

Universidad de San Andrés



Universidad de
SanAndrés

BIG DATA

RESEARCH PROPOSAL

**En la encrucijada entre la inocencia y la
responsabilidad: efectos del plan ENIA en el embarazo
adolescente en Argentina**

Profesor: Noelia Romero

Tutor: Victoria Oubiña

Alumnos: Solana Cucher, Victoria Rosino, Florencia Ruiz.

1. Introducción

En la encrucijada entre la inocencia y la responsabilidad, el fenómeno del embarazo adolescente emerge como un desafío social y personal. La maternidad precoz no solo conlleva riesgos para la salud, sino que también pone en peligro las trayectorias de vida y las perspectivas de integración social de las niñas y adolescentes. Dentro de este complejo panorama, Argentina se presenta como un escenario esperanzador: el embarazo en la adolescencia en el país bajó un 60% en la última década. Las cifras oficiales muestran que los nacidos vivos de madres adolescentes en 2013 fueron 117.386, mientras que para 2023 fueron solo 47.630. En este sentido, “son 70 mil embarazos menos, 70 mil niñas y adolescentes con más oportunidades” (Iglesias, 2023). La actual tendencia a la baja que indican las estadísticas oficiales (SENAF, 2022) nos lleva a preguntarnos cuáles son las principales causas de este fenómeno. En particular, nos interesa averiguar si los programas de asistencia social implementados durante 2017-2019 contribuyeron a reducir la fecundidad y la natalidad adolescente. En 2017 el Estado decidió ubicar en primer plano la problemática del embarazo precoz y estructurar una respuesta integral de política pública. Como resultado surgió el Plan Nacional de Prevención del Embarazo No Intencional en la Adolescencia (ENIA) que garantiza la libertad de elección, el acceso y calidad en la provisión de anticonceptivos de largo plazo (como el implante subdérmico y DIU), el fortalecimiento de las políticas para la prevención del abuso y la violencia sexual y el acceso a la Interrupción Voluntaria del Embarazo (IVE).

No obstante, a pesar de la mejora en términos de embarazo adolescente observada en los últimos años, la problemática sigue vigente y es de vital importancia entender sus causas, sobre todo por las consecuencias que podría tener el embarazo

precoz en la profundización de la pobreza y desigualdad. Podemos destacar tres razones principales por las cuales esta temática requiere especial consideración (MECCT, 2019a), que son las que motivan la presente investigación. En primer lugar, el embarazo adolescente es prácticamente un sinónimo de inequidad social. Este fenómeno posee un marcado gradiente socioeconómico, que afecta de manera indiscriminada a mujeres de estratos más vulnerables. En segundo lugar, la mayor parte de los embarazos adolescentes pueden ser catalogados como embarazos involuntarios. En particular, el 58.9% de los embarazos de adolescentes entre 15 a 19 años y el 83.4% de los embarazos de las menores de 15 años son reportados como no intencionales (PNSIA, 2018). Por último, la fecundidad adolescente trae emparejada múltiples consecuencias para las futuras madres. Entre ellas podemos destacar el abandono escolar, la inducción a la unión en parejas inestables y la exposición a abortos.

Por estos motivos, consideramos relevante entender cuáles son los mecanismos efectivos para reducir el embarazo adolescente. En particular, la pregunta de investigación que guía esta propuesta es si el Plan ENIA contribuyó a bajar la tasa de embarazo adolescente tardío¹. Entre los servicios brindados por el programa podemos destacar la educación sexual integral (ESI), las asesorías de salud integral en escuelas secundarias, la provisión de métodos anticonceptivos, las consejerías de salud sexual, el acceso a la justicia y más. Además, el plan no solo ofrece servicios, sino también capacitaciones a los prestadores de salud y a las instituciones educativas. Este programa fue diseñado en el año 2017 y su implementación comenzó a mediados del año 2018. Actualmente, la iniciativa se ejecuta en 36 departamentos ubicados en 12

¹ Según el Ministerio de Salud y Desarrollo Social de la Nación Argentina, se define al embarazo adolescente tardío como la fecundidad de mujeres entre 15 y 19 años.

provincias priorizadas según dos criterios: la incidencia del embarazo no intencional en la adolescencia y la densidad de población en el departamento en cuestión. Las provincias seleccionadas son: Catamarca, Chaco, Corrientes, Entre Ríos, Formosa, Jujuy, La Rioja, Misiones, Salta, Santiago del Estero, Tucumán y Buenos Aires.

