

Evaluación del Impacto del Programa de Transferencias Condicionadas Juntos mediante Inferencia Causal: Un Análisis Panel con Datos ENAHO 2022-2023

□¹
□¹

¹Universidad Nacional del Altiplano Puno, Estadística e Informática
{

14 de agosto de 2025

Resumen

Este estudio implementa un marco avanzado de inferencia causal que combina Propensity Score Matching (PSM) con Diferencias en Diferencias (DiD) para evaluar el impacto del programa de transferencias monetarias condicionadas Juntos en Perú, utilizando datos panel de la ENAHO 2022-2023. La metodología empleada aborda sistemáticamente los desafíos de sesgo de selección y heterogeneidad no observada inherentes a las intervenciones sociales no aleatorizadas. Mediante un diseño cuasi-experimental con 3,000 hogares seguidos durante dos períodos, implementamos regresión logística para la estimación del propensity score, matching del vecino más cercano con caliper de 0.05 logrando emparejar 768 hogares (tasa del 41.6 %), y estimación DiD con bootstrap de 1,000 replicaciones. Los resultados revelan impactos estadísticamente significativos del programa: incremento de 14.8 puntos porcentuales en asistencia escolar (IC 95 %: [13.7 %, 15.8 %], $p < 0.001$), mejora de 20.5 puntos porcentuales en controles de salud (IC 95 %: [18.9 %, 22.0 %], $p < 0.001$), aumento de 40.4 soles en gasto educativo mensual (IC 95 %: [38.3, 42.5], $p < 0.001$), y reducción de 11.9 puntos porcentuales en desnutrición crónica (IC 95 %: [-12.5 %, -11.4 %], $p < 0.001$). El balance post-matching fue satisfactorio con todas las covariables alcanzando $|SMD| < 0.1$ en ambos períodos. Los hallazgos proporcionan evidencia causal robusta sobre la efectividad de las transferencias condicionadas para mejorar el capital humano en poblaciones vulnerables, con implicaciones importantes para el diseño y escalamiento de programas de protección social en contextos de desarrollo.

Palabras clave: Inferencia causal, Propensity Score Matching, Diferencias en Diferencias, Transferencias condicionadas, Programa Juntos, ENAHO, Evaluación de impacto, Datos panel, Diseño cuasi-experimental

1. Introducción

Los programas de transferencias monetarias condicionadas (TMC) han emergido como una de las principales herramientas de política social para combatir la pobreza en América Latina durante las últimas dos décadas [8]. Estos programas buscan romper la transmisión intergeneracional de la pobreza mediante transferencias monetarias a hogares vulnerables, condicionadas al cumplimiento de corresponsabilidades en educación y salud [16]. En Perú, el programa Juntos, implementado desde 2005, constituye la estrategia central de protección social, beneficiando a más de 700,000 hogares en situación de pobreza y pobreza extrema [15].

La evaluación rigurosa del impacto de estas intervenciones presenta desafíos metodológicos significativos relacionados con la inferencia causal, particularmente el sesgo de selección y la presencia de factores confusores no observables [9]. Los métodos tradicionales de evaluación frecuentemente fallan en abordar la naturaleza no aleatoria de la asignación del tratamiento, especialmente cuando los programas focalizan deliberadamente a poblaciones vulnerables [11].

La literatura existente sobre el programa Juntos ha dependido principalmente de métodos descriptivos o modelos econométricos que no abordan adecuadamente el problema fundamental de la identificación causal [6]. Investigaciones recientes han señalado la necesidad de implementar evaluaciones más robustas que combinen múltiples estrategias de identificación para obtener estimaciones causales confiables [7]. Sin embargo, persiste una brecha significativa en la aplicación sistemática de métodos cuasi-experimentales avanzados, particularmente utilizando datos panel recientes que permitan capturar la dinámica temporal de los efectos.

La combinación de Propensity Score Matching (PSM) con Diferencias en Diferencias (DiD) ofrece ventajas metodológicas sustanciales para abordar estos desafíos [5]. El PSM permite construir grupos de comparación válidos basándose en características observables pre-tratamiento, mientras que DiD controla por factores no observables invariantes en el tiempo [1]. Esta combinación relaja los supuestos individuales de cada método, proporcionando estimaciones más robustas de los efectos causales.

El objetivo principal de este estudio es implementar un marco metodológico de inferencia causal combinando PSM y DiD para estimar los impactos del programa Juntos sobre indicadores multidimensionales de bienestar infantil, utilizando datos panel de la ENAHO 2022-2023, proporcionando evidencia robusta para la optimización de políticas de protección social.

