2024 Capacity Development on

IMPACT EVALUATION

April 26, 2024 | 1:00 PM

Pasig

Regression Discontinuity Design

Aleli D. Kraft

Professor, UP School of Economics

Outline

- Introduction
- Conditions and Intuition RDD
- Types of RDD
- Checking the validity of RDD designs
- Estimation approach and methods
- Limitations of RDD design
- Checklist for RDD
- Applications

What is RDD?

- Regression Continuity Design (RDD) → an impact evaluation method that can be used for programs with
  - Continuous eligibility index – e.g., predicted income based on assets, test scores, age for pensions
    - Also called running or forcing variable
  - Defined cutoff score – threshold below (above) which units are eligible for program and ineligible otherwise

History of RDD

- Thistlewaite and Campbell (1960)
  - estimated the effect that receiving a national merit scholarship award on student's ability to obtain additional college scholarship
  - National merit scholarship – given if students' test score exceeds a threshold
- Statisticians have developed and refined RD methods
  - Trochim (1984), van der Klaauw (2002), Hahn, Todd and van der Klaauw (2001), Lee and Lemiux (2010)
  - Developed parametric and semiparametric methods
  - Rigorous treatment of identification and estimation in these models

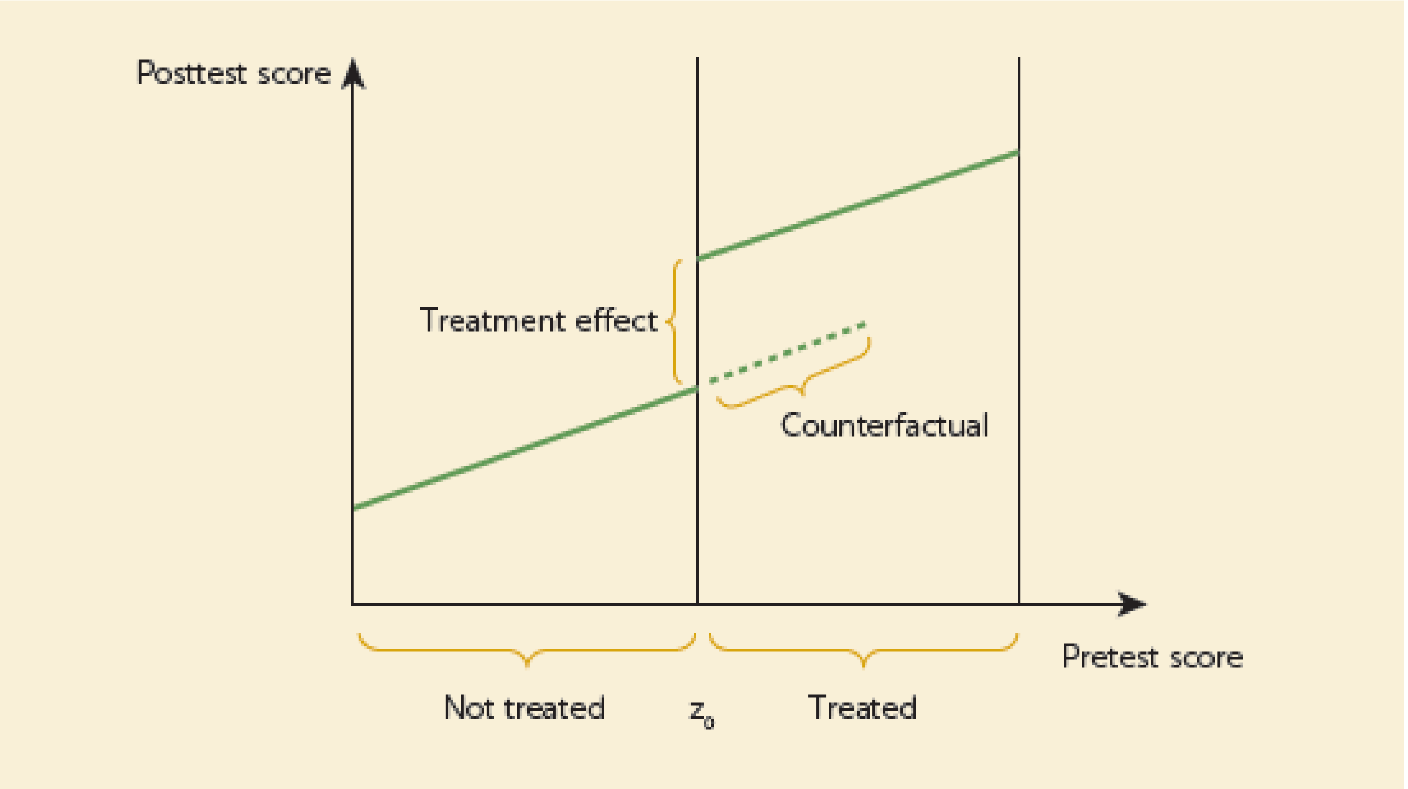
Conditions to apply an RDD

- Continuous eligibility index
  - population of interest can be ranked based on the index
- Clearly defined cutoff score
  - Point above or below which a population is classified as eligible
  - Unique to program of interest
- Score cannot be manipulated
  - Renders the treatment as "locally randomized"

Intuition of the RDD

- RDD impact around the eligibility cutoff – difference between average outcomes for units on the treated side of the cutoff and the average outcome in the untreated side of the cutoff

FIGURE 14.1 The intuition for regression discontinuity estimation



Intuition of the RDD

- Ineligible units close enough to the cutoff used as a comparison group to establish the counterfactual

- Similarity of the eligible and ineligible units imply that the program is the plausible reason for the difference in outcomes

- Since comparison group may just be above (below) the cutoff, impact of RDD is only valid “locally” → Local average treatment effect (LATE)

