalidad con la información socio-demográfica del censo, algo raramente intentado en países menos desarrollados. El seguimiento es posible en Costa Rica gracias a la feliz confluencia de dos circunstancias excepcionales: (1) un número único de identificación (el número de cédula) propio de cada costarricense y (2) un excelente registro computarizado de defunciones que maneja el Tribunal Supremo de Elecciones en un cuasi-registro permanente de la población costarricense.

Específicamente, este artículo describe el diseño de la muestra y el procedimiento para obtener la identificación de las personas de la boleta del censo (nombres) y su empate con las bases de datos del registro civil, incluyendo la base de datos de defunciones. Además se presenta un análisis demográfico de la muestra y sus patrones de mortalidad.

Cabe indicar que tanto los datos del censo como los de las estadísticas vitales, ambos del Instituto Nacional de Estadística y Censos (INEC), no contienen variable alguna que identifique a la persona a la cual pertenecen, en atención de la confidencialidad de la información que demanda la legislación del secreto estadístico. Los archivos de datos con identificaciones son los del Registro Civil, que administrativamente depende del Tribunal Supremo de Elecciones (TSE). Para fines científicos es necesario realizar la identificación (mediante el número de cédula que es único en Costa Rica y que se asigna al registro del nacimiento o naturalización) para enlazar la información de distintas fuentes y, por ejemplo, estudiar los factores asociados a la mortalidad. En acatamiento de la ley de estadística, fue personal del INEC el único que tuvo acceso a las boletas del censo de 1984 y a los nombres escritos en ellas. El archivo de datos resultante al final del proceso, y que servirá para los análisis, no tiene variable alguna que permita la identificación de las personas.

Bases de datos nacionales para seguir longitudinalmente la mortalidad de la población existen solamente en los pocos países desarrollados que tienen un registro completo y confiable de la población con sus características sociodemográficas. En los Estados Unidos existe desde 1983 un “estudio longitudinal de la mortalidad” (NLMS por sus siglas en inglés) (Rogot et al., 1983) que se asemeja al presente estudio en Costa Rica. Tiene como punto de partida varias encuestas de la Oficina del Censo con una línea de base de más de un millón de personas de 25 y más años de edad enlazadas al “Índice nacional de defunciones” que lleva el “*National Center for Health Statistics (NCHS)*”. El seguimiento del NLMS se inicia en 1979 y continúa hasta el presente. Esta base de datos ha sido fuente de innumerables análisis de los determinantes de la mortalidad y de la esperanza de vida (por ejemplo: Sortie, Baklund & Keller 1995).

## 2. DATOS Y PROCEDIMIENTOS

### *Diseño de la Muestra*

*Población de interés:* costarricenses por nacimiento o naturalización empadronados en el censo de 1984 y con edades 30 años o más: N= 753,363. (Notar que se excluyen los extranjeros porque al carecer de número único de identificación no es posible darles seguimiento en el registro de defunciones del TSE).

*Tamaño deseado de la muestra:* n = 20,000.

*Supuestos para seleccionar la muestra:* 4% de extranjeros que deberán descartarse y 5% de pérdida por deficiencias en la información de la boleta censal. Tamaño muestral inicial, incluyendo extranjeros: 21,900; sin extranjeros: 21,000.

*Fracción muestral:* 0.028 ó 1/35.7

*Diseño muestral:* por conglomerados en una única etapa. Se seleccionaron “segmentos censales” completos. Todos los costarricenses con 30 años y más de edad del segmento seleccionado son parte de la muestra. El segmento censal es una unidad de enumeración censal de aproximadamente 60 viviendas. El censo de 1984 tiene 10,817 segmentos censales. El promedio de adultos de 30 años y más por segmento es 69.6.

*Número de segmentos o conglomerados:* 302

*Tipo de selección:* sistemática con arranque aleatorio con los segmentos ordenados geográficamente e iso-probabilidad de selección. Luego de elegir al azar un segmento censal de arranque de entre los primeros 35, se seleccionaron sistemáticamente los siguientes con intervalo de selección de 35.69.

*Muestra resultante:*

- 302 segmentos o conglomerados
- 21,930 adultos de 30 y más años de edad
- 21,161 costarricenses.
- Muestra autoponderada

*Efecto de diseño:* No se dispone a priori de información sobre el grado de homogeneidad dentro de los conglomerados de la principal variable del estudio (mortalidad de los adultos) para estimar el efecto del diseño en ella. Sin embargo, de una manera preliminar hemos estimado un efecto de diseño de 1.98 en el riesgo de morir de los hijos tenidos por las mujeres, dato que está disponible en el censo. Es muy probable que la homogeneidad dentro del conglomerado sea menor para la mortalidad de adultos, por lo que se espera que el efecto de diseño en la estimación de la tasa de mortalidad de adultos sea menor a 1.5. Se recomienda que en el análisis de los datos de esta muestra se estimen en lo posible errores estándar corregidos por el efecto de conglomeración, para lo cual deberá utilizarse la variable identificadora del segmento.

### *Identificación de las personas en la muestra del censo*

Para identificar a las personas de la muestra del censo de 1984, fue necesario obtener sus nombres de las boletas correspondientes. Estas boletas se encuentran en el Archivo Nacional; dos terceras partes de ellas están en papel y el resto en microfilm.

Se preparó un programa para la captura de la información en el computador que permitió a los digitadores del INEC transcribir los nombres de las boletas. El programa realiza chequeos de consistencia para detectar errores de digitación. Este programa tiene las siguientes características:

- Selección del segmento: La muestra seleccionada tiene la identificación censal de los individuos elegidos (que son todos los de un determinado segmento). Las boletas del censo están archivadas por segmento. El programa le permite al digitador seleccionar el segmento que va a trabajar y le da un listado de las personas para quienes debe obtener los nombres de la boleta censal.
- Para evitar errores en la asignación del nombre a cada caso, el programa obliga al digitador a introducir el código de la variable con la relación de parentesco de la persona el cual debe coincidir con el valor correspondiente al campo de la base de datos de la muestra.
- Se verificó el 100% de los datos digitados introduciendo nuevamente -por otro digitador- cada nombre, utilizando el siguiente algoritmo:
  - Cuando el verificador ingresa el nombre al sistema se compara la exactitud con el nombre registrado por el digitador.
  - Si los nombres son idénticos se continúa con la siguiente persona.
  - Si los nombres son diferentes (aunque sea una letra) se obliga al verificador a tomar una decisión:
    - Aceptar el nuevo nombre pues el error lo había cometido el digitador
    - Corregir el nombre digitado por él pues estaba mal.

Fuera cual fuese la decisión, quedaron registrados ambos nombres para evaluar la cantidad de casos en los cuales el verificador tuvo que realizar la corrección. De la totalidad de la muestra de 21.161 personas, en el 3.5% de los casos fue necesario realizar la corrección por parte del verificador. Nótese que al tratarse de texto (no numérico) esta cifra es muy baja y se debe principalmente a diferentes interpretaciones de una misma letra confusa hechas entre el digitador y el verificador.