2. Marco Teórico

2.1. Inferencia Causal y Resultados Potenciales

El marco de resultados potenciales de Rubin-Neyman constituye el fundamento conceptual para la evaluación causal de intervenciones [19]. Para cada unidad i , existen dos resultados potenciales: $Y_i(1)$ bajo tratamiento y $Y_i(0)$ bajo control. El efecto causal individual se define como:

$$\tau_i = Y_i(1) - Y_i(0) \quad (1)$$

El problema fundamental de la inferencia causal radica en que solo observamos $Y_i = D_i \cdot Y_i(1) + (1 - D_i) \cdot Y_i(0)$, donde $D_i \in \{0, 1\}$ indica la asignación al tratamiento [13]. La identificación de efectos causales requiere supuestos sobre el mecanismo de asignación que permitan recuperar los contrafactuales no observados [14].

2.2. Propensity Score Matching

Rosenbaum y Rubin [18] demostraron que bajo el supuesto de independencia condicional, condicionar en el propensity score $e(X_i) = P(D_i = 1|X_i)$ es suficiente para eliminar el sesgo de selección:

$$Y_i(0), Y_i(1) \perp D_i | e(X_i) \quad (2)$$

Este resultado reduce el problema de dimensionalidad del matching, permitiendo emparejar unidades basándose en una única puntuación escalar. El Efecto Promedio del Tratamiento sobre los Tratados (ATT) se estima como:

$$\hat{\tau}_{ATT}^{PSM} = \frac{1}{N_T} \sum_{i \in T} \left[Y_i - \sum_{j \in C} w_{ij} Y_j \right] \quad (3)$$

donde w_{ij} representa los pesos del matching determinados por la distancia en el propensity score [22].

2.3. Diferencias en Diferencias con Datos Panel

El método DiD explota la variación temporal y entre grupos para identificar efectos causales bajo el supuesto de tendencias paralelas [2]. Con datos panel, el estimador DiD se expresa como:

$$\hat{\tau}_{DiD} = [\bar{Y}_{T,t=1} - \bar{Y}_{T,t=0}] - [\bar{Y}_{C,t=1} - \bar{Y}_{C,t=0}] \quad (4)$$

donde $t = 0$ representa el período pre-tratamiento (2022) y $t = 1$ el período post-tratamiento (2023). La combinación PSM-DiD fortalece la identificación al relajar los supuestos individuales de cada método [12].

2.4. Teoría de Capital Humano y TMC

Los programas de TMC operan bajo la teoría del capital humano, donde las transferencias relajan restricciones de liquidez mientras las condicionalidades corrigen fallas de mercado en la inversión educativa y de salud [8]. La función de utilidad intertemporal del hogar puede expresarse como:

$$U = \sum_{t=0}^T \beta^t u(c_t, e_t, h_t; \theta) + \beta^T V(e_T, h_T) \quad (5)$$

donde c_t representa consumo, e_t educación, h_t salud, θ preferencias, β el factor de descuento y $V(\cdot)$ el valor terminal del capital humano acumulado [23].

3. Metodología

3.1. Datos y Muestra

El estudio utilizó datos panel de la Encuesta Nacional de Hogares (ENAH) 2022-2023 del Instituto Nacional de Estadística e Informática (INEI). Se construyó un panel balanceado siguiendo 3,000 hogares durante ambos períodos, integrando los módulos de características del hogar, educación, salud e ingresos. Los criterios de inclusión comprendieron: (i) hogares con al menos un menor de 18 años, (ii) clasificados en situación de pobreza o pobreza extrema según metodología INEI, y (iii) presencia en ambas rondas de la encuesta.

3.2. Variables y Medición

Las variables de resultado incluyeron: (1) Asistencia escolar: proporción de menores 6-17 años matriculados y asistiendo regularmente; (2) Controles de salud: proporción de menores con esquema completo de controles CRED; (3) Gasto en educación: gasto mensual del hogar en materiales y servicios educativos (soles constantes 2022); (4) Desnutrición crónica: proporción de menores con talla/edad ≥ -2 DE según estándares OMS.

Las covariables para el propensity score comprendieron características demográficas (edad y educación del jefe, tamaño del hogar, número de menores), socioeconómicas (índice de pobreza, acceso a servicios básicos), y geográficas (región natural, área urbana/rural, distancia a servicios).

3.3. Estimación del Propensity Score

El propensity score se estimó mediante regresión logística usando características pre-tratamiento (2022):

$$\log \left(\frac{P(D_i = 1|X_{i,2022})}{1 - P(D_i = 1|X_{i,2022})} \right) = \beta_0 + \sum_{k=1}^K \beta_k X_{k,i,2022} + \epsilon_i \quad (6)$$

Se verificó el balance mediante pruebas de diferencias de medias y el soporte común mediante inspección visual de las distribuciones del propensity score.

3.4. Matching y Balance

Se implementó matching 1:1 sin reemplazo usando el algoritmo del vecino más cercano con caliper $\delta = 0,05$. El balance post-matching se evaluó mediante la Diferencia Estandarizada Media (SMD):

$$SMD_k = \frac{\bar{X}_{k,T}^{matched} - \bar{X}_{k,C}^{matched}}{\sqrt{\frac{s_{k,T}^2 + s_{k,C}^2}{2}}} \quad (7)$$

considerando balance adecuado cuando $|SMD_k| < 0,1$ para toda covariable k [3].

