

# Identificación de factores educativos y hábitos sexuales asociados con la prevalencia de HSV-2 y VIH en adolescentes a través de redes neuronales.

**Facultad de Ciencias, UNAM**

David Vega Mena  
davidvega@ciencias.unam.mx

27 de mayo de 2024

## 1. Introducción.

El Virus de Inmunodeficiencia Humana (VIH) sigue siendo uno de los desafíos más apremiantes para la salud global, afectando a millones de personas en todo el mundo. A pesar de los avances en la investigación y el tratamiento, la prevalencia del VIH, especialmente entre los adolescentes, plantea preocupaciones significativas. La transmisión del VIH en esta población vulnerable puede tener consecuencias duraderas para la salud individual y la salud pública en general.

Los adolescentes enfrentan una serie de desafíos únicos en relación con la prevención del VIH, incluyendo la exploración de su sexualidad, la falta de educación sexual integral y la presión social. Los adolescentes, al encontrarse en una etapa crucial de desarrollo de hábitos y costumbres, representan una oportunidad valiosa para concientizar sobre los riesgos asociados a prácticas de riesgo relacionadas con la propagación del virus VIH. Abordar esta conciencia desde la prevención no solo puede influir en la reducción de la incidencia del virus, sino también en la preparación para afrontar la pandemia del VIH de manera más efectiva. Ante este escenario, las campañas de prevención del VIH desempeñan un papel crucial en la educación, concientización y promoción de comportamientos saludables entre los adolescentes.

En este contexto resulta imperante el evaluar la efectividad de los esfuerzos realizados con este propósito. Se hace necesario determinar si las acciones emprendidas mediante las campañas de prevención del VIH y educación sexual se ejecutan con un enfoque adecuado y preciso, que pueda traducirse en movimientos tanto eficaces como eficientes en términos de la prevalencia de la enfermedad. Asimismo, es esencial identificar las características particulares de las campañas que contribuyen a un impacto más significativo en la población objetivo, con

la intención de perfeccionar las estrategias implementadas en cada demografía específica.

Para enfrentarse a esta problemática se buscará profundizar en tres cuestiones principales:

1. ¿Cuáles son los efectos en los hábitos sexuales y educación sexual en la prevalencia de VIH?
2. ¿Cuáles son las combinaciones de los conocimientos y conductas sexuales que pueden disminuir efectivamente la prevalencia de la enfermedad y en qué medida estos deben de mejorar a través de las campañas para conseguir tal objetivo?
3. ¿Es efectiva la educación sexual, por si misma, para disminuir la prevalencia de la enfermedad?

## **2. Objetivos.**

1. Determinar factores de conducta y educación sexual que impactan en mayor medida la prevalencia del VIH en adolescentes.
2. Encontrar qué enfoques de las campañas de prevención de VIH en adolescentes resultan más efectivos en la reducción de la prevalencia del virus.

## **3. Justificación.**

Este proyecto cobra una relevancia fundamental en el ámbito de la salud pública al aspirar a mejorar la efectividad de las campañas de prevención del VIH, en el ámbito de la disminución de prevalencia del virus. Al identificar las estrategias más eficientes, se planea contribuir significativamente a la optimización de recursos, permitiendo dirigir inversiones hacia enfoques que generen un impacto máximo.

## **4. Antecedentes.**

La propagación del Virus de Inmunodeficiencia Humana (VIH) es un fenómeno complejo influenciado por diversos factores de riesgo que impactan en distintas comunidades y poblaciones. La Organización Mundial de la Salud destaca detalladamente los factores que aumentan la vulnerabilidad ante el VIH. Entre estos se encuentran elementos clave como comportamientos sexuales de riesgo, falta de acceso a la educación sexual, niveles socioeconómicos, y la prevalencia de otras infecciones de transmisión sexual.

Con un enfoque en hábitos sexuales, de acuerdo con Douglas B. Kirby, B.A. Laris y Lori A. Roller en su trabajo "Sex and HIV Education Programs: Their

Impact on Sexual Behaviors of Young People Throughout the World”, donde realizan una revisión de 83 estudios que miden el impacto de programas de educación sexual y VIH en adolescentes menores a 25 años al redor del mundo, dos tercios de ellos mejoraron significativamente uno o más comportamientos sexuales como cantidad de parejas sexuales, uso de preservativo, relaciones sexuales de riesgo, número de parejas o tiempo de espera a la primera relación sexual, alineado en principio con la disminución de incidencia de VIH en estas poblaciones.

Nuevamente el estudio de Douglas B. Kirby, B.A. Laris y Lori A. Rolleria destaca que a nivel mundial los programas de educación sexual y VIH/ITS se perciben como una solución parcial a estos problemas y señalan que, aunque ha habido numerosas revisiones previas sobre el impacto de los programas de educación sexual y VIH en el comportamiento, estas revisiones suelen estar limitadas a áreas geográficas específicas como los Estados Unidos o el África subsahariana. El trabajo considera que algunas de estas revisiones pueden considerarse algo desactualizadas, o no analizan en profundidad las características de los programas efectivos.

El artículo *Sex and HIV Education Programs: Their Impact on Sexual Behaviors of Young People Throughout the World* asegura que: “en general, estos estudios (programas de prevención enfocados para adolescentes en relación a hábitos sexuales saludables) indican fuertemente que estos programas tenían muchas más probabilidades de tener un impacto positivo en el comportamiento (sexual) que un impacto negativo. En los 83 estudios en total, dos tercios (65 %) tuvieron un impacto positivo significativo en uno o más de estos comportamientos sexuales o resultados, mientras que solo el 7 % tuvo un impacto negativo significativo en uno o más de estos comportamientos o resultados”.

Por otro lado en el artículo *School-based interventions for preventing HIV, sexually transmitted infections, and pregnancy in adolescents* se encuentra que los programas educativos sobre salud sexual y reproductiva: “[...] Los programas educativos evaluados no tuvieron un efecto demostrable en la prevalencia del VIH”.

Ambos artículos son meta análisis de campañas, estudios y programas de prevención en los que se evalúan trabajos con los siguientes criterios de selección:

### **Sex and HIV Education Programs: Their Impact on Sexual Behaviors of Young People Throughout the World:**

#### **Identificación de estudios de evaluación:**

- Ser un programa de educación sexual o de VIH basado en un plan de estudios y en grupos.
- Centrarse en adolescentes o adultos jóvenes de entre 9 y 24 años.

### **Los métodos de investigación debían:**

- Incluir un diseño experimental con grupos de intervención y comparación, y recopilación de datos pre y post intervención.
- Tener un tamaño de muestra de al menos 100
- Medir el impacto del programa en uno o más de los siguientes comportamientos sexuales: inicio de relaciones sexuales, frecuencia de relaciones sexuales, número de parejas sexuales; uso de condones o anticonceptivos de manera más general; medidas compuestas de riesgo sexual (por ejemplo, frecuencia de relaciones sexuales sin protección); tasas de embarazo; tasas de natalidad; y tasas de ITS.
- Medir el impacto en aquellos comportamientos que pueden cambiar rápidamente (es decir, frecuencia de relaciones sexuales, número de parejas sexuales, uso de condones, uso de anticonceptivos o toma de riesgos sexuales) durante al menos 3 meses, o medir el impacto en aquellos comportamientos o resultados que cambian menos rápidamente (es decir, inicio de relaciones sexuales, tasas de embarazo o tasas de ITS) durante al menos 6 meses.

### **School-based interventions for preventing HIV, sexually transmitted infections, and pregnancy in adolescents:**

#### **Tipos de estudios:**

Se incluyeron ensayos controlados aleatorizados (ECA), tanto aleatorizados individualmente como aleatorizados por conglomerados.

#### **Tipos de participantes:**

Se consideraron adolescentes, definidos como personas de 10 a 19 años, que asistían a la escuela primaria, secundaria o preparatoria en el momento de la intervención. En aquellos países donde los niños comenzaban la escuela a una edad más avanzada o donde las poblaciones escolares a veces incluían jóvenes mayores de 20 años, se incluyeron los estudios en los que la mayoría de los participantes (más del 50 %) eran adolescentes.

#### **Tipos de intervenciones:**

Se incluyeron intervenciones destinadas a reducir el riesgo de VIH, otras infecciones de transmisión sexual (ITS) o embarazos entre adolescentes, llevadas a cabo principalmente en escuelas o vinculadas a ellas, con o sin un componente comunitario. Algunas intervenciones consistían en programas educativos basados en currículos impartidos principalmente por adultos (maestros u otros adultos) o por compañeros (educadores entre pares), y otras incluían características

adicionales para modificar el entorno escolar o comunitario, como cambios en las políticas escolares o mejoras en los servicios de salud. También se consideraron intervenciones que buscaban alentar a los adolescentes a permanecer en la escuela mediante la provisión de incentivos, ya sean en efectivo u otros materiales.

