



Gobierno de  
**México**



# Integración de información a través de la vinculación probabilística y construcción de cohortes administrativas

---

Dra. Lina Sofía Palacio Mejía

Dr. Juan Eugenio Hernández Ávila

Dra. Mariana Álvarez Aceves

Unidad de Inteligencia en Salud Pública

Instituto Nacional de Salud Pública

[lpalacio@insp.mx](mailto:lpalacio@insp.mx)





# Introducción

- Los estudios longitudinales son esenciales para hacer un seguimiento de los desenlaces en salud, identificar los factores de riesgo e informar las políticas de salud pública
- Sin embargo, el seguimiento de cohortes prospectivas en tiempo real requiere muchos recursos.
- Una alternativa costo-eficiente es la construcción de cohortes administrativas a través de la vinculación de datos probabilísticos



# Transformación digital

- Herramienta que apoya la atención a la salud, que permite generar un nuevo paradigma que pone en el centro al individuo
  - Información interoperable en diversos aspectos de la atención médica: Farmacia, imagenología, laboratorio, historia clínica, entre otras
  - Un mecanismo para incorporar información poblacional, personas que pueden no estar dentro de un padrón de usuarios y/o un expediente clínico electrónico, sería a través de la vinculación probabilística de datos de encuestas y registros de salud para construir cohortes administrativas



# **Factores de riesgo nutricionales de mortalidad prematura y utilización hospitalaria: Un análisis longitudinal vinculando datos administrativos y de encuestas**

## **Objetivo del proyecto**

Analizar las asociaciones entre los factores de riesgo y desenlaces específicos en salud, mediante el análisis de una cohorte administrativa construida por medio de la vinculación probabilística de información recolectada en las Encuestas Nacionales de Salud y Nutrición (ENSANUT 2006, 2012 y 2018) y mortalidad (2006 - 2023).

# Metodología



Gobierno de  
**México**





## Recolección de datos

- ENSANUT 2006, 2012, 2018
  - Adultos de 20 años o más con datos de salud
  - Muestra representativa de la población mexicana con datos sociodemográficos, clínicos, nutricionales, dietéticos, de estilo de vida y acceso a servicios de salud
- Sistema estadístico y epidemiológico de las defunciones (SEED) 2006-2023
- Estadísticas de mortalidad del INEGI 2018-2023



## Vinculación de datos ENSANUT - SEED

Para dar seguimiento a los individuos encuestados en las ENSANUT se construyó una cohorte administrativa a través de la vinculación probabilística de los registros individuales de las ENSANUT con el sistema estadístico y epidemiológico de defunciones (SEED), utilizando un algoritmo de vinculación probabilística desarrollado por el Instituto Nacional de Salud Pública, basado en el método de Fellegi-Sunter.



# Fuentes de datos



## ENSANUT 2006

Muestra representativa de 47,152 hogares y 206,700 individuos (45,241 adultos).



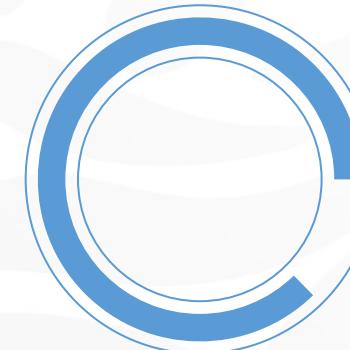
## ENSANUT 2012

Muestra representativa de 50,528 hogares y 194,923 individuos (46,277 adultos).



## ENSANUT 2018

Muestra representativa de 44,612 hogares y 158,044 individuos (43,070 adultos).