El nombre de 474 personas (2.2% de la muestra) no fue posible obtenerlo de las boletas debido a causas como boletas extraviadas, nombres faltantes en la boleta y nombres ilegibles

### *Búsqueda de los nombres en los datos del Registro Civil*

Los nombres de las personas en la muestra se empataron con la base de datos de poco más de 4 millones y medio de personas (vivas y muertas) almacenadas en el Registro Civil, en la etapa

posiblemente más compleja de este proceso. A continuación se describen las características de este empate.

Las variables disponibles en la muestra para realizar la búsqueda fueron:

- Sexo
- Nombre completo (Nombres, primer apellido, segundo apellido)
- Rango de fechas posibles de nacimiento de acuerdo con la edad declarada en el censo de 10 de junio de 1984 (ejemplo: una persona de 40 años en el censo pudo haber nacido entre el 11 de junio de 1943 y el 10 de junio de 1944)
- Provincia de nacimiento

Fue necesario elaborar un algoritmo para separar el nombre recolectado en las tres partes requeridas para el proceso de búsqueda: Nombre, Apellido1 y Apellido2. En este proceso se detectó uno de los mayores inconvenientes de todo el trabajo: muchas mujeres casadas utilizaron el primer apellido del cónyuge a la hora de declarar su propio nombre, sustituyendo en la mayoría de los casos su segundo apellido por el apellido del cónyuge precedido de la preposición “de”.

El problema de utilizar el apellido del cónyuge se encontró en 4.588 mujeres de la muestra recolectada, lo que corresponde a cerca de un 44% y de estas 3.844 (37%) no declararon su segundo apellido. Debido a que se conocía de antemano que los procesos de búsqueda en el Registro Civil sin un segundo apellido son de menor calidad fue necesario desarrollar una estrategia de contingencia utilizando dos características muy importantes:

- Los datos disponibles del Registro Civil incluyen (en la mayoría de los casos) el nombre del padre y de la madre
- En esa época una gran mayoría de las mujeres de 30+ años eran casadas y tenían hijos

Entonces, para las mujeres sin segundo apellido declarado que tenían “cónyuge” como relación de parentesco en el censo, se buscó en los datos de ese mismo hogar a la persona con relación de parentesco: “jefe” y se utilizó su nombre de la siguiente manera:

Se realizó una búsqueda en los datos del registro civil de todas las personas (hijos) que tuvieran los nombres del padre y de la madre congruentes con los nombres de la mujer (sin segundo apellido) y de su cónyuge (jefe). Para este conjunto de “hijos” se comparó el segundo apellido de la madre y en el caso de que fuera igual para todos se asumió que éste correspondía al segundo apellido de la mujer en cuestión. Este procedimiento, que tomó más de 50 horas seguidas de proceso de cómputo, permitió recuperar el segundo apellido de cerca de 2.300 mujeres casadas.

Otras actividades para mejorar la efectividad del proceso de búsqueda en los datos del Registro Civil fueron:

- Revisión visual de la descomposición del nombre en: nombre, apellido1, apellido2 pues existen excepciones difíciles de sistematizar como por ejemplo:  
“MARIA DEL MAR GONZALEZ DE LA MATA”

- Corrección de apellidos. Debido a que todas las personas en la muestra son nacionales, sus apellidos deben estar en la base de datos del Registro Civil. Se identificaron los apellidos de la muestra no registrados y se sustituyeron por los correctos en los casos evidentes.

Con los nombres depurados, se siguieron los siguientes pasos en el proceso de empate:

1. Con la base de datos de las personas en el Registro Civil generamos un archivo invertido de triadas (secuencias de tres letras del nombre). Este archivo invertido, es una especie de índice que contiene un registro por cada triada con la identificación de todas las personas cuyo nombre contiene a la triada (Baeza-Yates & Riveiro-Neto 1999).
2. Durante el proceso de búsqueda, el nombre de cada persona en la muestra fue dividido en triadas que se ordenaron quedando al principio las triadas más poco frecuentes. Con el archivo invertido se seleccionaron todas las personas en el Registro Civil que tuvieran la primera triada del nombre de la persona en la muestra luego se hizo la intersección con el conjunto de personas con la segunda, tercera y sucesivas triadas.
3. A cada persona del Registro Civil seleccionada de esta forma se le asignó un indicador del grado de similitud del nombre. Este indicador recibió una ponderación de 70% en el empate. Un segundo indicador, con ponderación de 20%, fue la similitud entre la fecha de nacimiento en el registro y la edad en el censo. Un tercer indicador, con una ponderación de 10%, fue la concordancia del primer dígito de la cédula con la provincia de nacimiento.
4. Si la persona con el puntaje máximo no alcanza un parámetro mínimo predeterminado (60), se elimina la primera triada (más poco frecuente) y se repite el proceso desde punto 2.
5. Al final, el caso con el puntaje más alto se identifica como posible empate

De un primer procesamiento se obtuvo la identificación segura de 13.488 empates. La posibilidad de un empate equivocado en este primer grupo es muy remota (Rogot, Sorlie & Johnson 1986) por lo que los denominamos empates “óptimos”. Previamente, 733 casos habían sido descartados por tener nombres incompletos imposibles de empatar (por ejemplo: MARIA LOPEZ) lo que redujo la muestra a 19.954 casos, de los cuales se identificó el 68% en este primer proceso. El criterio de identificación de un “empate óptimo” fue que el puntaje de similitud fuera superior a 80% y que el posible caso siguiente más cercano estuviera al menos 10 puntos por debajo. Adicionalmente se agregaron los casos con puntaje superior a 70% y con una diferencia de no menos de 15 puntos con el caso siguiente.

En una segunda ronda de empates se utilizó información del cantón en que las personas estaban registradas en el padrón electoral de 1990. Este dato se compara con el cantón de residencia en la muestra del censo. De este proceso y con criterios de similitud similares a los ya mencionados se logró identificar a 2.828 personas (14%).

Los posibles empates restantes (3.638) fueron revisados por un asistente para determinar si éste era obvio visualmente (aunque no lo había sido para los algoritmos de cómputo), aceptándose 1.431 (7%) empates adicionales.

Los casos restantes se reprocesaron incluyendo los tres empates más probables y revisando visualmente si alguno era obvio al ojo humano. Así se identificaron 511 (3%) empates más. Y en un proceso final en el computador se empataron 222 casos con la información del primer apellido solamente, la edad, sexo y el *distrito* de residencia declarado en el censo y de empadronamiento para votar en 1990.

Fue necesario eliminar del proceso a 76 casos a los cuales se le había asignado identificaciones repetidas (dos casos en la muestra con el mismo número de cédula). La principal causa de este problema se debe a que una misma persona fue empadronada en el censo en más de una ocasión. Esto fue comprobado pues en algunas, ambos casos formaban parte de hogares distintos de la misma vivienda o tenía una ubicación geográfica muy cercana.

En total se logró identificar en el Registro a 18,500 casos en números redondos o el 87% de la muestra inicial (93% de la muestra procesable). En total quedaron sin identificar alrededor de 2,700 personas, incluyendo 1,200 para quienes, o no apareció la boleta censal (500), o ésta venía sin nombre (700). Los casos restantes (1,500) son de muy difícil identificación en el Registro Civil, principalmente por ser nombres inventados, falsos nacionales u homónimos múltiples.