### **Tipos de medidas de resultado:**

Se evaluaron resultados clínicos/biológicos como la prevalencia de VIH, ITS y embarazos, así como resultados autoinformados conductuales, incluyendo el uso de condones masculinos en la primera y en la relación sexual más reciente, así como la incidencia de iniciación sexual.

Dada la similitud en los estudios analizados en ambos trabajos, resulta interesante observar las conclusiones aparentemente contradictorias entre ellos. Mientras que el primer estudio asegura que las campañas destinadas a mejorar los hábitos sexuales y las conductas de riesgo tienen efectos positivos, el segundo muestra cómo estos efectos son insignificantes en la prevalencia de la enfermedad. Por tanto, surge la interrogante de si los comportamientos sexuales y la educación sexual están fuertemente vinculados a la prevalencia del VIH y, en caso afirmativo, cuáles son los más relevantes para reducir la prevalencia de la enfermedad. Asimismo, es crucial estimar en qué medida estos comportamientos deben mejorar a través de los programas de educación, en caso de ser posible, y si estas mejoras pueden traducirse en una diferencia significativa en la disminución de la prevalencia del virus.

A continuación se presenta el gráfico de bosque del artículo *School-based interventions for preventing HIV, sexually transmitted infections, and pregnancy in adolescents* que compara los coeficientes de riesgo para grupos intervenidos y no intervenidos con programas de educación.

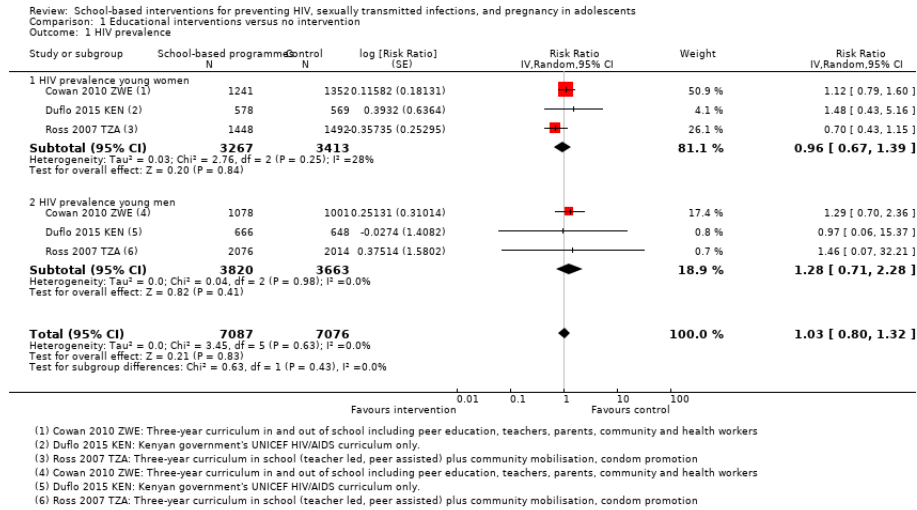


Figura 1: Coeficiente de riesgo por estudio de grupos intervenidos con programas de educación sexual contra grupos no intervenidos.

Por motivos de accesibilidad, la información que se utilizará para el trabajo es aquella incluida en los estudios Cowan 2010 y Duflo 2015 analizados dentro del trabajo *School-based interventions for preventing HIV, sexually transmitted infections, and pregnancy in adolescents*. Nuevamente, por cuestiones de accesibilidad a los datos, la información de los individuos evaluados del trabajo de Cowan se consiguió como descripción demográfica, mientras que para el trabajo de Duflo se pudo acceder a los datos de cada individuo incluido.

## 5. Métodos.

Dado que se pretende identificar cuáles son los factores educativos y conocimientos en materia de educación sexual que están fuertemente vinculados con la prevalencia del VIH en adolescentes y en qué medida estos deben mejorar en una población específica para conseguir una disminución significativa de la prevalencia es necesario poder estimar esta estadística de una población dadas sus características de conocimientos y hábitos.

Así, se consideran cuatro aspectos para conseguirlo:

1. Definir cuáles son los conocimientos, hábitos sexuales y características adicionales que determinan un individuo de una población.
2. Capacidad de simulación de poblaciones no intervenidas, de control.
3. Simulación de efectos en educación y hábitos sexuales.
4. Estimación de prevalencia de enfermedad dada una población.

Con lo anterior, se tomaron algunas de las características del trabajo de Cowan, de manera que los datos del trabajo de Duflo se homologaron para que ambas poblaciones tuvieran las mismas características.

Así, las características seleccionadas fueron:

Sexo, Edad, Nivel de educación, Alguna vez casado, Adquisición de VIH, Autoeficacia en la realización de pruebas de VIH, Autoeficacia en el uso del condón, Control sobre las parejas sexuales, Sexo seguro y uso de condones, Alguna vez ha tenido relaciones sexuales, Dos o más parejas sexuales a lo largo de la vida, Inicio de la actividad sexual a los 17 años o antes, Dos o más parejas sexuales en los últimos 12 meses, No utilizó condón en la última relación sexual, Buscó tratamiento para síntomas de ETS, Prevalencia de secreción genital, Prevalencia de verrugas o llagas genitales, Derechos dentro del matrimonio, Ha reportado embarazo actual o pasado, Prevalencia de cualquier síntoma de ETS, Infección por VIH e Infección por HSV-2.

Todas ellas categóricas y binarias en función de la presencia o ausencia de alguna característica, conocimiento y desconocimiento en algún tema, a excepción de Sexo, Edad y Nivel de educación donde existen las clases hombre y mujer, 18; 19-20 y 21-22 y Ninguna o primaria; F1-F2 y F3-F4 respectivamente.

El proceso de homologación de características de los datos de la población de Duflo a las covariables seleccionadas se realizó mediante la reorganización de las respuestas de los participantes, asignando la característica en función de los aciertos en esa materia.

Las preguntas escogidas para la homologación de características fueron:

- |                   |                      |
|-------------------|----------------------|
| ■ sex             | ■ Qb252agreement     |
| ■ Utreat          | ■ Qb253agreement     |
| ■ HIVtreat        | ■ Qb261decisionkids  |
| ■ Qa38dateofbirth | ■ Qb268educomparison |
| ■ Qb110           | ■ Qb378              |
| ■ Qb112edulevel   | ■ Qb379preventhiv    |
| ■ Qb248agreement  | ■ Qb380howpreventhiv |
| ■ Qb249agreement  | ■ Qb381              |
| ■ Qb250agreement  | ■ Qb383              |
| ■ Qb251agreement  | ■ Qb384              |

- Qb388
- Qb394
- Qb3102getcondoms
- Qb3104agreewithstatement
- Qb3105agreewithstatement
- Qb3106agreewithstatement
- Qb3107agreewithstatement
- Qb3108agreewithstatement
- Qb4114maritalstatus
- Qb4117hadsex
- Qb4118sexoneperson
- Qb4119age1stsex
- Qb41206months
- Qb4121sexpatner
- Qb4123usecondom
- Qb4126usedcondom
- Qb4127lastsexusecondom
- Qb4130
- Qb4132whodecides
- Qb5135changebehavior
- Qb5136
- Qb6139givenbirth
- Qb6161pregnant
- Qb6163wantedpregnancy
- Qb6164
- Qb6165
- Qb6202sti
- Qb6203discharge
- Qb6204
- Qb6206seektreatment
- Qb6208informpatner
- Qb6209stopsex
- Qb6210usecondom
- hivpositive
- hsv2positive

Mismas que se pueden encontrar dentro de la página de replicación del trabajo de Duflo. La homologación de características mantuvo la estratificación de grupo de control e intervenido por el programa analizado en ese trabajo.

La simulación de poblaciones se realizó creando una matriz de datos, donde cada fila representa un individuo y cada columna una covariable. Inicialmente vacío, se asignaron individuos hombres y mujeres y su diagnóstico de infección de HSV-2 de acuerdo a la descripción proporcionada por el artículo de Cowan.