## SEED

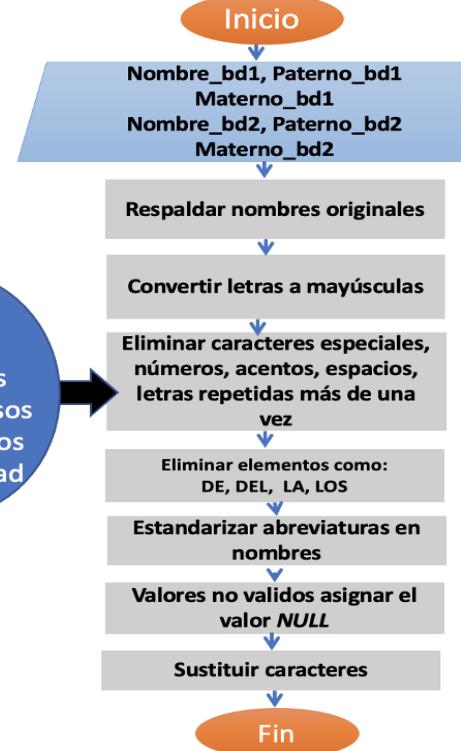
Sistema Estadístico y Epidemiológico de las Defunciones de la Secretaría de Salud con 12.1 millones de registros de defunciones de 2006 a 2023.



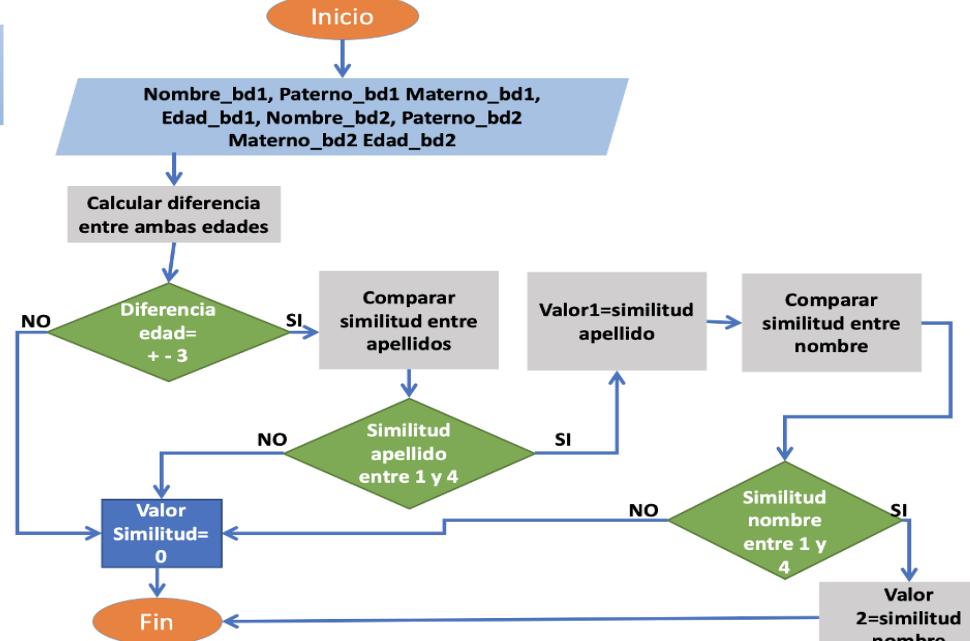
# Vinculación probabilística

## Limpieza y estandarización de nombres

**Entradas**  
Datos de encuestas  
BD de egresos hospitalarios  
Y mortalidad



## Función de similitud de cadena de caracteres



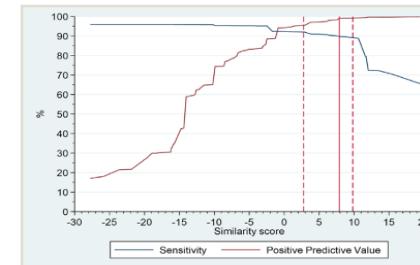
## Variables de decisión y score

Apellido	Nombre	Score
1	1	1
1	2	0.94
1	3	0.88
1	4	0.81
2	1	0.75
2	2	0.69
2	3	0.63
2	4	0.56
3	1	0.5
3	2	0.44
3	3	0.38
3	4	0.31
4	1	0.25
4	2	0.19
4	3	0.13
4	4	0.06
0	0	0