### *Identificación de las defunciones*

Para todas las defunciones del país de 1984 a 2007 se empató el registro en los archivos de defunciones del INEC con el del Registro Civil (que contiene el número de cédula). Esto con el propósito de contar con la información de la causa de defunción que existe solo en el registro del INEC, el cual, por otra parte, no tiene número de cédula. Para este empate se contó con las siguientes variables:

- Fecha de defunción
- Sexo
- Edad en años cumplidos al morir
- Lugar de muerte (hospital o distrito de residencia)

En Costa Rica mueren un poco más de 40 personas diariamente por lo que es posible utilizar las variables anteriores para obtener una alta efectividad en el empate entre ambas fuentes de datos. Sin embargo, hay que tener en cuenta en este proceso que los datos pueden contener errores y omisiones, como en la fecha y edad de la defunción (INEC) y en el lugar de la defunción del Registro Civil. También hay que considerar que algunos grupos de edad y lugares de residencia son más frecuentes que otros lo que hace que las probabilidades de éxito del empate varíen.

Este proceso requirió primero determinar la “unicidad” de los casos para estas cuatro variables; es decir, las defunciones que tienen una combinación única de las cuatro variables. Si la unicidad se da en las dos fuentes de datos el empate es óptimo. En los casos en que esto no ocurre, se analizó la unicidad únicamente con 3 variables, dejando de lado el lugar de muerte que tiene una importante cantidad de casos ignorados en el Registro Civil, especialmente antes de 1990.

Posteriormente, para los casos no empadronados se probaron variaciones en la edad y en la fecha de defunción, tratando de encontrar casos idénticos también respetando la unicidad. Esto es, se buscó empates con casos con uno, dos o tres años de diferencia en la edad y lo mismo con la

fecha de defunción con 1, 2 ó 3 días de diferencia. De esta forma se amplió significativamente la cantidad de casos empadronados.

En vista de que un buen número de defunciones del Registro Civil no tenían la información del lugar de muerte, se hizo un empate final usando en su lugar el distrito de empadronamiento para votar en un procedimiento similar al arriba descrito.

#### *Defunciones en la muestra*

Con base en el número único de identificación (la cédula) se empadronaron los individuos de la muestra con la base de datos de defunciones del Registro Civil ocurridas hasta el 31 de diciembre de 2007. Se identificaron 5.143 individuos fallecidos, o el 28% de los 18.480 con identificación. Para 112 de estos fallecimientos el registro no tiene la información de la fecha de defunción.

#### *Empate de la muestra identificada con las defunciones del INEC*

Finalmente, se realizó el empate entre los datos de la muestra y el archivo con las defunciones del INEC. Del total de las 5,143 defunciones en la muestra, se empadronaron con los archivos del INEC a 4,618 o el 90%. Las 525 no empadronadas incluyen a las 112 para las que no se conoce la fecha de defunción.

#### *Imputaciones y correcciones*

Para los 112 fallecidos cuya fecha de muerte se desconoce, se imputó al azar una fecha de muerte posterior a la del último padrón electoral en que aparece el individuo pero no mayor al 31 de diciembre de 2007, fecha de cierre del seguimiento de defunciones. Por ejemplo, para un individuo que aparece empadronado para votar en 2006, que ha fallecido y que no tiene fecha de muerte, se obtuvo al azar un número entre 0 y 761, el que se sumó a la fecha 30 de noviembre de 2005 (el padrón para las elecciones del primer domingo de febrero suele cerrarse el 30 de noviembre del año previo). Si en el ejemplo ese número fue 25, la fecha imputada de muerte es 25 de diciembre de 2005. Usar estas fechas imputadas es una alternativa superior a excluir del análisis estos 112 individuos que se sabe han fallecido, exclusión que resultaría en un sesgo hacia abajo de las estimaciones de mortalidad.

Para los individuos en la muestra con identificación (número de cédula) y que no han fallecido al 31/12/2007 (n=13,337) se analizó sus apariciones en los cinco padrones electorales de 1990 a 2006. La gran mayoría (93.2%) aparece empadronada en todos los cinco padrones, como debe ser, y un 4.4% adicional aparece en cuatro padrones. De hecho, si el Registro fuese perfecto y no hubiese emigración, todos los individuos que no han fallecido deberían estar empadronados para votar en las elecciones de 2006. Se determinó que 419 que no han fallecido no aparecen en el padrón de 2006. A estos individuos se les imputó al azar una fecha de pérdida de seguimiento en el intervalo entre el último padrón en que aparecen y el padrón en que ya no están. Por ejemplo a un individuo que aparece empadronado en 1994 pero no en 1998 y años siguientes, se le imputó al azar la fecha 21 de marzo de 1995, que se encuentra entre 1 de diciembre de 1993 y 30 de noviembre de 1997. Se recomienda usar estas fechas de pérdida de seguimiento como "censura" de los 419 individuos que no aparecen como fallecidos pero que han desaparecido de los padrones electorales. No hacerlo resulta en tasas de mortalidad muy bajas especialmente a partir de los 95 años de edad.

Finalmente, para el seguimiento de la mortalidad en la muestra se tomó la decisión de efectuarlo luego de un periodo de gracia de 90 días posterior al censo, o sea a partir del 9 de setiembre de 1984. Esto significa la pérdida de 46 individuos que fallecieron en ese lapso. Esta decisión se tomó ante la posibilidad de que el censo sub-enumere a individuos gravemente enfermos y, por ende, próximos a morir. Si tal sesgo existiera en el censo, las estimaciones de mortalidad en los meses inmediatamente posteriores al censo estarían sesgadas hacia abajo. Este sesgo se evita con la decisión de iniciar la observación tres meses después del censo.

## 4. RESULTADOS

A continuación se presentan los resultados de este muestreo con respecto a cuatro aspectos:

1. El rendimiento de los dos empates efectuados
2. La representatividad de la muestra
3. Posible efecto de empates erróneos
4. Error muestral y efecto de diseño *a posteriori*.
5. Calidad del registro de defunciones
6. Análisis demográfico de los patrones de mortalidad

### *Rendimiento de los empates y de la muestra*

Como ya se indicó, el 87% de la muestra original del censo de 1984 se logró empatar con el registro civil utilizando los nombres y otras características de las personas, para un tamaño muestral final de 18,480 personas con identificación. A su vez, el 92% de las defunciones detectadas en el registro fue posible empatarlas con las estadísticas vitales del INEC para tener la causa de muerte. La tabla 1 permite analizar si estas tasas de empate varían entre grupos geográficos y socio-demográficos. La proporción identificada (pareo 1) es menor entre las mujeres (84% vs. 90% de los hombres), debido al problema ya anotado de que en el censo de 1984 muchas mujeres consignaron el apellido de su esposo en lugar de su segundo apellido de soltera. Cuando no se tienen los dos apellidos, el empate a menudo no puede efectuarse debido a que, por ejemplo, hay muchas “María López”. En parte, ésta es también la razón por la que en personas con 80 y más años de edad cae la tasa de empates, pues a mayores edades hay mayor predominio de mujeres. En todo caso, la tasa de empate menor entre individuos nacidos en el Siglo XIX probablemente también ocurre porque el sistema de registro civil de Costa Rica, iniciado en 1883, no estaba completamente desarrollado en esa época, especialmente en las zonas de frontera como las regiones Brunca y Norte. También entre las personas sin educación, y probablemente analfabetas, la tasa de empate cae a 83%, debido a las limitaciones de estas personas para reportar correctamente sus nombres y apellidos.

