

Covid 19: Análisis de regresión (22/7/2020)

Miguel Messmacher Linartas¹
Director del Centro ITAM Covid-19

Los factores que determinan si una persona fallece por Covid parecen ser muy diversos. Por ello es que en semanas anteriores se revisó como cambiaba la tasa de letalidad -que es similar a la probabilidad de fallecer- dependiendo de si una persona tenía ciertas comorbilidades o si tenía cierta edad. Sin embargo, muchos de estos factores pueden estar relacionados entre sí. Por ejemplo, la obesidad tiende a estar relacionada con diabetes o hipertensión, y la diabetes tiende a ser más prevalente en personas de mayor edad. Por ello, para poder diferenciar el efecto de las distintas variables y obtener su contribución independiente, es necesario realizar un análisis de regresión.

El objetivo de este análisis es ver cómo distintas características de los pacientes afectan la probabilidad de que fallezcan con Covid-19.² Para este tipo de estudios, es común utilizar un modelo conocido como probit que nos permite identificar qué variables están relacionadas con la probabilidad de que ocurra un evento (en este caso un fallecimiento), así como la dirección y magnitud de esa relación.

El análisis se realizó empleando la base de datos diarios de la Secretaría de Salud (SSA) (<https://coronavirus.gob.mx/datos/#DownZCSV>), con corte al 14 de julio de 2020. Dado que no sabemos si las últimas personas en ser incorporadas a la base podrían fallecer transcurrido cierto tiempo, la base se filtra para dejar sólo a pacientes que mostraron síntomas cuando menos 15 días antes de la fecha en que se publica la base.³ Al filtrar por esto, el número de pacientes incluidos en la base es de 296,024. Se mantiene un número muy elevado de observaciones a pesar de haber filtrado 15 días de datos.

Para el análisis se utilizaron las siguientes variables incluidas en la base de la SSA:

1. **Sexo:** Variable indicadora de sexo, 0 si es hombre, 1 si es mujer.
2. **Rangos de edad:** Variable indicadora por cada rango de edad, exceptuando el rango de edad <20 años. La variable tiene un valor de 0 si el paciente no tiene una edad incluida en el rango, 1 si sí.
3. **Comorbilidades:** Para las siguientes condiciones, un valor de 0 indica si el paciente no padece la condición y 1 si padece la condición.
 - Diabetes
 - EPOC
 - Asma
 - Inmunosupresión
 - Hipertensión
 - Cardiovascular
 - Obesidad

¹ En la elaboración de esta nota colaboraron Efraín Rodríguez y Emilio Esperón.

² La información disponible nos permite saber si una persona tenía Covid-19 y falleció. No sabemos si falleció por causa de Covid-19 u otra razón.

³ De nuestro análisis previo, encontramos que el 75% de las personas que fallecían lo hacían a los 15 días o menos de haber mostrado síntomas.

- Renal crónica
 - Tabaquismo
 - Otra comorbilidad
4. **Entidad federativa:** Variable indicadora por entidad federativa, exceptuando la Ciudad de México. La variable tiene un valor de 1 si el paciente acudió a una unidad médica en la entidad federativa, 0 si no.
 5. **Sector de unidad médica:** Variable indicadora por sector de unidad médica (por ejemplo, IMSS, ISSSTE, unidades de la Secretaría de Salud), donde la variable tiene valor de 1 si el paciente acudió a una unidad médica perteneciente a este sector, 0 si no. Se exceptúa el sector SSA (Secretaría de Salud).
 6. **Interacciones:** Se reportan los resultados de dos tipos de interacciones, sexo/comorbilidad y sexo/rango de edad. En este caso, la variable de interacción es 1 si es una mujer con comorbilidad en el caso de la interacción sexo/comorbilidad. En el segundo caso, la variable de interacción es 1 si es una mujer con una edad incluida el rango de edad.⁴