Types of RDD

- Sharp design
  - Treatment status is deterministic and discontinuous function of a covariate
  - The treatment variable P_i depends deterministically on a variable Z_i

$$P_i = f(Z_i)$$

  - The $f(Z_i)$ changes discontinuously at a known point z_0 from a value of zero to a value of one

$$f(Z_i) = 0 \; if \; Z_i \leq z_0 \; and \; f(Z_i) = 1 \; if \; Z_i > z_0$$

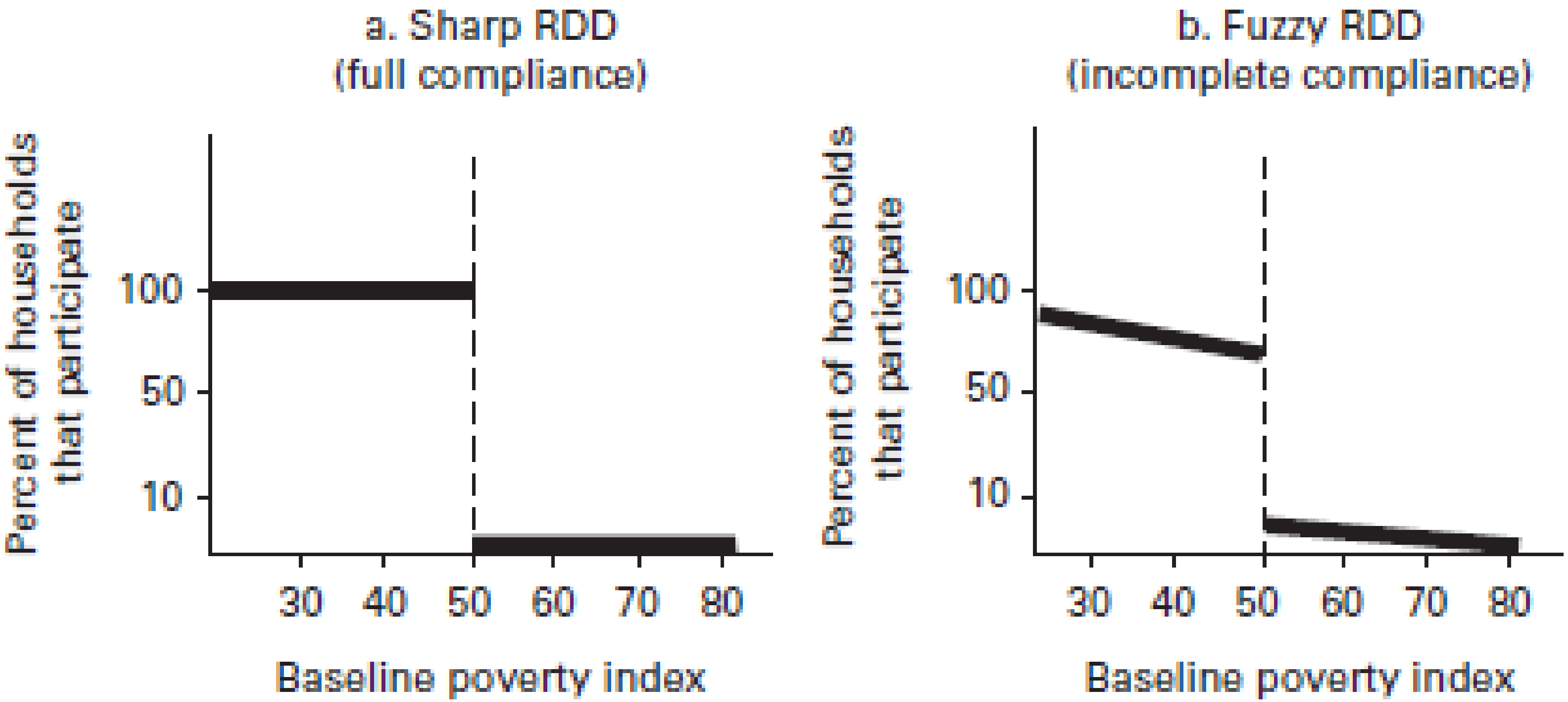
Types of RDD

- Fuzzy design
  - The treatment P_i is a random variable given (conditional on) Z_i.
  - Factors other than Z also influence program participation.
  - The expected value of P_i conditional on Z equals the probability that P_i is equal to 1 conditional on Z_i.

$$E[P|Z_i] = Prob[P_i = 1|Z_i] \equiv f(Z_i)$$

  - Under a fuzzy RD design, the conditional probability $Prob[P_i = 1|Z_i$ is discontinuous at z_0 a known point.