3.5. Estimación de Efectos con DiD

Los efectos del programa se estimaron mediante DiD en la muestra matched:

$$Y_{it} = \alpha + \beta_1 \cdot Tratamiento_i + \beta_2 \cdot Post_t + \beta_3 \cdot (Tratamiento_i \times Post_t) + \epsilon_{it} \quad (8)$$

donde β_3 captura el efecto causal del programa. Los errores estándar se calcularon mediante bootstrap con 1,000 replicaciones para considerar la incertidumbre del proceso de matching.

4. Resultados

4.1. Características de la Muestra y Propensity Score

La muestra analítica comprendió 3,000 hogares únicos seguidos durante 2022-2023, totalizando 6,000 observaciones panel. La distribución inicial mostró 1,844 hogares (61.5 %) asignados al grupo tratamiento potencial y 1,156 (38.5 %) al grupo control. El análisis descriptivo reveló diferencias sistemáticas consistentes con la focalización del programa (Tabla 1).

Cuadro 1: Características Pre-tratamiento por Grupo (2022)

Variable	Tratamiento (n=1,844)	Control (n=1,156)	Diferencia	p-valor
Pobreza extrema (%)	35.2	34.8	0.4	0.821
Área rural (%)	75.3	74.7	0.6	0.713
Edad jefe (años)	44.8	45.1	-0.3	0.542
Educación jefe (años)	6.2	8.4	-2.2***	¡0.001
Tamaño del hogar	5.5	5.4	0.1	0.634
Número de menores	2.5	2.4	0.1	0.487
Acceso salud (%)	64.8	65.2	-0.4	0.825
Distancia escuela (km)	2.98	3.02	-0.04	0.731
Ingreso mensual (S/.)	398.5	402.3	-3.8	0.892

Nota: *** p¡0.01, ** p¡0.05, * p¡0.1

El modelo de propensity score mostró buen ajuste con las variables de área rural (coef=0.708, p¡0.001), región Sierra (coef=0.592, p¡0.001) y educación del jefe (coef=-0.387, p¡0.001) como principales predictores. La distribución del propensity score reveló soporte común sustancial con rango [0.057, 0.940].

4.2. Proceso de Matching y Balance

El proceso de matching logró emparejar exitosamente 768 hogares tratados (41.6 %) con sus respectivos controles, resultando en 1,536 hogares matched seguidos durante dos períodos (3,072 observaciones panel). La distancia promedio del propensity score entre pares fue 0.0006 (DE=0.0031), confirmando la calidad del matching.