El empate de las defunciones (pareo 2) fue más exitoso. Se repite una tasa menor en la región Norte y en las personas sin educación, aunque la diferencia no es tanta como en el empate 1. También se observa que hubo mayor dificultad para empatar defunciones ocurridas antes de 1990, debido a que en esa época las defunciones del registro a menudo no contienen el dato del lugar de ocurrencia del deceso. Por último, se observa que 24 personas fallecieron antes de 1984. Esto obviamente es imposible pues son personas empadronadas en el censo de junio de 1984. Se debe, probablemente, a errores en el pareo 1. Es decir, personas que fueron

incorrectamente ligadas a un número de cédula que no es de ellas. El reducido número de personas en esta situación es, sin embargo, alentador de que este es un error muy poco frecuente.

La tabla 2 resume el rendimiento de la muestrapor edades. Los 18.400 individuos de la muestra fueron observados un total de 373,000 años-persona. La observación termina en la fecha de la muerte, el fin de periodo de observación (diciembre de 2007) o la fecha de pérdida d seguimiento arriba mencionada. La tasa anual de mortalidad del grupo es de 13.6 por mil. En el análisis de tasas por edad conviene tener presente dos notas de cautela:

1. Las estimaciones son menos estables a edades extremas. En los jóvenes (antes de la edad 50) porque la muerte es un evento raro. A edades avanzadas (95 y más años) porque se tienen números reducidos (menos de mil) de años persona de exposición.
2. Las estimaciones para los grupos más jóvenes corresponden a un pasado más distante. En particular, para los menores de 40 años las estimaciones reflejan la mortalidad de los 1980s. A partir de los 50 años de edad, las estimaciones de mortalidad corresponden en promedio a finales de la década de 1990 (tabla 2).

#### *Representatividad de la muestra*

La existencia de diferencias significativas en las tasas de empate es indicativo de sesgos. Es pues conveniente ponderar la muestra para corregir estos sesgos. La base de datos contiene un juego de ponderadores que reproducen en la muestra la distribución censal por sexo, grupos decenales de edad, región y educación (ninguna vs. resto). Estos ponderadores están normalizados de manera que los tabulados ponderados reproducen el  $n = 18,480$  de la muestra. La variable de ponderación tiene media de 1.00 y desviación estándar de 0.14. En análisis multivariables en que se controle el efecto de estas cuatro variables es innecesario usar los ponderadores.

Una manera de valorar la representatividad de la muestra es comparando sus tasas estimadas de mortalidad con las de las tablas de vida nacionales. La Gráfica 1 hace esa comparación, considerando las tablas de vida de Costa Rica del quinquenio inicial de la observación, 1985-90, (INEC & CCP 2008) y el cercano al final, 2000-05 (Rosero-Bixby & Collado 2008). Es de esperarse que las tasas de la muestra se ubiquen entre estas dos series nacionales: a edades más jóvenes más cerca de las de la tabla de vida de 1985-90 y a edades mayores más cerca de la tabla de 2000-05. La Gráfica muestra que las tasas de mortalidad estimadas con la muestra reflejan bien la mortalidad nacional por edad.

La gráfica 1 también corrobora lo indicado anteriormente: la menor estabilidad de las estimaciones para edades extremas reflejada en intervalos de confianza más grandes.

Es interesante notar también que las tasas de mortalidad por edad pueden representarse bastante bien con una función de Gompertz, especialmente después de la edad 45. Las tasas a edades más jóvenes se apartan de Gompertz debido probablemente a que corresponden a una época anterior y, por ende, de mayor mortalidad.

### *Empates erróneos*

La posibilidad de que empates erróneos hayan introducido sesgos en la estimación se evalúa y descarta con la Gráfica 2-A. El razonamiento es el siguiente. Si las personas se identificaron incorrectamente se les estará asignando características que no son de ellas. En un caso extremo, si se la asignación fuese al azar, esas características no deberían presentar asociación con la mortalidad. En la realidad, el empate erróneo de algunos individuos, sesgaría hacia la unidad las tasas relativas de muerte según características de interés. La grafica compara las tasas relativas de mortalidad según educación (estimadas con regresión de Cox y controladas por muchas otras variables) de toda la muestra con las tasas del grupo de individuos para quienes no hay dudas de la calidad del pareo (el grupo denominado de pareo “óptimo”). Se observa que si bien en los grupos extremos de educación hay un cierto movimiento de la estimación hacia una tasa relativa de la unidad, el posible sesgo correspondiente no es importante: las dos curvas muestran esencialmente el mismo efecto de la educación sobre al mortalidad.

### *Efectos de diseño*

Se calcularon efectos de diseño a posteriori, mediante comparación de los errores estándar que se tendrían con una muestra aleatoria simple (MAS) y los errores estándar estimados con el procedimiento “jackknife” para corregir conglomeración.

La tasa de mortalidad de 13.6 por mil observada en la muestra presenta un error estándar asumiendo MAS de 0.19. El estimador “robusto” de este error estándar considerando el efecto de conglomeración resultó 0.26. Vale decir que la tasa por mil antes citada está en un intervalo de confianza de aproximadamente más menos cinco décimas. Y el efecto de diseño DEFT en el error estándar de este estimador es del orden de 1.35.

Los efectos de diseño en los estimadores de efectos de características socio-económicas o demográficas sobre la mortalidad, es decir, en los parámetros de regresiones multivariables, son mucho menores. En regresiones de Cox del riesgo de morir con 33 variables explicatorias, el efecto de DEFT estimado como en el párrafo anterior supera 1.10 solamente en los errores estándar de tres parámetros: tener teléfono (1.12) en la vivienda, residir en la región Chorotega (1.20) y en la Pacífico Central (1.22). En conclusión el diseño muestral por conglomerados tiene consecuencias negligibles sobre la confiabilidad de los efectos en la mortalidad.

### *Calidad del registro de defunciones*

Estudios previos han mostrado que el registro de defunciones en Costa Rica es esencialmente completo (Rosero-Bixby 2008). Desde 1961 los anuarios demográficos de las Naciones unidas califican al sistema de estadísticas vitales de Costa Rica –establecido en 1883- como “completo” (United Nations 1961). Hill y colaboradores caracterizan a Costa Rica como uno de los 11 países en desarrollo cuyas estadísticas de los registros vitales eran alrededor de 1995 tanto completas (registro de al menos el 90 de nacimientos y defunciones) como confiables (producen estimaciones similares a las basadas en censos o encuestas) (Hill, Pande, & Mahy 1999). Un estudio de seguimiento longitudinal de 3.000 personas adultas mayores (estudio CRELES: <http://ccp.ucr.ac.cr/creles/index.htm>) localizó el 98% de los registros de defunción de las personas que fueron identificadas en el campo como fallecidas. Es probable que algunas

defunciones no localizadas se registren tardíamente, pero aún si tal cosa no ocurriera, el subregistro de defunciones de Costa Rica, según este seguimiento es negligible y no afecta las estimaciones de la mortalidad..