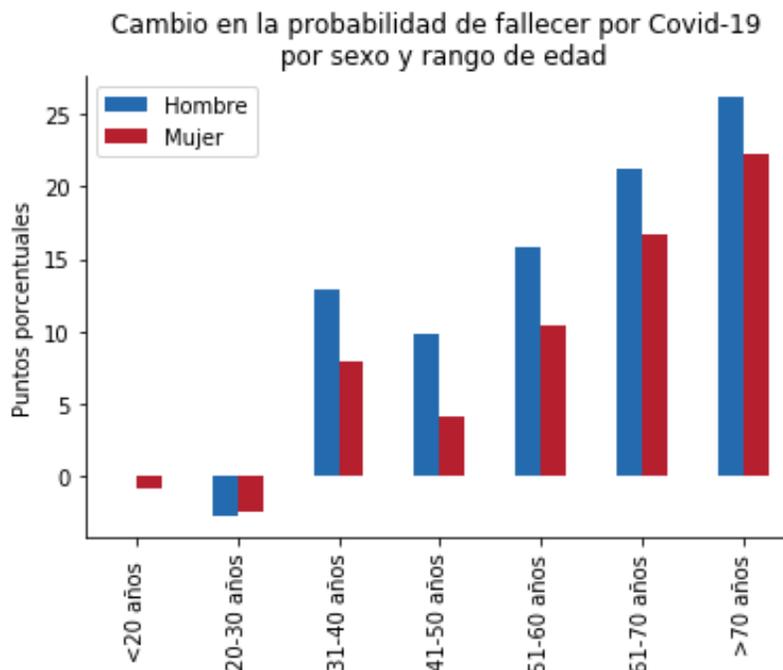
Los resultados completos de la regresión se reportan en un anexo al final del documento. Lo que reportamos a continuación son los efectos marginales de cada una de las variables mencionadas. Estos efectos marginales son el efecto que tiene una variable sobre la probabilidad de defunción de un paciente con Covid-19. Estos cambios son con respecto a ser un hombre de menos de 20 años sin comorbilidades en la Ciudad de México, que acudió a una unidad médica de la SSA.

Lo primero que reportamos son los efectos marginales por edad y sexo de los pacientes. Esos se presentan en la figura 1, y deben leerse como la diferencia que habría en la probabilidad con respecto a un paciente hombre menor a 20 años. Para el rango de 20-30 años, vemos una ligera disminución en la probabilidad de fallecer con Covid-19, pero luego las probabilidades aumentan con la edad a partir de la siguiente década de edad. Esto no debería ser sorpresa, al ser consistente con lo que habíamos encontrado viendo sólo la tasa de letalidad por edad en una nota anterior. Sin embargo, vale la pena notar que esas probabilidades son más bajas que las tasas de letalidad. Eso es porque la edad tiende a venir acompañada de comorbilidades, y la tasa de letalidad mezcla ambos efectos. A pesar de lo anterior, los aumentos son muy elevados, de forma que la probabilidad de fallecer con Covid-19 si se tiene más de 70 años aumenta en 26.1 puntos porcentuales.

Encontramos también que las mujeres tienen una menor probabilidad de fallecer con Covid-19 en todos los rangos de edad. Sin embargo, la diferencia cuando se es menor a 30 años es muy pequeña y no es estadísticamente significativa. La diferencia se vuelve grande proporcionalmente entre 30 y 60 años, y después a partir de la década de 60 a 70 años se vuelve proporcionalmente más pequeña, si bien sigue siendo significativa.

⁴ También se analizó la interacción entre comorbilidad y rango de edad. No se encontró evidencia estadísticamente significativa de que el efecto de las distintas comorbilidades cambiase con la edad.

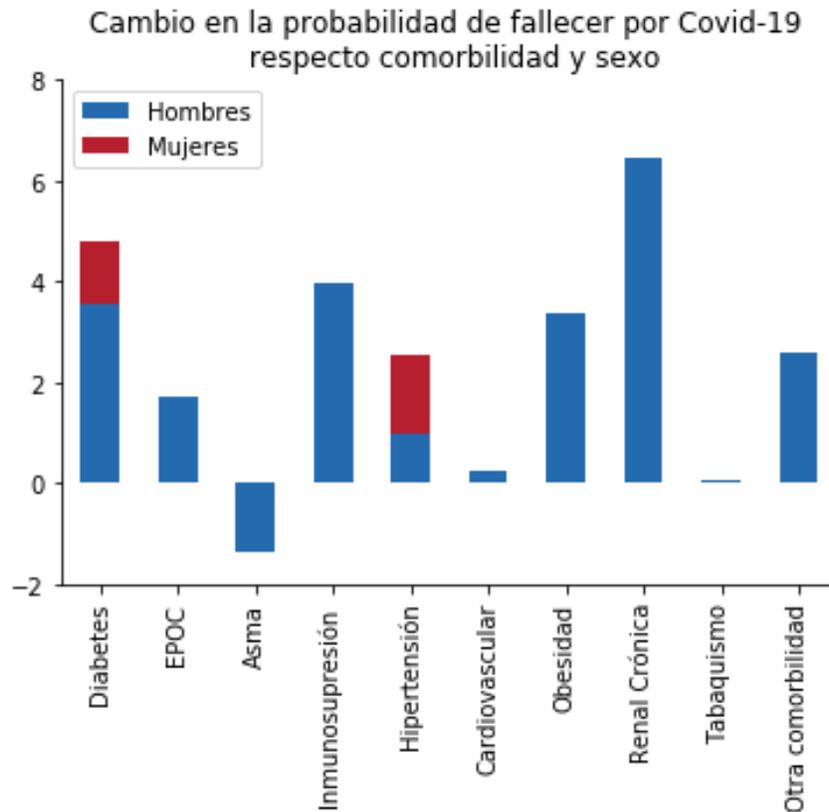
Figura 1.