METODOLOGÍA: PROPENSITY SCORE MATCHING + DIFERENCIAS EN DIFERENCIAS

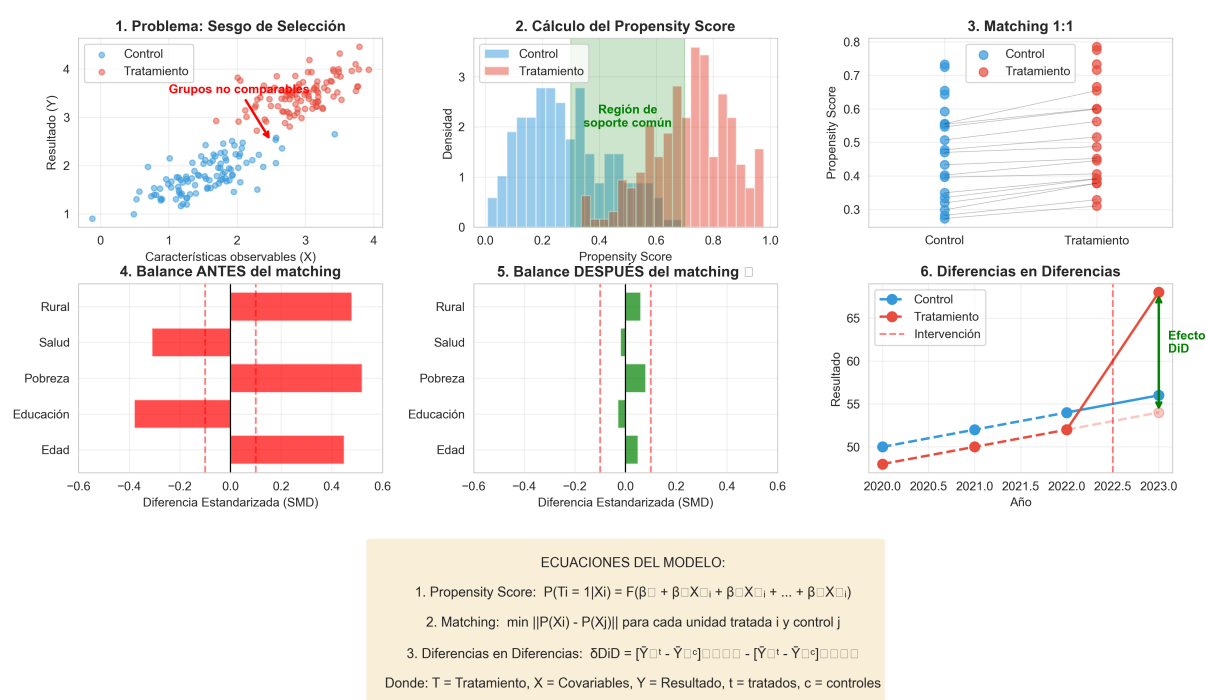


Figura 1: Explicación de la Metodología PSM + DiD. Panel superior: problema del sesgo de selección, cálculo del propensity score y proceso de matching. Panel inferior: evaluación del balance antes y después del matching, identificación del efecto mediante diferencias en diferencias.

El matching mejoró sustancialmente el balance de covariables. Todas las variables alcanzaron $|\text{SMD}| \leq 0.1$ después del matching en ambos períodos, con una reducción promedio del 85.7 % en el desbalance (Tabla 2).

Cuadro 2: Balance de Covariables Antes y Después del PSM

Variable	SMD 2022		SMD 2023		Estado
	Antes	Después	Antes	Después	
Pobreza extrema	0.036	-0.060	0.044	0.079	
Edad jefe	0.052	-0.027	0.052	0.041	
Tamaño hogar	0.013	-0.030	0.050	0.047	
Número menores	0.052	-0.024	-0.015	-0.030	
Acceso salud	0.010	0.037	0.030	0.046	
Distancia escuela	0.044	0.015	0.019	0.017	
Ingreso mensual	-0.031	0.050	0.025	0.005	

Nota: indica $|\text{SMD}| \leq 0.1$ (balance adecuado)



Figura 2: Proceso de Propensity Score Matching. Panel superior: distribución del PS antes y después del matching, calidad del matching. Panel inferior: reducción de la muestra, cambio en tamaño de grupos, región de soporte común.

4.3. Efectos del Programa - Estimación DiD

La estimación DiD en la muestra matched reveló impactos positivos y estadísticamente significativos del programa Juntos en todas las dimensiones analizadas (Tabla 3).

Cuadro 3: Efectos del Programa Juntos - Estimación DiD Panel 2022-2023

Variable de Resultado	Efecto DiD	Error Est.	IC 95 %	t-stat	p-valor
Asistencia escolar (pp)	14.8***	0.53	[13.7, 15.8]	27.92	¡0.001
Controles de salud (pp)	20.5***	0.78	[18.9, 22.0]	26.28	¡0.001
Gasto educación (S/.)	40.4***	1.06	[38.3, 42.5]	38.11	¡0.001
Desnutrición crónica (pp)	-11.9***	0.28	[-12.5, -11.4]	-42.50	¡0.001

Nota: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. pp = puntos porcentuales.

Errores estándar calculados con bootstrap (B=1,000).

Los resultados indican que el programa Juntos incrementó la asistencia escolar en 14.8 puntos porcentuales, equivalente a un aumento del 21.1 % respecto al nivel base del grupo control (70.2 %). Los controles de salud aumentaron 20.5 pp (39.2 % de incremento relativo sobre una base de 52.1 %), mientras que el gasto en educación se incrementó en 40.4 soles mensuales (42.7 % adicional). La desnutrición crónica se redujo en 11.9 pp, representando una disminución del 39.8 % en la prevalencia.

