



Big Data

Profesora: Romero, María Noelia

Tutora: Oubiña, Victoria

Propuesta Final

**Alumnos: Cucher Maximiliano, Saucedo Federico & Soares
Gache Manuel**

Fecha de Entrega: 05/07/2024

Introducción

En el presente trabajo nos proponemos responder dos preguntas fundamentales: ¿Cuáles son los factores que afectan la vulnerabilidad de los hombres en la Ciudad de Buenos Aires y cómo influyen en sus resultados a lo largo de la vida? y ¿En qué medida difieren estos factores y sus consecuencias en comparación con los de las mujeres? A lo largo de esta investigación, abordaremos dos temas específicos. En primer lugar, realizaremos un análisis similar al de Cucher, Rosino, Ruiz y Tomassi (2023) para identificar los factores que influyen en la vulnerabilidad de los hombres y cómo esta vulnerabilidad impacta en diversos resultados importantes a lo largo de la vida. Esta investigación es novedosa, ya que, a pesar de la abundante literatura que investiga los aspectos de vulnerabilidad de las mujeres, son escasos los trabajos que se centran en los hombres.

Nuestra motivación surge de la necesidad de llenar este vacío en la literatura existente. Mientras que numerosos estudios han examinado las vulnerabilidades y las brechas de género que afectan a las mujeres, poco se sabe sobre cómo estas dinámicas se manifiestan en la población masculina. Comprender estos aspectos es crucial para desarrollar políticas públicas más equitativas y efectivas que aborden las necesidades de ambos géneros.

En segundo lugar, compararemos nuestros resultados con los obtenidos por las autoras mencionadas para identificar posibles brechas de género en los factores que provocan vulnerabilidad y las consecuencias de dicha vulnerabilidad. De este modo, contribuiremos a la extensa literatura sobre la brecha de género, proporcionando una visión más completa y equilibrada de las desigualdades que afectan tanto a hombres como a mujeres.

Para llevar a cabo nuestro análisis, utilizaremos la Encuesta Demográfica Retrospectiva (EDER CABA 2019 - Buenos Aires), que recoge información sobre procesos sociodemográficos en la Ciudad de Buenos Aires durante la segunda mitad del siglo XX y el inicio del siglo XXI. Esta base de datos nos permitirá realizar un análisis comparativo

detallado de las trayectorias de vida de hombres y mujeres con al menos un hijo, identificando los factores clave que influyen en su vulnerabilidad.

Al investigar y comparar la vulnerabilidad de hombres y mujeres, esperamos proporcionar nuevas perspectivas y datos que puedan informar mejor las políticas de igualdad de género y contribuir a la creación de intervenciones más eficaces para reducir las desigualdades en diferentes ámbitos de la vida.

Base de Datos

La base de datos que se utilizará será la Encuesta Demográfica Retrospectiva ([EDER CABA 2019 - Buenos Aires](#)) la cual se realizó en el año 2019 la cual recolecta información sobre la naturaleza de procesos sociodemográficos de la población de la Ciudad de Buenos Aires durante la segunda mitad del siglo XX y el inicio del siglo XXI. La base de datos proporciona información de tres cohortes diferentes. La primera cohorte su nacimiento data de 1948-1952 (para el año en el que se realiza la encuesta sus edades serían entre 67 y 71 años). La segunda cohorte comprende individuos cuyo nacimiento se encuentra entre 1968-1972 (entre 47 y 51 años para 2019) y la última cohorte nacieron entre 1978-1982 (para el año en el que se realiza la encuesta, los individuos pertenecientes a esta cohorte tienen entre 37 y 41 años). En la muestra entre las 3 cohortes hay en total 1.978 padrés; sin embargo, nuestra población de interés serán aquellos padres con al menos un hijo dado que queremos realizar un análisis comparativo contra las trayectorias conyugales de las mujeres, las cuales son 490 observaciones.