Asumiendo independencia de las covariables, la asignación del resto de características se realizó de manera aleatoria aproximando la probabilidad de tener cierta característica dado el diagnóstico, estratificado por el grupo; control e intervención; y el sexo; hombre o mujer; utilizando el teorema de Bayes.

$$P(C|D) = \frac{P(C|D)P(D)}{P(C)}$$

donde:



- $P(C|D)$  es la probabilidad de tener cierta característica dado el diagnóstico de HSV-2.
- $P(D)$  es la probabilidad frecuentista a priori de tener el diagnóstico, de acuerdo al estrato (condición de sexo y grupo al que pertenece), determinada en el trabajo de Cowan.
- $P(C)$  es la probabilidad frecuentista a priori de tener la característica, de acuerdo al estrato (condición de sexo y grupo al que pertenece), determinada en el trabajo de Cowan.
- $P(D|C)$  es la probabilidad condicional de tener el diagnóstico dada la característica, estimada de manera frecuentista y estratificada a través de los datos de Duflo.

De este modo, se intenta aproximar la estructura real de los datos utilizados en el trabajo de Cowan, aproximando poblaciones de control e intervenidas por ese trabajo.

La simulación de los efectos en covariables de educación o hábitos sexuales se realiza de manera aleatoria, modificando con cierta probabilidad las covariables de interés.

La estimación de la prevalencia de la enfermedad en las poblaciones intervenidas simuladas se realiza a través de la proporción de infectados en la población, donde el diagnóstico se obtiene a través de una red neuronal perceptrón multicapa de clasificación que asigna el diagnóstico de cada persona en la población en función de las características de cada sujeto.

La red tiene una arquitectura conformada por 5 capas ocultas completamente conectadas estructuradas de la siguiente manera: 26 neuronas en la capa de entrada, por ser las características de la población en formato *one hot*, 70 en la primera capa oculta, 140 en la segunda, 280 en la tercera, 140 en la cuarta, 70 en la quinta y 1 de salida, siendo esta la probabilidad de tener la enfermedad. Todas ellas utilizan la función de activación sigmoide y la función de pérdida entropía cruzada binaria.

Se escogió el diagnóstico de HSV-2 como la variable de respuesta del modelo dada la prevalencia extremadamente baja de infecciones de VIH dentro de los datos disponibles.

Se probó el modelo entrenado con los datos del trabajo de Duflo y entrenado con datos simulados, en ambos casos con 140,000 iteraciones sobre el conjunto de entrenamiento. Los conjuntos de entrenamiento se construyeron con una selección estratificada de datos de la variable de respuesta, diagnóstico de HSV-2, realizando un balanceo de clases para obtener una población de entrenamiento 40 % infectados y 60 % sanos. El modelo seleccionado para la estimación de

prevalencias se escogió en función del desempeño de la matriz de confusión en ambos conjuntos de datos (reales y simulados) para los conjuntos de prueba.

La detección de los factores asociados con la prevalencia se realizó a través de experimentos y comparación de prevalencia de cada uno con el cálculo original de la población de control.

## 6. Descripción de la base de datos.

Es fundamental conocer la estructura de los datos reales, pues implica la detección de sesgos en algunas características. Por este fin, se presentan gráficas de barras que muestran las características de la población comparando a las poblaciones de los grupos de control e intervenidos de los datos del trabajo de Duflo.

A continuación se muestran gráficas de barras que expresan la prevalencia de las clases en los datos del trabajo de Duflo estratificadas por grupos de control e intervención, por sexo en el grupo de control y por sexo en el grupo de intervención.

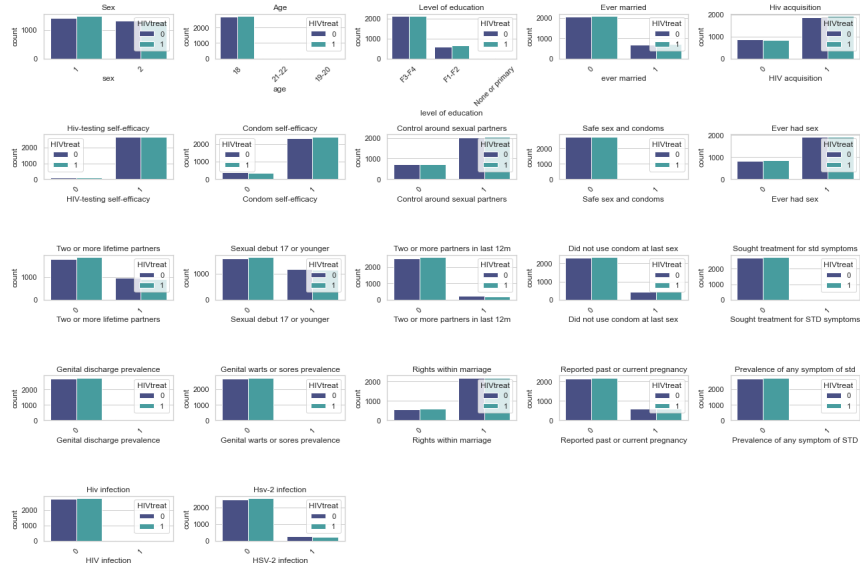


Figura 2: Duflo: Comparación de conteo de clases por grupo.

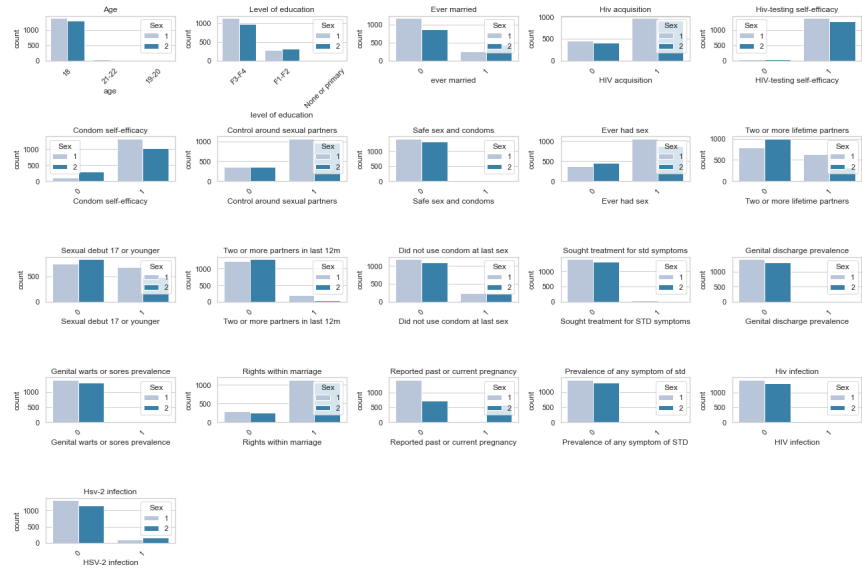


Figura 3: Duflo: Comparación de conteo de clases por sexo en el grupo de control.

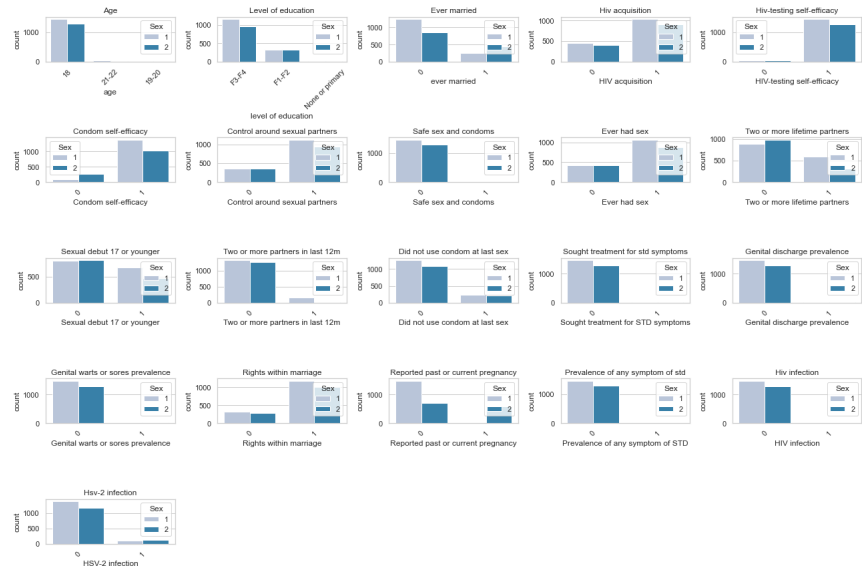


Figura 4: Duflo: Comparación de conteo de clases por sexo en el grupo intervenido.

Es fácilmente apreciable la existencia de una dominancia clara de algunas clases en el conjunto de datos, lo cual corresponde con un desafío difícil de tratar

en el entrenamiento de la red.