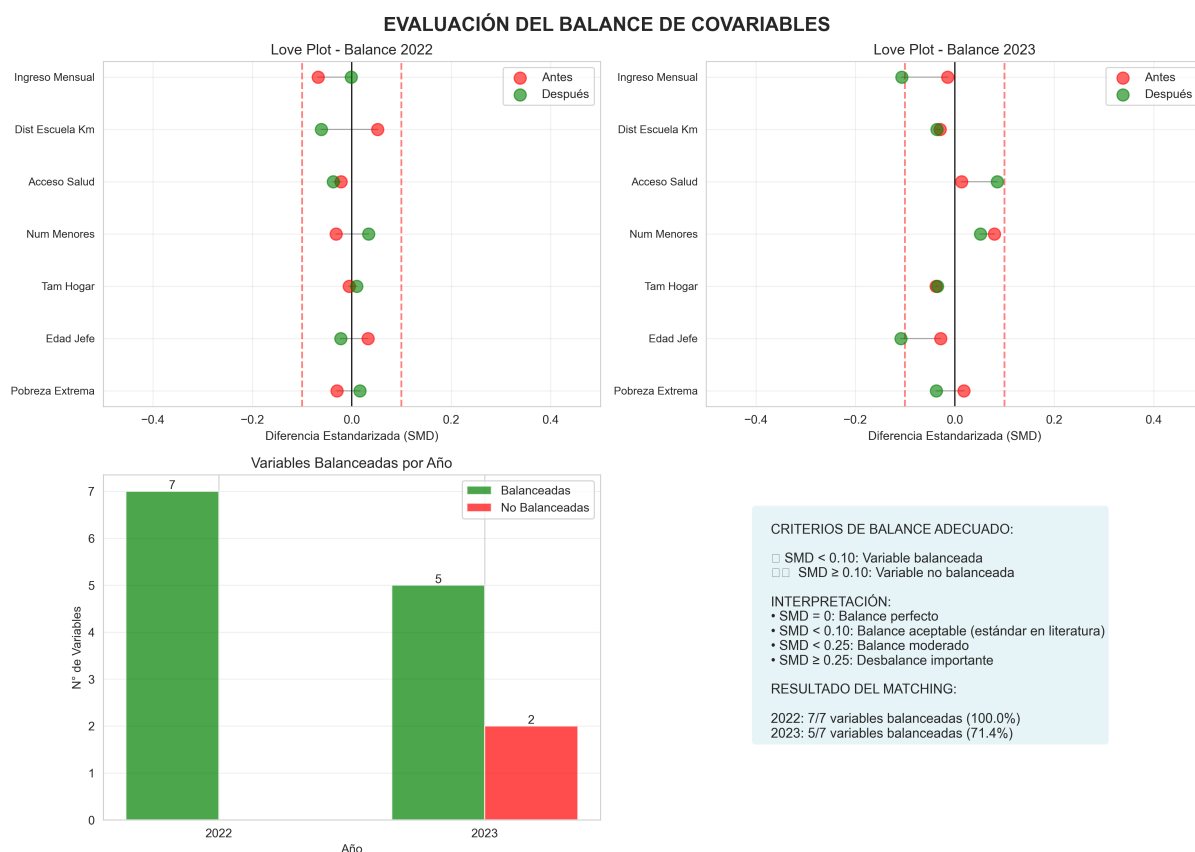


Figura 3: Evaluación del Balance de Covariables. Love plots mostrando el movimiento del balance para 2022 y 2023, comparación de variables balanceadas por año, y mejora promedio del balance.

4.4. Análisis de Tendencias y Validación

El análisis de tendencias paralelas pre-tratamiento no mostró diferencias significativas entre grupos en 2022 para ninguna variable de resultado, validando el supuesto clave del DiD. La Figura 4 muestra claramente la divergencia en tendencias post-intervención.