Algo relevante a comentar es la manera en la que están contruidos los datos; la manera en la que se realiza es retrospectiva. Es decir, en el año 2019 los individuos reportan su situación en cada momento del tiempo en lo referente a sus hijos y cónyuges¹. Al igual que Cucher et al., al realizar la limpieza de la base de datos y quedarnos con nuestro grupo de interés,

¹ Incluso en aquellos casos en los que el cónyuge no era parte de la vida del individuo particular o su hijo todavía no había nacido.

transformaremos la base de datos a un formato *cross-section* y cada observación representará la trayectoria de la vida de un único individuo²

La información proporcionada por la EDER se puede sintetizar como se presenta en la Tabla 1. Estos datos describen, de manera similar al caso de las mujeres, los diferentes momentos de la vida de los hombres con al menos un hijo en CABA. En particular, permiten caracterizar su familia de origen, a ellos mismos y el entorno en el que nacieron y crecieron sus hijos en términos de educación, mercado laboral, vivienda, convivencia y salud. Como se puede observar en la Tabla 1, contamos con información sobre la educación de las mujeres de la muestra y la de sus padres. En cuanto a las características relacionadas con el mercado laboral, se dispone de datos sobre la cantidad y proporción de años que el hombre estuvo empleado y trabajando en el sector informal. Las proporciones se calcularon sobre el total de años que el hombre aparece en la encuesta (desde su nacimiento hasta 2019). Además, se puede determinar si los padres de las mujeres estaban empleados cuando ellas tenían 14 años y quién era el principal sostén económico del hogar en ese momento.

Además, sobre las características de las casas tenemos la cantidad de años durante su niñez (desde el nacimiento hasta los 12 años) que el hombre vivió en una vivienda deficitaria (definición según INDEC) y durante su adultez (toman de 21 y 40 años). En lo referente a las características a las características de la familia de origen con las que cohabitan la encuesta posee cantidad de años que vivió con la madre, con los padres y con parejas de los padres y madres hasta los 18 años. Con respecto al hijo, tenemos la cantidad de años que el hijo vivió con la madre.

En lo referente a la categoría de salud, tenemos los período de inicio y fin de uso de anticonceptivos para el caso de los hombres. El criterio que utilizan Cucher et al. para contabilizar esta proporción para las cohortes 2 y 3 realizaron cantidad de años de uso de

² El motivo por el cual realizamos esto lo destacaremos en la sección de Conclusiones y limitaciones.

métodos anticonceptivos desde su primer relación sexual hasta el momento de la encuesta mientras que para primer cohorte hasta el momento de la menopausia (51 años) para el caso de los hombres la denominación análoga es “andropausia”.

Por último, en la siguiente sección explicaremos la manera en la que se construyen con la encuesta EDER el *fragility and vulnerability index*.

	VARIABLES RELATED TO FAMILY OF ORIGIN	WOMEN OWN VARIABLES	CHILDREN VARIABLES
Education	Education father Education mother	Education	
Labor market	Was your father working when you were 14? Was your mother working when you were 14? Main breadwinner at 14	# years employed # years informal employment	
Housing	Poor housing during childhood	Poor housing during adulthood	
Cohabitation during childhood	# years lived with mother # years lived with father #years lived with stepparent		# years lived with mother
Cohort		Cohort: 1, 2, 3.	
Health practices		Use of contraceptives	
Conyugal and family trajectories		Y_1 : multipartner fertility Y_2 : (n)ever married Y_3 : many children Y_4 : teen birth	$Y_i \begin{cases} 0 & \text{if no} \\ 1 & \text{if yes} \end{cases}$
Index		Fragility Index: $\sum_{i=1}^4 Y_i$	
Early childhood environment		V_1 : Absent mother V_2 : Uneducated mother V_3 : Poor housing V_4 : Fragility Index	$V_i \in [0,1]$
		Vulnerability Index: $\sum_{i=1}^4 V_i$	

Tabla 1: Variables de la Encuesta Demográfica Retrospectiva. Fuente: Cucher, Rosino, Ruiz & Tommasi (2023).