La mayoría de las relaciones entre las covariables incluidas tienen un coeficiente de contingencia bajo, lo que expresa cierta independencia entre las covariables analizadas.

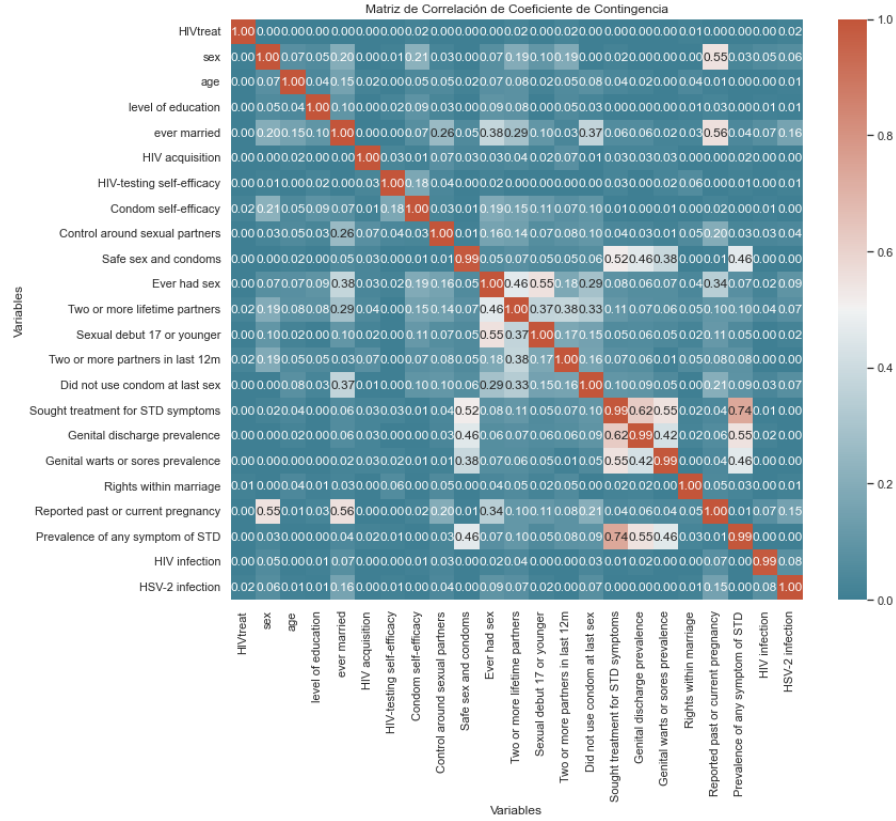


Figura 5: Dufflo: Matriz de contingencias.

Sin embargo, sí existen covariables relacionadas y que hacen sentido. Tal es el caso de la prevalencia de ITS con búsqueda de tratamiento, descarga genital y presencia de heridas genitales; haber tenido relaciones sexuales con dos o más parejas sexuales a lo largo de la vida, haberlas tenido antes de los 17 y ligeramente con no haber usado condón la última vez y tener más de dos parejas en 12 meses o bien estar casado con haber tenido relaciones sexuales, haber presentado embarazo o no usar condón en la última relación sexual.

Del mismo modo, se presenta el resumen para los datos simulados, correspondientes al total de 10 simulaciones.

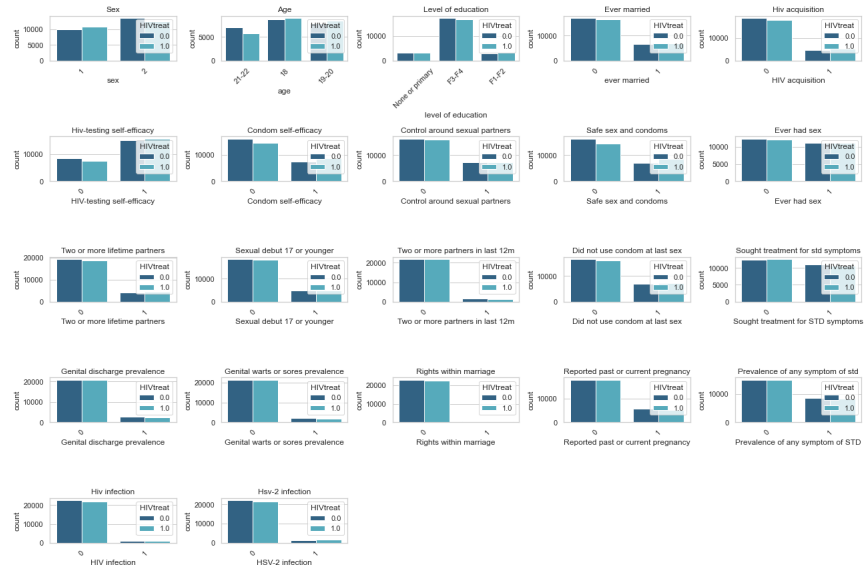


Figura 6: Simulación: Comparación de conteo de clases por grupo.

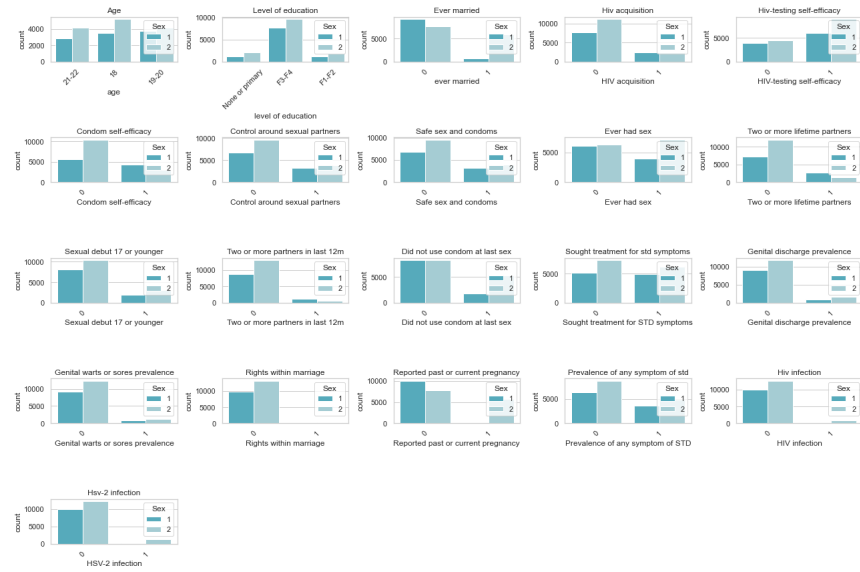


Figura 7: Simulación: Comparación de conteo de clases por sexo en el grupo de control.

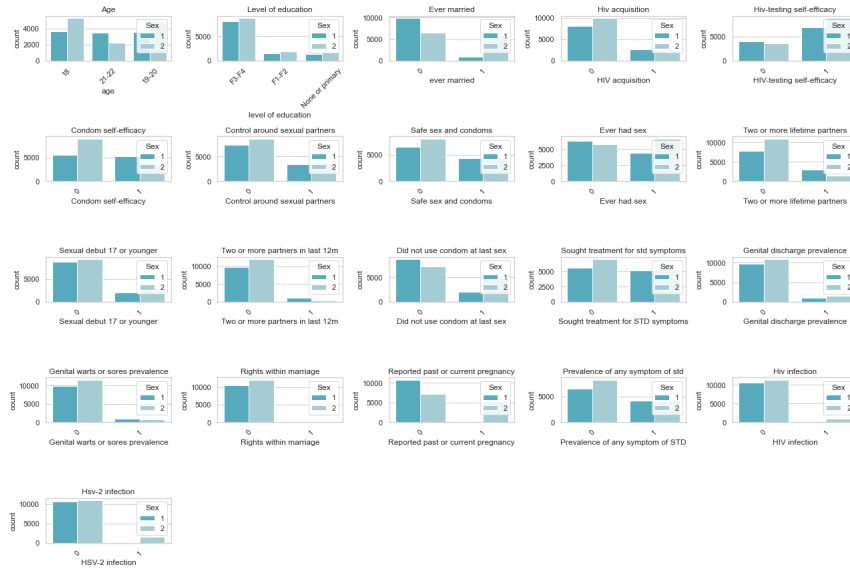


Figura 8: Simulación: Comparación de conteo de clases por sexo en el grupo intervenido.