Para desarrollar nuestra propuesta de investigación, continuaremos de la siguiente manera. En la sección II realizaremos una breve revisión bibliográfica que nos ayudará a comprender las contribuciones de nuestra idea. En la sección III presentaremos la base de datos a utilizar. En la sección IV describiremos la metodología a emplear para la estimación de los efectos del Plan ENIA en el embarazo adolescente en Argentina. En último lugar, concluimos en la sección V.

2. Literatura previa

En el contexto de la investigación sobre la maternidad precoz y las políticas adoptadas para combatirla, resulta importante conocer qué aportes existen en la literatura y qué interrogantes faltan responder. A continuación, repasamos los hallazgos en relación con el embarazo adolescente, *machine learning* y las distintas aproximaciones causales.

Existen algunos trabajos que intentan identificar causalmente los efectos de planes de prevención del embarazo adolescente y la mayoría de ellos fueron realizados para Estados Unidos. Por ejemplo, el artículo de Bennett (2005), mediante una revisión bibliográfica sobre *randomized control trails* (RCTs), compara los efectos de distintos programas escolares en el embarazo precoz en Estados Unidos. El resultado obtenido es que, para reducir la tasa de embarazo adolescente, lo más eficiente es mejorar el uso y acceso a anticonceptivos entre los jóvenes. Esta conclusión favorece nuestra actual investigación, en tanto uno de los objetivos del plan ENIA fue mejorar el acceso a métodos de anticoncepción. La evidencia para Argentina, que comenzó a desarrollarse

recientemente, es todavía escasa. Por ejemplo, Roig (2023) emplea un modelo de diferencias en diferencias para estudiar el efecto causal del plan ENIA y encuentra que las tasas de natalidad se redujeron en aproximadamente un 10%. Los efectos comienzan sólo después de nueve meses de la adopción, la duración esperada de un embarazo. Sin embargo, parecería que aún no se han hecho esfuerzos por involucrar las técnicas de *machine learning* para identificar cuestiones relacionadas al embarazo adolescente.

Algunos estudios descriptivos fueron realizados en América Latina y, más en particular, en Argentina. Por ejemplo, el trabajo de Azevedo et al. (2012) encuentra que las adolescentes que tienen mayores probabilidades de quedar embarazadas son aquellas que tienen menor nivel educativo, provienen de familias pobres y viven en zonas periféricas. En Argentina en especial, también se observa este gradiente socioeconómico (Edo, Thailinger y Tommasi, 2021). Además, Edo, Talingher y Tomassi (2021) mencionan que a partir de 2015 podemos observar una importante caída en la fecundidad adolescente (21%). En el caso particular de Argentina, la caída en la fecundidad adolescente fue mayor en regiones más desarrolladas del país donde las tasas ya eran relativamente menores (Rofman, 2020).

En resumen, la contribución de este trabajo radica, por un lado, en combinar técnicas de *machine learning* y econometría aplicada para identificar causalidad en torno al embarazo adolescente. Por otro lado, aportaríamos evidencia para la Argentina sobre una temática relevante y poco estudiada en nuestro país.

3. Base de datos

Trabajaremos con la Encuesta de Indicadores Múltiples por Conglomerados (MICS, por sus siglas en inglés) llevada a cabo por UNICEF con el objetivo de proporcionar información actualizada para evaluar la situación de los niños, adolescentes y mujeres

en Argentina. Esta encuesta fue realizada en Argentina en dos ocasiones: en 2012, antes de la implementación del Plan ENIA, y en 2020, luego de la implementación de la política. La MICS de 2012 cuenta con cuatro cuestionarios: (1) Hogares, (2) Miembros del hogar, (3) Mujeres 15 a 49 años, (4) Niños menores a 5 años. En la MICS de 2020, a estos cuatro cuestionarios se agrega uno más para niños entre 0 y 5 años.