La figura 2 presenta los efectos marginales de las distintas comorbilidades. También analizamos si había diferencias por sexo, pero sólo encontramos diferencias para diabetes e hipertensión, lo cual se discute más adelante. Al igual que en el caso de la edad, el aumento en la probabilidad de fallecer es menor que la simple tasa de letalidad que presentamos en la nota de hace dos semanas sobre comorbilidades. De nuevo, eso se debe a que la tasa de letalidad estaba mezclando el efecto de factores que están relacionados. En cuanto a la gravedad de las condiciones, hay algunos cambios en las conclusiones con respecto al simple análisis de tasas de letalidad por tipo de comorbilidad.

Algo que se mantiene igual es que la insuficiencia renal es la comorbilidad que lleva al mayor aumento en la probabilidad de fallecer. En contraste, el efecto de EPOC, que era el otro factor asociado con muy alta letalidad, aparece como mucho menor aunque sigue siendo positivo. Eso se debe a que en realidad parecía estar capturando la edad del paciente. En segundo nivel en cuanto al aumento que generan en la probabilidad de fallecer están diabetes, inmunosupresión, obesidad e hipertensión, todos con efectos independientes dado que ahora estamos controlando por si una persona con obesidad tiene o no diabetes. Con respecto a diferencias entre mujeres y hombres, no encontramos ninguna diferencia significativa para la mayoría de las comorbilidades (en cuyo caso sólo usamos un color para la columna), pero se encontró que la probabilidad de fallecer con Covid si se padece de diabetes o hipertensión parece ser mayor para mujeres (la probabilidad total para mujeres es la suma de la parte azul y roja de esas dos columnas). Si bien no se encontró evidencia estadísticamente significativa de que otras condiciones cardiovasculares aumenten la probabilidad de fallecer con Covid-19, es importante notar que otros problemas cardiovasculares tienden a estar asociados con hipertensión y obesidad, que son factores de riesgo elevado.

Figura 2.



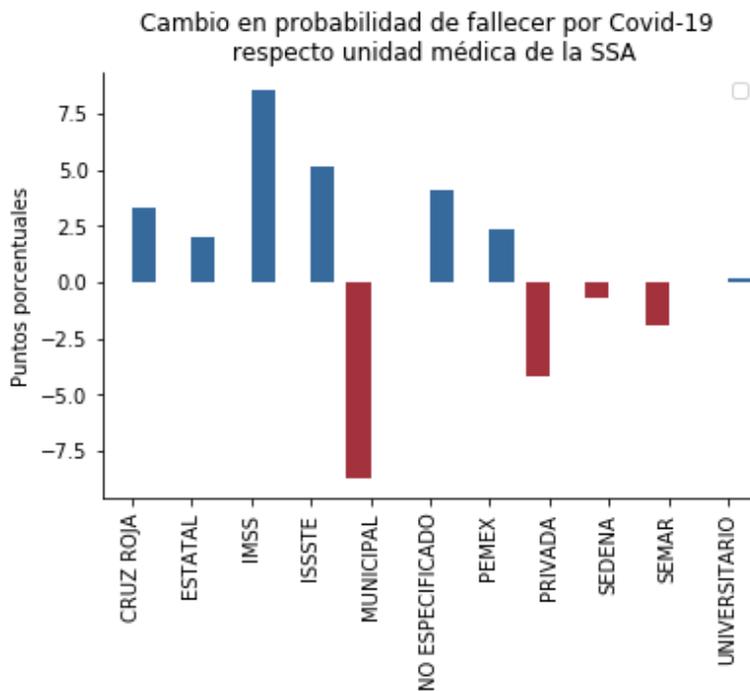
Los resultados por el tipo de sector de la unidad médica que brindó la primera atención al paciente se muestran en la figura 3. En este caso, los resultados deben leerse como el cambio con respecto a la probabilidad en caso de que el paciente hubiera sido atendido en una unidad médica de la Secretaría de Salud (SSA).

