

# Cada quien es diferente: El efecto de la salud sobre los ingresos laborales

Gabriel Martínez

## Resumen

Se estudia el posible efecto causal del estado de salud hacia los ingresos laborales. En un modelo económico las personas deciden moverse hacia estados de mejora en salud o de no mejora. La decisión afecta los ingresos laborales. Se exploran modelos que plantean otras relaciones de causalidad, más simples, para evaluar la consistencia de las estimaciones. Se concluye que hay una heterogeneidad sustancial en la relación entre salud e ingresos laborales. Por ello, los modelos que miden efectos promedio son poco informativos. Las personas con mayor propensión a mejorar en salud tienen las mayores ganancias en ingresos laborales de hacerlo, pero los intervalos de confianza estimados para los efectos marginales de mejorar en salud son grandes.

**Palabras clave:** salud, salarios, trabajo, obesidad, hipertensión, diabetes, México

**Clasificación JEL:** I15, J31

Everyone is different: health effect on labor income in Mexico

## Abstract

The possible causal effect of health status towards labor income is studied. In an economic model people decide to move towards states of improvement in health or no improvement. The decision affects labor income. We explore models that pose other simpler causal relationships to evaluate the consistency of estimates. It is concluded that there is substantial heterogeneity in the relationship between health and labor income. Therefore, models that measure average effects are not informative. People with greater propensity to improve on health have the highest gains in labor income to do so, but the confidence intervals estimated for the marginal effects of improving on health are large.

**Keywords:** health, wages, work, obesity, hypertension, diabetes, Mexico

**JEL:** I15, J31

## 1. Introducción

Esta investigación busca identificar la existencia y magnitud de una relación causal y exógena del estado de salud a los ingresos laborales. La relación entre salud y trabajo es compleja pues presumimos relaciones mutuas de causalidad. Además, esas relaciones pueden no ser iguales para todas las personas debido a diferencias en la capacidad individual para mantener buena salud o para trabajar bajo una condición deteriorada de salud. Para algunas personas, mejor salud puede motivar menor trabajo en el mercado; para otras, mayor volumen de trabajo puede inducir mejora o deterioro de salud. Para enfrentar esta situación de causalidad compleja, aplicamos sistemáticamente supuestos específicos de causalidad, de manera que podamos evaluar la consistencia de las estimaciones para llegar a un conjunto preferido.

Empleamos un modelo de conducta bien especificado para llegar a un conjunto de efectos estimados de la salud sobre los ingresos, los cuales comparamos con resultados de modelos más simples. Esos modelos simples no contemplan la posibilidad de autoselección de la población hacia algún estado de salud. Ese es el caso de las especificaciones de mínimos cuadrados ordinarios (MCO) y variables instrumentales (VI), así como de modelos en diferencias y de diferencia en diferencias.

Hemos identificado dos series de encuestas que proveen información útil, la Encuesta Nacional de Nutrición y Salud (ENSANUT; INSP, 2012) y la Encuesta Nacional sobre Niveles de Vida de los Hogares (ENNVH; Rubalcava y Teruel, 2006, 2008 y 2013). La primera consiste de cortes transversales repetidos, mientras que la segunda construye información longitudinal. En este artículo usamos la ENNVH por sus propiedades longitudinales.

La sección 2 discute las especificaciones empleadas para estudiar la relación entre salud e ingresos. En la sección 3 describimos y discutimos los datos empleados. La sección 4 contiene los resultados, y la sección 5 argumenta acerca de cuáles son las especificaciones preferidas.

## 2. Marco teórico

Nuestro interés principal radica en identificar el impacto que el estado de salud tiene sobre los ingresos laborales. La literatura señala a la posible causalidad mutua entre salud y trabajo, así como a la posible existencia de otros factores que provocan resultados en los ámbitos de la salud y del trabajo (Deaton, 2006 y Weil, 2015). Algunos de esos resultados apuntan a que la relación puede ser de largo plazo, pasando de condiciones en la infancia o el embarazo hasta la edad adulta, e incluso incluyendo ajustes que toman generaciones en llevarse a término (por ejemplo, el aumento en la estatura al mejorar la oferta nutricional). Ante esta complejidad, queremos mantener un marco de referencia que nos permita comparar de forma consistente distintas especificaciones y llegar a conclusiones sobre el ámbito de validez de las estimaciones.

¿Cuál es el efecto de la salud en determinar los ingresos laborales de las personas? Para estimar el impacto de una variable es necesario expresar la hipótesis de conducta que genera los valores de la variable dependiente y la forma en que esa conducta puede afectar la medición del impacto de la variable de interés. El estado de salud es valorado por las personas, por lo cual los economistas estudian la demanda por servicios de salud como una demanda derivada, involucrándose un costo de lograr un mejor estado de salud. Tam-

bién es posible que el estado de salud sea en parte determinado por el entorno social y económico, y que sea en algún grado como un evento exógeno a la persona. Esta distinción es de relevancia para el diseño de política, pues en la medida en que se identifiquen intervenciones de bajo costo que afecten positivamente la salud, las estimaciones sobre el impacto de la salud sobre los ingresos pueden servir para valorar los beneficios de las intervenciones. En nuestra aplicación, el individuo toma decisiones sobre su estado de salud y hay costos, en general no observables por el analista, que afectan esa decisión.

El modelo económico planteado para medir la relación entre estado de salud e ingresos laborales es el modelo generalizado de Roy (MGR; Heckman y Vytlacil, 2007). Este es un paradigma usado en la economía aplicada por su parsimonia y por incorporar hipótesis de decisión en forma explícita. Las personas deciden su participación en uno de dos “sectores” de acuerdo con las ganancias que esperan tener de estar en alguno de ellos. Es un modelo que se puede emplear en un enfoque de evaluación de programas pues la participación en un “tratamiento” se interpreta como la participación en uno de dos sectores (población tratada y no tratada), sea ésta totalmente exógena o que involucre elementos de decisión. En nuestra aplicación, las personas deciden moverse hacia un mejor estado de salud o no hacerlo; conservamos el lenguaje de “tratamiento” por ser este convencional, pero, en forma más general, ese tratamiento no es aplicado externamente, sino que refleja la decisión de la persona de llevar a cabo las acciones que producen mejor salud. En esa decisión típicamente juegan factores fuera de control de la persona (como la edad), pero también decisiones individuales que tienen costo, como pueden ser cambiar de peso corporal, educarse más o efectuar otras acciones que no son medidas por las encuestas. También caben naturalmente en este marco las intervenciones de política, como puede ser un nuevo programa público de seguro o servicio de salud.

La salud de cada persona es determinada en buena medida por la tecnología disponible en un punto en el tiempo y por su entorno social, en particular el costo de acceso a los servicios, que a su vez es afectado por la localización de la persona y el entorno de protección social. Esto no contradice que, condicionada por ese entorno que determina costos de mantener un estado de salud, la persona decida llevar a cabo acciones que elevan o disminuyen la probabilidad de ocupar cierto estado de salud. El tema es complejo porque la definición misma de salud

es compleja. Por ejemplo, la diabetes, el cáncer y la hipertensión son en ocasiones etiquetadas como enfermedades “silenciosas”, de manera que una persona afectada puede percibir un daño bajo o nulo a su salud siendo que el mismo es grande. En consecuencia, la intensidad del trabajo de la persona puede ser alta aun cuando su cuerpo ya tenga un daño en la etapa “silenciosa” de la enfermedad. Por otro lado, los marcadores objetivos de salud que emplean encuestas como la ENNVIH tienen una capacidad limitada de predicción sobre la conducta futura. Una referencia reciente acerca de esto es el estudio de Jones, Molitor y Reif (2018), quienes concluyen que los programas de bienestar en las empresas (los cuales incluyen mediciones de marcadores de salud) no tienen efecto sobre el gasto médico, la conducta de salud del individuo, la productividad o el estado auto-declarado de salud, sino que más bien sirven como un mecanismo de filtrado a los empleadores. Es decir, no hay razones para suponer que mediciones de hipertensión, azúcar, isquemia u otras medidas físicas y objetivas de riesgo de enfermedad sean poderosas para pronosticar cambios de conducta, tales como el esfuerzo laboral o la percepción propia del estado de salud.

En el marco del modelo tenemos como variables endógenas la mejora en salud y el ingreso de la persona. La disponibilidad de datos longitudinales nos permite la definición de variables de mejora en salud, así como explorar resultados con base en los niveles y cambios en el ingreso a lo largo del tiempo. El ingreso laboral ( $Y_{ij}$ ) depende de si la persona se ha movido hacia un mejor estado de salud. El subíndice  $j$  toma valores de 1 y 0, y se refiere al sector de participación (buena o mala salud en algunos modelos, o mejora o no mejora de salud en otras de nuestras estimaciones), mientras que el subíndice  $i$  se refiere a las personas. En la ecuación 1 la función indicador  $D$  toma el valor de 1 si la persona tiene buena (o mejor) salud y cero en caso contrario.

$$Y_i = D_i Y_{i1} + (1 - D_i) Y_{i0}, \quad i = 1, \dots, n. \quad (1)$$

El ingreso en cada estado es función de un conjunto de variables exógenas ( $X_i$ ) y del estado de salud. Así, el ingreso en la ecuación (2) es función de la educación y otras variables, así como de un término de error que cumple con los supuestos del modelo de MCO (de aquí en adelante, las variables exógenas y los coeficientes corresponden a vectores):

$$Y_{ij} = X_i \beta_{ij} + s_{ij} + \varepsilon_{ij} \quad j = 0, 1. \quad (2)$$

El estado de salud ( $s_{ij}$ ) es función de un conjunto de características personales y comunitarias exógenas ( $X_{si}$ ), así como de una variable no observada  $U_{ij}$ , la cual incide sobre la salud (ecuación 3). Esta última variable no es un error estadístico, sino una habilidad de la persona para tener mejor estado de salud; observamos que es específica al sector en que participa la persona (en esta aplicación definido por el estado de salud). Esta variable no observada refleja elementos idiosincráticos, entre los que pueden estar el costo en esfuerzo físico y la propensión biológica de la persona de mejorar su salud.

$$s_{ij} = \gamma_{ij}X_{si} + U_{ij}, \quad (3)$$

No estamos especificando que la salud misma pueda ser función directa del ingreso, pero al usar el modelo de Roy hay una relación implícita, ya que el ingreso depende del sector elegido por la persona. Los individuos mejoran su estado de salud cuando se cumplen condiciones para ello, de acuerdo con unaregladecisiónqueincluyeleeffectosobreel ingreso de la salud y que tiene como argumentos el conjunto de variables exógenas  $Z_i$  y también posiblemente las variables en  $X$ . Estas afectan a la variable latente  $D^*$  (ecuación 4). La variable  $V$  tiene una distribución arbitraria en la población que denotaremos por  $F$ , de manera que  $F(V)$  es la probabilidad acumulada de que la persona mejore en salud, es decir, es el score de propensión  $P(Z)$ ; un mayor valor de  $V$  significa que es menos probable que la persona mejore en salud. Podemos notar que las variables no observables  $U_{ij}$  inciden sobre la regla de decisión, y pueden estar correlacionadas con la  $V$ .

$$\begin{aligned} I^* &= \mu_{10}(X_i Z_i) - V_i \\ D_i &= 1 \text{ si } D_i^* > 0, \\ D_i &= 0 \text{ en otro caso} \end{aligned} \quad (4)$$

También podemos observar que:

$$Y_{i1} - Y_{i0} = [\beta_{i1} - \beta_{i0}]X_i + [\gamma_{i1} - \gamma_{i0}]X_{si} + [U_{i1} - U_{i0}] + [\varepsilon_{i1} - \varepsilon_{i0}] \quad (5)$$

En resumen, la buena (o mejora) en salud resulta de la interacción de elementos exógenos y de la decisión del trabajador. Las variables  $Z$  pueden incluir a las variables exógenas  $X$  y  $X_s$  y los coeficientes de las ecuaciones 2 y 3 pueden ser fijos o aleatorios.