Residencia	Score
Coincidencia	Score
Estado, municipio y localidad	1
Estado, municipio	0.67
Solo estado	0.33
Diferente estado	0
SEXO	
Condición	Score
Mismo sexo	1
Sexo ND en una BD	1
Sexo ND en ninguna BD	1
Sexo diferente	0

**Salidas**  
BD para cohortes administrativas

$$SCORE_{edad} = \frac{1}{|edad_{BD1} - edad_{BD2}|^3 + 1}$$



Registros individuales  
(ENSANUT)  
Hospitalizaciones  
(SAEH)  
Defunciones (SEED)

Vinculación de bases de datos para la construcción de cohortes administrativas

ENSANUT → SAEH → SEED

ENSANUT	SAEH	SEED
base	seguimiento	desenlace
Censura en el último año con datos de mortalidad disponibles		

# Vinculación ENSANUT 2006 y 2012 – SEED 2006-2023



Gobierno de  
**México**



Las ENSANUT 2006 y 2012 se vincularon con registros de mortalidad del SEED del 2006 al 2023 siguiendo la metodología descrita, dentro del INSP.

Se utilizaron como variables de vinculación el nombre, apellidos, sexo, fecha de nacimiento, entidad, municipio y localidad de residencia.

Los procedimientos de bloqueo se realizaron en SQL-Server Standard y el algoritmo de EM fue programado en Stata 14.

Posteriormente el 47,7% de las 5.417 coincidencias de la ENSANUT-SEED 2006 se sometieron a revisión manual.



## Vinculación ENSANUT 2018 – Defunciones INEGI 2018-2023

La ENSANUT 2018 se vinculó con los registros de mortalidad del Instituto Nacional de Estadística y Geografía (INEGI) del 2018 al 2023 **dentro del laboratorio de microdatos del INEGI.**

Se utilizaron las variables de vinculación: el nombre, apellidos, sexo, día, mes y año de nacimiento y entidad de residencia.

Los procedimientos de bloqueo y el algoritmo de EM fueron programados en R versión 4.2.

Posteriormente el 35,8% de 4.815 coincidencias de la ENSANUT-SEED 2006 se sometieron a revisión manual.