En los datos simulados se presenta prevalencia de algunas clases, aunque en cierta medida, este conjunto parece estar mejor balanceado en general. Los sesgos hacia alguna de ellas estén causados probablemente por la probabilidad condicional estimada de los datos de Duflo puesto que estos tienen una tendencia clara en algunos casos. Pese a que en general los datos descritos en el trabajo de Cowan parecen estar balanceados en cierta medida, la simulación no puede evitar la prevalencia de algunas clases. Precisamente por el desbalance que existe con las variables de interés: diagnósticos de VIH y HSV-2, se optó por seleccionar HSV-2 como la variable de respuesta, intentando así mejorar el desempeño de la red tras realizar el entrenamiento.

La tabla de contingencias refleja pérdida de la relación existente entre algunas covariables que, aunque beneficioso para el entrenamiento de la red, se podría explorar modificar el proceso de simulación condicionando la probabilidad a algunos cofactores relacionados, además del diagnóstico, pues el supuesto de independencia se hace notar en los resultados de la simulación. Se podría explorar la simulación utilizando covariables que cumplan en mayor medida el supuesto de independencia.

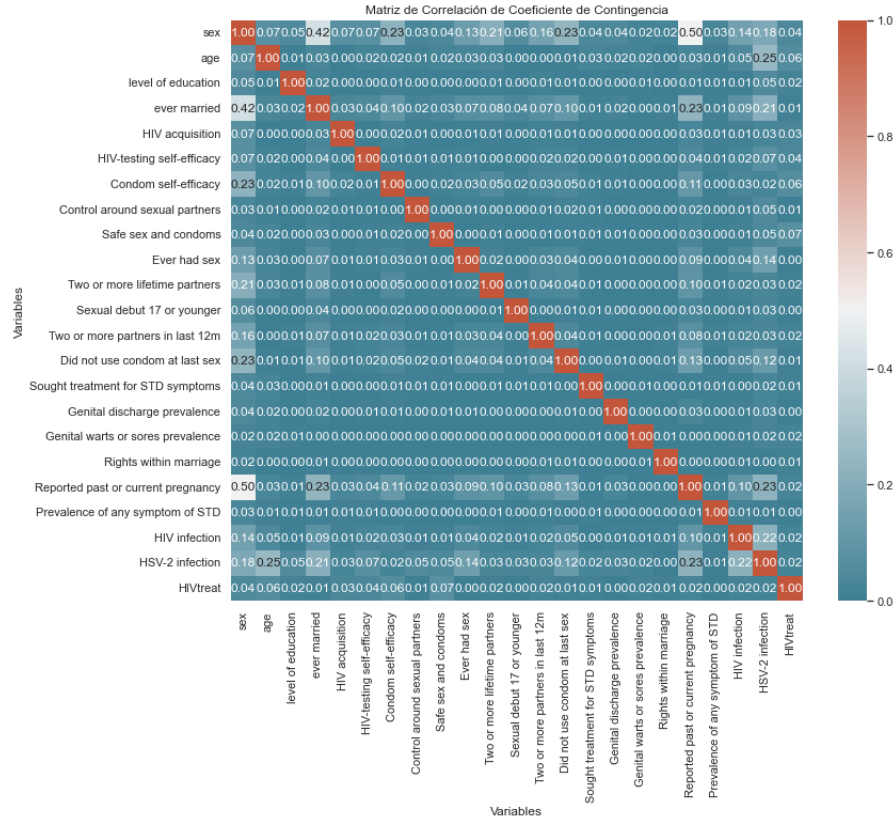


Figura 9: Simulación: Matriz de contingencias.

Aún así, el método parece conservar parte de la estructura de correlación relacionada con el sexo, el diagnóstico y la situación marital.

## 7. Desempeño de los modelos.

Se realizó el entrenamiento de la misma arquitectura de red utilizando datos reales y datos simulados. El desempeño de los mismos se evaluó comparando precisión, sensibilidad, especificidad, puntuación F1 y la matriz de confusión obtenidas en los conjuntos de datos de entrenamiento y prueba de ambos tipos, simulados y reales.

### 7.1. Entrenado con datos reales.

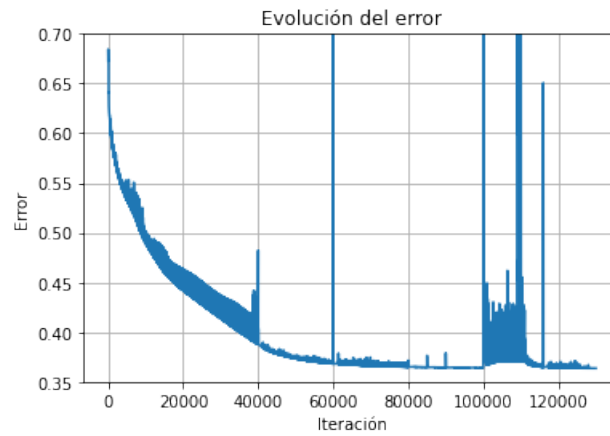


Figura 10: Evolución del error del modelo entrenado con datos reales.

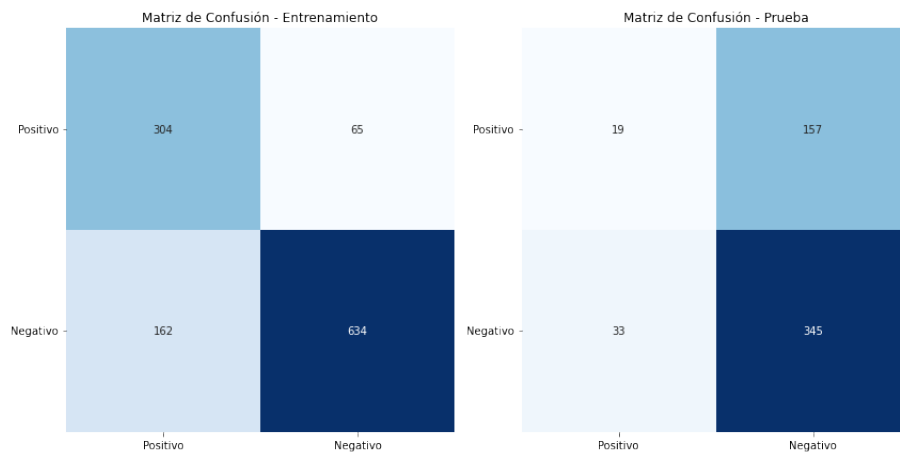


Figura 11: Matriz de confusión con datos reales.

Métricas del conjunto de entrenamiento:

- Precisión (Accuracy): 0.80
- Sensibilidad (Recall): 0.65
- Especificidad: 0.91
- Puntuación F1: 0.72



Métricas del conjunto de prueba:

- Precisión (Accuracy): 0.66
- Sensibilidad (Recall): 0.37
- Especificidad: 0.69
- Puntuación F1: 0.47

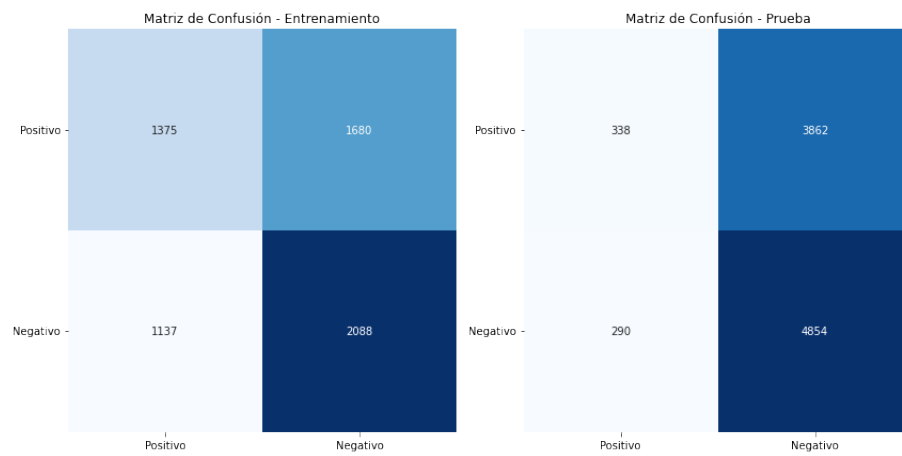


Figura 12: Matriz de confusión con datos simulados.