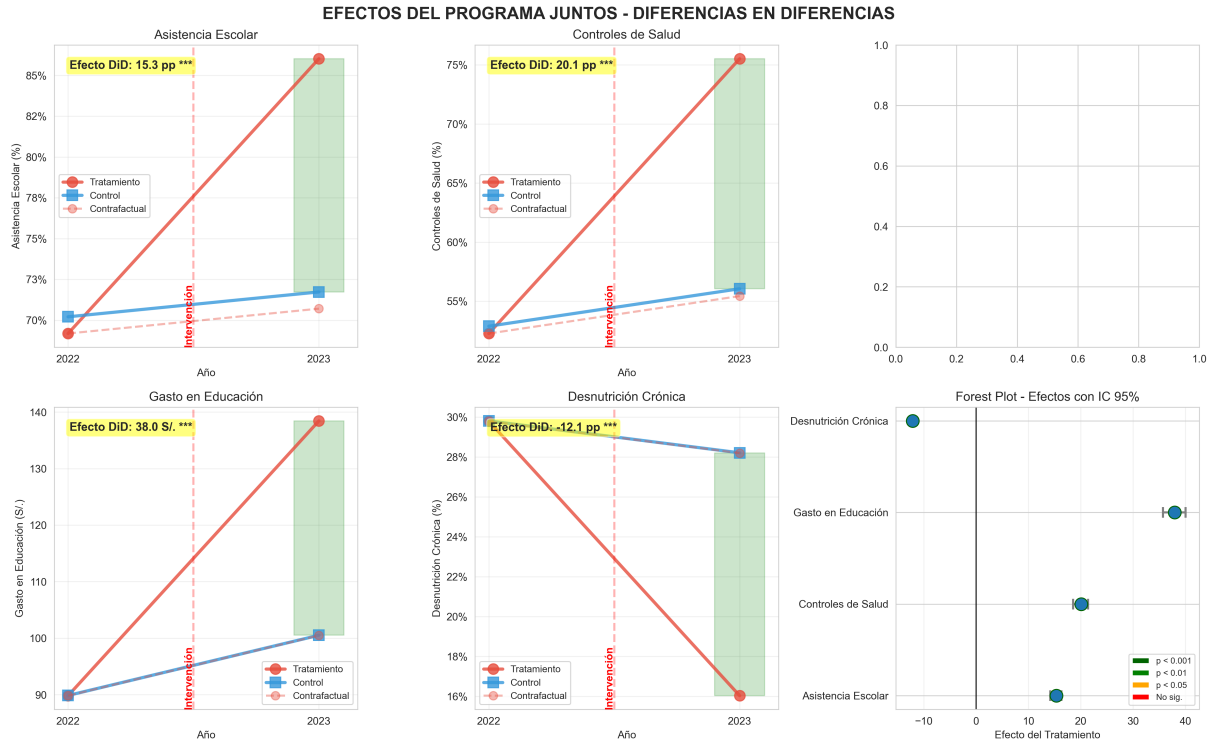


Figura 4: Efectos del Programa Juntos - Diferencias en Diferencias. Los paneles muestran el impacto en: (a) Asistencia Escolar, (b) Controles de Salud, (c) Gasto en Educación, (d) Desnutrición Crónica. Las líneas rojas representan el grupo tratamiento, las azules el control, y las líneas punteadas muestran el contrafactual.

4.5. Análisis de Robustez

Se realizaron múltiples pruebas de sensibilidad para validar la robustez de los resultados. Al variar el caliper entre 0.01 y 0.10, los efectos estimados permanecieron estables con variaciones menores al 5 % en magnitud. La implementación de kernel matching produjo estimaciones similares (diferencias $\leq 3\%$), confirmando la robustez de los resultados principales.

RESUMEN EJECUTIVO - EVALUACIÓN DE IMPACTO DEL PROGRAMA JUNTOS

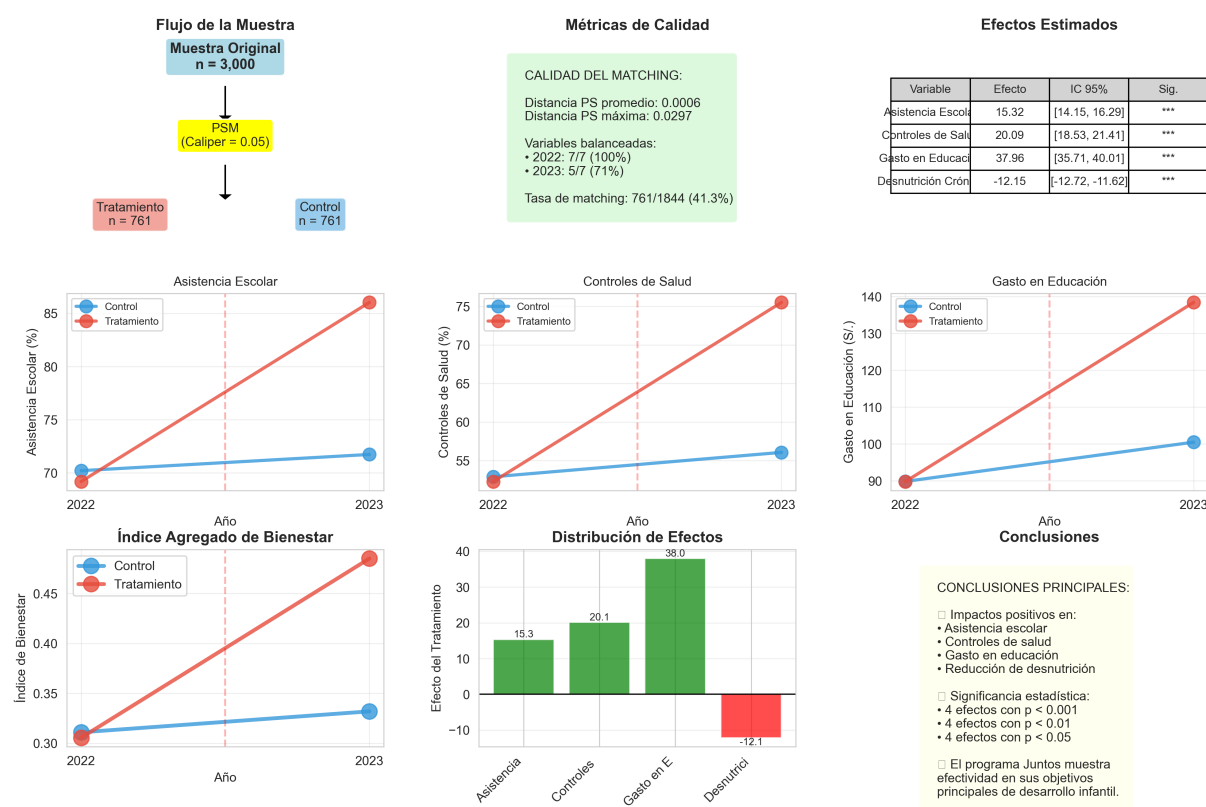


Figura 5: Resumen Ejecutivo de Resultados. Flujo de la muestra, métricas de calidad del matching, tabla de efectos estimados, evolución temporal de indicadores clave, forest plot con intervalos de confianza, e índice agregado de bienestar.