Para este trabajo, emplearemos el cuestionario de mujeres entre 15 y 49 años que contiene información sobre los antecedentes de cada individuo de la muestra, sus embarazos y si fueron deseados, el uso de anticonceptivos, la salud de las madres durante y luego el embarazo, su comportamiento sexual, matrimonio/unión, su consumo de tabaco/alcohol, entre otros datos. De este cuestionario, tomaremos únicamente a las mujeres entre 15 y 19 años al momento de la encuesta. De esta manera, nuestra muestra estará conformada por las adolescentes de la MICS 2012 más las de la MICS 2020. El *outcome* de interés es una variable binaria que indica si la mujer está embarazada o tuvo un hijo antes. En cuanto a las variables explicativas, podemos estadísticas descriptivas de las más relevantes, en la Tabla 1.

En primer lugar, podemos notar que nuestra muestra se encuentra balanceada en términos de la edad porque, en promedio, las mujeres poseen 17 años (tenemos en cuenta adolescentes entre 15 y 19 años). Luego, una variable relevante para explicar el embarazo precoz es la educación que reciben las jóvenes, entendiendo que el atraso o la inasistencia al colegio puede ser un síntoma de carencias más profundas (como ser familiares, alimentarias o de ingresos). Dado que en la variable de asistencia a una institución educativa el 1 implica que actualmente asiste y el 2 que no asiste ahora pero sí lo hizo, la media y la mediana indican que la mayoría de las adolescentes de nuestra muestra se encuentra en el sistema educativo.

Tabla 1 - Estadísticas descriptivas

	Media	Mediana	Desvío Estándar	Mínimo	Máximo
Edad	17.10	17.00	1.42	15.00	19.00
Asistencia al colegio	1.26	1.00	0.44	1.00	2.00
Grado más alto asistido	3.70	4.00	1.74	1.00	9.00
Cobertura médica	1.54	2.00	0.50	1.00	2.00
Aborto	1.98	2.00	0.13	1.00	2.00
Conoce el preservativo	1.02	1.00	0.13	1.00	2.00
Conoce las pastillas anticonceptivas	1.02	1.00	0.14	1.00	2.00
Convivencia con la pareja	2.90	3.00	0.31	1.00	3.00
Edad a la que tuvo relaciones por primera vez	9.15	14.00	7.62	0.00	19.00
AUH	1.64	2.00	0.65	1.00	3.00
Educación del jefe/a del hogar	0.43	0.00	0.65	0.00	2.00
Percentil de riqueza	4.65	4.00	2.83	1.00	10.00

En relación con el grado más alto asistido, observamos que, en promedio, las mujeres se encuentran casi en cuarto grado. Si contrastamos este dato con la media de las edades de la muestra (17 años), vemos que existe cierta repitencia ya que, por la edad promedio, lo adecuado sería que las mujeres de la muestra estén, en promedio, en el último año del secundario. Por otra parte, consideramos que poseer cobertura médica es un indicador socioeconómico importante. Según la mediana, al menos la mitad de las mujeres no poseen obra social de ningún tipo (2).

Otro tipo de variables son aquellas referidas al ámbito sexual y de la intimidad. Primero, podemos señalar que la mayoría de las adolescentes de nuestros datos no ha sufrido un aborto (2). Luego, podemos destacar que la mayoría tiene conocimiento sobre los métodos anticonceptivos más comunes (preservativos y pastillas anticonceptivas). Por otro lado, se les pregunta a las mujeres encuestadas si conviven con su pareja o marido. Vemos que, en promedio, la variable toma un valor de 2,9 lo

que significa que las mujeres no se encuentran ni unidas ni casadas (3). Por último, podemos analizar la edad a la que las jóvenes tuvieron relaciones sexuales por primera vez. Si bien el promedio es sorprendentemente bajo (9.15 años), el desvío estándar es alto (7.62). Por lo tanto, un mejor indicador sería la mediana que nos indica que al menos el 50% de las mujeres tuvieron relaciones íntimas a los 14 años o menos.

En último lugar, podemos resaltar algunas variables que revelan las condiciones socioeconómicas de las jóvenes. La primera de ellas es si el hogar en el que habitan recibe Asignación Universal por Hijo (AUH). El hecho de cobrar esta asistencia refleja cierta fragilidad social debido a los requisitos de la AUH (como informalidad o desempleo). Tanto la media como la mediana indicarían que la mayoría de los hogares no reciben AUH. Luego, la máxima educación alcanzada por el jefe del hogar es un buen predictor tanto de la educación que puede alcanzar su descendencia como también de las condiciones de vida en el hogar. Como vemos en la Tabla 1, la media y mediana son cercanas a 0, lo que significa que la muestra está cercana al secundario incompleto. Finalmente, el percentil de ingreso en el que se clasifica el hogar en el que vive la mujer indica que, en promedio, las mujeres pertenecen al percentil 5 (aproximadamente). Lo que refleja, una vez más, que nuestra muestra se encuentra balanceada.