Métricas del conjunto de entrenamiento:

- Precisión (Accuracy): 0.55
- Sensibilidad (Recall): 0.55
- Especificidad: 0.55
- Puntuación F1: 0.55

Métricas del conjunto de prueba:

- Precisión (Accuracy): 0.55
- Sensibilidad (Recall): 0.54
- Especificidad: 0.57
- Puntuación F1: 0.55

## 7.2. Entrenado con datos simulados.

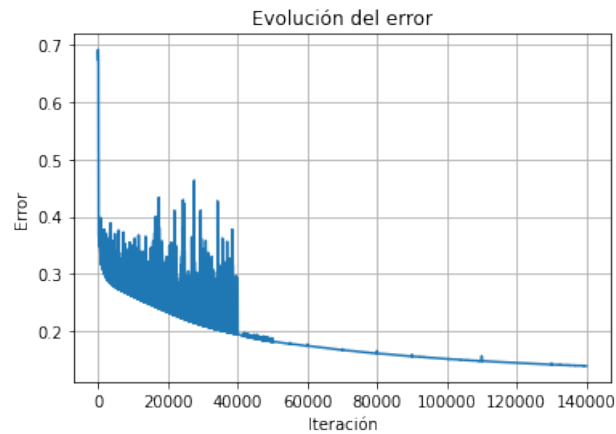


Figura 13: Evolución del error del modelo entrenado con datos simulados.

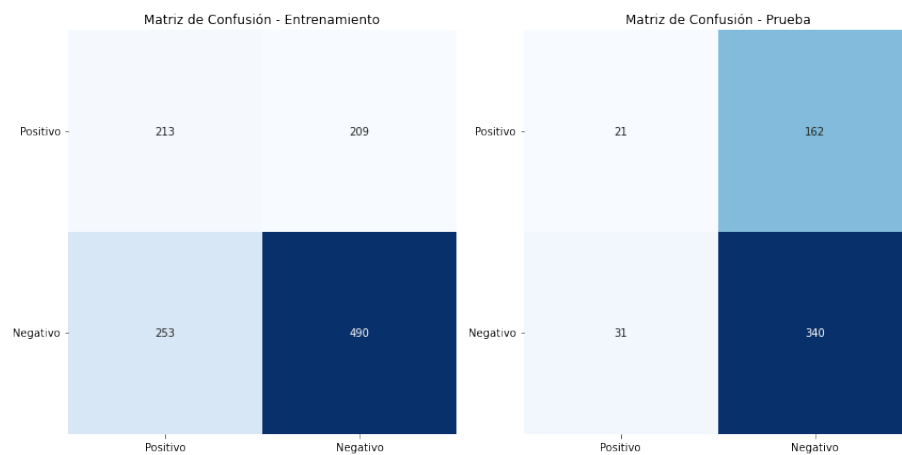


Figura 14: Matriz de confusión con datos reales.

Métricas del conjunto de entrenamiento:

- Precisión (Accuracy): 0.60
- Sensibilidad (Recall): 0.46
- Especificidad: 0.70
- Puntuación F1: 0.52

Métricas del conjunto de prueba:

- Precisión (Accuracy): 0.65
- Sensibilidad (Recall): 0.40
- Especificidad: 0.68
- Puntuación F1: 0.50

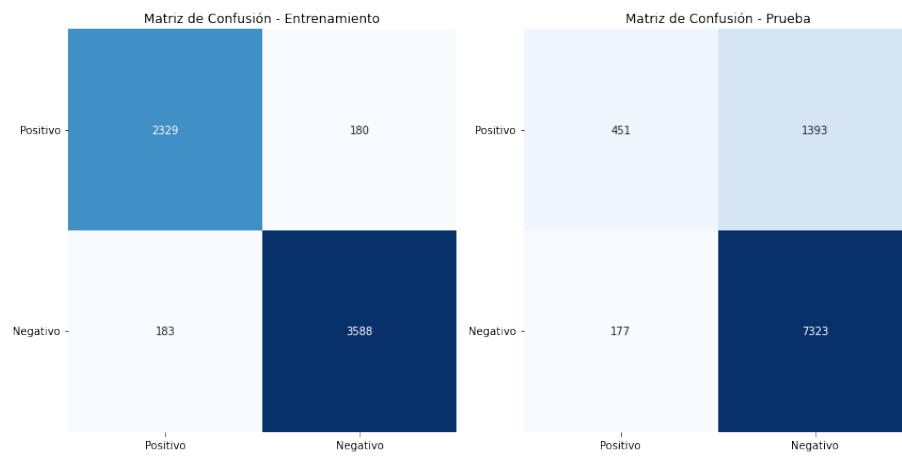


Figura 15: Matriz de confusión con datos simulados.

Métricas del conjunto de entrenamiento:

- Precisión (Accuracy): 0.94
- Sensibilidad (Recall): 0.93
- Especificidad: 0.95
- Puntuación F1: 0.93

Métricas del conjunto de prueba:

- Precisión (Accuracy): 0.83
- Sensibilidad (Recall): 0.72
- Especificidad: 0.84
- Puntuación F1: 0.77

En general, ambos modelos tienen un desempeño moderado-bajo en la clasificación de datos de prueba reales. Es posible que el desempeño obtenido por el conjunto de datos reales se deba a un posible sobreajuste para el conjunto de entrenamiento pues es evidente la caída en las métricas de evaluación en generalización.

Por otro lado, el modelo entrenado con datos simulados tiene un buen desempeño en los conjuntos de datos simulados de entrenamiento y prueba, superando en líneas generales al modelo entrenado con datos reales en la clasificación del conjunto de prueba de datos reales.

El buen desempeño obtenido del modelo de datos simulados para generalizar refleja consistencia entre simulaciones, y por tanto en el método de simulación, puesto que no caen sus métricas de manera abrupta siendo que los datos pertenecen a varias simulaciones diferentes. Sin embargo, al enfrentarse a datos reales pierde su capacidad observada en los conjuntos simulados, mostrando un área de oportunidad para conseguir simulaciones más representativas de lo que puede suceder en una población real.

Se puede explorar simulación a través de introducir condicionalidad a más de una característica cuando existe dependencia observada de covariables. Otro posible factor causante de la caída de desempeño es la diferencia que podrían haber tenido las poblaciones originales del conjunto de datos del trabajo de Cowan y del trabajo del Duflo, pues al utilizar estimaciones de ambas poblaciones podrían existir dificultades para representar las características originales de cualquiera de ellas. Así, se podría explorar el método de simulación utilizando únicamente estimaciones de la población de datos del trabajo de Duflo.

Considerando lo anterior y el hecho de que se utilizarán experimentos de poblaciones de control simuladas para la identificación de factores asociados a la prevalencia de HSV-2, para este fin, se opta por utilizar el modelo entrenado con datos simulados.

## 8. Experimentos.

La identificación de factores asociados con la prevalencia de HSV-2 se hará a través de experimentos simulados donde se mide la proporción de diagnósticos positivos proporcionados por la red.

Para la ejecución de los experimentos se utilizó una población simulada a partir de 5 simulaciones de control. En cada experimento se utiliza la misma población inicial y se modifican algunas covariables para expresar el efecto del experimento.

El desempeño del modelo en la población de control utilizada es similar a aquellas utilizadas en entrenamiento y prueba pese a ser ajenas.

- Precisión (Accuracy): 0.8124
- Sensibilidad (Recall): 0.71
- Especificidad: 0.82
- Puntuación F1: 0.7556
- VP/P : 0.21

Es igualmente importante agregar como referencia que el modelo clasifica con convicción a un individuo con características no relacionadas con el HSV-2 con un diagnóstico negativo, mientras que clasifica a un individuo con las características totalmente relacionadas con el HSV-2 con un diagnóstico positivo.

Es altamente probable que la prevalencia de HSV-2 esté sobreestimada dada la dificultad de detección de falsos positivos. En la experimentación no se puede evaluar la efectividad del modelo, pero se espera que los resultados de prevalencia sean proporcionales a los que un escenario que cuente con la información arrojaría.

En los siguiente experimentos se simula modificando las covariables de interés con probabilidad  $p$ , para  $p$  de 0 a 1. Las covariables educativas de interés son: conocimiento de transmisión de VIH, conocimiento sobre pruebas de VIH, conocimientos de uso del condón, control sobre pareja sexual, sexo seguro y condones, búsqueda de tratamiento por síntomas de ITS y derechos maritales. Las covariables de interés para hábitos sexuales son: debut sexual antes de los 17, dos o más parejas en 12 meses, no uso del condón en la última relación sexual y reporte de embarazo.

### 8.1. Efecto de la educación por si sola.

En la siguiente gráfica se expresa la evolución de prevalencia de acuerdo a una mejora uniforme del  $p\%$  en materia de educación sexual, es decir, que bajo el efecto de una campaña el  $p\%$  aprenda sobre cada uno de los temas.