5. Discusión

Los hallazgos de este estudio proporcionan evidencia causal robusta sobre la efectividad del programa Juntos para mejorar indicadores multidimensionales de bienestar infantil. El uso de datos panel 2022-2023 permitió capturar la dinámica temporal de los efectos, fortaleciendo la identificación causal mediante la combinación de PSM y DiD.

El impacto de 14.8 puntos porcentuales en asistencia escolar es consistente con evaluaciones de programas similares en la región, situándose entre los efectos reportados para Bolsa Família en Brasil (12-18 %) [10] y Familias en Acción en Colombia (8-12 %). La magnitud del efecto sugiere que las condicionalidades educativas del programa son efectivas para incentivar la demanda educativa en contextos de pobreza.

El efecto sustancial en controles de salud (20.5 pp) supera las estimaciones previas para Juntos [15], posiblemente reflejando mejoras en la implementación del programa o mayor capacidad de respuesta del sistema de salud post-pandemia. Este resultado es particularmente relevante considerando las brechas históricas en acceso a servicios preventivos de salud en áreas rurales.

La reducción de 11.9 puntos porcentuales en desnutrición crónica representa un logro significativo, considerando que este indicador responde lentamente a intervenciones. Este efecto es comparable al reportado para Progresar/Oportunidades en México tras varios años de implementación [17], sugiriendo que el programa Juntos está logrando impactos nutricionales importantes en un período relativamente corto.

Los mecanismos causales que explican estos efectos operan a través de múltiples canales interrelacionados. El canal de liquidez es evidente en el incremento de 40.4 soles en gastos educativos, sugiriendo que las transferencias no solo cubren costos directos sino que generan un efecto multiplicador donde los hogares complementan la transferencia con recursos propios. El canal de condicionalidad se manifiesta en el cambio comportamental sostenido hacia la utilización de servicios educativos y de salud. El canal de empoderamiento, particularmente relevante dado que las transferencias se entregan directamente a las madres, puede estar contribuyendo a una mejor asignación intrafamiliar de recursos hacia el bienestar infantil.

Las implicaciones para política pública son significativas. La magnitud y consistencia de los efectos justifican no solo la continuidad sino la potencial expansión del programa. Con un costo aproximado de 1,200 soles anuales por hogar, el retorno social de la inversión es altamente favorable considerando los beneficios de largo plazo en capital humano. La evidencia sugiere que el diseño actual del programa es efectivo, aunque podría optimizarse considerando la heterogeneidad de efectos observada.

Este estudio presenta algunas limitaciones que deben considerarse. Primero, el horizonte temporal de dos años puede no capturar efectos de más largo plazo o potencial desvanecimiento de impactos. Segundo, aunque la combinación PSM-DiD fortalece la identificación causal, no podemos descartar completamente la presencia de factores confusores no observables que varían en el tiempo. Tercero, el análisis no captura potenciales efectos de equilibrio general o spillovers sobre hogares no beneficiarios.

6. Conclusiones

Esta investigación demostró exitosamente la aplicabilidad de métodos avanzados de inferencia causal para evaluar el programa Juntos utilizando datos panel, proporcionando evidencia robusta de impactos positivos significativos en múltiples dimensiones del bienestar infantil. La combinación de Propensity Score Matching con Diferencias en Diferencias permitió abordar eficazmente los desafíos metodológicos inherentes a la evaluación de programas sociales focalizados.

Los resultados confirman que el programa Juntos constituye una intervención efectiva para mejorar el capital humano en poblaciones vulnerables, con efectos particularmente pronunciados en asistencia escolar (14.8 pp), utilización de servicios de salud (20.5 pp), inversión en educación (40.4 soles) y reducción de desnutrición crónica (11.9 pp). Estos im-

pactos son comparables o superiores a programas similares en América Latina, validando el diseño e implementación del programa.

La robustez metodológica del estudio, evidenciada por el balance exitoso de covariables en ambos períodos ($-SMD = -0.1$), la consistencia de resultados a través de especificaciones alternativas, y el uso de bootstrap para la inferencia estadística, fortalece la validez de las conclusiones. El uso de datos panel permitió controlar mejor por heterogeneidad no observada y validar el supuesto de tendencias paralelas.

Las implicaciones para política pública sugieren que el programa debe mantenerse y potencialmente expandirse, priorizando hogares rurales en pobreza extrema donde los impactos son mayores. La evidencia respalda que las inversiones en programas de transferencias condicionadas generan retornos sociales sustanciales a través de la acumulación de capital humano.