Como se indicó anteriormente, 419 individuos de la muestra (2%) ni tienen registro de defunción ni aparecen en el padrón electoral más reciente. Hemos recomendado incluir a estos individuos en el análisis como “pérdidas de seguimiento”, es decir, como observaciones con “censura” en una fecha que hemos imputado al azar entre la última aparición en el padrón y el primer padrón en que no aparecen. Nos hemos inclinado a creer que estos son individuos que en su mayoría emigraron de Costa Rica. Sin embargo, cabe la posibilidad de que hayan desaparecido del padrón porque fallecieron sin que se registre la defunción. Con esta posibilidad en mente, hemos hecho un análisis de sensitividad a dos escenarios: (1) todos estos 419 individuos fallecieron en la fecha en que se imputó la pérdida de seguimiento y (2) todos ellos están con vida a la fecha de cierre del estudio (diciembre de 2007). El efecto de estas dos simulaciones sobre el nivel general de la mortalidad es obvio: en el primer escenario se tendrá una mayor mortalidad y en el segundo una menor mortalidad. Ya se mencionó también que la distorsión que causa este grupo es sobre todo en la mortalidad de edad avanzada. El análisis de sensitividad se centra en el efecto que podría tener este grupo de personas sobre las estimaciones de las gradientes socioeconómicas de la mortalidad, principal propósito de la creación de esta muestra.

La gráfica 2-B muestra las gradientes de la mortalidad relativa según educación. Se compara la línea incluyendo los individuos considerados como casos censurados (recomendada) con las simulaciones extremas de todos fallecidos o todos vivos. El ejercicio demuestra que prácticamente no hay efecto alguno en la gradiente por educación se considera a estas observaciones perdidas como censuras o como todas vivas a diciembre de 2007. La gradiente tampoco cambiaría mucho si todas estas personas habrían fallecido, y el cambio sería un efecto ligeramente más pronunciado de la educación; por ejemplo, las personas con educación universitaria tendrían una mortalidad relativa de 0.68 (en vez de 0.71) con el grupo de primaria incompleta como referencia. Por su parte, los efectos regionales y de residencia urbano-rural no se modifican en este análisis de sensitividad. En suma, las gradientes socioeconómicas de la mortalidad en esta muestra no son sensibles a la posibilidad extrema de que todas las pérdidas de seguimiento se deban a defunciones no registradas.

Uno problema grave en el estudio de la mortalidad a edades avanzadas es la exageración de la edad en censos y encuestas que infla los denominadores y reduce las tasas a edades altas. Por ello en esta muestra no se utiliza el dato del censo sino que la edad se calcula con la fecha de nacimiento del Registro Civil. No hay razones para pensar que en el Registro conste una fecha de nacimiento anterior a la verdadera que resultaría en exageración de la edad. Además, el sistema de registro de Costa Rica y el número de identificación asignado a cada individuo para toda la vida (el número de cédula) permite un chequeo adicional de la calidad del dato de la edad. Esta identificación contiene información del número secuencial del libro y la página en que consta el asiento del nacimiento. La numeración de estos libros permite identificar el año de registro del nacimiento, el cual solo puede ser mayor o igual al del nacimiento. En algunos casos existen registros tardíos o muy tardíos, por ejemplo alguien nacido en 1910 y registrado en un libro de 1950. Las personas registradas en un libro del mismo año en que nacieron es imposible que adolezcan exageración de la edad (podrían haber nacido varios años antes y declarado erróneamente que nacieron en el año de registro, pero esto produciría reducción de la edad,

nunca exageración), pues es imposible ir atrás en el tiempo y registrarse en un libro de años anteriores al del nacimiento. En cambio, los registros tardíos podrían tener errores de la fecha de nacimiento en cualquier dirección. Para evaluar esta posibilidad se creó una variable que indica si el nacimiento se registró oportunamente y, en consecuencia, está libre de la posibilidad de exageración de la edad. Como se verá más adelante (tabla 3), las personas con registro oportuno tienen una mortalidad 11% mayor, pero el efecto de la edad no difiere significativamente para este grupo. Las personas con registro tardío tienen, entonces, una menor mortalidad, esto podría ser un efecto genuino (muchas de ellas tienen registro tardío porque nacieron en el extranjero y cabe la posibilidad de que éste sea simplemente un efecto del “migrante saludable”) o podría originarse en que sistemáticamente se aumentaron la edad y por tanto son más jóvenes de lo que dice su certificado de nacimiento. Tan solo el 13% de la muestra no aparece como registrada oportunamente, porcentaje que es mayor (27%) entre los nacidos antes de 1900.

### *Patrón demográfico de la mortalidad*

Como se mostró la gráfica 1, una función de Gompertz describe bien la mortalidad de la muestra, sobre todo a partir de los 45 años de edad. Gompertz es una función sencilla de dos parámetros, utilizada para describir la mortalidad de la especie humana en edades adultas, en la que la tasa de mortalidad,  $m_x$ , aumenta exponencialmente con la edad  $x$  (o linealmente en los logaritmos)

$$m_x = \exp(\lambda + \gamma x)$$

- $\lambda$  es una constante que indica el nivel de la mortalidad (intersección a la edad  $x = 0$ );
- $\gamma$  mide el efecto de la edad: aumento instantáneo de  $m_x$  con cada año de aumento de  $x$ .

Otra forma de escribir esta función es con la relación multiplicativa:

$$m_x = L G^{(x-30)}$$

- $L$  es la tasa de mortalidad a la edad 30; y
- $G$  es el efecto de la edad medido por la tasa de mortalidad relativa con cada año que se aparta de la edad 30, así para  $\gamma = 0.1$ ,  $G$  es 1.1052 lo que constituye el efecto multiplicativo de un año de edad

Al modelo se le pueden agregar los efectos de un vector de variables  $z$ , en cuyo caso deben considerarse dos efectos para cada una de estas variables: (1) uno sobre el nivel, que denominaremos  $A_z$ , y (2) otro sobre la pendiente, que denominaremos  $B_z$ . Los parámetros se estiman con regresión (“streg” en el software Stata usado para el presente análisis), debiendo incluirse para cada variable  $z$  el valor de esta variable y una segunda variable de interacción entre ésta y la edad (producto de  $z$  por  $x$ ). En símbolos:

$$m_x = (L A_z) (G B_z)^{(x-30)}$$

Con los datos de la muestra llevados al formato de “tiempo de supervivencia” y usando el software Stata10 (Statacorp 2007), estimamos los efectos generales  $L$  y  $G$  y los efectos

modificadores A y B para las variables demográficas sexo (ser varón) y año calendario (tiempo t) y para tres características de la muestra: (1) haber nacido en marzo o abril, que estudios preliminares han mostrado que elevan el riesgo de morir por ciertas enfermedades; (2) haber sido inscrito en el año de nacimiento (que como se vio minimiza la posibilidad de que haya exageración de la edad); y (3) ser un empate “óptimo” tal y como se definió en una sección anterior. Si los efectos de interacción estadística no fueron estadísticamente significativos, se excluyeron del modelo de regresión para hacerlo más parsimonioso. Los resultados se muestran en la tabla 3.