Un elemento crucial en la evaluación empírica es la heterogeneidad. Ausencia de heterogeneidad en el efecto significa que los coeficientes  $\gamma_i = \gamma$ , y ausencia de he-

terogeneidad en la selección significa que  $\mu_i = \mu$ . En ausencia de heterogeneidad en el tratamiento, la ecuación 4' depende principalmente de  $[U_{i1} - U_{i0}]$ , la diferencia en habilidades no observables para generar ingreso en cada sector (es decir, en este argumento permanece heterogeneidad en los no observables). Es posible que una estimación por MCO de las  $\gamma$  esté sesgada pues la hipótesis de que las  $U_{ij}$  resumen beneficios y costos para el individuo de mejorar o no su estado de salud, incluye la posible correlación con el valor de otras variables exógenas  $X$  y  $X_s$ . Por otro lado, la ecuación de selección puede corregir ese sesgo si algunos elementos de  $Z$  afectan la probabilidad de mejorar la salud de forma independiente a los factores que determinan el ingreso. También podemos notar que el modelo de regresiones intercambiables (*switching regressions*) nos permite estimar coeficientes específicos a cada sector, y si los coeficientes y la varianza de los errores son iguales entre sectores en la ecuación (2) tenemos el modelo de efectos de tratamiento (Brave y Walstrum, 2014). Nuestras estimaciones incluyen resultados de MCO, VI, diferencias, diferencias en diferencias, regresiones intercambiables y modelo de Roy con variables instrumentales locales.

En nuestro análisis encontramos las siguientes situaciones. Primero, los términos  $U_{ij}$  generan heterogeneidad del efecto sobre los ingresos según el sector que se elija, y corresponden al parámetro de habilidad por sector en el modelo de Roy. Segundo, el modelo no restringe las relaciones de covarianza entre  $U_0$ ,  $U_1$  y  $V$ . Por ejemplo, una persona puede tener valores relativamente elevados de  $U_0$  y  $U_1$  (correlación positiva de las habilidades, pero para otra puede darse lo opuesto (de manera que tiene  $U_0$  alta y  $U_1$  baja o viceversa). Estas covarianzas, así como las varianzas de habilidades dentro de cada sector pueden derivar en patrones distintos de selección entre poblaciones. Por ejemplo, personas de alta habilidad se definen como aquellas para quienes mejorar en salud lleva a un incremento importante en ingresos en el sector que llamaremos 1 (es decir, en  $Y_1$ , pues el sector 1 es el de mejora en salud). Si el mercado ofrece una alta varianza en  $Y_1$ , ese sector atrae a personas de alta habilidad para mejorar en salud ( $U_{ij}$ ). Por el contrario, personas de baja  $U_1$  se van al sector de baja varianza, pues ello les garantiza un ingreso cercano el promedio de ese sector; es decir, personas que mejoran poco su ingreso al mejorar en salud pueden optar por no mejorar su salud para no incurrir en el costo de hacerlo (costo que en general no observamos directamente).

Para la estimación de modelos con heterogeneidad esencial, Heckman y Vytlacil (2001) expresan la ecuación (6). Por brevedad, incluimos los coeficientes  $\gamma$  en las  $\beta$  en los siguientes párrafos de esta sección. En la ecuación 6, las  $\alpha$  son interceptos de las ecuaciones sectoriales, de manera que los coeficientes estimados se pueden interpretar como los del sector de no tratamiento (no mejora en salud en nuestro caso;  $(\alpha_0 + x\beta_0)$ , más el valor esperado de las ganancias por recibir tratamiento  $((\alpha_1 - \alpha_0)p + \{x(\beta_1 - \beta_0)\}p)$ , más una función  $K(p)$  que se escribe en la ecuación 7. Esta última mide el valor esperado de la ganancia en tratamiento (el segundo término del lado derecho de la  $K(p)$ ), más el valor de la habilidad no observada en ausencia de tratamiento (la  $U_0$ ).

$$E\{Y|X = x, P(Z) = p\} = \alpha_0 + x\beta_0 + (\alpha_1 - \alpha_0)p + \{x(\beta_1 - \beta_0)\}p + K(p) \quad (6)$$

$$K(p) = E\{U_0|P(z) = p\} + E\{U_1 - U_0|P(z) = p\}p \quad (7)$$

La estimación del modelo regresiones cambiantes y del modelo MGR incluyen el coeficiente para la variable  $K(p)$ , cuya derivada  $(\partial K(p)/\partial p)$  incide sobre el efecto marginal de tratamiento.

Los efectos marginales de tratamiento para el modelo paramétrico descansan en el supuesto de distribución normal de los residuales:

$$MTE(X = x, U_D = u_D) = (\alpha_1 - \alpha_0) + x(\beta_1 - \beta_0) + (\rho_1 - \rho_0)\Phi^{-1}(u_D)$$

El modelo semiparamétrico, de variables instrumentales locales: (i) estima el score de propensión,  $P(Z)$ ; (ii) regresa localmente  $X$ ,  $XxP(Z)$  y  $Y$  sobre  $P(Z)$  y calcula los residuales  $\hat{e}_Y, \hat{e}_X, \hat{e}_{X \times P(Z)}$ ; (iii) regresa  $\hat{e}_Y$  sobre  $\hat{e}_X, \hat{e}_{X \times P(Z)}$  para obtener los valores de  $\beta_0$  y  $(\beta_1 - \beta_0)$ ; (iv) regresar los residuales obtenidos en el paso anterior sobre  $P(Z)$  para llegar a la estimación de  $\partial K(p)/\partial p$  y del efecto marginal de tratamiento.

Podemos notar que, si las  $U$  son aproximadamente constantes para cada persona e iguales entre sectores (o la persona no cambia de sector), podemos obtener las diferencias mediante la ecuación (5) entre puntos en el tiempo separados por un intervalo de duración  $\tau$ , para eliminar el efecto de la heterogeneidad. No encontramos una razón teórica para invocar este tipo de estacionariedad, que podría justificar el uso de estimadores en diferencias o de diferencias en diferencias con fines de identificación.

Dentro de este marco planteamos distintas restricciones de causalidad para investigar la sensibilidad del ingreso laboral al estado de salud. A continuación, defi-

nimos cinco especificaciones que estiman la relación de interés. Cada una involucra supuestos específicos de causalidad, y si bien el modelo de Roy es el más general conceptualmente, lo que determinará nuestra decisión sobre el modelo preferido es la consistencia de los resultados y el análisis económico sobre su validez.

*Modelo I. El modelo de salud exógena.* Todas las variables del lado derecho en la ecuación (2) son exógenas y no hay selección entre sectores. En forma equivalente, en la ecuación (3) no hay heterogeneidad entre la población en la variable  $U$  y las variables del lado derecho de la ecuación (3) son exógenas. El término  $\varepsilon_i$  es un error que cumple con los supuestos del modelo de MCO. Podemos estimar la regresión con sólo las variables de la ecuación (2).

*Modelo II. El modelo de salud endógena.* Ahora consideramos la posibilidad de que la variable no observada  $U$  esté correlacionada con algunos regresores. Por ejemplo, personas con distinto nivel de educación pueden tener distinto grado de acceso a los servicios de salud y por ello tener información diferente sobre sus condiciones. Para atender este problema estimamos dos modelos de variables instrumentales. El primero es uno de mínimos cuadrados en dos etapas (2SLS), y el otro es un modelo de método generalizado de momentos (GMM). En ellos la variable a instrumentar es el estado de salud. Estos modelos mantienen la especificación de la ecuación (2), suponiendo que hay un sólo sector, y usando a la ecuación (3) como especificación de la primera etapa.

*Modelo III. El modelo en diferencias.* Una implicación de la ecuación (2) y de suponer que los regresores son exógenos (por ejemplo, como el en Modelo I), es que bajo los supuestos mantenidos debe darse valores estadísticamente similares de los coeficientes entre las estimaciones con variables expresadas en niveles o en diferencias. Es decir, podemos estimar la ecuación vía MCO o IV y comparar los resultados con los modelos correspondientes estimados con las variables en niveles. Aún si hay heterogeneidad individual, mientras esta sea aproximadamente constante a lo largo del tiempo para cada persona, el modelo produce estimadores consistentes. Alternativamente, si la heterogeneidad es importante y se correlaciona con alguna de las variables  $X$ , en el caso de MCO observaremos diferencias entre las estimaciones, mientras que en el caso de VI podemos observar un resultado similar si contamos con un instrumento válido.

*Modelo IV. El modelo de regresiones cambiantes (switching regressions, SR).* Hemos explorado la posibilidad de que la condición de salud sea una variable endógena por medio del método de variables instrumentales. Con ello hemos explorado el cálculo del efecto promedio. Ahora nos preguntamos cómo la endogeneidad y la heterogeneidad pueden incidir sobre el efecto. Para ello requerimos de un modelo económico que plantee la condición de salud como resultado de una elección individual. Esta especificación y la siguiente hacen eso; ambas son versiones del modelo de Roy.

En el modelo de regresiones cambiantes permitimos que los coeficientes asociados a las variables exógenas  $X$  (es decir,  $\beta$  y  $\gamma$ ) sean específicos a cada sector. El efecto de la variable de interés se estima como la diferencia promedio (efecto promedio de tratamiento) de los valores proyectados de ingreso para toda la muestra en cada sector. Este plantea que las personas eligen entre dos estados, que en nuestra aplicación definimos como la mejoría en salud (que lleva al ingreso  $Y_{i1}$ ) y la no-mejoría en salud (que lleva al ingreso  $Y_{i0}$ ). La forma en que aplicamos el modelo es la siguiente; la persona decide entrar en cierto estado ("tratamiento") de acuerdo con las ventajas que ello le reporta en comparación a no hacerlo. Para nosotros es de interés que la persona se mueva hacia cierto estado de salud. La primera etapa del modelo es un probit que estima la ecuación (4); es el mismo en este modelo y en el siguiente.