Cuadro 1 Cantidad de personas a encuestar en cada zona del Marco de Viviendas particulares generales, según sexo y cohorte de pertenencia. Ciudad de Buenos Aires. Año 2019

Cohorte	Zona y sexo									
	Total	Norte			Centro			Sur		
		Total	Varón	Mujer	Total	Varón	Mujer	Total	Varón	Mujer
Total	3803	855	441	414	1786	926	860	1162	611	551
1948-1952	1132	272	131	141	513	249	264	347	183	164
1968-1972	1266	278	158	120	596	302	294	392	194	198
1978-1982	1405	305	152	153	677	375	302	423	234	189

Tabla 2: Cantidad de personas a encuestar. Fuente: [EDER CABA 2019 - Buenos Aires](#)

Metodología

En esta línea, realizamos similar a Cucher et al. Un análisis de las trayectorias conyugales, pero con respecto a los hombres en CABA. Llevaremos a cabo un análisis de cluster, la cual es una técnica de aprendizaje no supervisado para agrupar observaciones

homogéneas/similares entre sí y heterogéneos entre diferentes cluster los cuales son disjuntos. Es decir, homogéneos entre cluster y heterogéneos fuera del cluster. Los cluster están determinados a partir de cuatro variables dummies. En primer lugar, *teen birth* la cual vale 1 en caso de que la pareja del hombre tuvo un embarazo a los 19 años o antes y cero en caso contrario; *many children's* que vale 1 si el hombre tuvo 4 hijos o más y cero en caso contrario. *Never married* la cual toca el valor uno si el hombre nunca se casó y 0 en caso contrario y *multipartner fertility* que vale 1 si el hombre tuvo hijos con más de una pareja y 0 en caso contrario.

Siguiendo la metodología de los autores, utilizan las *never married* y *multipartner fertility* para analizar el margen extensivo en el sentido de distanciar a los hombres que tuvieron hijos con una pareja de aquellas que tuvieron con muchas. En esta línea, también consideraban relevante diferenciar de las que no se casaron nunca de las que se casaron una vez.

Realizando más énfasis en la metodologías, dado que usamos variables *dummies* utilizaremos el algoritmo de cluster *k-means* (Huang, 1997) el cual formalmente se define de la siguiente manera:

$$d(X, Y) = \sum_{j=1}^m \delta(x_j, y_j) \quad (1)$$

donde $\delta(x_j, y_j) = 1[x_j \neq y_j]$. A menor distancia significa que las observaciones presentan menos discrepancias entre sí. Para construir los clusters, el algoritmo K-modas minimiza la siguiente función de costos:

$$C = \sum_{w=1}^k \sum_{i=1}^n y_{i,w} d(X_i, Q_w) \quad (2)$$

donde k indica el número de cluster, n refiere a las observaciones, $y_{i,w}$ es un elemento de la matriz de partición $Y_{n \times k}$, X_i es un vector de objetos categóricos del individuo i , Q_w es el vector nombrado en (1). Es decir, el hombre será asignado al cluster w si se encuentra a menor distancia en las cuatro variables.

Dado que este algoritmo se especifica previamente la cantidad de cluster las autoras llegan a que la la cantidad óptima de cluster son 8 o 16 (lo realizan mediante *cross-validation*) y optar por realizar 8 cluster dado que 16 hace que la cantidad de observaciones por cluster disminuya mucho.

Intentaremos que los clusters sean lo más similar al de las autoras posibles para lograr compararlos posteriormente. Esta parte es clave para el análisis, ya que si encontramos que los clusters son muy diferentes, no lograremos compararlos con los clusters de las mujeres. También realizaremos un índice de fragilidad construido a partir de la sumatoria de las cuatro

variables con las que formamos los clusters, familias, ningún casamiento, hijos, y embarazo adolescente.