La variable  $z$  sexo se introdujo en el modelo de regresión como una indicatriz o *dummy*, acompañada de una segunda variable producto de ésta por ( $x-30$ ). Los coeficientes de la regresión para estas dos variables, debidamente exponentiados, estiman los efectos A y B. Los valores estimados para la variable *dummy* varón = 1 fueron 1.97 y 0.99 (tabla 3). Esto quiere decir que la mortalidad de los hombres a la edad 30 es 97% mayor que la de las mujeres y el efecto de la edad (la pendiente de la curva) es menor (1% menor con cada año) que el de las mujeres, es decir que la curva masculina aumenta más lentamente con la edad, o que hay una convergencia entre los sexos a edades avanzadas, o que la edad causa menos estragos en los hombres que en las mujeres.

El modelo de la tabla 3 registra también la tendencia de la mortalidad en el tiempo  $t$ . El modelo estima que el nivel general de la mortalidad disminuyó a razón de 1.9% por año calendario, pero esta caída fue menor entre los varones (1.0% por año). Además, en el periodo 1984-89 el efecto de la edad sobre la mortalidad es menos intenso.

Por otra parte, las personas nacidas en marzo o abril presentan una mortalidad 6% mayor. Este efecto no es significativo en este modelo, pero conviene tenerlo presente pues cobra mayor fuerza en otros modelos por causa de muerte o con variables adicionales.

Las personas cuyo nacimiento se inscribió oportunamente hace muchos años tienen una mortalidad 11.5% mayor, resultado ya comentado en la sección de calidad de la información; en tanto que los empates “óptimos” tienen una mortalidad 20% menor a la edad 30 pero como la inclinación de la curva por edad es mayor, a edades avanzadas esta ventaja tiende a desaparecer.

Como ejemplo de utilización de los resultados del modelo, en la gráfica 3 se muestran las tasas por sexo y edad para enero de 2003 (punto medio del periodo 2000-2005), para los empates óptimos, los nacimientos registrados oportunamente y nacimientos distribuidos uniformemente en el año (1/12 en cada mes). Los dos parámetros de Gomperz de este grupo derivados de la tabla 3 son:

$$\begin{array}{lll} \text{Hombres:} & A = 0.000698 & B = 1.0921 \\ \text{Mujeres:} & A = 0.000301 & B = 1.1056 \end{array}$$

La gráfica 3 muestra también las tasas observadas en toda la muestra en el periodo 1999-2006 (centrado en enero 1 de 2003), con el correspondiente intervalo de confianza para grupos de edad de 10 años de amplitud. La gráfica confirma el buen ajuste de los datos observados a la función de Gompertz.

La gráfica 3 incluye también datos comparativos de moralidad para el periodo 2000-2005 de los EEUU (Arias 2000-2005) y Japón (*Ministry of Health, Labour and Welfare* 2009). Las diferencias no son grandes entre estos dos países y la muestra costarricense, tanto en el nivel como en la pendiente de las curvas. Sin embargo, un examen más cuidadoso muestra que entre los hombres, la mortalidad en los EEUU es mayor, mientras que entre las mujeres, destaca Japón por su menor mortalidad. A manera de resumen, la esperanza de vida en años a la edad exacta 50, derivada de las series de tasas de mortalidad por edad es como sigue en estas tres poblaciones:

	Hombres	Mujeres
Costa Rica 1999-2006	30.3	33.1
EEUU 2000-2005	28.4	32.3
Japón 2000-2005	30.3	36.4

Los hombres costarricenses tienen a los 50 años de edad una esperanza de vida similar que los japoneses y dos años mayor que los estadounidenses. Entre las mujeres, la esperanza de vida en Costa Rica es superior en un año que la de Estados Unidos, pero inferior en tres años que Japón. La brecha de esperanza de vida entre los sexos es menor en Costa Rica (2.8 años a la edad 50) que en los EEUU (3.9 años) o en Japón (6.1 años)

## 5. DISCUSIÓN

Se ha descrito el procedimiento seguido en el “estudio longitudinal de mortalidad de adultos costarricenses” (ELMAC). El estudio consiste en una muestra de aproximadamente 20.000 costarricenses de 30 o más años de edad del censo de población de 1984, a los que se les da seguimiento hasta fines de 2007.

Se seleccionó una muestra probabilística de todos los costarricenses con 30 y más años de edad de 302 conglomerados del censo elegidos aleatoriamente. El Instituto Nacional de Estadística y Censos (INEC) digitalizó los nombres de la boleta censal de los individuos en la muestra. Procesos informáticos depuraron los nombres y los enlazaron al número único de identificación, el número de cédula, del Registro Civil. El empate se logró hacer en el 87% de la muestra. Los individuos identificados se siguieron en las bases de datos del Registro Civil para establecer su sobrevivencia. Se identificaron más de 5.000 fallecidos entre el censo de 1984 y diciembre de 2007. Procesos informáticos adicionales enlazaron 92% de estas defunciones con las de las estadísticas vitales del INEC que contienen el dato de la causa de defunción.

El 13% de individuos para los que no fue posible hacer el empate tienen ciertas peculiaridades. Los sesgos que podría introducir esta pérdida de información se corrigieron introduciendo ponderadores que hacen que la muestra reproduzca la distribución censal por sexo, edad , educación y residencia urbana o rural.

El patrón de mortalidad de la muestra, que comprende 373.000 personas-años de observación, reproduce bien las tasas de las tablas de mortalidad del país calculadas de manera independiente

en estudios previos. Dicho patrón puede modelarse bien con una función de Gompertz, especialmente después de los 40 años de edad. La muestra confirma la baja mortalidad de los adultos costarricenses, especialmente de los varones que a los 50 años de edad presentan una esperanza de vida similar a la del Japón que suele ser considerada la más alta del mundo.

El principal propósito de este estudio longitudinal es determinar factores socioeconómicos asociados a la mortalidad de los adultos en análisis a publicar en el futuro. En el presente estudio se efectuaron análisis de sensibilidad de las gradientes socioeconómicas a dos imputaciones que debieron efectuarse para conformar la muestra (empates con el registro civil para individuos cuyos nombres y edades no coincidían en un 100% y pérdidas de seguimiento de individuos que desparecieron de los padrones electorales). Se determinó que dichas gradientes no se ven afectadas por las imputaciones efectuadas y supuestos adoptados.

En conclusión, el ELMAC permite hacer análisis robustos de los determinantes socioeconómicos de la mortalidad de los adultos costarricenses, una oportunidad que es única para un país en desarrollo, ya que hasta la fecha prácticamente todos los análisis de este tipo, o se han efectuado en países desarrollados, o se han concentrado en la mortalidad de la niñez con datos de censos o encuestas.