El modelo de regresiones cambiantes es paramétrico en un sentido adicional, ya que supone una distribución normal para las variables  $U_{i0}$ ,  $U_{i1}$  y  $V_i$ . Esto restringe a las funciones de efecto marginal de tratamiento a ser continuas y a comportarse normalmente. Se producen parámetros específicos a cada sector que son fijos dentro del sector. Cornelissen *et al.* (2016) hacen una exposición general de estos modelos.

*Modelo V. El modelo generalizado de Roy (MGR).* El modelo de Roy permite modelar la relación de interés en forma heterogénea. Buscamos una función que describe el efecto marginal de tratamiento entre la población en función de la probabilidad de haber mejorado en salud (del score de propensión, propensity score). Esa es la probabilidad de seleccionarse hacia el tratamiento, condicional en las variables que determinan la decisión,  $P(X,Z)$ . El score de propensión se obtiene con el mismo modelo probit que el modelo de regresiones cambiantes.

Con este enfoque, suponemos que las ganancias en ingreso por estar más sano resumen todas las ganancias y costos para el individuo. Esto se debe a que las variables exógenas observables y no observables incluyen los costos y beneficios monetarios y subjetivos de lograr estar en un mejor estado de salud (Eisenhauer, Heckman y Vytlačil, 2015). La estimación que presentamos es semiparamétrica, lo que significa que se hace una estimación de VI a nivel local, en torno a valores del score de propensión. Esto es lo que permite calcular efectos marginales.

### 3. Datos

Nuestra fuente de datos es la Encuesta Nacional sobre Niveles de Vida de los Hogares (ENNVIIH). Esta es una encuesta longitudinal que ha producido tres rondas: 2002, 2005-2006 y 2009-2012. Nuestra regla de selección se basa en la edad y en la disponibilidad de información. Consideramos sólo a personas que hayan tenido de 25 a 54 años al momento de la encuesta (en cualquiera de las rondas), y que tengan información en las variables de interés. Es decir, podemos tener personas nacidas aproximadamente entre 1947 y 1986. La selección por edad se debe a que nos interesa la relación entre salud y trabajo, y los jóvenes y las personas cercanas a la edad de retiro pueden tener conductas distintas a los del grupo seleccionado que afectan al ingreso. Por supuesto, estudiar cómo los adultos mayores dejan de trabajar por causas de salud es importante, pero no es el objeto de esta investigación. Para los jóvenes es posible que antes de los 25 las decisiones de trabajo dependan principalmente de las decisiones de educación, que tampoco son nuestro objeto de interés en esta ocasión.

Todas las tablas y gráficas corresponden a personas entre 25 y 54 años al momento de la ronda correspondiente de la ENNVIIH.

La Tabla 1 describe las variables empleadas. La Tabla 2 muestra estadísticas de las variables, por sexo y por estatus laboral. Los modelos empleados pueden tener menos observaciones que las contenidas en estas tablas porque para algunas variables, como las mediciones de salud, la encuesta no provee información para todos los individuos. También, los modelos que emplean datos de más de una ronda están sujetos a abandono (*attrition*). No hemos explorado si el abandono de la encuesta genera sesgo en nuestras estimaciones.

Tabla 1  
Descripción y notas sobre variables

*Variable dependiente.*

Ingreso mensual en trabajo principal, pesos. En los modelos que emplean los valores en más de un periodo, se ajustan por el Índice Nacional de Precios al Consumidor en julio del año correspondiente.

*Variable endógena*

Salud declarada (*salud\_declara*): valores de 1 a 5; 1 = muy buena, 5 = muy mala.

Diferencia en salud declarada (*dfsalud*)

Mejora en salud: 1 si Diferencia en salud declarada < 1 (si la variable de salud auto-declarada en la ronda 3 menos el valor en la ronda 1 es cero o negativo y si la salud declarada en la ronda 3 es de regular a muy buena); 0 en otro caso.

Trabajo: trabajó o hizo una actividad que ayudó al gasto del hogar.

*Variables exógenas*

Educación: se mide por nivel terminado. Se asignaron los años que corresponden a la terminación de ciclos oficiales (grado universitario se asignaron 16 años, y a posgrado 20 años)

Horas: semanales en trabajo principal.

Afiliación a la seguridad social: incluye a todos los organismos federales y estatales; no incluye al Seguro Popular

Discapacidad (*discap*): respuesta de la persona a la pregunta de si tiene alguna discapacidad.

IMC: índice de masa corporal

Desviación de IMC: IMC - 25.

Exceso de IMC (*ximcmas*): IMC - 25 si IMC > 25, y cero en otro caso. También empleamos esta variable al cuadrado (*ximcmas2*).

Promedio de 8 limitaciones de movilidad (*salud\_discap*): cada una se califica de 1 = fácilmente a 5 = no lo podría hacer. Las actividades son: poder llevar una cubeta pesada, caminar 5 km, inclinarse, arrodillarse, subir escaleras, vestirse sin ayuda, pararse de una silla, bañarse sin ayuda, levantarse del suelo.

*Instrumentos*

Hipertensión, Tensión medida dentro del rango y Tensión medida arriba del rango: mediciones directas efectuadas por la encuesta.

Declaración de Diabetes o Hipertensión: por parte de la persona entrevistada.

Estado mental: promedio de 21 variables. Los valores van de 1 = sí, algunas veces, a 4 = no. Se refieren a si la persona se ha sentido triste, ha llorado, ha dormido mal, ha despertado sin ánimo, le cuesta concentrarse, ha disminuido ganas de comer, se ha obsesionado, ha disminuido su interés sexual, siente menor remordimiento, siente presión en el pecho, se ha sentido nerviosa, se ha sentido cansada, se ha sentido pesimista, ha tenido dolor de cabeza-nuca, se ha sentido irritada, se ha sentido insegura, es menos útil a su familia, tiene miedo de algunas cosas, ha sentido deseos de morir, ha perdido el interés en las cosas o se ha sentido sola.

Enfermedades crónicas de padre o madre: declaración de la persona.

Tabla 2  
Estadística descriptiva, mujeres y hombres de 25 a 54 años de edad por estatus laboral en ENNVIH-3

Variable	Mujeres				Hombres			
	Trabajan (n = 2,288)		No trabajan (n = 4,312)		Trabajan (n = 5,000)		No trabajan (n=1150)	
	Media	DE	Media	DE	Media	DE	Media	DE
Trabajo, R3	1	0	0	0	1	0	0	0
Ingreso laboral mensual, R3	2,736	4,883	62	697	2,946	4,767	317	1,836
Horas totales, R3	38.69	19.25	1.35	7.14	46.1	18.36	2.72	10.4
Seguridad social, R3	0.52	0.5	0.32	0.47	0.41	0.49	0.19	0.4
Unión marital, R3	0.56	0.5	0.79	0.41	0.8	0.4	0.53	0.5
Discapacidad, R3	0.01	0.1	0.02	0.12	0.01	0.1	0.04	0.19
IMC - 25	3.84	5.58	4.36	5.9	2.99	4.66	3.11	5.07
(IMC - 25) si IMC > 25	4.44	4.88	4.96	5.16	3.62	3.88	3.82	4.21
Embarazada, R3	0.02	0.15	0.03	0.16				
Declaración de salud, R3	2.44	0.7	2.56	0.7	2.38	0.69	2.54	0.74
Mejoró salud entre R1 y R3	0.62	0.49	0.59	0.49	0.55	0.5	0.32	0.47
Padre tuvo enfermedad crónica	0.3	0.46	0.27	0.45	0.26	0.44	0.16	0.37
Madre tuvo enfermedad crónica	0.33	0.47	0.33	0.47	0.29	0.45	0.18	0.39
Discapacidad de salud, R3	0.27	0.64	0.46	0.9	0.29	0.58	0.34	0.79
Tensión alta	0.12	0.33	0.15	0.36	0.2	0.4	0.18	0.38
Declara hipertensión	0.11	0.32	0.13	0.34	0.07	0.26	0.07	0.25
Declara diabetes	0.06	0.24	0.09	0.28	0.07	0.25	0.06	0.24
Estado mental, R3	0.37	0.4	0.39	0.42	0.22	0.31	0.33	0.39

Nota: R1 y R3 se refiere a la ronda de la ENNVIH

Una pregunta prevalente en la literatura es cómo medir la salud. Algunas encuestas, como la ENNVIH, buscan combinar auto-declaraciones de la persona y mediciones directas o posiblemente menos subjetivas del estado de salud. Por ejemplo, se efectúan mediciones de presión arterial y glucosa. Sobre la auto-declaración, la encuesta pregunta sobre el estado de salud actual, sobre si se ha dado un deterioro o mejoras desde hace un último año, y si ha habido un evento grave en los últimos cuatro años, así como preguntas específicas acerca de la presencia de alguna enfermedad específica. En la Tabla 3 hemos escrito las correlaciones en variables de la ronda más reciente, incluyendo también la variable que hemos construido sobre estado mental. La variable de algún problema grave en últimos cuatro años es la que generalmente tiene menor correlación con las demás, lo que posiblemente denota que se trata de problemas de baja prevalencia pero que generan daño elevado. Por otro lado, la salud auto-declarada actual tiene una correlación más alta (entre .17 y .24) con un empeoramiento en el último año, el indicador de estado mental, la diabetes y la hipertensión. Diabetes e hipertensión tienen una correlación de 0.17.

Los resultados presentados a continuación utilizan a la variable de salud auto-declarada como principal variable dependiente. Dependiendo de la especificación, empleamos el valor de la variable en la tercera ronda de la encuesta, en cada una de las tres rondas, o en la forma de una variable binaria. En la ENNVIH, la auto-declaración del estado de salud toma valores de 1 si la persona considera tener mejor salud, y 5 si considera tener una salud

peor. En los modelos MCO, IV, diferencias, y diferencias en diferencias, empleamos la variable sin transformarla. Para los modelos que involucran una selección por sector construimos una variable binaria y hemos experimentado con dos transformaciones. En un caso, la variable binaria mide “buena salud” si la variable de salud auto-declarada es menor a 3; en el otro, asignamos un valor de 1 a “salud mejora” si el valor se mantuvo igual o mejoró entre la primera y la tercera ronda, y de 0 en otro caso. Esta transformación es necesaria porque en esos modelos sólo podemos emplear variables dependientes binarias.

Podemos señalar que probamos los modelos con las variables que miden deterioro de salud durante el último año y el estado mental, y en general los resultados estadísticos fueron menos significativos. La variable de deterioro en salud es también una variable auto-declarada de salud. La segunda, el “estado mental” es el promedio de 21 preguntas sobre percepciones de la persona sobre su estado mental, las cuales no son en general propiamente mediciones de salud (por ejemplo, si tiene problemas de concentración, si está triste, si le duele la cabeza y otras); esta variable sí es utilizada como una variable exógena en el modelo de regresión.