Vale la pena hacer un par de comentarios antes de ver los resultados. Primero, estas unidades médicas son las que dieron. Segundo, hay un problema muy conocido en el análisis de evaluaciones de servicios de salud que es que pacientes muy enfermos pueden buscar asistir al mejor hospital posible o con el mejor médico. Eso puede hacer que parezca que se mueren muchos pacientes de un muy buen médico u hospital. El ejemplo contrario podría explicar porque unidades médicas de gobiernos municipales son las que aparecen asociadas con la probabilidad más baja de fallecer. Es posible que personas con síntomas ligeros y casos de poca gravedad hubieran asistido a una unidad municipal, mientras que casos de mayor gravedad asistían a otro tipo de unidades médicas. No tenemos evidencia definitiva en un sentido u otro, pero se debe tener cuidado con interpretar todos estos resultados a nivel de sector médico.

Pasando a los resultados, se encontró que la probabilidad de que un paciente con Covid-19 fallezca aumenta alrededor de 8 puntos porcentuales en unidades médicas del IMSS y 5 puntos porcentuales

en unidades médicas del ISSSTE, de nuevo con respecto a unidades médicas de la Secretaría de Salud. En el caso de unidades la Cruz Roja, gobiernos estatales, Pemex, Sedena y Semar se encontraron ligeras diferencias al alza o a la baja, pero son menores y generalmente no significativas. Las unidades del sector privado y el sector municipal están asociados a una disminución de 4 y 9 puntos porcentuales en la probabilidad de fallecer, si bien el sector municipal parecería especialmente sensible a uno de los problemas mencionados en el párrafo anterior.

Figura 3.



Analizamos también si el número de días entre haber observado síntomas y haber ido a la unidad médica parecía afectar la probabilidad de fallecer. La hipótesis era que si uno retrasaba el tiempo en ir al médico podría aumentar esa probabilidad. Sin embargo, nuestra regresión resultó en un coeficiente negativo y significativo, lo contrario de lo esperado. Creemos que lo que está sucediendo es que pacientes con síntomas más graves acuden más rápido al médico, y son quienes tienen mayor probabilidad de fallecer. En cambio, pacientes con síntomas ligeros toman más tiempo para ir al médico y tienen menor probabilidad de fallecer. Este resultado no debe interpretarse de ninguna manera como que esperar para ir a una unidad médica reduce la probabilidad de fallecer.⁵

⁵ Idealmente, uno quisiera poder tener pacientes con idénticos síntomas y ver si entonces esperar más tiempo para ir al médico aumenta o reduce la probabilidad de fallecer.

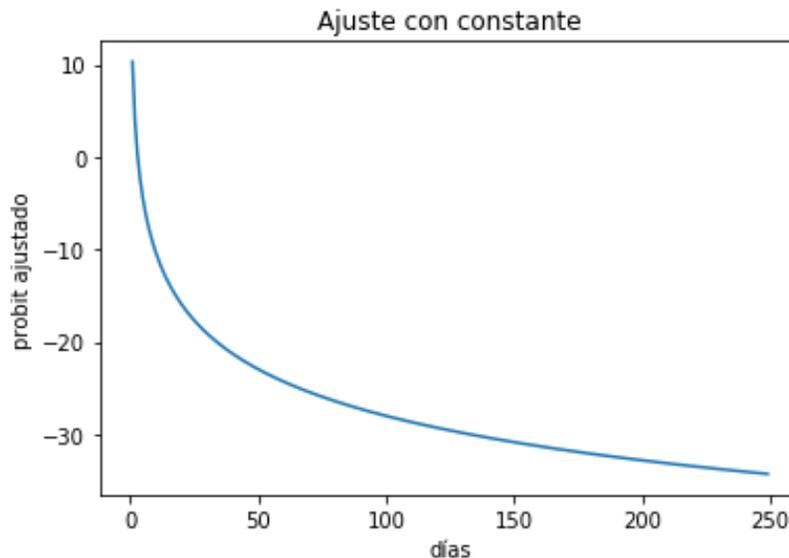
cuadrado de ese logaritmo. Este dato se obtuvo de la base de datos diarios de COVID-19 de la SSA.