## 6. REFERENCIAS

- Baeza-Yates R. & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. New York: Adisson Wesley.
- Hill, K., Pande, R., & Mahy, M. (1999). Trends in Child Mortality in the Developing World 1960-1996. New York: UNICEF.
- INEC & CCP [Instituto Nacional de Estadística y Censos and Centro Centroamericano de Población] (2008). Estimaciones y proyecciones de población por sexo y edad (cifras actualizadas) : 1950-2050. San José, Costa Rica: Publicaciones INE.
- Ministry of Health, Labour and Welfare (Statistics and Information Department). (2009) Documento electrónico consultado el 29/08/2009 en: <http://www.mhlw.go.jp/english/database/db-hw/vs02.html>
- Rogot, E. Sorlie, P. & Johnson , N. J. (1986). Probabilistic methods in matching census samples to the National Death Index. *Journal of Chronical Diseases*, 39(9): 719-734.
- Rogot, E.; Sorlie, P. D.; Johnson, N. J.; Glover, C. S. & Treasure, D. W. (1988). A mortality study of one million persons by demographic, social and economic factors: 1979-1981 follow up. NIH Publication No. 88-2896. National Institutes of Health, Bethesda, MD.
- Rosero-Bixby, L. & Collado-Chaves, a. (2008) Tablas de mortalidad, jubilación e invalidez, Costa Rica 2000-2005. *Población y Salud en Mesomaérica* 6(1): Artículo 5. Revista electrónica en: <http://ccp.ucr.ac.cr/index.htm>
- Rosero-Bixby, L (2008). The Exceptionally High Life Expectancy of Costa Rican Nonagenarians. *Demography* 45(3): 673-691.
- Sorlie, P. D.; Baklund, E. & Keller, J. B. (1995). US mortality by economic, demographic and social characteristics: the National Longitudinal Mortality Study. *American Journal of Public Health*, 85: 949-956.
- Statacorp. (2007). *Stata Statistical Software: release 10*. College Station, Texas: Stata Corporation.
- United Nations. (1961). Demographic Yearbook. New York: United Nations.
- Arias, E. (2000-2005) United States Life Tables. En National Vital Statistics Report, National Center for Health Statistics. Documentos electrónicos disponibles en: [http://www.cdc.gov/nchs/products/life\\_tables.htm#life](http://www.cdc.gov/nchs/products/life_tables.htm#life)

## RECONOCIMIENTOS

Investigación apoyada por donaciones de Wellcome Trust (grant N. 072406) y del *Center for the Economics and Demography of Aging* at the University of California at Berkeley (NIA grant P30 AG01283). El Instituto Nacional de Estadística y Censos (INEC) y Los Archivos Nacionales colaboraron con este proyecto en la obtención de los nombres de las personas de la muestra del censo de 1984. William Vargas, funcionario del INEC, supervisó el proceso de digitación de nombres. El Tribunal Supremo de Elecciones proporcionó la base de datos de personas y de defunciones del Registro Civil, así como los padrones electorales 1990 a 2006. El INEC proporcionó las bases de micro-datos de las estadísticas vitales de defunciones.

**Tabla 1. Rendimiento de los dos tipos de empates**

Variables	Pareo 1 (N)	(N)	Pareo 2 (N)	(N)
<b>Total (N)</b>	<b>87.3% (18,480)</b>	<b>(21,161)</b>	<b>91.5% (4,618)</b>	<b>(5,045)</b>
<b><u>Sexo</u></b>				
Hombres	90.4%	(10,452)	91.1%	(2,781)
Mujeres	<b>84.3%</b>	(10,709)	92.0%	(2,264)
<b><u>Edad</u></b>				
30-9	87.6%	(7,973)	89.6%	(559)
40-9	87.8%	(5,181)	91.2%	(714)
50-9	88.3%	(3,646)	90.7%	(1,080)
60-9	87.3%	(2,418)	91.7%	(1,294)
70-9	86.0%	(1,360)	93.1%	(1,009)
80-9	<b>78.3%</b>	(503)	93.1%	(347)
90 y +	<b>64.6%</b>	(79)	90.5%	(42)
<b><u>Área</u></b>				
Urbana	86.7%	(10,262)	92.4%	(2,511)
Rural	87.9%	(10,899)	<b>90.7%</b>	(2,534)
<b><u>región</u></b>				
Metro San José	85.6%	(6,096)	91.2%	(1,471)
Central resto	90.2%	(8,198)	91.3%	(2,089)
Chorotega	86.3%	(1,802)	93.0%	(398)
Pacífico-centro	89.4%	(1,154)	93.4%	(256)
Brunca	<b>83.3%</b>	(1,711)	93.3%	(343)
Norte	<b>81.1%</b>	(1,364)	<b>89.2%</b>	(314)
Atlántica	89.5%	(836)	91.4%	(174)
<b><u>Educación</u></b>				
Ninguna	<b>82.6%</b>	(2,878)	<b>89.1%</b>	(978)
Primaria incom.	88.3%	(8,433)	91.7%	(2,504)
Primaria completa	88.4%	(4,810)	92.5%	(877)
Secundaria incomp	86.6%	(1,754)	92.8%	(250)
Secundaria completa	86.3%	(1,306)	94.4%	(214)
Post secundaria	89.0%	(1,980)	92.8%	(222)
<b><u>Año defunción</u></b>				
Antes de 1984		0.0%	(24)	
1984		<b>81.8%</b>	(77)	
1985-9		<b>89.6%</b>	(691)	
1990-4		92.1%	(970)	
1995-9		93.7%	(1,237)	
2000-4		90.6%	(1,263)	
2005-7		94.5%	(783)	

Pareo 1 = entre censo y Registro Civil

Pareo 2 = entre defunciones del Registro y de las Estadísticas Vitales del INEC

**Tabla 2. Resultados de la muestra para la estimación de tasas de mortalidad por grupos de edad de cinco años**

Edad	Defunciones	Años-persona	Fecha media de exposición	Tasa mortalidad	
				Observada	Ponderada
30-34	12	9,039.6	5-Feb-86	0.00133	0.00132
35-39	29	25,830.7	11-Jun-88	0.00112	0.00116
40-44	104	39,395.4	18-Mar-91	0.00264	0.00262
45-49	143	50,290.9	8-Feb-94	0.00284	0.00286
50-54	250	57,900.5	21-Jan-97	0.00432	0.00428
55-59	350	51,513.9	30-Nov-97	0.00679	0.00672
60-64	409	41,870.5	25-Oct-97	0.00977	0.00973
65-69	514	33,154.0	10-Dec-97	0.01550	0.01537
70-74	603	25,402.7	10-Dec-97	0.02374	0.02358
75-79	760	18,194.7	14-Jan-98	0.04177	0.04172
80-84	736	11,522.7	26-Feb-98	0.06387	0.06399
85-89	623	5,924.4	23-Jun-98	0.10516	0.10521
90-94	374	2,257.0	26-Feb-99	0.16571	0.16767
95-99	134	516.9	6-May-99	0.25924	0.25023
100-4	28	102.2	5-Apr-99	0.27397	0.28146
105+	7	25.1	14-Dec-01	0.27888	0.26436
Total	5,076	372,941.2	14-Aug-95	0.01361	0.01363