4. Metodología

El objetivo principal de este trabajo será estimar el efecto causal de la exposición al plan ENIA de prevención del embarazo adolescente y, para ello, nos basaremos en técnicas de *machine learning*. En particular, proponemos estimar la efectividad de la política mediante el modelo *casual random forest* desarrollado por Wager & Athey (2018), que extiende el algoritmo tradicional de bosques aleatorios para poder identificar un coeficiente β de forma casual en los nodos terminales.

El efecto del tratamiento (T_i) condicional en x se define como:

$$\tau(x) = E[Y_i^{(1)} - Y_i^{(0)} | X_i = x] \quad (1)$$

donde $Y_i^{(1)}$ e $Y_i^{(0)}$ son los *outcomes* potenciales del individuo i como tratado y control respectivamente, X_i es un vector de características del individuo i y τ es el efecto del tratamiento. El problema con (1) es que, para cada persona, solo podemos observar uno de los dos resultados potenciales $Y_i^{(0)}$ e $Y_i^{(1)}$. Por lo tanto, no se puede entrenar un modelo de aprendizaje automático para la ecuación (1). Con el objetivo de solucionar tal problema, Wager & Athey (2018) proponen el modelo denominado *causal random forest*. Este consiste, a grandes rasgos, en tres pasos: (1) construir árboles de la misma manera que en el algoritmo tradicional de bosques aleatorios, (2) estimar el efecto causal del tratamiento en los nodos terminales de cada árbol (3) promediar las estimaciones a lo largo de las B muestras Bootstrap tomadas en el paso (1). Cabe destacar que (1) requiere tomar B muestras Bootstrap del conjunto de datos de entrenamiento para armar B árboles. Para realizar las particiones, se tiene en cuenta un conjunto aleatorio m de variables dentro de los p predictores disponibles.

El supuesto de identificación detrás de *causal random forest* es que las últimas “hojas” del árbol son lo suficientemente pequeñas como para que los pares (Y_i, T_i) actúen como si provinieran de un experimento aleatorio. En el caso particular de nuestro trabajo, cada nodo terminal estaría formado por observaciones de mujeres entre 15 y 19 años que son muy parecidas en sus observables (por la forma en la que se construyó el árbol). Solo se diferencian en un aspecto (central para la identificación): algunas de ellas pertenecen a la cohorte que tenía entre 15 y 19 años en 2012 (observaciones de la MICS 2012), mientras que otras pertenecen a la cohorte que tenía entre 15 y 19 años en 2020 (observaciones de la MICS 2020). Sin embargo, la similitud

de las adolescentes de una misma hoja permite suponer que el hecho de que algunas hayan sido expuestas al plan ENIA y otras no es completamente aleatorio. Entonces, en cada árbol, podemos estimar el efecto del tratamiento como:

$$\hat{\tau}(x) = \frac{1}{|\{i: W_i=1, X_i \in L\}|} \sum_{\{i: W_i=1, X_i \in L\}} Y_i - \frac{1}{|\{i: W_i=0, X_i \in L\}|} \sum_{\{i: W_i=0, X_i \in L\}} Y_i \quad (2)$$

donde L representa las hojas o nodos terminales del árbol. Como habíamos mencionado, luego promediamos los $\hat{\tau}_b(x)$ de manera que el efecto del tratamiento estimado por el bosque aleatorio causal resulta ser:

$$\hat{\tau}(x) = B^{-1} \sum_{b=1}^B \hat{\tau}_b(x)$$

Además, esta metodología requiere una condición fundamental: los árboles individuales deben satisfacer el supuesto de “honestidad”. Es decir, cada observación de entrenamiento i solo puede usar el *outcome* para estimar el efecto del tratamiento usando (2), o bien para realizar las particiones del árbol, pero no para ambas tareas. Para que esta condición sea satisfecha, nos basaremos en el algoritmo llamado “árbol de doble muestra” que, como sostienen Wager & Athey (2018), cumple con la condición de honestidad. Este algoritmo divide los datos de entrenamiento en dos partes: una (denominada J) es utilizada para realizar las particiones del árbol, mientras que la otra (denominada I) se reserva para realizar la estimación dentro de la hoja.