En apoyo a la decisión de emplear la variable de salud auto-declarada podemos señalar a evidencia de que esta medición tiene una correlación importante con el estado real y eventos futuros de mortalidad (ver, por ejemplo, McGee *et al.*, 1999, Miilunpalo *et al.*, 1997 y Tamayo-Fonseca, *et al.*, 2013). Jylhä (2009) propone que para entender por qué la auto-declaración es un predictor bueno

Tabla 3  
Coeficientes de correlación de variables relacionadas con salud, ENNVIH-3

	Ha empeorado la salud en último año	Salud declarada actual	Ha tenido evento grave de salud en últimos 4 años	Promedio de 21 variables de estado mental	Diabetes	Hipertensión
Ha empeorado la salud en el último año	1					
Salud declarada actual	0.24	1				
Ha tenido evento grave de salud en los últimos 4 años	0	0.06	1			
Promedio de 21 variables de estado mental	0.11	0.19	0.07	1		
Diabetes	0.07	0.17	0.07	0.08	1	
Hipertensión	0.05	0.17	0.07	0.1	0.12	1

de la salud futura y de la mortalidad las preguntas clave son: qué es lo que sabe la persona y cómo sabe lo que sabe. Propone que hay información única provista por sensaciones del cuerpo que está disponible sólo al individuo, sensaciones que reflejan disfunciones fisiológicas tales como procesos inflamatorios. Por otro lado, la auto-declaración está sujeta a condicionamientos culturales y elementos sociales y económicos (por ejemplo, el acceso a servicios de salud e información). Sin embargo, como indica la misma autora, hay una amplia literatura confirmando que, a pesar de tales condicionamientos, la auto-declaración provee información sobre el futuro que no es revelada por mediciones "objetivas" del estado de salud, tales como exámenes clínicos y diagnósticos médicos.

A continuación, describimos estadísticamente las variables dependientes y algunas de las principales variables independientes e instrumentales empleadas. En la Tabla 4 podemos ver que, en promedio, la salud auto-declarada no cambia entre las rondas de la encuesta, lo que

es previsible pues no hay razones para esperar un deterioro o una mejora general en una población nacional en un intervalo de menos de una década. Por otro lado, sí hay cambios en la posición de las personas en la distribución. En los extremos hay quienes mejoran (valores de -3) o empeoran mucho (valores hasta 4), y en casi todos los casos los percentiles 25 y 75 tienen valores de -1 y 1, respectivamente, lo que significa que al menos la mitad de la población declaró un cambio en su estado entre las encuestas. Cabe señalar que, en el estudio citado de Milunpalo *et al.* (1997), el estado de salud auto-declarado es útil para predecir visitas al médico hasta un año adelante, y mortalidad general hasta más de 10 años adelante, pero menos para predecir el inicio de enfermedades crónicas, las cuales tienen baja prevalencia; en opinión de esos autores la auto-declaración del estado de salud es una buena aproximación al estado real de salud de las personas de edad media, que son la población objetivo de este estudio.

Tabla 4  
Cambios en las variables de salud autodeclarada entre rondas

Variable	n	Media	S.D.	Cuantiles				
				Min	0.25	Mdn	0.75	Max
<i>Mujeres que trabajan</i>								
dfsald_1	2107	0.03	0.79	-3	0	0	0	4
dfsald_2	2070	0.1	0.83	-3	0	0	1	3
<i>Mujeres que no trabajan</i>								
dfsald_1	3809	0	0.81	-3	0	0	0	3
dfsald_2	3872	0.08	0.83	-3	0	0	1	3
<i>Hombres que trabajan</i>								
dfsald_1	3587	-0.01	0.78	-3	0	0	0	3
dfsald_2	3575	0.11	0.81	-3	0	0	1	3
<i>Hombres que no trabajan</i>								
dfsald_1	660	0.05	0.8	-3	0	0	0	3
dfsald_2	701	0.16	0.83	-3	0	0	1	3

dfsald\_1: diferencia entre la ENNVih-3 y la ENNVih-1.

dfsald\_2: diferencia entre la ENNVih-3 y la ENNVih-2.

En las Gráficas 1 a 6 se describen algunos elementos que juegan un papel en las estimaciones de la siguiente sección. Las Gráficas 1.h y 1.m muestran las distribuciones de ingresos mensuales de hombres y mujeres en las tres rondas; para estas gráficas hemos ajustado los valores por inflación a valores de 2009. Los valores modales están debajo de \$5,000 para ambos sexos; para ponerlo en perspectiva, el salario mínimo mensual en 2009 era \$1,616 pesos. La moda de ingresos en estas gráficas está en torno a 3 salarios mínimos o un poco más (para efecto visual, quitamos observaciones de más de \$30,000 pues generan una cola larga, pero sí usamos esos datos en el análisis de regresión). También vemos que el ingreso modal es inferior para las mujeres (Gráfica 1.a).

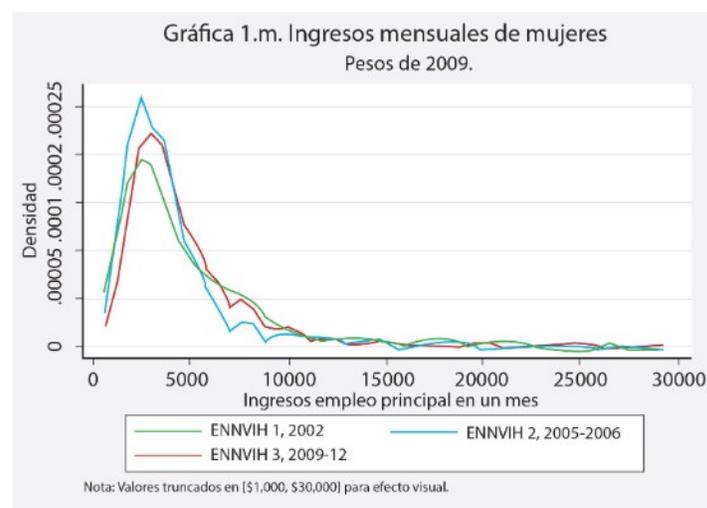
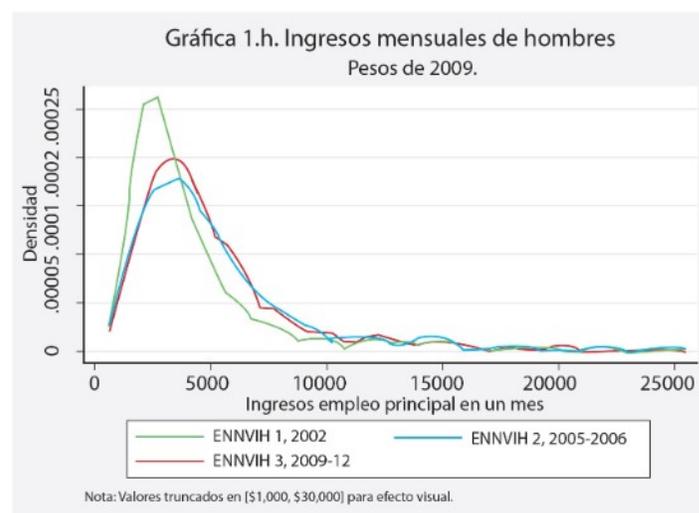
La auto-declaración de salud es predominantemente “buena” o regular” para ambos sexos (Gráficas 2.h y 2.m). Los hombres trabajadores tienen más eventos “buenos” y “regulares” y menos “malos” que los no-trabajadores (Gráfica 2.th); para las mujeres, ese patrón no es visible gráficamente (Gráfica 2.tm). Las Gráficas 3.h y 3.m muestran el IMC para trabajadores y no trabajadores, por sexo. Para ambos sexos, las distribuciones para no trabajadores se inclinan hacia mayor exceso de IMC; es decir, los trabajadores son más delgados que los no trabajadores.