Por otro lado, intentaremos ampliar el análisis de las autoras. Con el objetivo de entender qué variables son las más importantes a la hora de provocar ciertos eventos como un embarazo adolescente, o que una persona tenga hijos con muchas personas, realizaremos distintos *random trees* utilizando la estrategia de *bagging*. De esta manera podremos saber qué variables son las que tienen una mayor importancia en predecir/provocar esos sucesos. Esto es primordial en un escenario como el de Argentina en donde la política pública tiene recursos finitos, ya que podemos concentrar nuestros esfuerzos en políticas públicas ya sea para contener a las mujeres que cumplan esas variables, o prevenir que esas variables sucedan.

Siguiendo con nuestro enfoque de comparar las brechas entre hombres y mujeres, realizaremos este mismo proceso para los hombres, para identificar si la importancia de las variables para predecir los mismos sucesos son similares.

Literatura previa

La literatura de brechas de géneros ha sido un tema relevante y ha sido foco de atención de muchos autores relevantes. Uno de los epicentros de este estudio es el análisis de las brechas de género en el mercado laboral. Goldin (2006) afirma que la participación de las mujeres en el mercado laboral ha aumentado significativamente en las últimas décadas, un fenómeno que ella describe como la “revolución silenciosa”. Esta evolución refleja una mayor participación laboral femenina, y cambios en la toma de decisiones respecto al trabajo y la identidad profesional de las mujeres. Sin embargo, persisten grandes brechas en términos de participación laboral, horas trabajadas, tipos de ocupaciones y acceso a puestos jerárquicos. Bertrand (2020) destaca que aunque la participación laboral femenina ha crecido, las mujeres siguen subrepresentadas en puestos jerárquicos y tienden a trabajar en sectores menos remunerados.

En el año 2015, en América Latina³ hubo una gran brecha en la cantidad de horas trabajadas y la participación laboral entre hombres y mujeres. Estas brechas disminuyen a medida de que aumenta el nivel educativo y parecieran desaparecer para las personas más educadas. En Argentina en 2017 uno de cada cuatro hombres trabaja a tiempo parcial, mientras que una proporción mucho más alta de mujeres lo hacen, una de cada dos (Marchionni et al, 2018).

³ DGEMyEL - MTEySS, en base a datos de la EPH-INDEC.

Además, hay una gran brecha en las ocupaciones laborales de las mujeres y los varones. En Argentina (2017) las mujeres suelen estar centradas en servicios y varios, mientras que los hombres están centrados en actividades de comercialización, extractivas y de construcción e industrial⁴. A la vez hay una correlación negativa entre empleos de mayor participación femenina y salarios, es decir, las mujeres participan en empleos que ofrecen en promedio menores salarios⁵.

Pareciera también haber un “techo de cristal” que impide que las mujeres lleguen a puestos jerárquicos. En Argentina, mientras que la participación de mujeres en puestos altamente calificados (profesionales y técnicos) es levemente superiores a la de los hombres (6%), en los empleos de mayor jerarquía muestran grandes brechas (Marchionni et al, 2018). A la vez, hay una proporción muy baja de mujeres en puestos de dirección del Estado⁶ y en cargos sindicales⁷.

Pareciera también haber “pisos pegajosos” siendo en Argentina en 2018⁸ las mujeres sobrerrepresentadas en empleos de baja calidad, en especial sobre las mujeres poco educadas. En promedio, las mujeres reciben salarios 17% menores a los hombres con misma educación, edad, estado civil, número de hijos y área de residencia. En la Argentina, esa brecha es del 9% (Marchionni et al, 2018).

Está claro que hay una brecha en el mercado laboral, pero habría que analizar qué es lo que provoca esta diferencia. Pueden ser brechas en educación, por la crianza y la naturaleza, brechas en la distribución de roles en el hogar por el efecto de la maternidad, o simplemente por preferencias que difieren a la de los hombres (Marchionni et al, 2018).