Figure 6.3 Compliance with Assignment

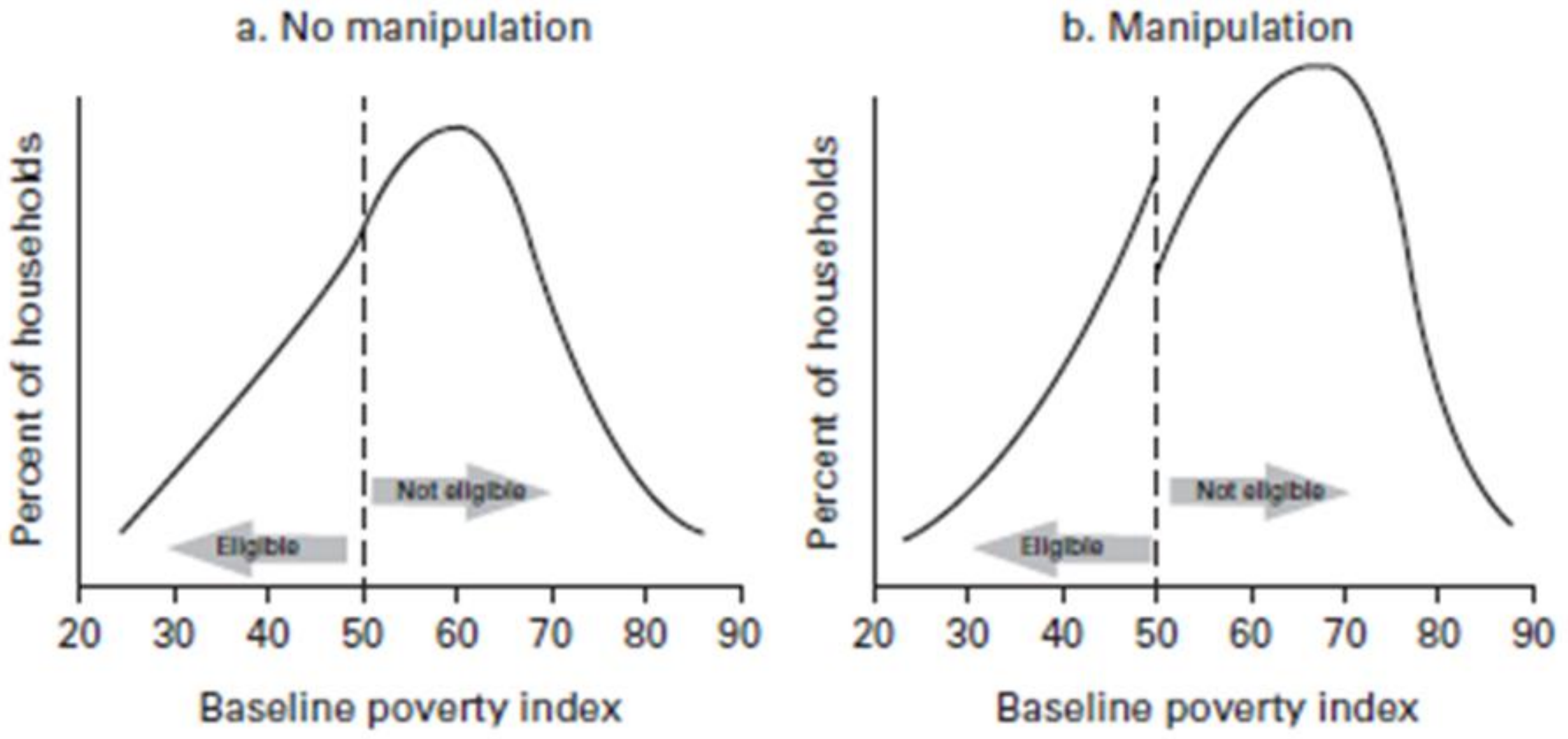


Checking for conditions of RDD

- Eligibility index should not be manipulated

- Tests
  - Look at distribution of eligibility index around the cutoff.
  - Plot eligibility index against the outcome variable and check for no "jumps" around the cutoff line

Figure 6.4 Manipulation of the Eligibility Index



Checking for assumptions of RDD

- Key assumption – people around the cutoff are comparable
- Examine the distribution of baseline covariates around the cutoff value
  - Test for statistically different means of baseline characteristics within some interval above and below the cutoff.
  - Use sharp RD estimation methods using baseline characteristics as the outcome variable

RDD estimation approach

- For both the shard and fuzzy designs, the following ratio identifies the treatment effect at $Z = z_o$

$$\hat{\Delta}_{FRD} = \frac{\hat{Y}^+ - \hat{Y}^-}{\hat{P}^+ - \hat{P}^-}$$

  - Where :
    - $\hat{Y}^+ = \lim_{e \to 0^+} \{E[Y_i | Z_i = z_o + e]$
    - $\hat{Y}^- = \lim_{e \to 0^+} \{E[Y_i | Z_i = z_o - e]$
    - $\hat{P}^+ = \lim_{e \to 0^+} \{E[P_i | Z_i = z_o + e]$
    - $\hat{P}^- = \lim_{e \to 0^+} \{E[P_i | Z_i = z_o - e]$
- For sharp design, probability of treatment goes from zero to one, denominator is one, only numerator needs to be estimated

Estimation methods

- Local means approach
  - Nonparametric estimator
  - Estimates the limits by taking averages over the Y values and the P values within a specified distance of the boundary points called the bandwidth, h.

Estimation methods

- Local means approach
  - Given the bandwidths above and below the cutoff point (h_+ and h_- , respectively), the limits are estimated by:
    - $\hat{Y}^+ = \frac{\sum_{i=1}^{n} Y_i \times 1(z_0 < Z_i < z_0 + h_+)}{\sum_{i=1}^{n} 1(z_0 < Z_i < z_0 + h_+)}$
    - $\hat{Y}^- = \frac{\sum_{i=1}^{n} Y_i \times 1(z_0 - h_- < Z_i < z_0)}{\sum_{i=1}^{n} 1(z_0 - h_- < Z_i < z_0)}$
    - $\hat{P}^+ = \frac{\sum_{i=1}^{n} P_i \times 1(z_0 < Z_i < z_0 + h_+)}{\sum_{i=1}^{n} 1(z_0 < Z_i < z_0 + h_+)}$
    - $\hat{P}^- = \frac{\sum_{i=1}^{n} P_i \times 1(z_0 - h_- < Z_i < z_0)}{\sum_{i=1}^{n} 1(z_0 - h_- < Z_i < z_0)}$
  - Four terms represent simple means of either Y or P for the observations that are sufficiently close to the cutoff point.

Estimation methods

- Local linear regression approach
    - Local linear regressions fit separate linear regressions on the right and left hand side of the cutoff value which reduces bias at the boundary points relative to simple averaging.

Estimation methods

- How can the bandwidths be chosen?
  - Hahn, Todd and van der Klaauw use leave-one-out cross validation to select the optimal bandwidth of the RDD estimator
  - Imbens and Kalyanaram developed data dependent optimal plug-in bandwidth estimator for the RDD estimator based on the local linear regression approach.
- Standard errors can be obtained using bootstrap methods.
  - Construct bootstrap samples, which are data subsamples (with replacement) of the original observations
  - RDD estimates are obtained for each of the bootstrap samples
  - Standard errors can be estimated from the empirical variance over the bootstrap estimates (excluding the estimate that is based on the original data set).