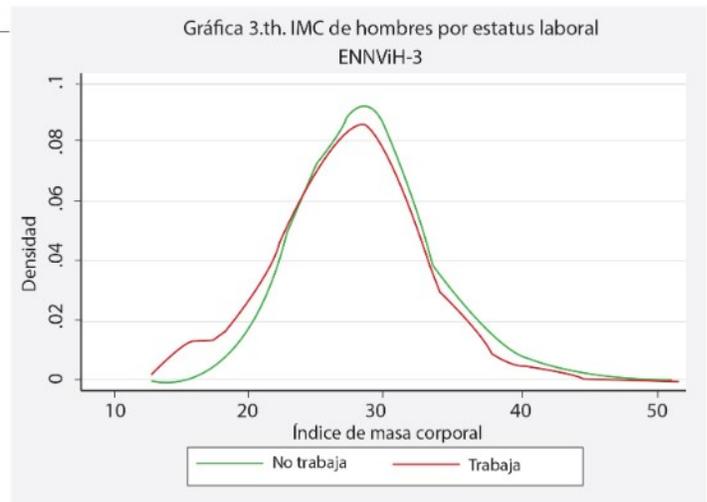
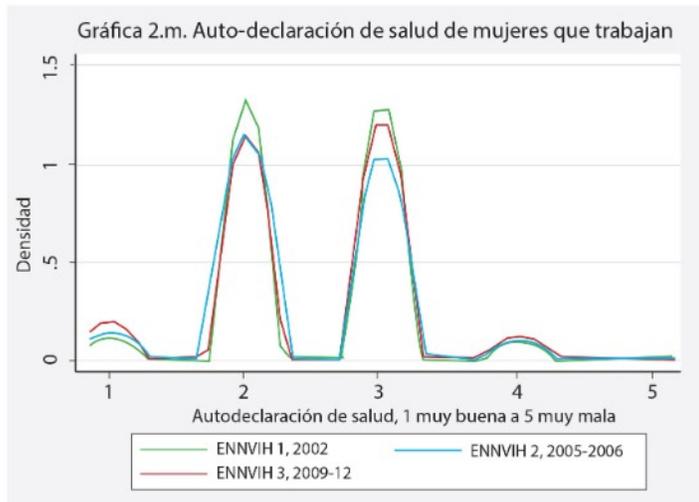
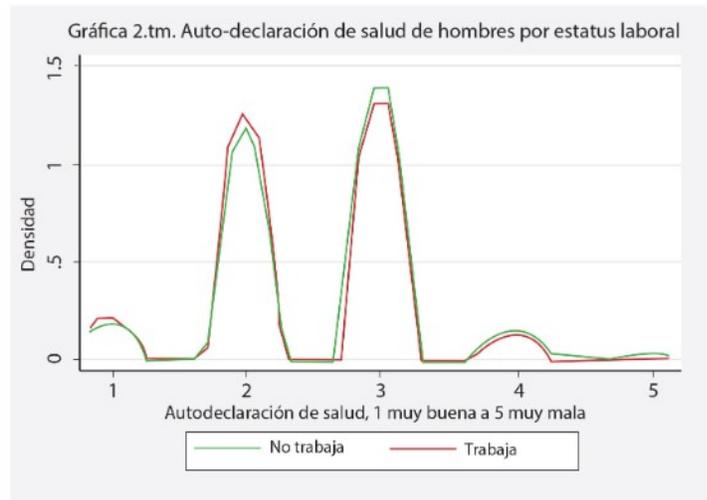
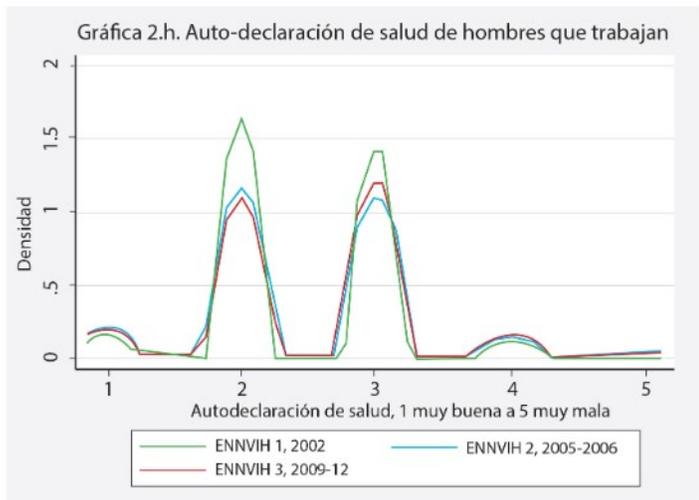
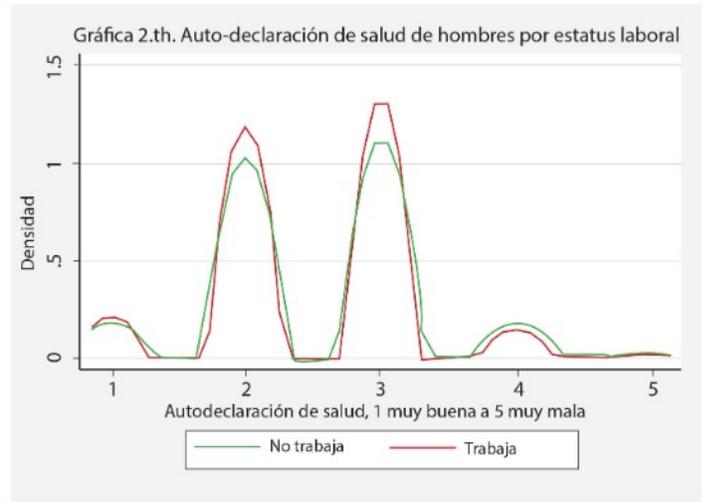
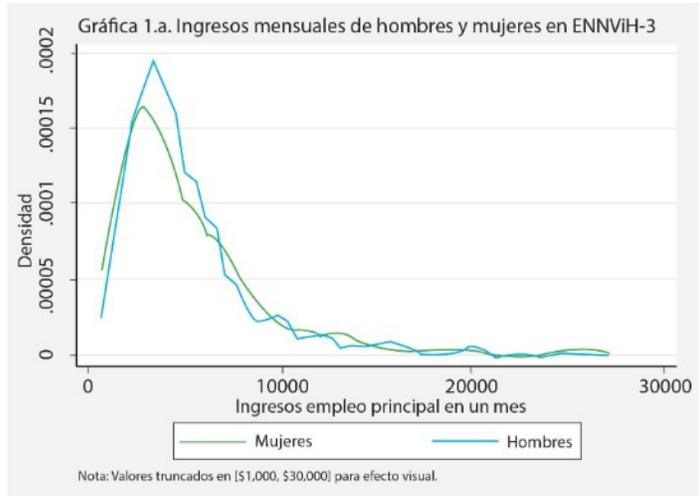
Las horas de trabajo en México son largas. Usualmente se ocupa uno de los primeros lugares en comparaciones internacionales; según el reporte de empleo de la OCDE (2016), en 2015 los asalariados trabajaron 2,346 horas al año, o cerca de 47 horas por 50 semanas. La moda para hombres y mujeres en la ENNViH está un poco debajo de las 50 horas (Gráficas 4.h y 4.m). La densidad de horas trabajadas de mujeres tiene colas más gordas hacia el margen de trabajo de tiempo parcial (es decir, trabajan menos de tiempo completo).

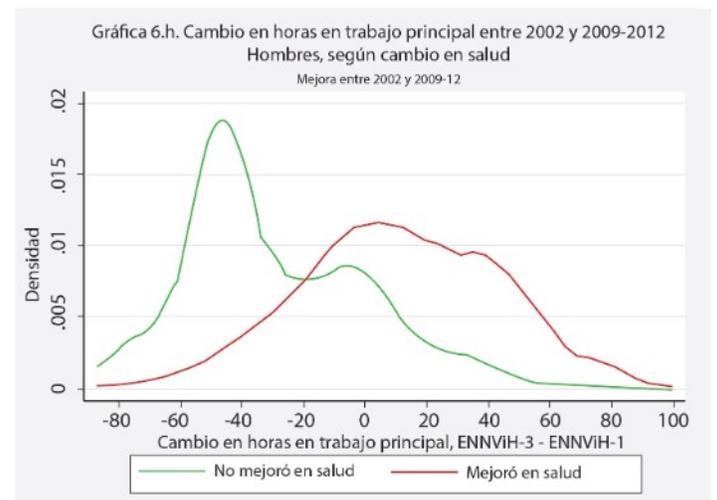
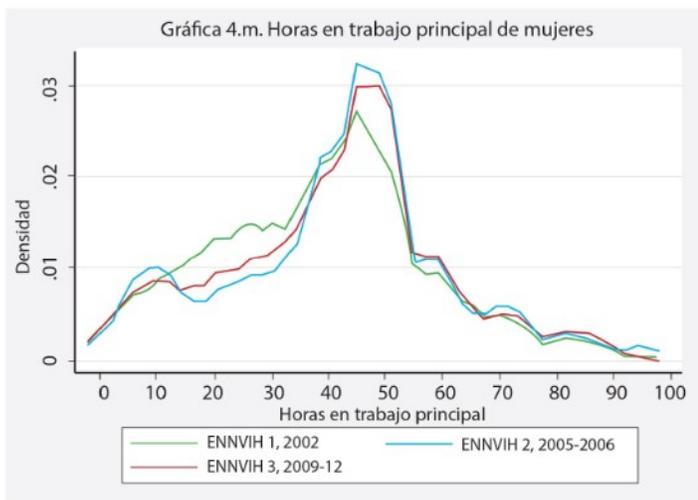
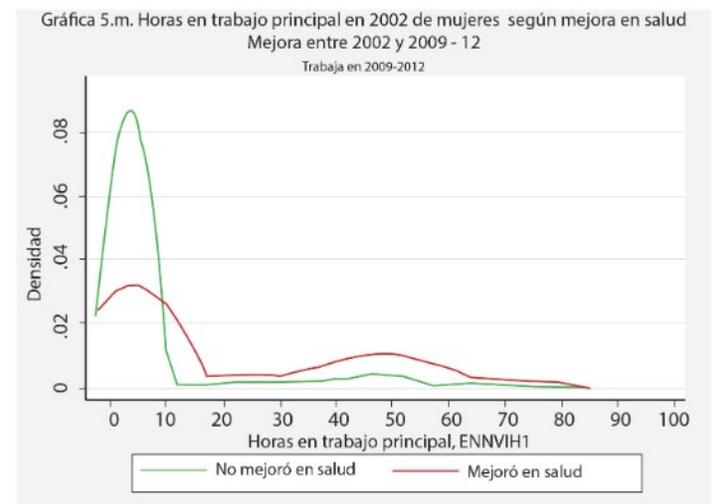
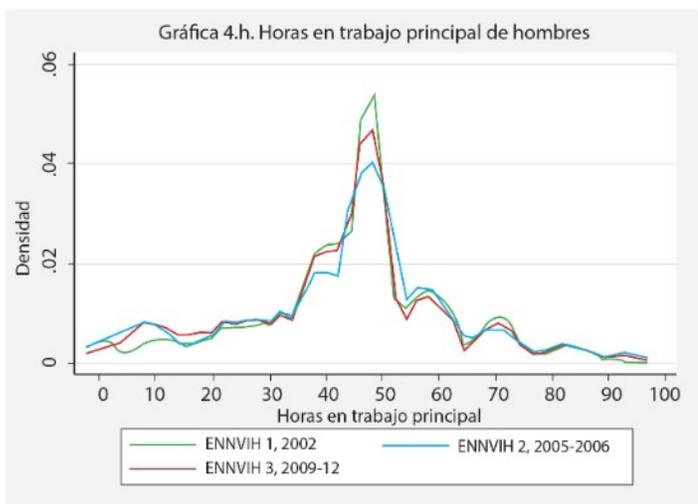
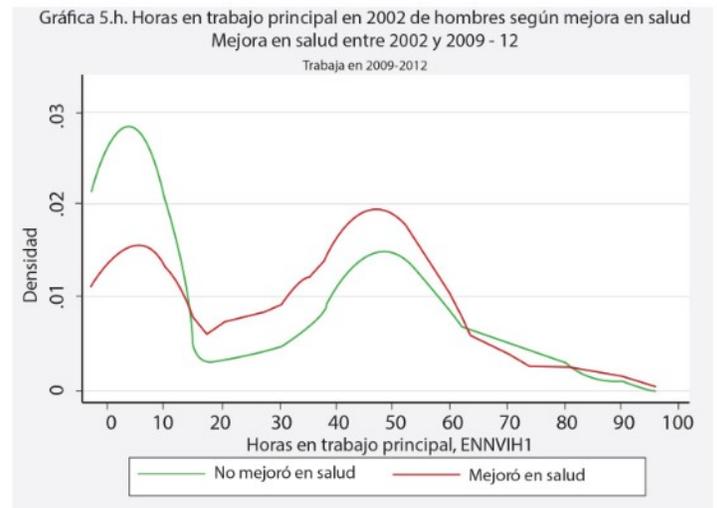
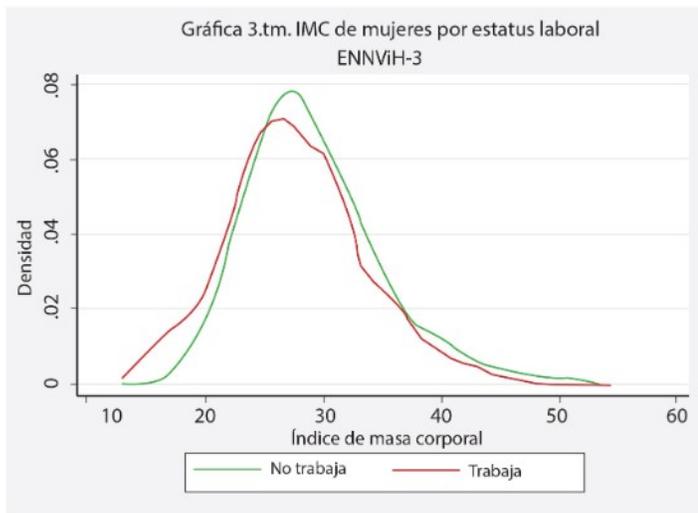
Al combinar información de horas y mejora en salud, y horas trabajadas en la primera ronda de la encuesta (esto es en 2002) encontramos un patrón peculiar para personas que trabajaban en 2009-2012 (es decir, en la tercera ronda de entrevistas): para quienes mejoró la salud se dieron más horas de trabajo en la primera ronda, tanto para hombres como para mujeres (Gráficas 5.h y 5.m). El patrón es más marcado para las mujeres. Es decir, quienes trabajan se pueden dividir en dos, los que mejoraron y los que no mejoraron en salud. En ambos grupos tenemos quienes trabajaban de tiempo completo en el pasado y quienes no lo hacían (de ahí la distribución bimodal). Finalmente, quienes mejoraron en salud trabajaron más horas desde el periodo inicial (2002