# Resultados



Gobierno de  
**México**



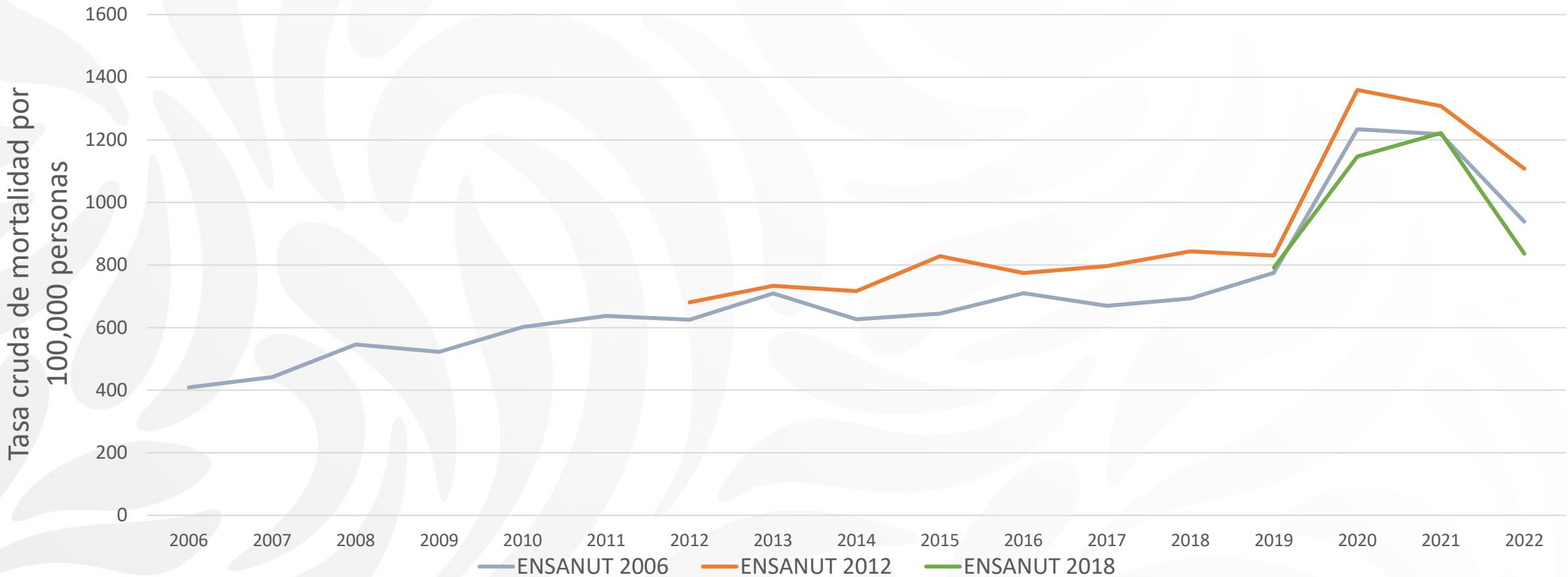


## Resultados

- Se creó una cohorte de 134,588 individuos de 20 años o más, con 11,928 defunciones identificadas (8.9%).
- ENSANUT 2006 (**45,241 adultos con información sobre salud y nutrición**), mediana de edad de 39 años, rango de 20 a 99)  
**5,407 (12%) defunciones** registradas en el periodo 2006 - 2023
- ENSANUT 2012 (**46,277 adultos con información sobre salud y nutrición**), mediana de edad de 41 años, rango de 20 a 114.  
**4,796 (10.4%) defunciones** registradas en el periodo 2012 - 2023
- ENSANUT 2018 (**43,070 adultos con información sobre salud y nutrición**)  
**1,696 (4%) defunciones** registradas en el periodo 2018 - 2023