Estimation methods

- Alternate methods:
- Single regression that includes observations both below and above the cutoff with an indicator variable for receiving treatment.

$$Y = \alpha + \tau P + \mathbf{X}'\boldsymbol{\beta} + \varepsilon$$

  - Imposes a specific functional form for the outcome
  - Makes use of all the data in the estimation
  - May be used to check if impact estimates would differ if only data near cutoff is used.

Estimation methods

- Variation: Flexible parametric specification for $g(\boldsymbol{X}) = E[Y_{0i}|\boldsymbol{X}_i]$ and added as a control function to the regression of Y on P.

- For fuzzy design: P on the control function-augmented regression equation is replaced by the first stage estimate of $E[P_i|Z]$

Estimation method

- Regression equation

$$Y_i = \alpha + \beta Z_i + \rho P_i + \eta_i,$$

  - where ρ is the causal effect of interest.
- P_i (the treatment indicator) - only correlated with Z_i, in fact a deterministic function of Z_i.
- RDD captures the causal effects by distinguishing the nonlinear and discontinuous function, $1(Z_i \leq z^*)$, from the smooth and linear function, Z_i.

Limitations

- Impact valid only in the neighborhood around the cutoff score
- Estimates cannot be generalized to units whose scores are far away from the cutoff
- May still be helpful to answer policy questions about expanding or cutting at the margin

Limitations

- Fewer observations can be used than in other methods that include all units
- Large samples are required to obtain statistical power
- Bandwidth should include sufficient number of observations yet maintain balance of characteristics of population
- Even with limitations, RDD yields unbiased estimates of the impact in the vicinity of the eligibility cutoff.

RDD Checklist

- Is the index continuous around the cutoff score at the time of the baseline?
- Is there any evidence of noncompliance with the rule that determines the eligibility for treatment? If there is noncompliance → combine RDD with an IV approach to correct for the "fuzzy" discontinuity
- Is there any evidence that the index scores may have been manipulated in order to qualify for the program?
- Is the cutoff unique to the program being evaluated, or is the cutoff used by other programs as well?

RDD Applications

Orbeta, Aniceto, Jr. , Kris Ann Melad and Nina Victoria Araos (2023), "Reassessing the Impact of the Pantawid Pamilyang Pilipino Program: Results of the Third Wave Impact Evaluation" PIDS Research Paper Series 2023-06

Pantawid Pamilya Pilipino Program (4Ps)

- Launched in in 2008, registering approximately 300,000 HH beneficiaries in its first year
- Currently has more than 4 million HH from almost all municipalities and provinces nationwide.
- Budget - 50 million in 2008 to 78 billion in 2017
  - (61 percent of DSWD’s budget and 0.5 percent of GDP in 2017)
- Covers approximately 60 percent of the poorest quintile HH