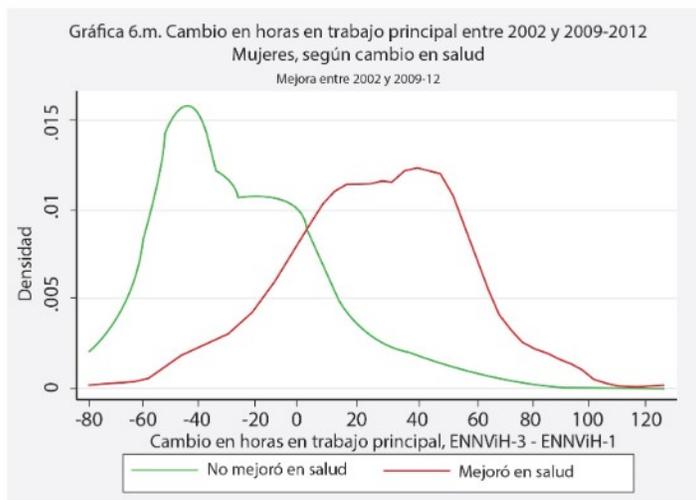
en este caso). Lo interesante es que se trata de las horas en el año inicial (2002), y los patrones se dan al considerar el cambio en salud hacia el futuro. Cabe señalar que a estas gráficas no entran quienes no trabajaban en 2009-2012 (sea que lo hayan hecho o no en 2002), de manera que el patrón bimodal se asocia a quienes no trabajaban en 2002 y sí en 2009-12; si empleamos los datos de quienes trabajaban en 2002, tenemos una distribución similar a la de corte transversal (es decir, Gráficas 4.h y 4.m).

Finalmente, las Gráficas 6.h y 6.m relacionan el cambio en horas trabajadas entre 2002 y 2009 con el cambio en el estado auto-declarado de salud. Para ambos sexos, el cambio en horas para quienes mejoraron en salud está claramente a la derecha de quienes no lo hicieron.









#### 4. Resultados

A lo largo de esta sección revisamos los resultados de cada modelo para llegar a una conclusión sobre el tamaño del efecto de la salud sobre los ingresos laborales

La Tabla 5 muestra los resultados del Modelo I, que supone que el estado de salud es una variable exógena. La muestra comprende sólo a trabajadores de la ronda 2009-12. En términos estadísticos, la variable de estado auto-declarado de salud no tiene un efecto que sea estadísticamente significativo. En este modelo y en todos los que siguen la educación muestra un efecto creciente sobre los ingresos y no lo comentaremos más.

Con respecto a otras variables, destacan los efectos de la estatura, la discapacidad y la participación en la seguridad social. Para facilitar la interpretación hemos añadido a esta tabla y a las subsiguientes algunos cálculos de elasticidades, márgenes predictivos y pruebas de hipótesis para variables y valores representativos. Los cálculos se condicionan por los valores promedio de las variables exógenas, o por valores a una y dos desviaciones estándar. Vemos que la elasticidad del ingreso con respecto al mal estado de salud, medida entre la media del estado de salud y los valores más altos de la muestra, va de -0.99 a -0.198 para hombres, y para mujeres de 0.246 a 0.654. Recordemos que los valores de la variable de salud auto-declarada van de 1 (muy buena) a 5 (muy mala), de manera que un coeficiente negativo indica pérdida de ingresos al deteriorarse la salud. Así, por ejemplo, para hombres, un deterioro en una desviación estándar es de 0.69, o 28% de estado de salud promedio, y aplicando las elasticidades en el rango indicado nos da un resultado de 27% de disminución del ingreso en la media, pero sólo 5% a una

desviación estándar (es decir, para personas de salud más mala el efecto se hace más chico). Otra aproximación la proporciona el sub-panel de márgenes predictivos, que muestra una diferencia de \$50 para hombres entre trabajadores con salud declarada al valor medio de la muestra y aquellos a una desviación estándar (igual a \$2996 - \$2936). Esto es, son efectos pequeños y poco significativos estadísticamente. Podemos señalar que en este primer modelo se aprecia ya una relación convexa entre IMC e ingreso; es decir, para excesos no muy grandes de IMC el efecto sobre los ingresos es pequeño, pero para excesos de peso grandes domina un efecto negativo.

El modelo II, de salud endógena, es el de variables instrumentales, el cual estimamos vía mínimos cuadrados en dos etapas y método generalizado de momentos. Las estimaciones se encuentran en la Tabla 5 para hombres y mujeres, respectivamente. La principal diferencia con el MCO es que la variables de interés, la auto-declaración de salud, muestra para ambos sexos efectos negativos y estadísticamente significativos. También podemos indicar que: (i) las estimaciones vía 2SLS o GMM son similares entre sí; (ii) los diagnósticos de la primera etapa y las pruebas de endogeneidad no permiten rechazar la hipótesis de que el instrumento es válido; y, (iii) las pruebas de sobre-identificación no permiten rechazar la hipótesis de que las variables en la primera etapa son exógenas, es decir, que el modelo está especificado de forma apropiada. Cuantitativamente, los impactos de la salud sobre los ingresos estimados por variables instrumentales son, para hombres, de \$527 (\$2962 - \$2435), usando como referencia una variación de una desviación estándar. Para mujeres, el cálculo análogo es de \$726.

Por brevedad no presentamos tablas específicas a los modelos en diferencias (que hemos listado con III), pero sí los incluimos en las tablas resumen.

La Tabla 6 muestra los resultados del modelo regresiones cambiantes y del MGR. Resaltamos que la variable de autodeclaración empleada en estos modelos difiere de los anteriores, pues es binaria e igual a 1 si hay mejora en salud entre la ronda 1 y la ronda 3 de la encuesta, y cero en caso contrario. También recordamos que ahora no estimamos un efecto paramétrico, sino efectos marginales que varían entre la población según la probabilidad individual de mejorar en salud.

Tabla 5  
Modelos de ingresos laborales de por sexo, MCO y VI 25 a 54 años de edad en cualquier ronda que trabajan en Ronda 3

	Hombres		Mujeres		Hombres		Mujeres	
	MCO	MCO	2SLS	GMM	2SLS	GMM	2SLS	GMM
Autodeclaración de salud, 1 muy buena a 5 muy mala	-86.55 (91.35)	61.58 (129)	-694.55 (321.52)*	-762.53 (315.05)*	-1007.42 (429.09)*	-1051.85 (337.17)**		
Horas totales asalariado	14.544002 (3.25)**	20.07 (5.68)**	13.98 (3.27)**	14.55 (3.27)**	18.57 (5.67)**	14.17 (4.74)**		
Educación, Ronda 3	-133.83 (63.93)*	-225.3 (81.62)**	-141.33 (63.99)*	-149.52 (63.08)*	-256.24 (79.45)**	-246.49 (76.99)**		
Educación al cuadrado, Ronda 3	16.23 (3.89)**	25.21 (4.53)**	15.6 (3.92)**	15.8 (3.85)**	24.7 (4.60)**	23.47 (4.46)**		
Seguridad social, Ronda 3	2246.73 (146.65)**	1713.53 (196.13)**	2202.43 (148.40)**	2253.09 (145.87)**	1662.24 (195.75)**	1564.47 (185.13)**		
En unión marital, Ronda 3	110.49 (193.46)	-112.29 (194.78)	153.14 (193.64)	109.98 (188.67)	-45.44 (206.16)	-94.17 (190.28)		
Discapacidad, Ronda 3	-917.1 (374.57)*	287.62 (505.3)	-712.65 (374.75)	-723.05 (372.7)	611.36 (484.95)	704.18 (490.27)		
Estatura, Ronda 3	28.9 (10.46)**	2.73 (17)	25.41 (10.60)*	26.87 (10.27)**	-1.63 (16.85)	-10.1 (16.2)		
Exceso de IMC, Ronda 3	-4.88 (39.26)	74.17 (46.06)	4.9 (39.71)	-5.83 (38.21)	92.61 (47.7)	49.95 (43.57)		
Exceso de IMC al cuadrado	-1.04 (2.52)	-4.07 (2.3)	-1.42 (2.55)	-0.76 (2.46)	-4.45 (2.32)	-2.24 (2.11)		
Embarazada		37.51 (363.43)			80.33 (381.54)	275.99 (372.71)		
Constante	-3536.06 (1,725.93)*	36.82 (2607.66)	-1396.44 (2018.31)	-1371.39 (1956.21)	3728.27 (2780.02)	5420.81 (2,558.98)*		
R2	0.12	0.14	0.11	0.11	0.12	0.11		
N	4383	2278	4379	4379	2275	2275		
<b>Márgenes predictivos</b>								
ximc3mas, ximc23mas a la media	3000	2734	3001	3005	2735	2653		
ximc3mas, ximc23mas a la media más 1 de	2926	2723	2945	2942	2778	2690		
ximc3mas, ximc23mas a la media más 2 de	2851	2711	2887	2878	2821	2727		
salud_discap = 0	3011	2753	3009	3013	2753	2658		
salud_discap = 1	2093	3041	2297	2290	3364	3362		
salud_declara a la media	2996	2759	2962	2962	2703	2606		
salud_declara a la media mas 1 de	2936	2974	2483	2435	1988	1860		
salud_declara a la media mas 2 de	2877	3190	2004	1909	1272	1113		
<b>Pruebas de hipótesis (Prob &gt; F)</b>								
ximc23mas, salud_declara3	0.5	1.64	4.79	5.86	7.42*	9.8**		
ximc3mas ximc23mas	0.87	1.57	1.24	1.2	3.85	1.31		
discap3, salud_declara3	3.63**	0.3	11.04**	12.65**	6.07*	10.5**		
<b>Estadísticos de la primera etapa</b>								
F robusta			62.2	33				
Prob > F			0.000**	0.000**				
<b>Sobreidentificación</b>								
Score chi2(10)			9.52		11.4			
valor p			0.14		0.07			
J de Hansen chi2(10)				9.52		11.4		
valor p				0.14		0.07		
<b>Endogeneidad</b>								
Score robusto chi2(1)			4.03		6.64			
valor p			0.04**		0.001			
Estadístico C chi2(1)				5.57		8.1		
valor p				0.01**		0.004		
Regresión robusta F(1,2755)			3.99		6.71			
valor p			0.04**		0.009			

\* p<0.05; \*\* p<0.01

Nota: los modelos incluyen variables de control por región geográfica, estatus de embarazo (para mujeres) y estatus indígena.

El efecto promedio (ate en la Tabla 6) es negativo en todos los casos. Esto mide el efecto que se tendría si toda la población mejorara en salud (se seleccionara al sector 1). En los mgr semiparamétricos, el ate es estadísticamente significativo. Sin embargo, la discusión sobre cómo la selección afecta la relación entre mejorar en salud y los ingresos está encaminada a plantear que tal efecto puede ser heterogéneo, diferente entre las personas.