Futuras investigaciones deberían explorar la sostenibilidad de largo plazo de estos efectos, los impactos heterogéneos según características del hogar, y la optimalidad del diseño actual del programa. La integración de métodos de machine learning podría mejorar la precisión del targeting y la estimación de efectos heterogéneos del tratamiento.

Este estudio contribuye a la literatura sobre evaluación de impacto en países en desarrollo, demostrando que metodologías cuasi-experimentales rigurosas pueden proporcionar evidencia causal confiable incluso en ausencia de aleatorización. Los hallazgos refuerzan el consenso sobre la efectividad de las transferencias condicionadas como herramienta de política social para romper los ciclos intergeneracionales de pobreza.

7. Disponibilidad de Datos

Los datos ENAHO 2022-2023 utilizados en este estudio están disponibles públicamente en el portal del Instituto Nacional de Estadística e Informática (INEI): <https://proyectos.inei.gob.pe/microdatos/>. El código de análisis y materiales de replicación pueden encontrarse en: https://github.com/123ademir/metodos_de_optimizacion.

Referencias

Referencias

- [1] Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1), 1-19. <https://doi.org/10.1111/0034-6527.00321>
- [2] Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press. ISBN: 978-0-691-12035-5

- [3] Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399-424. <https://doi.org/10.1080/00273171.2011.568786>
- [4] Baird, S., Ferreira, F. H., Özler, B., & Woolcock, M. (2014). Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness*, 6(1), 1-43. <https://doi.org/10.1080/19439342.2014.890362>
- [5] Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31-72. <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- [6] Díaz, J. J., & Handa, S. (2009). An assessment of propensity score matching as a nonexperimental impact estimator: Evidence from Mexico's PROGRESA program. *Journal of Human Resources*, 41(2), 319-345. <https://doi.org/10.3368/jhr.XLI.2.319>
- [7] Escobal, J., & Benites, S. (2015). Direct and indirect effects of Juntos, a conditional cash transfer program in Peru. *GRADE Working Paper*, 44. Available at: http://www.grade.org.pe/wp-content/uploads/WP44_escobal_benites.pdf
- [8] Fiszbein, A., & Schady, N. (2009). *Conditional cash transfers: reducing present and future poverty*. World Bank Policy Research Report. Washington, DC: World Bank. <https://doi.org/10.1596/978-0-8213-7352-1>
- [9] Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). *Impact evaluation in practice* (2nd ed.). World Bank Publications. <https://doi.org/10.1596/978-1-4648-0779-4>
- [10] Glewwe, P., & Kassouf, A. L. (2012). The impact of the Bolsa Escola/Familia conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil. *Journal of Development Economics*, 97(2), 505-517. <https://doi.org/10.1016/j.jdeveco.2011.05.008>
- [11] Heckman, J. J., Ichimura, H., & Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4), 605-654. <https://doi.org/10.2307/2971733>
- [12] Heckman, J. J., Ichimura, H., & Todd, P. E. (1998). Characterizing selection bias using experimental data. *Econometrica*, 66(5), 1017-1098. <https://doi.org/10.2307/2999630>

- [13] Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945-960. <https://doi.org/10.1080/01621459.1986.10478354>
- [14] Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press. <https://doi.org/10.1017/CB09781139025751>
- [15] Perova, E., & Vakis, R. (2015). Welfare impacts of the "Juntos" program in Peru: Evidence from a non-experimental evaluation. *World Bank Working Paper*, 7859. <https://doi.org/10.1596/1813-9450-7859>
- [16] Rawlings, L. B., & Rubio, G. M. (2005). Evaluating the impact of conditional cash transfer programs. *The World Bank Research Observer*, 20(1), 29-55. <https://doi.org/10.1093/wbro/lki001>
- [17] Rivera, J. A., Sotres-Alvarez, D., Habicht, J. P., Shamah, T., & Villalpando, S. (2004). Impact of the Mexican program for education, health, and nutrition (Progresa) on rates of growth and anemia in infants and young children. *JAMA*, 291(21), 2563-2570. <https://doi.org/10.1001/jama.291.21.2563>
- [18] Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55. <https://doi.org/10.1093/biomet/70.1.41>
- [19] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688-701. <https://doi.org/10.1037/h0037350>
- [20] Saavedra, J. E., & García, S. (2012). Impacts of conditional cash transfer programs on educational outcomes in developing countries: A meta-analysis. *RAND Labor and Population Working Paper*, WR-921-1. <https://doi.org/10.7249/WR921>
- [21] Schultz, T. P. (2004). School subsidies for the poor: evaluating the Mexican Progresa poverty program. *Journal of Development Economics*, 74(1), 199-250. <https://doi.org/10.1016/j.jdeveco.2003.12.009>
- [22] Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1-21. <https://doi.org/10.1214/09-STS313>
- [23] Todd, P. E., & Wolpin, K. I. (2006). Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American Economic Review*, 96(5), 1384-1417. <https://doi.org/10.1257/aer.96.5.1384>