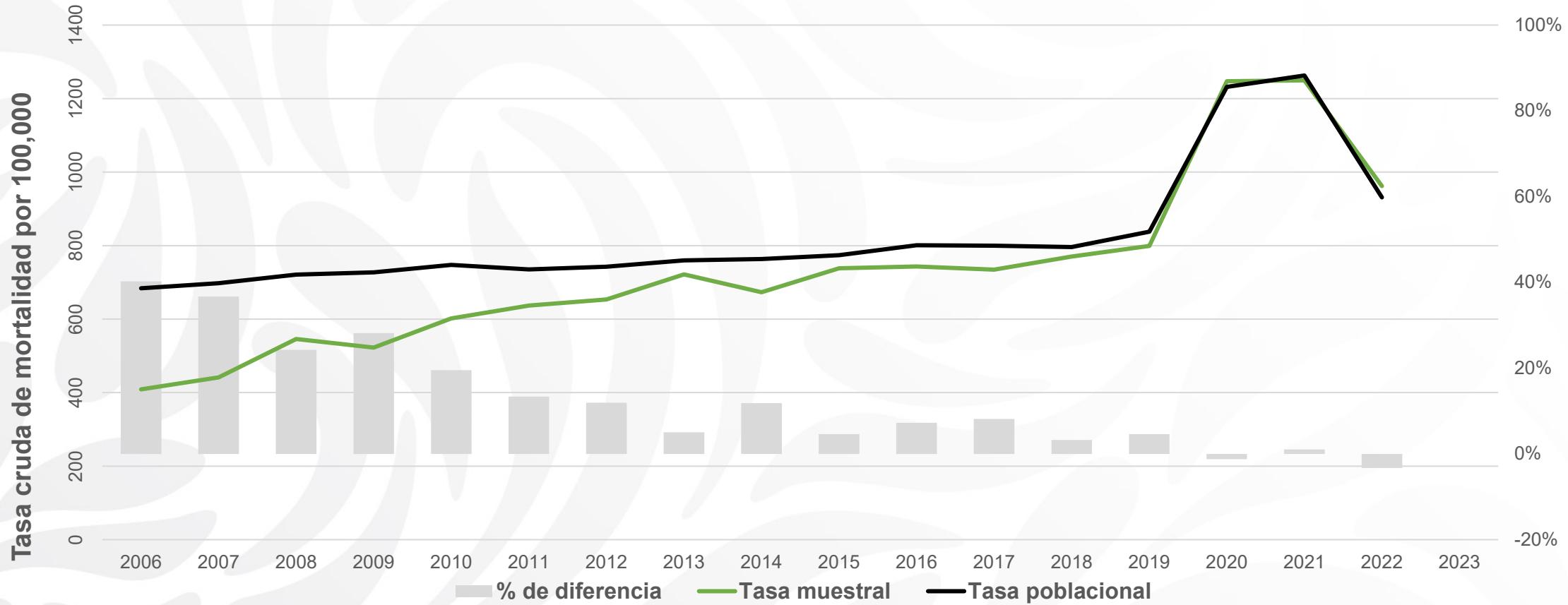


# Tasas crudas de mortalidad por todas las causas en mayores de 20 años





# Tasas crudas de mortalidad por todas las causas en mayores de 20 años



Tasa muestral incluye ENSANUT 2006, 2012 y 2018  
Tasa poblacional calculada con proyecciones del CONAPO

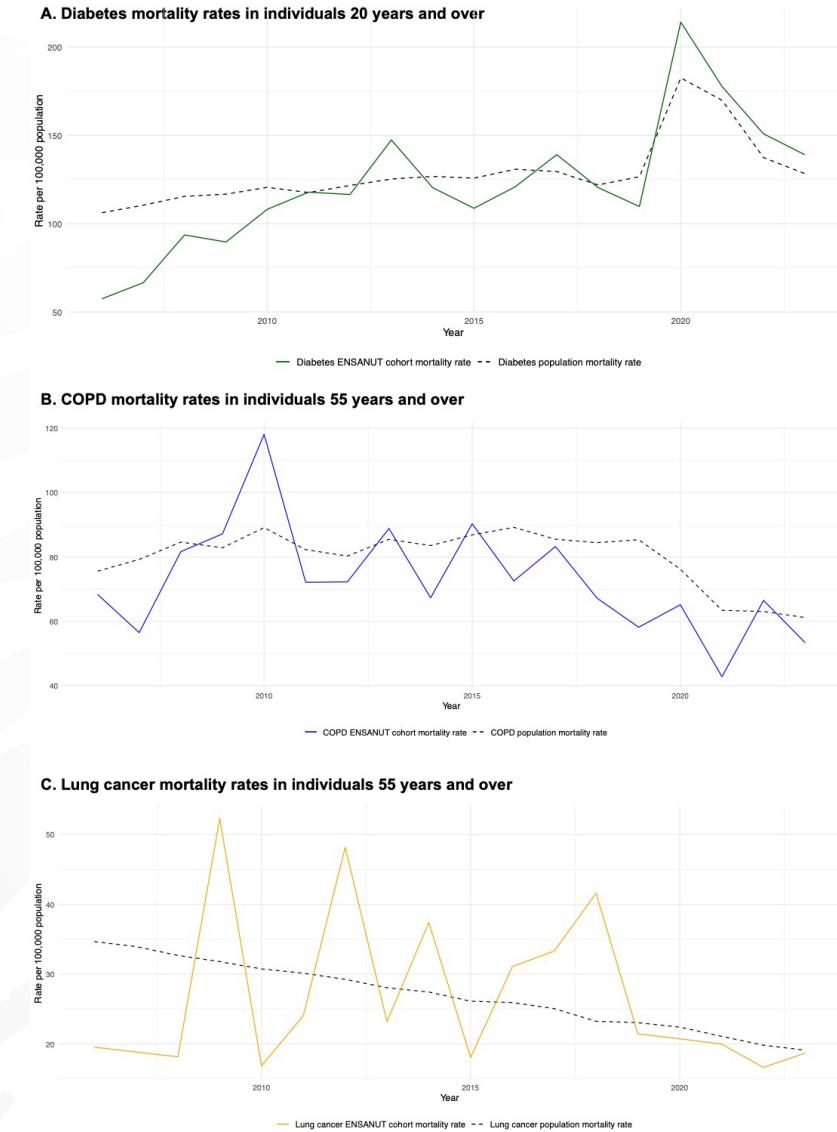


# Tasas de mortalidad por Diabetes, EPOC y cáncer de pulmón

La TBM por diabetes en la cohorte fue de 128.2 por 100,000 personas/año, en comparación con 129.9 por para la población mexicana en el mismo rango de edad, con una diferencia de 1.3%.