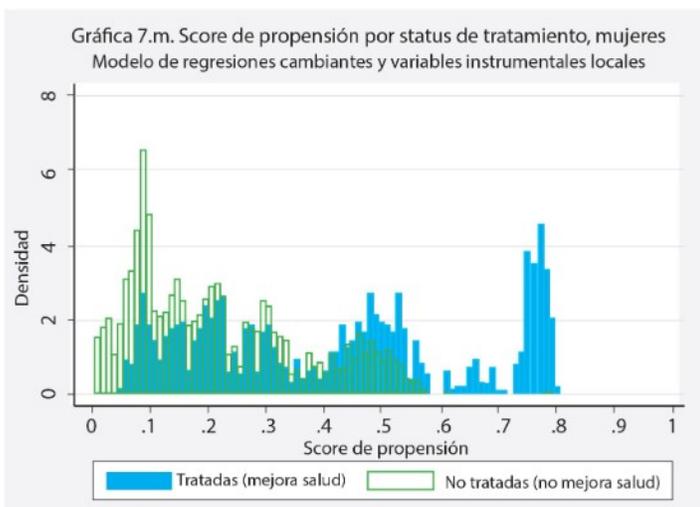
Para explorar el efecto de la heterogeneidad, presentamos primero, en las Gráficas 7.m y 7.h, los scores de propensión, el cálculo de la primera etapa mediante un modelo probit (es decir,  $P(X,Z)$ ). Claramente, no se cubre todo el intervalo (0,1), y no hay un traslape de frecuencias sobre los segmentos cubiertos del soporte. Esto nos preocupa porque describe una situación en que para alguna población la probabilidad de mejorar o de no mejorar no es uno para nadie. Además, para el modelo semiparamétrico el cálculo de variables instrumentales locales requiere de un número suficiente de observaciones en cada intervalo de  $P(Z)$ . De manera evidente, tenemos un modelo que nos da variación importante entre tratados y no tratados, pero posiblemente la información no es suficientemente rica para cumplir con los supuestos del modelo, en especial en la versión de variables instrumentales locales.

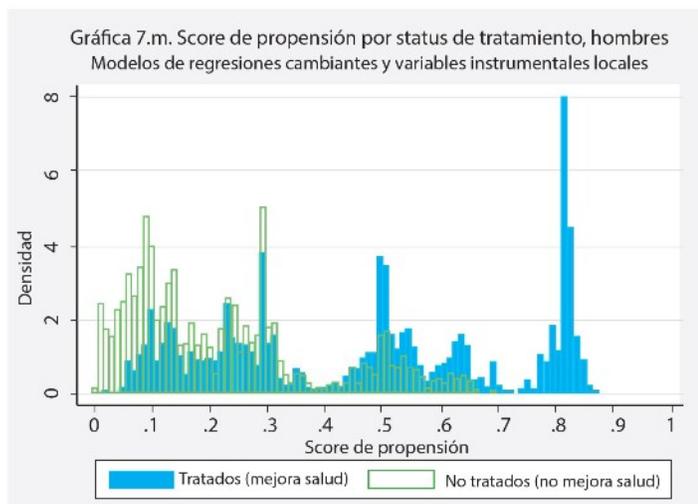
Los instrumentos empleados en la estimación del score de propensión son la presencia de enfermedades crónicas en los padres, la discapacidad, la presencia de hipertensión o diabetes, la edad y las variables de exceso de IMC (lineal y al cuadrado) interactuadas con la edad y dummies de edad. Las observaciones corresponden a personas que trabajan en la ronda 3 de la ENVIH. No mostramos los re-

sultados del probit, pero señalamos que la pseudo- $R^2$  es de aproximadamente 0.25 para ambos sexos y que son significativas las variables de exceso de peso interactuadas con edad, la presencia de enfermedades crónicas en padres, de diabetes o hipertensión y de discapacidad. Claramente, el soporte no cubre todo el intervalo de 0 a 1. En el modelo con LIV el MTE se estima sólo en rangos de soporte común de las estimaciones del score de propensión en la primera etapa (es decir, la función  $K$  se calcula sólo en los modelos paramétrico y semi-paramétrico con polinomio; es decir, no se calcula para el modelo LIV). También notamos que los scores de propensión no tienen pendientes suaves, sino que más bien parecen tener tres modas. Hemos probado con las variables de edad definidas año por año o agrupadas y este resultado trimodal se repite. Literalmente, significa que la población se segmenta en clusters de los que tienen altas probabilidades de mejorar, los que tienen muy pocas probabilidades, y un tercer grupo intermedio.

Los efectos marginales de tratamiento estimados para hombres con el modelo semiparamétrico muestran ganancias para los más propensos a mejorar en salud, y pérdidas para los menos propensos (Gráficas 8.m y 8.h). El eje vertical lo podemos leer como el retorno en dinero, y los valores hacia la izquierda del eje horizontal corresponden a los individuos con mayor propensión a mejorar su salud; hacia la derecha están las personas con menores scores de propensión a mejorar su salud. Si bien los errores de la estimación general producen intervalos de confianza que no nos permiten rechazar que los efectos son cero, si parece haber un patrón decreciente desde los menores valores de  $U_D$ . Esto significa que para quienes tienen mayor probabilidad de mejorar en salud, las ganancias en ingreso por hacerlo son mayores. El patrón es más claro para las mujeres. Sin embargo, los intervalos de confianza son grandes y no cubren todo el intervalo unitario. Esto se debe posiblemente a que la muestra es pequeña con relación a la necesaria para estimar estos modelos. También, las dificultades de medir el estado de salud pueden estar jugando un rol, que posiblemente no se resuelva aún si llegan a producirse bases de datos más grandes.

El modelo de variables instrumentales locales requiere que los datos tengan suficientes observaciones de personas tratadas y no tratadas, sobre el rango común del score de propensión. Esto se debe a que el método divide la muestra en bolas en torno a valores de las  $X$ , estima la  $p = P(Z)$ , modela el resultado (i.e. el ingreso de la persona) como función de  $p$ ,  $\hat{y}$  obtiene el efec-





to marginal como el cambio en ese ingreso predicho con respecto a  $p$ . Esto requiere que en cada celda haya suficiente variación de las  $X$ . Es decir, la propensión a mejorar en salud se da por las  $Z$ , y para cada valor estimado de esa propensión, hay observaciones de población tratada y no tratada, con variación en las  $X$  que permite estimar localmente parámetros para poder dibujar el efecto marginal al cambiar la propensión a mejorar en salud. Entonces, idealmente quisiéramos tener suficiente variación en las variables exógenas  $Z$  como para que se produzcan valores de  $p$  sobre todo el rango  $[0, 1]$  para las poblaciones tratada y no tratada, y que para cada segmento de valores de  $p$  se de variación en las  $X$  que afecta al ingreso. En breve, nuestros instrumentos no tienen suficiente variabilidad a nivel de todos los valores posibles de la probabilidad de mejorar en salud.

Al revisar los coeficientes estimados para las variables exógenas en el modelo regresiones cambiantes encontramos algunas diferencias importantes entre sectores: a) el efecto de la discapacidad es grande y negativo en general, pero lo es más para las mujeres tratadas (es decir, para quienes tienen mejor salud), para quienes es además significativo; b) el exceso de IMC no es significativo; c) la estatura tiene un efecto significativo e importante para los hombres; d) el efecto de las horas sobre los ingresos es grande y es significativo para las mujeres, además de que es casi tres veces más grande para las que mejoran en salud (en pesos, hay un aumento de \$21 por hora). Para

Nota: los modelos incluyen variables de control por región geográfica

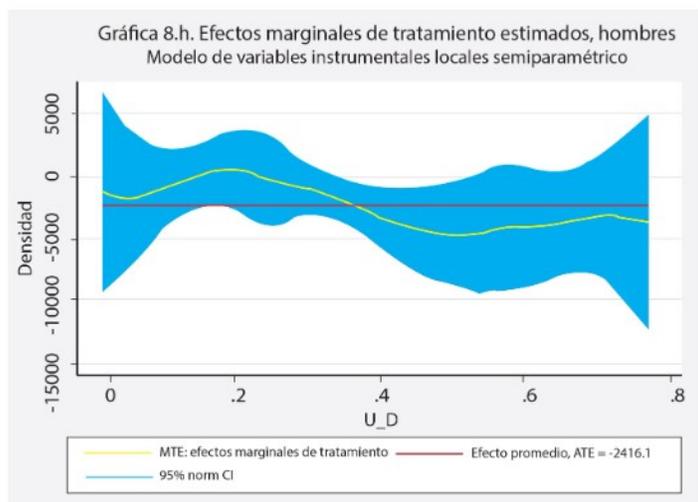
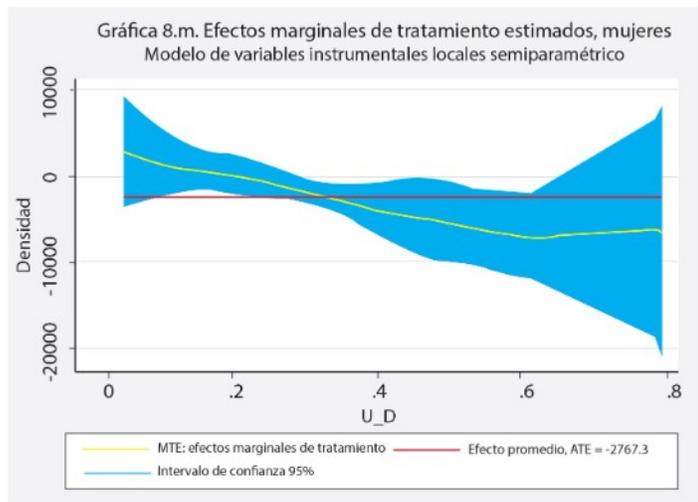
Tabla 6  
Modelo de ingresos laborales en ENVIH-3 Modelo generalizado de Roy-24 a 55 años de edad

	Hombres		Mujeres	
	Normal paramétrico	No paramétrico	Normal paramétrico	No paramétrico
<i>Tratados</i>				
Horas totales asalariado	10.562		32.257	
Educación, Ronda 3	-9.5 (113.7)		(9.453)** -277	
Educación al cuadrado, Ronda 3	23.39 (6.127)**		29.7 (6.117)**	
Seguridad social, Ronda 3	1409.5 (286.185)**		1383.9 (264.732)**	
Estatura	59.1 (17.788)**		33.1 (17.7)	
En unión marital, Ronda 3	341.5 (269.3)		84.1 (300.2)	
Discapacidad	-848.8 (510.6)		-2299.8 (857.508)**	
Indígena	46.3 (328.5)		-775.8 (328.946)*	
Estado mental	-638.9 (373.4)		-31.7 (460.7)	
Embarazo			-657.9 (340.8)	
Exceso de IMC	-68.2 (101.2)		16.5 (52.9)	
Exceso de IMC al cuadrado	2.1 (5.3)		1.8 (2.9)	
Ronda 2	1992.9 (450.916)**		2468.4 (389.074)**	
Ronda 3	4272.5 (600.598)**		4010.7 (563.496)**	
k	-987 (525.2)		-1695.5 (446.408)**	
Constante	-11367 (3,055.978)**		-8457.7 (2,960.053)**	