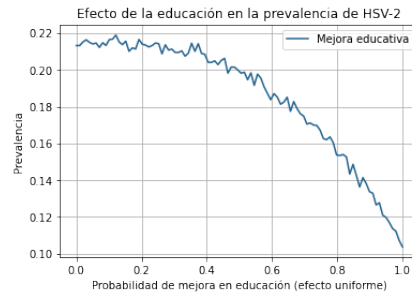


Figura 16: Efecto de la educación por si sola.

## 8.2. Efecto de mejores hábitos por si solos.

A continuación se expresa la evolución de prevalencia de acuerdo a una mejora uniforme del  $p\%$  en materia de hábitos sexuales.

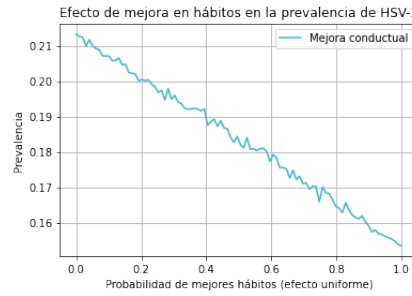


Figura 17: Efecto de mejores hábitos por si solos.

## 8.3. Efecto uniforme conjunto de educación y mejores hábitos.

Se representa la evolución de prevalencia de acuerdo a una mejora uniforme del  $p\%$  en materia de hábitos sexuales y educación simultáneamente.

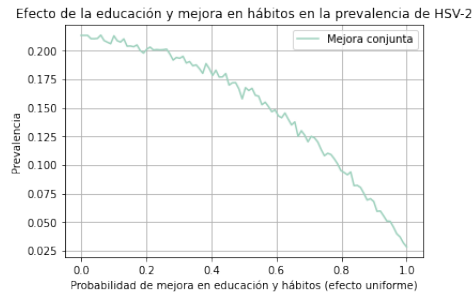


Figura 18: Efecto de mejores hábitos y conocimientos.

#### 8.4. Comparación: mejores hábitos, mejores conductas y mejora conjunta.

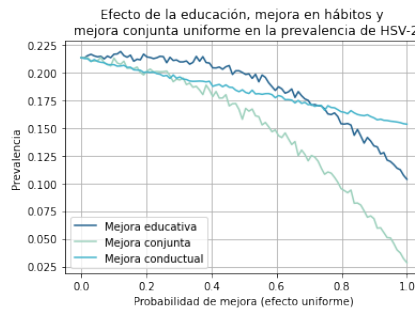


Figura 19: Comparación de mejores conductas, conocimientos y en conjunto.

Tal como es esperado, la mejora conjunta en la misma proporción en comportamientos sexuales y conocimientos disminuye la prevalencia de la enfermedad en gran medida, observando disminución considerable incluso en efectos conjuntos pequeños. La mejora conjunta supera ampliamente a la mejora individual en cualquiera de los dos aspectos, pero se encuentra cerca a la mejora individual en conducta para ligeros efectos de mejoría.

Por su parte, las mejoras individuales tienen un comportamiento interesante. En lo que se refiere a las mejoras educativas se presenta una situación en la que mejoras pequeñas tienen un efecto pequeño pero a medida que este aumenta el efecto crece significativamente.

La mejora conductual presenta una mejoría proporcional al efecto de tal mejora y se ve superado después de cierta tasa de efecto por la mejora educativa. La mejora conductual tiene un efecto constante y se presenta como una alternativa eficaz para disminuir la prevalencia rápidamente pues su efecto inicial se apro-

xima al efecto conjunto incluso hasta antes del 40 % de efectividad, pero este se ve superado ampliamente después de esta tasa y también por una mejora individual educativa al rededor del 75 %.

La tercia de experimentos proporciona información valiosa del margen de mejora de una población dadas sus características y presenta una idea de aquellos esfuerzos que se pueden realizar para conseguir disminuir la prevalencia de manera óptima.

### 8.5. Efecto del desconocimiento.

Resulta interesante pensar en las consecuencias que podría tener una población que tiene un déficit educativo al de la población de control, pensar en la estimación de prevalencia de una población que desconoce encierta proporción sobre educación sexual.

Para este fin se modifica de manera negativa con cierta probabilidad el conocimiento de los individuos de cada tema.

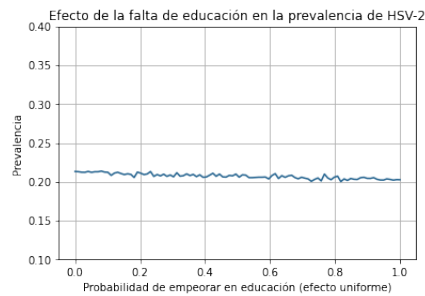


Figura 20: Efecto del desconocimiento.

Es interesante que el efecto del desconocimiento es prácticamente imperceptible, manteniendo la prevalencia constante, sugiriendo que la falta de educación por sí misma no es un factor que aumente la prevalencia de HSV-2.

De acuerdo al modelo, esto sitúa a la educación sexual como una herramienta exclusiva para disminuir la prevalencia de HSV-2 pero no como un argumento para justificar la prevalencia de la enfermedad en una población.

### 8.6. Efecto de peores hábitos.

Analizar el efecto de una población con peores hábitos sexuales que los de la población de control puede ayudar a entender la relevancia que tienen estos en la prevalencia de HSV-2.



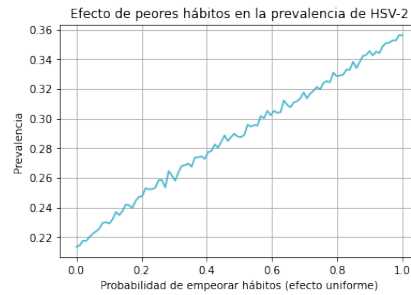


Figura 21: Efecto de peor conducta.

El efecto de peores comportamientos sexuales es dramático a comparación de la falta de conocimientos. Peores conductas sexuales, de acuerdo al modelo, sí están asociados con una prevalencia mucho mayor. Esto resulta interesante para comprender la situación de prevalencia de una población de interés y también para entender cómo factores sociales no relacionados con la educación sexual pero que están relacionados con conducta sexual impactan la prevalencia de la ITS.

### 8.7. Efectos de hábitos sexuales.

Se exploró la relevancia de cada hábito sexual para comprender cuáles están estrechamente vinculados con la prevalencia de la enfermedad.

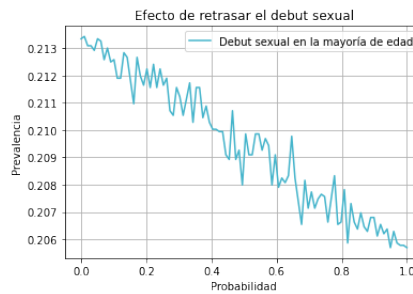


Figura 22: Efecto de retrasar el debut sexual.

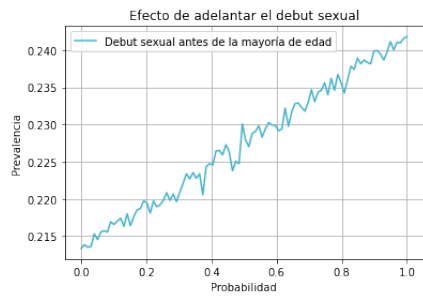


Figura 23: Efecto de adelantar el debut sexual.

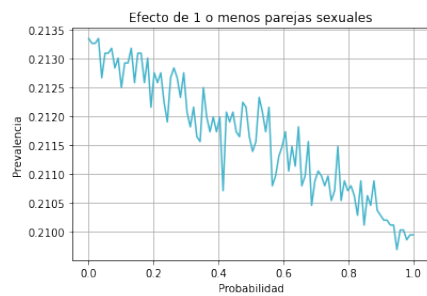


Figura 24: Efecto de 1 o menos parejas sexuales en 12 meses.

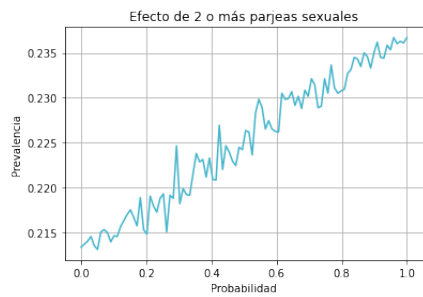


Figura 25: Efecto de 2 o más parejas sexuales en 12 meses.

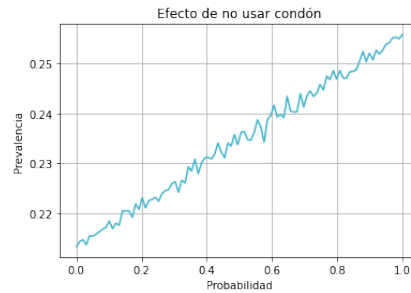


Figura 26: Efecto de no usar condón.