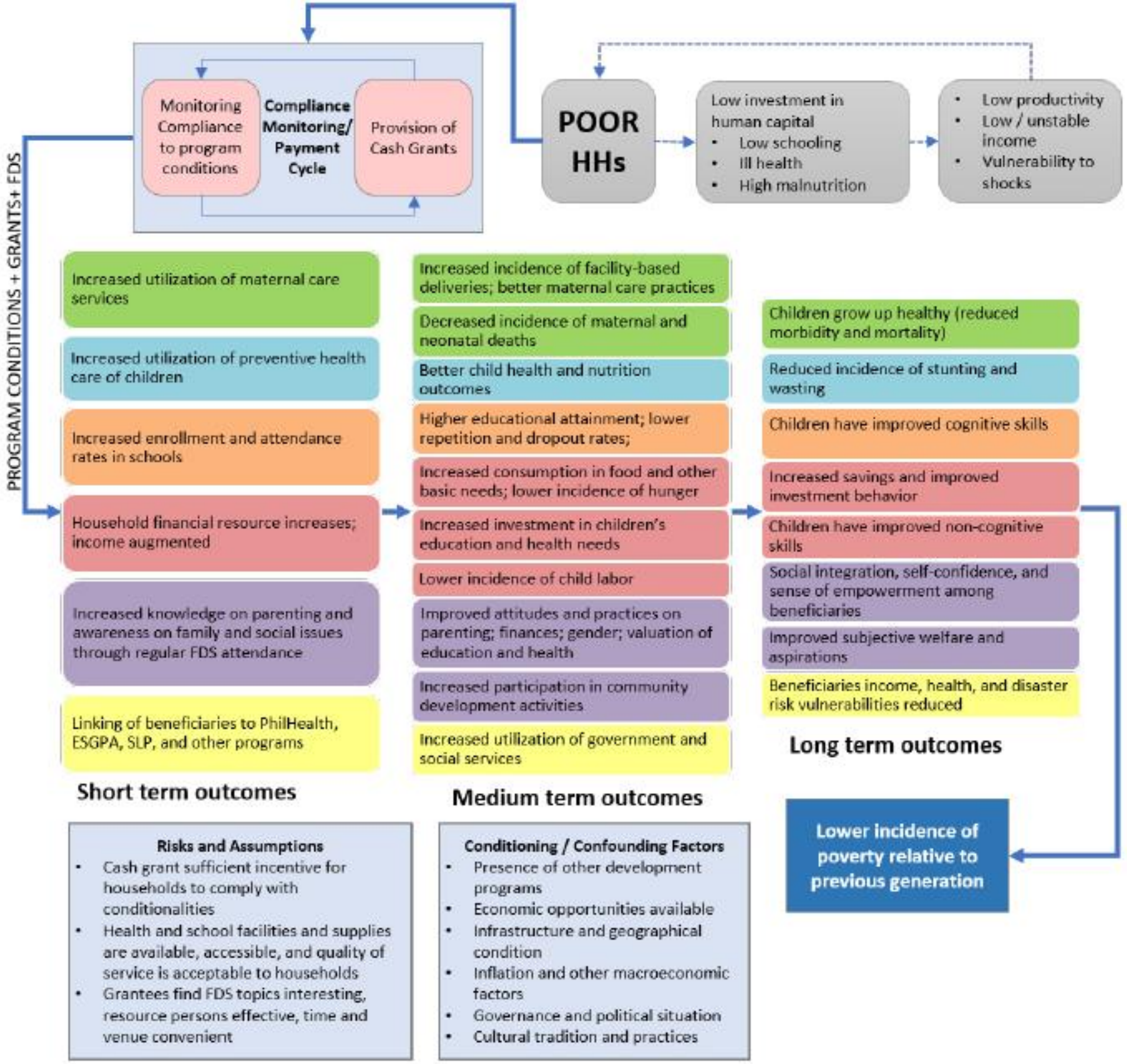
4Ps

- Has undergone rigorous evaluations as part of its M and E system.
  - 1st round in 2011, 2nd round in 2013
  - Findings from both rounds →
    - achieve keeping children in school and healthy through increased use of maternal and child health care services.
    - Mixed results for some outcomes and observed no program impact on total HH consumption and infant immunization
- 10th year of program implementation
  - IE conducted to reassess the program’s impact on short-term and intermediate outcomes
  - confirm mixed results of previous studies.
  - assess after 4P underwent design changes since the 1st and 2nd study

4Ps

- Program modifications
  - Extension of age coverage – DSWD extended the education grants to children up to 18 years (previous 0-14)
  - Education grant increased from 300 to 500 for children enrolled in HS
  - Change in exit policy -Lifted the five-year limit of program participation in 2015 so that beneficiary HH cease to receive program benefits when the last of their three children graduate from high school or reach 19 y.o., whichever comes first
  - Open selection of monitored children
  - Rice subsidy
    - HH receive this if they comply with either the education or health and FDS conditionalities.

Figure 1. *Pantawid Pamilya* program theory



HH = household; FDS = Family Development Sessions; TOC = theory of change; PhilHealth = Philippine Health Insurance Corporation; ESGPA = Expanded Student's Grants-in-Aid Program for Poverty Alleviation; SLP = Sustainable Livelihood Program

Source: Adapted from the program TOC prepared by the 4Ps Impact Evaluation Technical Working Group (18 August 2017)

Hypotheses

i. Maternal health
   i. H1: The Pantawid Pamilya promotes higher awareness and utilization of responsible parenthood interventions
   ii. H2: the PP promotes the utilization of maternal healthcare services
   iii. H3: PP mothers experience fewer problems during pregnancy and delivery

ii. Child health
   i. H4: PP increases utilization of health care services by children
   ii. H5: PP participation improves the childcare practices of parents
   iii. H6: PP children have better nutrition and health outcomes

i. Education and child labor
   i. H7: PP increases the school participation of children
   ii. H8: PP results in improve education outcomes of children (lower drop-out rates, enrolled in age-appropriate education levels)
   iii. H9: PP reduces the incidence and time spent on child labor
   iv. H10: PP promotes higher investments in education

ii. Household consumption and income
   i. H11: PP increases HH consumption and income
   ii. H12: PP does not encourage dependency (beneficiaries are not expected to have a lower labor force participation rate and reduced time spent in work compared to non-beneficiaries)
   iii. H13: PP increases access to social services and utilization of government services and benefits

iii. Other behavioral outcomes
   i. H14: PP increases participation in community development activities
   ii. H15: PP promotes a better outlook on their children's current situation and future.

Methodology

- Analysis uses Regression Discontinuity Design RDD
- Running or forcing variable: HH whose PMT score was below the provincial poverty threshold with children 0-18 or pregnant members are eligible
- RDD assumes that near the cutoff, observations below or above the eligibility criteria are comparable, and assignment to treatment or comparison group is as if done randomly.
- A large jump in the outcome variable at the cutoff after the implementation of the intervention can be causally attributed to the intervention

Estimation strategy

- Program impacts estimated by sharp and fuzzy RDD
- Sharp RDD
    - Sharp RDD considers all HH below the cutoff as treated regardless of receipt of program benefits
    - With non-adherence to treatment assignment, analysis reports intent-to-treat (ITT) effects
    - ITT - unbiased effect of the intervention among all eligible HH regardless of adherence to the treatment assignment
- Fuzzy RDD
    - Reports the treatment on the treated (TOT) effects, considering compliance with the treatment assignment
    - Some HH waive their benefits and not participate while some HH who are not eligible maneuver their way to receive program benefits
    - To address the issue of non-compliance, and IV approach is used. Administrative information of the actual receipt of PP benefits determines who got the benefits, while the treatment assignment based on the eligibility criteria is used as an instrument

Estimation strategy

- Measure program impact using local linear regression models.
- Analysis uses the STATA package *rdrobust* (developed by Calonico et al in 204 and upgraded in 2016)
  - Allowed for data-driven bandwidth selection, cluster-robust options for variance estimation
  - Impact estimates and significant levels for sharp and fuzzy RD estimations presented in the report are based on this command.
- Means of outcomes for the treatment and comparison groups were computed by getting the predicted outcome values at the threshold using standard least-squares regression that replicates the conventional estimates of rdrobust.

Estimation strategy

- Sharp RD base estimation model :

$$Y_i = \beta_0 + \tau T_i + \beta_1 \bar{X}_i + \beta_2 T_i \bar{X}_i + \beta_n z + \varepsilon_i$$

  - $\bar{X}_i$ is the running variable, T is the treatment assignment and z are the other covariates in the model
  - The equation is estimated within the bandwidth h determined by rdrobust.
- For fuzzy RD, two-stage least squares estimation was used with the treatment assignment as the instrument for the actual receipt of benefits
- Impact of the program is estimated within three sets of bandwidths:
  - Coverage error rate (CER)-optimal bandwidths
  - Mean square error (MSE)-optimal bandwidth
  - Full sample bandwidth

Estimation strategy

- Impact estimates based on the MSE bandwidths, significance based on both CER and MSE-optimal bandwidths
- PMT scores were recentered at the cutoff to simplify interpretation of results
- Include
  - Municipal dummies for municipal fixed effects
  - Supply and baseline covariates were included to improve precision of estimates