Tabla 7

	Hombres		Mujeres	
	Normal paramétrico	No paramétrico	Normal paramétrico	No paramétrico
<i>No tratados</i>				
Horas totales asalariado	5.803 (3.5)		11.179 (4.6)*	
Educación, Ronda 3	-206.4 (67.393)**		-182.82 (86.5)*	
Educación al cuadrado, Ronda 3	26.61 (4.6)**		27.924 (5.3)**	
Seguridad social, Ronda 3	1695.3 (143.1)**		1172.426 (140.3)**	
Estatura	26 (11.5)*		-4.479 (12.9)	
En unión marital, Ronda 3	224 (193.2)		-85.858 (163.5)	
Discapacidad	-291.1 (570)		-231.05 (470)	
Indígena	-76.4 (187.8)		-285.431 (205.2)	
Estado mental	-299 (250)		-401.008 (266)	
Embarazo	-30.24 (53.123)		-679.636 (396.9)	
Exceso de IMC	-30.2 (53.1)		34.62 (-46.1)	
Exceso de IMC al cuadrado	0.5 (2.8)		-2.3 (-2.0)	
Ronda 2	2032.6 (184.2)**		1618.7 (222.3)**	
Ronda 3	4079.9 (298.6)**		3597.2 (354.3)**	
k	-188 (429.6)		-1002.15 (389.9)*	
Constante	-3650.1 (1869.9)		670.996 (2132.9)	
Mills ( $p1 - p0$ )	-798.9 (610.7)		-693.3 (-600.66)	
E(Y1-Y0)@X	-1267 (670)	-2416 (688)**	-2565 (556)**	-2767 (771)**
N	6407	6407	3570	3570
Variables y valores p (pruebas chi-2)				
Discapacidad y Estado mental	.046**	0.131	.017**	
Estatura, ximcmas ximc2mas	.000**	0.47	.038**	
Indígena, ximcmas ximc2mas	0.78	0.144	.001**	

\* p<0.05; \*\* p<0.01



la función  $k$ , que mide el valor esperado de la mejora, tenemos efectos negativos y significativos para mujeres

Estimamos también modelos en diferencias y de efectos fijos (usando MCO y VI), usando las características longitudinales de los datos. No presentamos el detalle de estos resultados porque no encontramos diferencias importantes con los modelos ya presentados, pero sí los incluimos en el resumen siguiente.

La Tabla 8 presenta un resumen de resultados. Tenemos ahí los coeficientes para las principales variables en la ecuación de ingreso (excepto educación). La variable relativa al estado de salud es significativa y negativa en los modelos de VI. Como hemos discutido arriba, estimamos que el efecto es heterogéneo, por lo cual estos efectos promedio pueden ser altamente dependientes de la muestra a la mano.

El resumen para variables distintas a nuestro principal interés es el siguiente: (i) la afiliación a la seguridad social tiene un signo consistentemente positivo; no exploramos aquí este importante tema, y sólo mencionamos que una evaluación detallada requiere considerar el efecto de los impuestos y de la calidad del empleo como diferencias equalizadoras; (ii) la discapacidad tiene un efecto consistentemente negativo; (iii) el embarazo tiene un efecto consistentemente negativo pero la significación estadística nos da en general; (iv) el embarazo tiene efectos también consistentemente negativos, pero posiblemente se requiere estudiar un modelo dinámico de la vida de la mujer para estudiar este tema (la variable empleada es la de embarazo al momento de la encuesta, que está altamente relacionada con la participación laboral actual y futura); (v) el exceso de IMC no parece tener un patrón claro de efecto sobre los ingresos; (v) el efecto de las horas es diferente entre tratados y no tratados y es un componente importante de la diferencia en ingresos.

## 5. Conclusiones

El estado de salud tiene una relación importante con el ingreso laboral. El estado de salud y los cambios en el mismo parecen tener un efecto causal sobre los ingresos, pero este es un efecto altamente heterogéneo entre la población. De los resultados obtenidos, consideramos que los estimadores de VI resumen mejor el efecto promedio de la mejora en salud. Sin embargo, la estimación de efectos marginales nos indica que el efecto es muy diferente entre la población.

Las estimaciones con base en el modelo de Roy indican que el efecto de la salud sobre los ingresos es altamente heterogéneo. Así, el efecto promedio estimado no es representativo de lo que está ocurriendo en la población, pues hay diferencias importantes entre las personas. Aparentemente tenemos que las personas con mayor propensión a mejorar en salud tienen las mayores ganancias en ingresos laborales de hacerlo. Sin embargo, los intervalos de confianza calculados son demasiado grandes.

Estimaciones negativas del efecto marginal significan que para algunas personas, mejorar la salud implica una pérdida en ingresos; es decir, hay un costo implícito de mejorar la salud. En analogía, en el ámbito de la educación sería bueno en algún sentido para todos tener más educación, pero para algunos puede ser preferible evitar mayor gasto y esfuerzo pues los retornos no los compen-

Tabla 8  
Efecto de estado de salud sobre ingresos laborales  
Resumen de resultados

	Hombres	Mujeres	Hombres	Mujeres
	Autodeclaración de salud		Embarazo	
MCO	-86.6	61.6	na	37.5
VI-GMM	-762.5**	-1051.9**	na	276
Modelo en diferencias (GMM)	-313.4	-1177*	na	-707
MGR-tratados	na	na	na	-657.9
MGR-no tratados	na	na	na	-679.6*
Efectos fijos de edad-IV	-502.1	-757.1*	na	-326.1
	IMC 25+		Seguridad social	
MCO	-4.9	74.2	2246.7**	1713.5**
VI-GMM	26.9	50	2253.1**	1564.5**
Modelo en diferencias (GMM)	29.1	-15	631.2	488.1
MGR-tratados	-68.2	16.5	1409.5**	1383.9
MGR-no tratados	-30.2	-2.3*	1695.3**	1172.4
Efectos fijos de edad-IV	na	na	1503.09**	1339.65**
	(IMC 25+)2		Discapacidad	
MCO	-1	-4.1	-917.1**	287.62
VI-GMM	-5.8	-2.2	-723.1**	704.18
Modelo en diferencias (GMM)	-6.4	6.5	-205.8	-505.9
MGR-tratados	2.1	1.8	-848.8	-2299.77**
MGR-no tratados	0.5	-2.3*	-291.1	-231.05
Efectos fijos de edad-IV	na	na	-204.9	-504.88
	Estatura		Horas	
MCO	28.9**	2.7	14.5**	20.07**
VI-GMM	26.9**	-10.1	14.6**	14.17**
Modelo en diferencias (GMM)	-3.6	12.5	12.6**	8.5*
MGR-tratados	59.1**	33.1	10.6	32.26**
MGR-no tratados	26*	-4.5	5.8	11.18
Efectos fijos de edad-IV	43.9**	-1.5	11**	12.39**

na. No aplica.

san. En nuestras (imprecisas) estimaciones, las personas menos propensas a mejorar su salud tienen efectos negativos en ingresos de hacerlo.

En conclusión, si parece haber una relación importante entre el estado de salud y los ingresos de las personas, pero esta es heterogénea, y es diferente entre hombres y entre mujeres. Es necesario profundizar en las causas que provocan un cambio en el estado de salud y la forma en que afectan la participación laboral, elemento en que no hemos profundizado en este documento.



## Referencias

- Brave, S. y T. Walstrum (2014), "Estimating marginal treatment effect using parametric and semiparametric methods", *The Stata Journal*, 14(1), pp. 191-217.
- Cornelissen, T., C. Dustmann, A. Raute, y U. Schönberg (2016), "From LATE to MTE: Alternative methods for the evaluation of policy interventions", *Labour Economics*, 41, pp. 47-60.
- Deaton, A. (2006), "The Great Escape: A Review of Robert Fogel's The Escape from Hunger and Premature Death, 1700-2100", *Journal of Economic Literature*, 44(1), pp. 106-114.
- Eisenhauer, P., J. J. Heckman y E. Vytlačil (2015), "The Generalized Roy Model and the Cost-Benefit Analysis of Social Programs", *Journal of Political Economy*, 123, pp. 413-443.
- Heckman, J.J. y E.J. Vytlačil (2001). "Local instrumental variables". En: *Nonlinear Statistical Modelling: Proceedings of the Thirteenth Annual International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, ed. C. Hsiao, K. Morimune y J.L. Powell, 1-46. Nueva York: Cambridge University Press.
- Heckman, J.J. y E.J. Vytlačil (2007). "Econometric evaluation of social programs, Part II: using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments". En: *Handbook of Econometrics, Volume 6B*, Capítulo 71, ed. Heckman, J.J. y Leamer, E.E., Elsevier B.V.
- INSP (2012), "Encuesta Nacional de Nutrición y Salud 2012", Cuernavaca: Instituto Nacional de Salud Pública.
- Jylhä, M. (2009), "What is self-rated health and why does it predict mortality? Towards a unified conceptual model", *Social Science & Medicine*, 69, pp. 307-316.
- McGee, D. L., Y. Liao, G. Cao, y R. Cooper (1999), "Self-reported Health Status and Mortality in a Multiethnic US Cohort", *American Journal of Epidemiology*, 149(1), pp. 41-46.
- Miilunpalo, S., P. Oja, M. Pasanen y H. Urponen (1997), "Self-rated health status as a health measure: The predictive value of self-reported health status on the use of physician services and on mortality in the working-age population", *Journal of Clinical Epidemiology*, 50, pp. 517-528.
- Rubalcava, L. y G. Teruel (2006), "Encuesta Nacional sobre Niveles de Vida de los Hogares, Primera Ronda, Documento de Trabajo", [www.ennvih-mxfls.org](http://www.ennvih-mxfls.org).
- \_\_\_\_\_. (2008), "Encuesta Nacional sobre Niveles de Vida de los Hogares, Segunda Ronda, Documento de Trabajo", [www.ennvih-mxfls.org](http://www.ennvih-mxfls.org).
- \_\_\_\_\_. (2013), "Encuesta Nacional sobre Niveles de Vida de los Hogares, Tercera Ronda, Documento de Trabajo", [www.ennvih-mxfls.org](http://www.ennvih-mxfls.org).
- Tamayo-Fonseca, N., J.A. Quesada, A. Nolasco, I. Melchor, J. Moncho, P. Pereyra-Zamora, R. López, J. Calabuig, X. Barber (2013), "Self-rated health and mortality: a follow-up study of a Spanish population", *Public Health*, 127, pp. 1097-1104.
- Weil, D. N. (2015), "A Review of Angus Deaton's The Great Escape: Health, Wealth, and the Origins of Inequality", *Journal of Economic Literature*, 53(1), pp. 102-114.