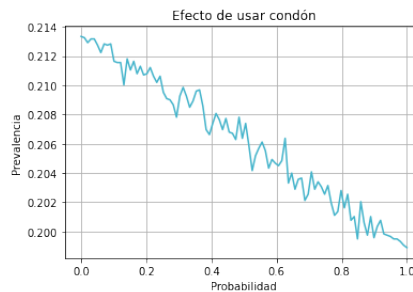


Figura 27: Efecto de usar condón.

De primera instancia parece que los comportamientos sexuales por separado no tienen un impacto significativo. Una población adoptando un único comportamiento sexual no disminuye la prevalencia de la enfermedad significativamente, no obstante, una población que presente peores hábitos o comportamientos en un único aspecto sí puede presentar una prevalencia mayor, especialmente si presentan un debut sexual adelantado, tienen más parejas sexuales o no utilizan el condón. Estos tres aspectos parecen ser los que están vinculados fuertemente con la prevalencia de la ITS, sin dejar precisamente atrás a la prevalencia que tiene una población con más de dos parejas sexuales en 12 meses.

En general, se puede apreciar que el modelo no considera que una mejora en un único aspecto disminuya la prevalencia de manera significativa pero que una población con un único comportamiento peor sí aumenta su prevalencia. De tal modo, un buen hábito sexual no parece hacer mucha diferencia mientras que un mal hábito sexual sí.

## 8.8. Efecto de la abstinencia.

La abstinencia, pensada como no haber tenido relaciones sexuales, podría ser potencialmente efectiva en la disminución de prevalencia de HSV-2 puesto

que, si un individuo no ha tenido relaciones sexuales entonces su comportamiento adquiere, a causa de la ausencia de actividad sexual, características como menos de 2 parejas en 12 meses, negativo a no uso de condón en la última relación, negativo a embarazo, negativo a dos o más parejas sexuales históricas y debut sexual después de los 18. Al tener influencia en varios parámetros, la abstinencia podría ser uno de los factores que podrían disminuir la prevalencia de manera significativa.

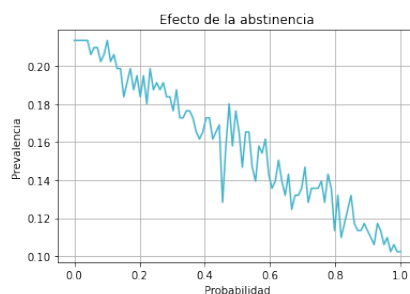


Figura 28: Efecto de la abstinencia.

Las simulaciones muestran cómo el efecto de la abstinencia en la población logra disminuir rápidamente la prevalencia de la enfermedad disminuyéndola hasta en un 50 %.

## 9. Efecto conjunto de mejoras independientes en educación y hábitos sexuales.

Es posible que mejorar conocimientos sea, en general, una tarea más sencilla que mejorar los hábitos sexuales de una población. Por tanto, se plantea cómo afecta una mejora a diferente tasa en cada uno de los aspectos con la intención de encontrar cuáles son las mejores combinaciones en cuanto a la disminución de prevalencia.

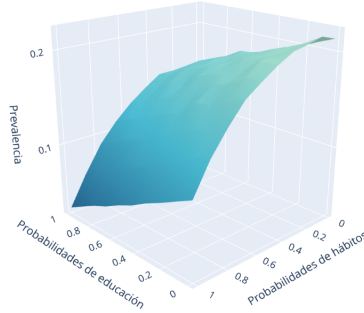


Figura 29: Efecto de mejora conjunta independiente.

La superficie de prevalencia parece estar conformada en general por la suma de los efectos de mejora en hábitos sexuales y mejora en educación, no parece existir una interacción más allá de lo que consiguen mejorar por separado.

Esta aproximación refleja que es prioridad mejorar los hábitos sexuales de una población pues disminuye la prevalencia en mayor medida que la educación sexual. Por otro lado, en caso de existir dificultad para mejorar los hábitos a partir de cierta proporción, se puede aproximar la tasa de mejora de la educación sexual necesaria para llegar a una cota esperada.

Gracias a este enfoque es posible comprender en qué medida un proyecto puede disminuir la prevalencia de la infección de acuerdo a sus pretenciones, además de servir de guía para conseguir combinaciones más eficientes en caso de poder asociar la mejora de cada parámetro con alguna función de costo a optimizar.

## 10. Conclusión.

En este trabajo, se ha demostrado el uso eficaz de la simulación de poblaciones y el entrenamiento de redes neuronales multicapa para entender los factores que influyen en la prevalencia de HSV-2. La metodología incluyó la simulación de datos basados en los estudios de Duflo y Cowan, asumiendo independencia entre covariables y utilizando el teorema de Bayes para asignar probabilidades condicionales.

La simulación de datos ha proporcionado una herramienta valiosa para entender cómo diferentes factores afectan la prevalencia de HSV-2. Los resultados muestran que el modelo entrenado con datos simulados tiene un mejor desempeño en la generalización comparado con el modelo entrenado con datos reales, aunque ambos modelos mostraron un desempeño moderado-bajo en la clasificación de datos de prueba reales. Esto sugiere que, aunque la simulación conserva parte

de la estructura de correlación observada en los datos reales, aún hay un margen significativo para mejorar la representatividad de las simulaciones.

Un hallazgo clave es la importancia de la educación sexual y los hábitos sexuales en la prevalencia de HSV-2. Las simulaciones indican que la mejora conjunta en educación sexual y hábitos sexuales es significativamente más efectiva en reducir la prevalencia de la enfermedad que las mejoras individuales en cualquiera de estos aspectos. Específicamente, una mejora en los conocimientos sexuales, aunque inicialmente tiene un impacto modesto, se vuelve considerablemente más efectiva a medida que aumenta. Por otro lado, la mejora en los hábitos sexuales muestra un efecto más constante y significativo, especialmente en los primeros niveles de intervención.

La simulación de peores condiciones educativas reveló que la falta de educación sexual por sí sola no incrementa significativamente la prevalencia de HSV-2, sugiriendo que la educación sexual actúa más como un mecanismo de mitigación que como un factor determinante en la prevalencia de la enfermedad. En contraste, los peores hábitos sexuales tienen un impacto dramático, aumentando considerablemente la prevalencia de HSV-2. Esto subraya la importancia de los comportamientos sexuales seguros y su influencia directa en la propagación de la infección.

En la exploración de hábitos sexuales específicos, se encontró que el debut sexual temprano, tener múltiples parejas sexuales en un período corto y no usar condón son factores fuertemente asociados con una mayor prevalencia de HSV-2. Mejorar un solo aspecto de los comportamientos sexuales no tiene un impacto significativo en la disminución de la prevalencia, mientras que empeorar un solo aspecto puede aumentar considerablemente la prevalencia.

Los resultados enfatizan la importancia de intervenciones combinadas que aborden tanto la educación sexual como la promoción de hábitos sexuales seguros. Las políticas públicas y los programas de salud deben enfocarse en estrategias integrales que no solo aumenten el conocimiento sobre las ITS, sino que también fomenten comportamientos sexuales responsables. La simulación de datos y el uso de modelos predictivos ofrecen una herramienta poderosa para planificar y evaluar la efectividad de dichas intervenciones, proporcionando una base sólida para la toma de decisiones informadas en la lucha contra las infecciones de transmisión sexual.

## 11. Referencias.

- Kirby, D. B., Laris, B. A., & Roller, L. A. (2007). Sex and HIV education programs: Their impact on sexual behaviors of young people throughout the world. *Journal of Adolescent Health*, **40**(3), 206–217.  
<https://doi.org/10.1016/j.jadohealth.2006.11.143>
- Cowan, F. M., Pascoe, S. J. S., Langhaug, L. F., Mavhu, W., Chidiya, S., Jaffar, S., ... Hayes, R. J. (2010). The Regai Dzive Shiri Project: results of a randomised trial of an HIV prevention intervention for Zimbabwean youth. *AIDS*, **24**(16), 2541–2552. doi: 10.1097/QAD.0b013e32833e77c9
- Duflo, E., Dupas, P., & Kremer, M. (2015). Education, HIV, and Early Fertility: Experimental Evidence from Kenya. *American Economic Review*, **105**(9), 2757–2797. doi: 10.1257/aer.20121607  
Enlace de la página de replicación: <https://www.openicpsr.org/openicpsr/project/112899/version/V1/view>