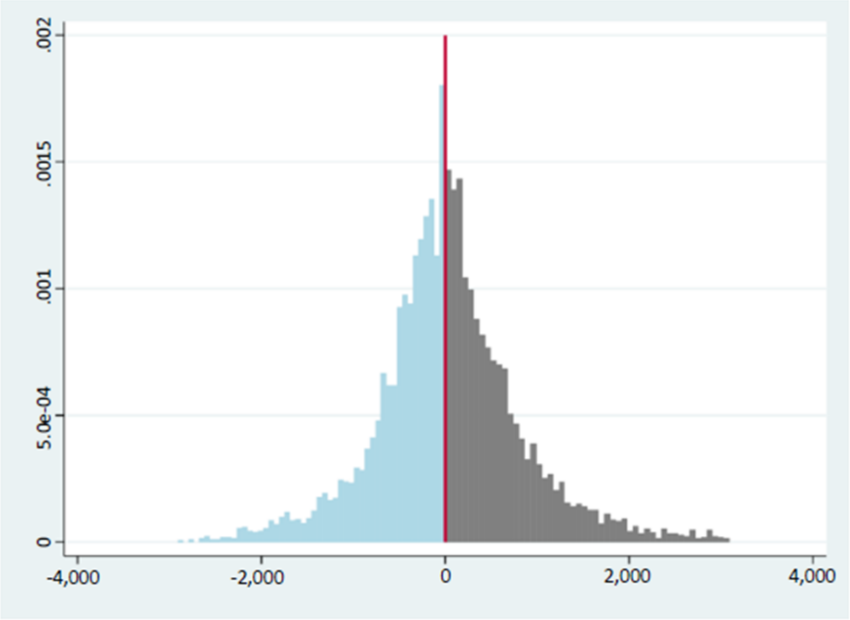
IE Wave 3 Data

- HH with at least 2 years exposure to program of those registered from 2008 to 2014
- Municipalities with at least 20 bgs having at least 30 HH
- HH sampled based on proximity to cutoff for poverty thresholds
- Special SWS survey conducted from Nov. 17- Jan 2018
- New questions - Income of households, Access to government services, Coping mechanisms due to difficulties, Community involvement , Social integration, Access to information, Perceptions of non-4p of the program, Food hygiene and positive disciplining, Quality of health services received, Questions on grit and parent-child relations

Validity of assumptions

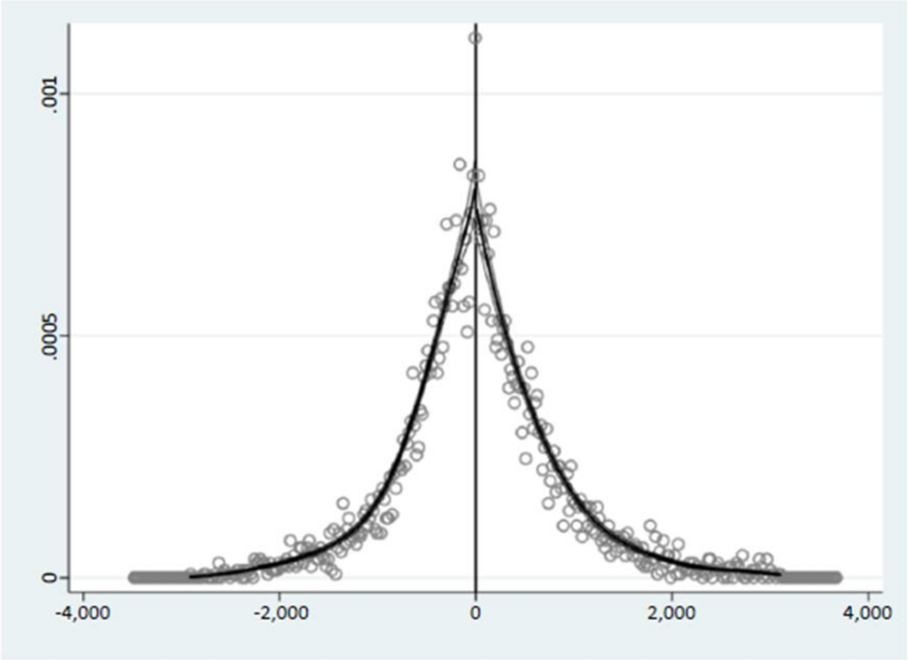
- Beneficiaries should not have any influence on treatment assignment

Figure 1. Distribution of the running variable



Source: Authors' computations using Listahanan 1 data of the sample households

Figure 2. McCrary density test



Source: Authors' computations using Listahanan 1 data of the sample households

Validity of assumptions

- HH to left and right near cutoff are comparable in terms of key baseline characteristics

Table 3. Test for discontinuity of baseline covariates

Outcomes		Bandwidths		
		CER Optimal	MSE Optimal	Sample
Natural logarithm of family size	impact	-0.06***	-0.05**	-0.03***
	se	0.02	0.02	0.01
	p-value	0.006	0.013	0.008
	non-Pantawid	1.64	1.64	1.62
	number of obs.	3,294	4,087	6,763
No. of children 0–5 years old	impact	0.03	0.04	0.01
	se	0.05	0.05	0.03
	p-value	0.665	0.466	0.326
	non-Pantawid	0.63	0.61	0.62
	number of obs.	2,881	3,590	6,763
No. of children 6–14 years old	impact	0.00	0.03	0.04
	se	0.07	0.06	0.04
	p-value	0.976	0.719	0.522
	non-Pantawid	1.12	1.09	1.14
	number of obs.	3,436	4,217	6,763
No. of children 15–18 years old	impact	-0.17***	-0.14***	-0.06***
	se	0.05	0.04	0.03
	p-value	0.001	0.003	0.005
	non-Pantawid	0.48	0.51	0.52
	number of obs.	2,663	3,333	6,763