- Camas, es el número de camas hospitalarias por cada 100 mil habitantes en cada estado. La idea era capturar la disponibilidad de servicios hospitalarios, donde hubiéramos esperado que mayor número de camas llevara a menor probabilidad de fallecer. El dato se obtuvo del Catálogo CLUES de la SSA.
- Sin salud, representa el porcentaje de la población de cada estado que no cuenta con acceso a los servicios de salud, de acuerdo con las estimaciones del CONEVAL, realizadas con los Módulos de Condiciones Socioeconómicas de la ENIGH 2018. Posibles efectos mixtos: hubiéramos esperado que mientras más grande fuera el porcentaje de la población sin acceso a servicios de salud, fuera mayor la probabilidad de fallecer; pero es posible que esas personas fallezcan en sus hogares y no sean capturadas por las estadísticas de la SSA.
- Densidad, representa la densidad de población de cada estado de acuerdo con la Encuesta Intercensal 2015. Podría haber distintos efectos: mayor densidad podría facilitar el contagio, pero también podría implicar mejor acceso a servicios médicos.
- Población urbana, representa el porcentaje de la población de cada estado que vive en el área urbana, de acuerdo con la Encuesta Nacional de la Dinámica Demográfica 2018. Igual que la anterior.
- Edad, representa la edad mediana de la población de cada estado de acuerdo con la Encuesta Intercensal 2015. Si no hubiéramos controlado por la edad de los pacientes, hubiéramos esperado que mayor edad llevaría a mayor probabilidad de fallecer. Sin embargo, dado que ya estamos controlando a nivel de pacientes, no esperábamos ver ningún efecto adicional.
- PIB per cápita, es el PIB per cápita no minero de cada estado en 2018. Para obtener el PIB no minero, al PIB total de cada estado se le restó el PIB por actividades mineras. El dato se obtuvo del Sistema de Cuentas Nacionales. Esperábamos que mayor PIB per cápita redujera la probabilidad de fallecer, principalmente por mejor acceso a servicios de salud.
- Pobreza, representa el porcentaje de la población de cada estado que vivía en condición de pobreza patrimonial en 2018, de acuerdo con las estimaciones del CONEVAL, realizadas con los Módulos de Condiciones Socioeconómicas de la ENIGH 2018. En sentido contrario del caso anterior, esperábamos que mayor pobreza aumentara la probabilidad de fallecer.
- Informalidad, representa el porcentaje de la PEA de cada estado que trabajaba en el sector informal, de acuerdo con los cálculos de la ENOE 2018. Similar al caso de pobreza.
- Escolaridad, representa el grado de escolaridad promedio de la población de cada estado de acuerdo con la Encuesta Intercensal 2015. Esperábamos que mayor escolaridad redujera la probabilidad de fallecer.
- Turismo, es el porcentaje que representan los servicios de alojamiento temporal y de preparación de alimentos y bebidas en el PIB total de cada estado, de acuerdo con los datos de 2018 del Sistema de Cuentas Nacionales. La idea era que la importancia de esas actividades en la producción del estado indicase que una cantidad alta de la población había visto afectados sus ingresos, por lo que habían visto disminuido su acceso a servicios de salud, o que no hubieran podido seguir medidas de aislamiento para limitar el contagio.

Los resultados detallados de la regresión se muestran en el anexo, pero aquí resumimos los mismos. De las 11 variables mencionadas, encontramos que las únicas que tienen coeficientes significativamente distintos de cero para explicar las diferencias observadas entre los estados son 2 variables: i) el nivel del PIB per cápita, y ii) el número de días transcurridos desde que el estado registro 10 muertes diarias. Los efectos son los siguientes:

- i) PIB per cápita: pacientes que asisten a unidades médicas en estados más ricos (con mayor PIB per cápita) tienen menor probabilidad de fallecer. Ello se debe probablemente a que los servicios son mejores y a que los pacientes tienen más recursos que les permiten recibir mejor atención. Este indicador probablemente captura algunos de los efectos que intentábamos capturar a través de otras variables como el número de camas o el porcentaje de población en pobreza.
- ii) El número de días transcurridos: encontramos que la probabilidad de fallecer ha venido bajando en el tiempo. Esa es una muy buena noticia en la que entraremos en mayor detalle en la nota de la próxima semana. El efecto estimado de los días transcurridos se presenta en la figura 5. Al inicio se da una disminución más fuerte en la probabilidad de fallecer, y si bien sigue bajando conforme van aumentando los días, esas disminuciones se van volviendo más pequeña. Es probable que ello se deba a que se ha venido dando un proceso de aprendizaje y descubrimiento en las unidades médicas, donde el personal médico va aprendiendo que tipo de tratamientos son los que mejor funcionan para cada tipo de paciente.