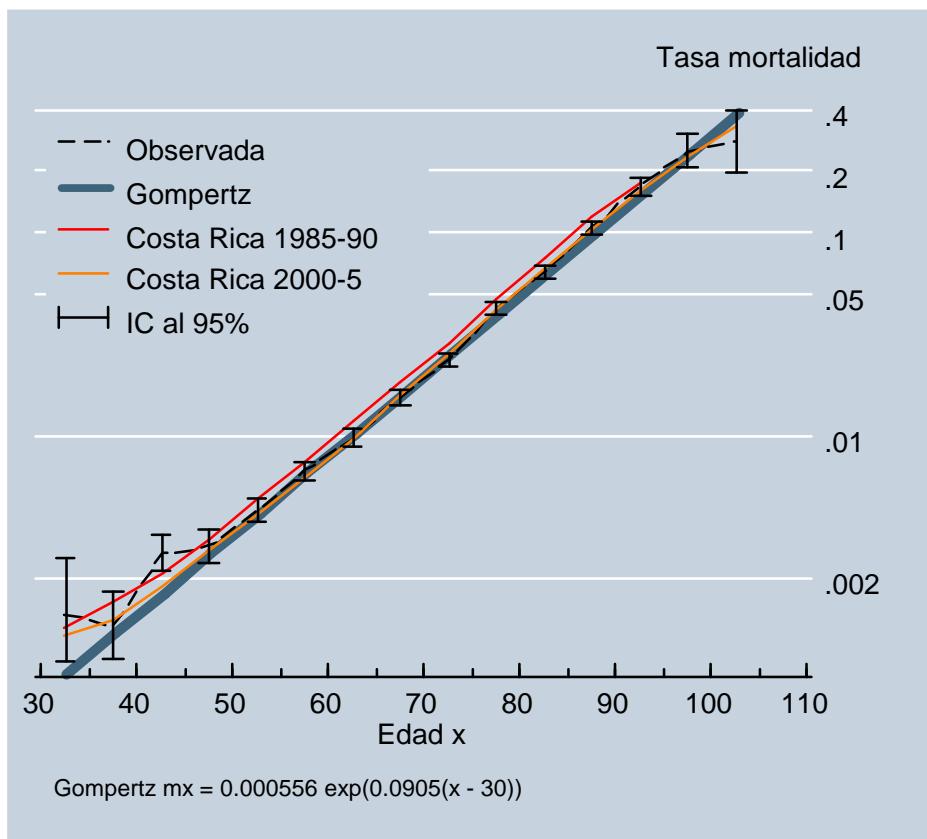
**Tbala 3. Modelo que describe la moralidad de la muestra con una función de Gompertz y variables demográficas. Costa Rica 1984-2007, edades 30 y más**

Efectos z	Tasa relativa	Interv. 95%	Confianza	Sig. P>z
<b>A. Nivel (a edad 30)</b>				
L	<b>0.00047</b>	0.00036	0.00061	0.00
Varón	<b>1.968</b>	1.592	2.433	0.00
t año (1984=0)	<b>0.981</b>	0.974	0.988	0.00
t * varón	<b>1.009</b>	1.001	1.018	0.04
Nacido mazo-abril	<b>1.061</b>	0.981	1.148	0.14
Registro oportuno	<b>1.115</b>	1.024	1.215	0.01
Empate óptimo	<b>0.801</b>	0.629	1.020	0.07
<b>B. Efecto edad (30=0)</b>				
G	<b>1.09961</b>	1.09301	1.10625	0.00
Si varón	<b>0.988</b>	0.983	0.992	0.00
Si periodo 1984-89	<b>0.995</b>	0.993	0.997	0.00
Si empate óptimo	<b>1.005</b>	1.000	1.011	0.06

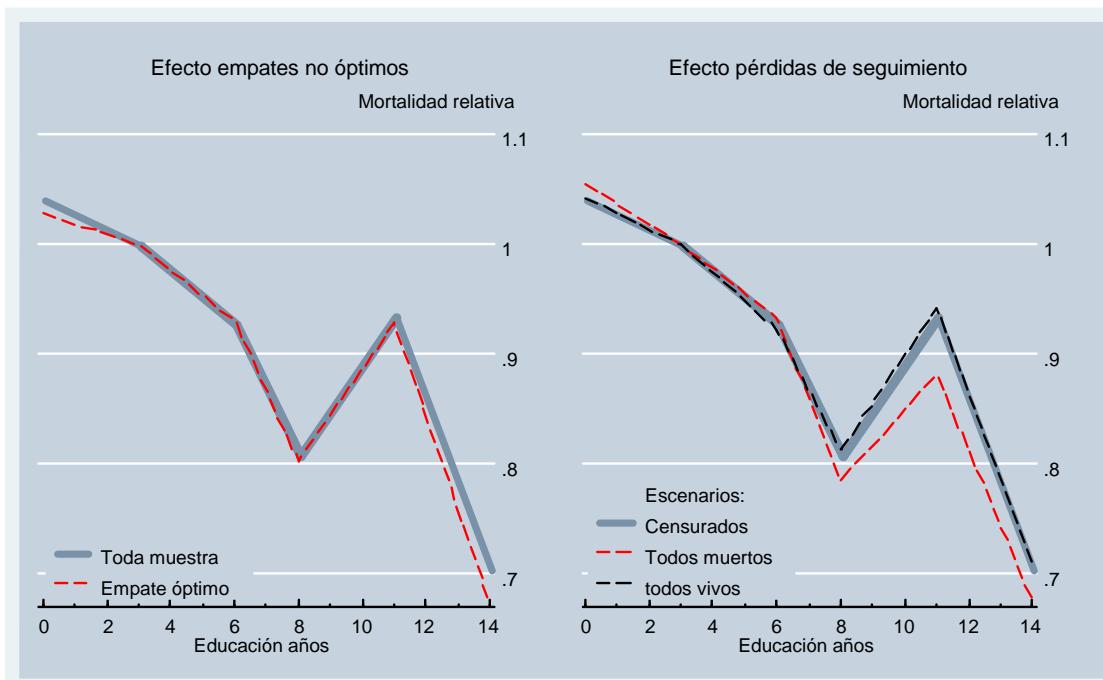
\* indica interacción estadística

El modelo:  $m_x = (L A_z) (G B_z)^{(x-30)}$

**Gráfica 1. Tasas de mortalidad por edad en la muestra y en las tablas de vida de Costa Rica 1985-90 y 2000-05**



**Grafica 2. Sensitividad del efecto de educación en la mortalidad a cambios en supuestos críticos de los datos**



Mortalidad relativa estimada con regresión de Cox controlando edad, sexo, región, ocupación, bienes en el hogar y condición de asegurado.

**Gráfica 3. Tasa de mortalidad 2000-05 observada y del modelo. Comparación con EEUU blancos y Japón para el mismo periodo**