Table 3 *(continued)*

Outcomes		Bandwidths		
		CER Optimal	MSE Optimal	Sample
Agricultural household	impact	-1.93	-1.46	-0.05
	se	2.81	2.55	1.96
	p-value	0.510	0.590	0.649
	non-Pantawid	3.95	3.19	0.52
	number of obs.	3,389	4,178	6,763
Availability of domestic help at household	impact	-0.11	-0.13	-0.07
	se	0.07	0.09	0.08
	p-value	0.121	0.148	0.226
	non-Pantawid	0.05	0.08	0.39
	number of obs.	3,133	3,880	6,763
Single household	impact	1.04	0.95	1.07
	se	0.74	0.65	0.39
	p-value	0.176	0.175	0.105
	non-Pantawid	2.76	2.73	1.87
	number of obs.	3,123	3,850	6,763
Roof made of light materials	impact	-0.04	-0.84	-1.47
	se	2.50	2.35	1.65
	p-value	0.956	0.691	0.833
	non-Pantawid	15.12	14.37	14.82
	number of obs.	3,052	3,775	6,763

Validity of assumptions

- At baseline, outcomes should not show discontinuity at the cutoff

Appendix 6. Means of outcome indicators by treatment assignment

Outcome	Treatment	Comparison	Obs.	T-test p value
Awareness of any modern FP method	0.9962	0.9956	5,138	0.7229
Ever use of any modern FP method	0.7857	0.7696	5,117	0.1681
Count of modern FP methods aware of	6.6062	6.2803	5,138	0.0000***
Count of modern FP methods ever used	1.3583	1.2827	5,138	0.0122**
Current users of any modern FP method	0.4611	0.4298	5,138	0.0243**
Contraceptive prevalence rate	0.4989	0.4683	4,594	0.0378**
At least one prenatal checkup	0.9742	0.9696	3,135	0.4289
At least 4 prenatal checkups	0.8210	0.7851	3,135	0.0115**
Frequency of prenatal checkup	6.2437	6.1121	3,047	0.2691
Prenatal care provided by skilled professional	0.9499	0.9462	3,176	0.6309
Prenatal care availed in health facility	0.9592	0.9467	3,174	0.0964*
Skilled birth attendance	0.8705	0.8698	2,932	0.9540
Skilled birth attendance by a doctor	0.3992	0.3976	2,932	0.9310
Skilled birth attendance by a midwife	0.4566	0.4596	2,932	0.8705
Skilled birth attendance by a nurse	0.0774	0.0724	2,932	0.6064
Facility-based delivery	0.8374	0.8324	2,937	0.7133
Postnatal check within 72 hours	0.4889	0.4950	2,415	0.7636
Postnatal check within 72 hours by a skilled professional	0.4679	0.4671	2,413	0.9700
Postnatal check within 24 hours	0.2965	0.3131	2,413	0.3777
Postnatal check within 24 hours by a skilled health professional	0.2965	0.3131	2,413	0.3777

Selected results

- Impact refers to estimated program impact at the threshold
- Non-pantawid is predicted mean of the outcome variable for non-treated observations above the poverty threshold under the sharp RD
- Results of fuzzy estimation consistent with sharp RD in terms of direction, higher in magnitude than shard RD

Hypothesis 6: PP children have better nutrition and health outcomes

Table 16. Nutrition and child health outcomes among children below 6 years old

Outcomes		Sharp RD			Fuzzy RD		
		CER Optimal	MSE Optimal	Sample	CER Optimal	MSE Optimal	Sample
Underweight	Impact	5.32	4.50	2.20	6.07	5.18	2.70
	Robust p-value	0.14	0.19	0.19	0.16	0.20	0.19
	Non-Pantawid	20.10	20.16	20.14	20.02	20.06	20.08
	Number of obs.	1,688	2,091	3,717	1,829	2,247	3,717
Severe underweight	Impact	1.22	1.04	0.96	1.94	1.29	1.18
	Robust p-value	0.60	0.66	0.45	0.49	0.63	0.46
	Non-Pantawid	5.09	4.91	4.70	4.91	4.95	4.67
	Number of obs.	1,705	2,095	3,717	1,584	2,012	3,717
Stunting	Impact	5.53*	5.60*	4.59**	8.69*	7.14*	5.61**
	Robust p-value	0.10	0.08	0.04	0.05	0.09	0.04
	Non-Pantawid	29.77	29.51	29.68	29.28	29.55	29.56
	Number of obs.	2,059	2,477	3,628	1,529	1,928	3,628
Severe stunting	Impact	5.34**	4.98**	3.06**	7.27**	6.42**	3.73**
	Robust p-value	0.04	0.04	0.02	0.03	0.04	0.02
	Non-Pantawid	8.16	8.47	9.30	7.53	8.05	9.22
	Number of obs.	1,770	2,156	3,628	1,445	1,825	3,628
Wasting	Impact	-1.17	-1.24	0.37	-1.67	-1.22	0.45
	Robust p-value	0.62	0.50	0.97	0.56	0.53	0.97
	Non-Pantawid	12.35	11.84	10.65	12.29	11.53	10.64
	Number of obs.	1,343	1,698	3,239	1,489	1,842	3,239
Severe wasting	Impact	-2.05	-1.83	-0.81	-3.29	-2.72	-0.98
	Robust p-value	0.19	0.17	0.33	0.19	0.20	0.34
	Non-Pantawid	4.23	4.07	3.65	5.04	4.52	3.67
	Number of obs.	1,949	2,340	3,239	1,338	1,689	3,239

RD = regression discontinuity; CER = coverage error rate; MSE = mean square error; obs. = observation
Notes: Treatment and control means are calculated using predicted values from replicating the rdrobust routine using least-squares regression. The p-value presented is from the robust version of the estimation that corrects for bias. P-values are rounded off to two decimal places. Asterisks reflect level of significance of the estimates: *** $p<0.01$, ** $p<0.05$, * $p<0.10$
Source: Authors' computations