En cuanto a la brecha en educación se puede comentar lo siguiente. Pareciera que la cantidad de años de educación no puede explicar la brecha en el mercado laboral. Las mujeres tienen tasas de participación más grandes en educación secundaria y terciaria (Marchionni et al, 2018). En cuanto a los aprendizajes, en las pruebas PISA que se realizan a niños de 15 años, se encontró que las mujeres tenían ventaja en lengua, mientras que los hombres en ciencias y matemáticas (Marchionni et al, 2018). Muchos estudios muestran sesgos en maestros y padres respecto a las creencias sobre las capacidades de las niñas y los niños.

⁴ DGEMyEL - MTEySS, en base a datos de la EPH-INDEC.

⁵ CIPPEC en base a INDEC (EPH).

⁶ GPS del Estado-CIPPEC.

⁷ CIPPEC en base a MTEySS (2017).

⁸ CIPPEC en base a INDEC (EPH).

Del mismo modo, también hay brecha de género en la elección de las carreras. Por ejemplo, el 25% de los estudiantes de Ingeniería y Ciencias Aplicadas son mujeres, mientras que en carreras como idiomas, ciencias de la salud y ciencias sociales, más del 70% de los estudiantes son mujeres. Por lo tanto, con lo que respecta a la educación, pareciera que solo la diferencia en aprendizaje y en elección de carreras pueden explicar las brechas en el mercado laboral.

Otro aspecto a tener en cuenta es la naturaleza de las mujeres. La evidencia muestra que las mujeres son más adversas al riesgo, obtienen peores resultados en ambientes competitivos y los evitan, negocian y piden menos salarios o condiciones, y tienen menos confianza en sí mismas (Rozada & Yeyati (2018); Dohmen et al, (2011); Gneezy et al (2003); Niederle & Vesterlund, (2007).

Si los resultados encontrados reflejan verdaderas preferencias de las mujeres no deberían ser considerados subóptimos (Bertrand, 2020). Se encuentra que las elecciones de las mujeres son moldeadas por fuertes estereotipos de género, o bien porque son internalizados, o bien porque es costoso desviarse de la conducta aceptada socialmente. Fernández (2007) encuentra que la cultura juega un gran papel también en las elecciones de las mujeres.

Por otro lado, las brechas en el hogar son también una gran explicación de las brechas de género. Marchioni, Gasparini y Edo (2018) encuentran que las mujeres dedican más tiempo a quehaceres domésticos y de cuidado.

El efecto de la maternidad pareciera ser una explicación fundamental en la brecha de género. Autores tales como Kleven et al. (2018), Rosenzweig et al (1980) y Kleven et al (2019) muestran mediante evidencia empírica que las mujeres madres participan menos del mercado laboral que los hombres y las mujeres sin hijos, cuando participan, participan menos horas, suelen optar por trabajos más flexibles, y lo hacen en puestos menos jerárquicos.

Conclusiones y limitaciones

En conclusión, esperamos encontrar que los factores que afectan la vulnerabilidad de los hombres en la Ciudad de Buenos Aires incluyen variables relacionadas con la educación, el mercado laboral, la vivienda, la convivencia familiar y la salud. Anticipamos que estas vulnerabilidades tendrán consecuencias significativas en diversos aspectos de la vida de los hombres, como la estabilidad laboral, la salud mental y física, y las relaciones familiares.

En cuanto a la comparación con las mujeres, prevemos identificar diferencias notables en los factores que provocan vulnerabilidad y en las consecuencias de dicha vulnerabilidad. Esperamos que, aunque algunos factores sean comunes a ambos géneros, las dinámicas y magnitudes de sus efectos varíen significativamente. Por ejemplo, es probable que las responsabilidades domésticas y el cuidado de los hijos tengan un impacto más profundo en la vulnerabilidad de las mujeres, mientras que factores relacionados con el empleo y la estabilidad económica puedan ser más determinantes para los hombres.

Con estos hallazgos, buscamos contribuir a una comprensión más completa de la brecha de género en términos de vulnerabilidad y sus efectos a lo largo de la vida, proporcionando datos que puedan informar políticas públicas más equitativas y efectivas.