Figura 5.
Efecto de días transcurridos sobre efecto estatal



ANEXO

A continuación, se presentan los cuadros con los resultados. Aunque estén por separado, los cuadros 1-5 son parte de la misma regresión.

Cuadro 1: Efectos marginales por comorbilidad/rangos de edad

Variable	dy/dx	Error Estándar	z	P> z
SEXO	-0.0078	0.010	-0.795	0.427
DÍAS ENTRE SÍNTOMAS/INGRESO	-0.0011***	0.000	-7.182	0.000
Rango Edades 20-30	-0.0275***	0.008	-3.605	0.000
Rango Edades 31-40	0.0293***	0.007	4.136	0.000
Rango Edades 41-50	0.0982***	0.007	14.098	0.000
Rango Edades 51-60	0.1585***	0.007	22.799	0.000
Rango Edades 61-70	0.2127***	0.007	30.394	0.000
Rango Edades >70	0.2614***	0.007	37.107	0.000
DIABETES	0.0354***	0.002	21.805	0.000
EPOC	0.0171***	0.004	4.119	0.000
ASMA	-0.0137	0.005	-2.623	0.009
INMUSUPR	0.0396***	0.005	7.710	0.000
HIPERTENSION	0.0098***	0.002	6.053	0.000
CARDIOVASCULAR	0.0023	0.004	0.635	0.525
OBESIDAD	0.0339***	0.002	20.526	0.000
RENAL_CRONICA	0.0646***	0.004	18.036	0.000
TABAQUISMO	0.0008	0.002	0.348	0.728
OTRA	0.0259***	0.004	6.986	0.000
Número de observaciones	296,024			

*** Se refiere un valor $p < 0.001$

Cuadro 2: Efectos marginales interacciones sexo/rangos de edad.
Aplican cuando el sexo es mujer

Variable	dy/dx	Error Estándar	z	P> z
SEXO_Rango Edades_20-30	-0.0238	0.011	-2.119	0.034
SEXO_Rango Edades_31-40	-0.0495***	0.010	-4.765	0.000
SEXO_Rango Edades_41-50	-0.0572***	0.010	-5.637	0.000
SEXO_Rango Edades_51-60	-0.0551***	0.010	-5.455	0.000
SEXO_Rango Edades_61-70	-0.0454***	0.010	-4.455	0.000
SEXO_Rango Edades_>70	-0.0388***	0.010	-3.775	0.000
Número de observaciones	296,024			

*** Se refiere un valor $p < 0.001$

Cuadro 3: Efectos marginales interacciones sexo/comorbilidades.
Aplican cuando el sexo es mujer

Variable	dy/dx	Error Estándar	z	P> z
SEXO_DIABETES	0.0127***	0.003	4.946	0.000
SEXO_EPOC	0.0081	0.006	1.308	0.191
SEXO_ASMA	0.0010	0.007	0.142	0.887
SEXO_INMUSUPR	0.0019	0.007	0.255	0.798
SEXO_HIPERTENSION	0.0155***	0.003	6.073	0.000
SEXO_CARDIOVASCULAR	0.0055	0.006	0.961	0.337
SEXO_OBESIDAD	-0.0039	0.003	-1.504	0.132
SEXO_RENAL_CRONICA	0.0096	0.006	1.725	0.085
SEXO_TABAQUISMO	-0.0167	0.005	-3.406	0.001
SEXO_OTRA_COM	-0.0064	0.005	-1.190	0.234
Número de observaciones	296,024			

*** Se refiere un valor $p < 0.001$

Cuadro 4: Efectos marginales por sector.
Con respecto a unidades médicas de la Secretaría de Salud (SSA)