La mortalidad por EPOC, la TBM media entre los individuos de 55 años o más fue de 69,8 por 100.000 años-persona en la cohorte y de 79,1 en la población general, con una diferencia media del 11,7%.

La TBM de cáncer de pulmón fue de 26,5 por 100.000 años-persona en la cohorte y de 26,1 en la población, con una media del 1,5%.





## Mortalidad por EPOC relacionada con el tabaco

Cohorte		Nunca fumó		Actualmente fumador		Dejo de Fumar	
		Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
Mexico (2006-2023)	Número de muertes por cancer de pulmón	49	94	39	14	77	21
	Tasa de mortalidad estandarizada por 100,000 años persona*	106.89	88.25	216.07	297.65	230.28	150.10
	RR	1.00	1.00	<b>2.02</b>	<b>3.37</b>	<b>2.15</b>	<b>1.70</b>
USA (1959-1965)	Número de muertes por cancer de pulmón	82	136	338	66	151	9
	Tasa de mortalidad estandarizada por 100,000 años persona*	39.42	15.70	227.64	45.74	168.23	43.32
	RR	1.00	1.00	<b>5.77</b>	<b>2.91</b>	<b>4.27</b>	<b>2.76</b>
USA (1982-1988)	Número de muertes por cancer de pulmón	94	192	462	367	740	273
	Tasa de mortalidad estandarizada por 100,000 años persona*	33.72	18.71	318.24	199.17	209.27	104.14
	RR	1.00	1.00	<b>9.44</b>	<b>10.65</b>	<b>6.21</b>	<b>5.57</b>
USA (2000-2010)	Número de muertes por cancer de pulmón	183	320	656	720	2639	1994
	Tasa de mortalidad estandarizada por 100,000 años persona*	18.56	16.09	497.62	312.92	138.57	103.40
	RR	1.00	1.00	<b>26.81</b>	<b>19.45</b>	<b>7.47</b>	<b>6.43</b>



## Conclusiones

- Logramos construir cohortes administrativas mediante la vinculación probabilística de las encuestas y las bases de datos de mortalidad

Las tasas muestrales presentan tendencias muy similares y a las tasas nacionales mostrando la consistencia del método,  
La variabilidad encontrada es resultado esperado debido al tamaño de muestra de las encuestas

- Logramos el análisis de factores de riesgos y desenlaces consistentes con la literatura

Tabaquismo  
Diabetes

# Conclusiones



Gobierno de  
**México**



- Nuestros resultados hacen una aportación importante a la transformación digital

Permite utilizar los datos existentes con una inversión mínima para tener resultados enfocados en la de persona y de forma oportuna

Nos ayudan a identificar mecanismos de interoperabilidad para ser planeados a futuro para que la información pueda ser integrada

Base fundamental para la implementación de otros programas



## Limitaciones y áreas de oportunidad

- Aun cuando la vinculación fue exitosa, existen áreas de mejora de la calidad de la información y oportunidad en el uso del CURP.

Las tasas muestrales de mortalidad resultantes de la vinculación de la ENSANUT 2006 presentan mayores diferencias contra las tasas nacionales

Calidad de los datos

Optimización de los algoritmos de vinculación probabilística

Incorporar supervisión humana en el caso de registros ubicados en la franja de incertidumbre

Explorar nuevas soluciones para la vinculación probabilística

“Nunca subestimes el poder de hacer algo pequeño hoy,  
que pueda cambiarlo todo mañana.”

Gracias



Gobierno de  
**México**