En cuanto a las limitaciones, una de las principales limitaciones que podríamos enfrentar es el sesgo de memoria, dado que la Encuesta Demográfica Retrospectiva (EDER) se basa en la memoria de los participantes para reportar eventos pasados, lo que puede introducir errores en los datos. Los individuos pueden olvidar detalles importantes o recordar eventos de manera inexacta, afectando la precisión de los datos recopilados.

Además, aunque la EDER proporciona una muestra significativa, los resultados pueden no ser completamente representativos de toda la población masculina de la Ciudad de Buenos Aires, especialmente si hay subgrupos importantes que están subrepresentados en la muestra. Esto podría limitar la generalización de nuestros hallazgos a otros contextos.

Al comparar los resultados de hombres y mujeres, es posible que enfrentemos dificultades para asegurar que los grupos sean completamente comparables en todos los aspectos relevantes, lo que podría afectar la validez de nuestras conclusiones sobre las diferencias de género. Asimismo, las variables disponibles en la EDER pueden no capturar todos los aspectos relevantes de la vulnerabilidad, especialmente aquellos relacionados con factores sociales y culturales más sutiles.

Asimismo, el hecho de que los padres respondan por cuestiones vinculadas a los hijos hacen que esas respuestas estén sesgadas por lo que nuevamente el *first best* sería que para las preguntas relacionadas a los hijos respondan ellos mismos. Por último, los resultados obtenidos de esta muestra específica de Buenos Aires pueden no ser completamente aplicables a otras regiones o contextos culturales, limitando la generalización de nuestras conclusiones.

A pesar de estas limitaciones, creemos que nuestra investigación proporcionará *insights* valiosos sobre la vulnerabilidad de los hombres y sus diferencias con respecto a las mujeres, contribuyendo de manera significativa a la literatura existente y ofreciendo información útil para la formulación de políticas públicas.

Bibliografía

Bertrand, M. (2020), Gender in the 21st century, American Economic Association Richard T. Ely Lecture. AEA Papers and Proceedings.

Cucher, S., Rosino, V., Ruiz F.m., & Tommasi M. (2023). Trayectorias Conyugales, estructuras familiares y vulnerabilidad social: Una Mirada Tres Generaciones De Mujeres la Ciudad de Buenos Aires. Working Paper.

Fernández, R. (2007). Women, work, and culture. Journal of the European Economic Association, 5(2-3), 305-332.

Gneezy, U., Haruvy, E., & Roth, A. E. (2003). Bargaining under a deadline: Evidence from the reverse ultimatum game. Games and Economic Behavior, 45(2), 347-368.

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. American economic review, 96(2), 1-21.

González Rozada y Levy Yeyati (2018), Do women ask for lower salaries? The supply side of the gender wage gap. Documento de Trabajo Universidad Torcuato Di Tella, 2018/4.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2018). Child penalties across countries: Evidence and explanations. In AEA Papers and Proceedings (Vol. 109, pp. 122-126). 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association.

Kleven, H., Landais, C., & Søgaaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4), 181-209.

Marchionni, M., Gasparini, L., & Edo, M. (2018). Brechas de Género En América Latina (CAF, Ed.). *Corporación Andina de Fomento*.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much?. *The quarterly journal of economics*, 122(3), 1067-1101.

Rosenzweig, M. R., & Wolpin, K. I. (1984). Heterogeneity, intrafamily distribution and child health (No. 463). Center Discussion Paper.

Rozada, M. G., & Yeyati, E. L. (2018). Do women ask for lower salaries? The supply side of the gender pay gap (No. 201804). Universidad Torcuato Di Tella.

DGEMyEL - MTEySS, en base a datos de la EPH-INDEC.

CIPPEC en base a INDEC (EPH). [CIPPEC](#)

GPS del Estado-CIPPEC.

CIPPEC en base a MTEySS (2017).