Variable	dy/dx	Error Estándar	z	P> z
SECTOR CRUZ ROJA	0.0329	0.053	0.617	0.537
SECTOR ESTATAL	0.0195***	0.004	5.553	0.000
SECTOR IMSS	0.0849***	0.001	73.043	0.000
SECTOR ISSSTE	0.0510***	0.002	22.123	0.000
SECTOR MUNICIPAL	-0.0875***	0.023	-3.858	0.000
SECTOR NO ESPECIFICADO	0.0411***	0.006	6.751	0.000
SECTOR PEMEX	0.0236***	0.004	5.485	0.000
SECTOR PRIVADA	-0.0420***	0.004	-10.556	0.000
SECTOR SEDENA	-0.0073	0.007	-1.000	0.317
SECTOR SEMAR	-0.0190	0.007	-2.678	0.007
SECTOR UNIVERSITARIO	0.0020	0.017	0.121	0.903
Número de observaciones			296,024	

*** Se refiere un valor $p < 0.001$

Cuadro 5: Efectos marginales por entidad federativa.
Con respecto a la Ciudad de México

Entidad Federativa	dy/dx	Error Estándar	z	P> z
AGUASCALIENTES	-0.0562***	0.007	-8.388	0.000
BAJA CALIFORNIA	0.0457***	0.003	17.735	0.000
BAJA CALIFORNIA SUR	-0.0614***	0.008	-7.284	0.000
CAMPECHE	-0.0147	0.005	-2.689	0.007
CHIAPAS	0.0062	0.004	1.565	0.118
CHIHUAHUA	0.0471***	0.004	11.217	0.000
COAHUILA	-0.0678***	0.005	-14.706	0.000
COLIMA	-0.0078	0.010	-0.810	0.418
DURANGO	-0.0417***	0.007	-6.182	0.000
GUANAJUATO	-0.0727***	0.004	-19.880	0.000
GUERRERO	0.0138***	0.003	4.191	0.000
HIDALGO	0.0389***	0.004	10.195	0.000
JALISCO	-0.0300***	0.003	-8.945	0.000
MÉXICO	-0.0196***	0.004	-4.900	0.000
MICHOACÁN	0.0403***	0.004	9.367	0.000
MORELOS	0.0353***	0.002	19.156	0.000
NAYARIT	-0.0225***	0.006	-3.533	0.000
NUEVO LEÓN	-0.0593***	0.004	-15.279	0.000
OAXACA	0.0008	0.004	0.226	0.821
PUEBLA	0.0195***	0.003	7.434	0.000
QUERÉTARO	0.0153	0.005	2.846	0.004
QUINTANA ROO	0.0325***	0.004	8.118	0.000
SAN LUIS POTOSÍ	-0.0340***	0.005	-6.224	0.000
SINALOA	0.0148***	0.003	5.248	0.000
SONORA	-0.0197***	0.003	-6.554	0.000
TABASCO	0.0120***	0.003	4.375	0.000
TAMAULIPAS	-0.0328***	0.004	-8.860	0.000
TLAXCALA	0.0179***	0.005	3.674	0.000
VERACRUZ	-0.0028	0.003	-1.098	0.272
YUCATÁN	-0.0177***	0.004	-4.252	0.000
ZACATECAS	-0.0139	0.008	-1.662	0.096
Número de observaciones		296,024		

*** Se refiere un valor $p < 0.001$

Cuadro 6. Variable dependiente: Efectos marginales por entidad federativa

	Modelo inicial	Modelo final
Constante	8.5675 * (-4.9138)	10.2863 ** (-3.866)
ln(Días desde 10 muertes)	-7.5892 (-4.823)	-9.5134 ** (-3.8257)
ln(Días desde 10 muertes)^2	0.2097 (-0.1266)	0.2599 ** (-0.0995)
ln(Camas)	0.0777 (-0.1577)	
ln(Densidad)	0.0223 (-0.0274)	
ln(Edad)	0.5111 (-0.5772)	
ln(Escolaridad)	0.3411 (-0.3921)	
ln(PIB per cápita)	-0.9604 * (-0.55)	-0.506 *** (-0.1593)
Pob. urbana	-0.0575 (-0.0875)	
Pobreza	0.0303 (-0.1219)	
Sin salud	-0.0629 (-0.3354)	
Informalidad	-0.0639 (-0.128)	
Turismo	0.1366 (-0.1369)	

Errores estándar en paréntesis.

* p<0.1, ** p<0.05, ***p<0.01