5. Conclusiones y limitaciones

La presente propuesta de investigación se centra en estudiar una de las problemáticas sociales centrales de la Argentina: el embarazo adolescente. En particular, buscamos entender cómo la maternidad precoz puede lograr reducirse a través de políticas públicas efectivas. Para ello, nos preguntamos si el Plan ENIA implementado a partir de 2017 en Argentina contribuyó a bajar la tasa de embarazo adolescente tardío.

En términos empíricos, el objetivo de este trabajo es estimar el efecto causal de la exposición al Plan ENIA de prevención del embarazo adolescente basándonos en técnicas de *machine learning*. En particular, proponemos estimar la efectividad de la política para las adolescentes mediante el modelo de *casual random forest* desarrollado por Wager & Athey (2018), que extiende el algoritmo tradicional de bosques aleatorios para poder identificar un coeficiente β de forma casual en los nodos terminales. Para llevar a cabo la estimación, proponemos utilizar las encuestas MICS realizadas en el país en dos ocasiones (2012 y 2020) ya que la política a estudiar se encuentra en el medio de las dos encuestas, lo que nos permite tener datos antes y después de la implementación del Plan ENIA. Además, las bases de datos correspondientes contienen la información necesaria sobre la maternidad adolescente y características relevantes de las mujeres que precisamos para las estimaciones.

Por último, cabe mencionar algunas limitaciones que deberían tenerse en cuenta al llevar a cabo la propuesta. En primer lugar, *casual random forest* requiere tomar B muestras bootstrap del conjunto de datos de entrenamiento para la construcción de los B árboles. Sin embargo, la cantidad de observaciones de las bases de datos propuestas puede ser limitada, generando un problema en la implementación del mecanismo de Bootstrap al tratarse de un método asintótico que requiere que el tamaño de la muestra sea lo suficientemente grande para que el Teorema Fundamental de la Estadística sea válido. En segundo lugar, la encuesta MICS de 2020 contiene algunas variables que no están en la encuesta realizada en 2012 (o viceversa). Si bien los principales predictores pueden derivarse de preguntas similares en ambas, debería prestarse especial atención a que exista homogeneidad en las variables que finalmente se utilizan para las estimaciones.

Bibliografía:

Azevedo, J., et al. (2012). Embarazo Adolescente y Oportunidades en América Latina y el Caribe. Sobre Maternidad Temprana, Pobreza y Logros Económicos. Washington: Banco Mundial.

Bennett, S. (2005). School-based teenage pregnancy prevention programs: A systematic review of randomized controlled trials. *Journal of Adolescent Health*, 36(1), pp.72-81.

Iglesias, M (10 de agosto de 2023). El embarazo en la adolescencia bajó el 60% en la última década: a qué se debe. *Clarín*. <https://www.clarin.com/10-98-23>

Ministerio de Educación, Cultura, Ciencia y Tecnología (2019a). Documento técnico N° 5: El embarazo y la maternidad en la adolescencia en la Argentina.

Ministerio de Educación, Cultura, Ciencia y Tecnología (2019b). Plan ENIA: Recorrido, logros y desafíos.

Molina, M (30 de julio de 2023). El ejemplo de Argentina en reducción de embarazos adolescentes es "innovador e inspirador". *Telam*. <https://www.telam.com.ar/30-07-23>

Roig, N. (2023). Adolescent Fertility and Reproductive Rights. Evidence From Argentina. . *Working paper University of Southern California*

Rofman, R. (2020). La reciente declinación de la fecundidad en Argentina. Una primera mirada a las tendencias, causas e impactos. Notas de Desarrollo Humano.

Secretaría Nacional de Niñez, Adolescencia y Familia (2022). Indicadores de niñez y adolescencia Fecundidad adolescente. <https://www.argentina.gob.ar/SENAF.pdf>

Tommasi, M., Edo, M., & Thailinger, A. (2021). Familia y desarrollo humano en el siglo XXI. Revisión de la literatura internacional y algunos apuntes para el caso argentino. Centro de Estudios para el Desarrollo Humano, Universidad de San Andrés.

Wager, S., & Athey, S. (2015). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113 (523), pp. 1228-1242.