Table 17. Nutrition and health outcomes among children 0–2 years old

Outcomes		Sharp RD			Fuzzy RD		
		CER Optimal	MSE Optimal	Sample	CER Optimal	MSE Optimal	Sample
Underweight	Impact	-4.75	-2.03	-0.41	-9.24	-5.22	-0.51
	Robust p-value	0.38	0.56	0.73	0.27	0.40	0.73
	Non-Pantawid	17.15	15.42	14.07	19.04	16.87	14.08
	Number of obs.	596	717	1,072	497	616	1,072
Severe underweight	Impact	-0.17	0.40	-0.10	0.85	-0.28	-0.12
	Robust p-value	0.95	0.94	0.49	0.88	0.97	0.49
	Non-Pantawid	4.00	3.61	4.03	3.82	3.93	4.04
	Number of obs.	516	635	1,072	437	549	1,072
Stunting	Impact	5.75	5.50	3.77	8.72	7.46	4.64
	Robust p-value	0.35	0.35	0.36	0.34	0.37	0.37
	Non-Pantawid	16.87	16.78	17.42	16.45	16.71	17.37
	Number of obs.	575	691	1,031	449	567	1,031
Severe stunting	Impact	3.33	3.87	1.47	5.62	4.13	1.80
	Robust p-value	0.43	0.34	0.40	0.39	0.49	0.40
	Non-Pantawid	5.73	5.65	6.62	5.19	5.69	6.60
	Number of obs.	583	700	1,031	447	567	1,031
Wasting	Impact	-7.88	-5.01	2.58	-8.48	-4.56	3.11
	Robust p-value	0.36	0.43	0.58	0.39	0.47	0.58
	Non-Pantawid	26.67	24.20	19.94	25.86	23.36	19.91
	Number of obs.	371	460	748	418	502	748
Severe wasting	Impact	-4.43	-4.49	-4.06	-5.12	-6.91	-4.90
	Robust p-value	0.47	0.47	0.28	0.42	0.24	0.29
	Non-Pantawid	8.83	8.62	9.41	8.62	9.38	9.47
	Number of obs.	363	452	748	426	510	748

RD = regression discontinuity; CER = coverage error rate; MSE = mean square error; obs. = observation
Notes: Treatment and control means are calculated using predicted values from replicating the rdrobust routine using least-squares regression. The p-value presented is from the robust version of the estimation that corrects for bias. P-values are rounded off to two decimal places. Asterisks reflect level of significance of the estimates: *** $p<0.01$, ** $p<0.05$, * $p<0.10$
Source: Authors' computations

Hypothesis 8: PP participation results in improved education outcomes of children

Table 26. Participation in extracurricular activities among school-aged children

Outcomes		Sharp RD			Fuzzy RD		
		CER Optimal	MSE Optimal	Sample	CER Optimal	MSE Optimal	Sample
Participation in any extracurricular activity in school	Impact	5.85**	4.09*	1.34	7.08**	5.86**	1.55
	Robust p-value	0.03	0.08	0.60	0.02	0.02	0.60
	Non-Pantawid	47.72	48.93	51.49	47.31	48.17	51.45
	Number of obs.	5,077	6,389	11,773	4,686	5,842	11,773
Count of extracurricular activities participated in school	Impact	0.04	0.05	-0.01	0.04	0.06	-0.01
	Robust p-value	0.59	0.42	0.72	0.62	0.43	0.72
	Non-Pantawid	1.10	1.10	1.18	1.10	1.10	1.18
	Number of obs.	4,425	5,581	11,773	4,330	5,469	11,773

RD = regression discontinuity; CER = coverage error rate; MSE = mean square error; obs. = observation
Notes: Treatment and control means are calculated using predicted values from replicating the rdrobust routine using least-squares regression. The p-value presented is from the robust version of the estimation that corrects for bias. P-values are rounded off to two decimal places. Asterisks reflect level of significance of the estimates: *** p<0.01, ** p<0.05, * p<0.10
Source: Authors' computations

Hypothesis 9: PP reduces the incidence and time spent on child labor

Table 27. Child labor among ages 10–14

Outcomes		Sharp RD			Fuzzy RD		
		CER Optimal	MSE Optimal	Sample	CER Optimal	MSE Optimal	Sample
At least 1 hour of work (with or without pay) last month	Impact	0.35	0.85	0.97	1.22	0.40	1.12
	Robust p-value	0.86	0.65	0.37	0.68	0.99	0.37
	Non-Pantawid	5.52	5.06	4.26	5.52	5.65	4.23
	Number of obs.	2,815	3,368	4,557	2,097	2,579	4,557
At least 1 hour of paid work last month	Impact	0.76	1.31	1.09	1.68	0.65	1.26
	Robust p-value	0.64	0.39	0.26	0.55	0.93	0.26
	Non-Pantawid	4.96	4.46	3.98	5.09	5.32	3.95
	Number of obs.	3,072	3,572	4,557	2,013	2,484	4,557
Number of days worked (with or without pay) last month	Impact	-0.40	-0.03	-0.05	-0.54	-0.17	-0.06
	Robust p-value	0.77	0.94	0.79	0.73	0.86	0.79
	Non-Pantawid	5.20	5.07	5.21	5.28	5.12	5.21
	Number of obs.	121	137	207	115	131	207
Worked with or without pay in the last 12 months	Impact	-0.20	-1.02	0.08	0.10	-0.83	0.10
	Robust p-value	0.85	0.53	0.85	0.93	0.61	0.85
	Non-Pantawid	8.85	8.84	6.99	8.85	8.90	6.98
	Number of obs.	2,174	2,670	4,560	1,990	2,451	4,560

RD = regression discontinuity; CER = coverage error rate; MSE = mean square error; obs. = observation
Notes: Treatment and control means are calculated using predicted values from replicating the rdrobust routine using least-squares regression. The p-value presented is from the robust version of the estimation that corrects for bias. P-values are rounded off to two decimal places. Asterisks reflect level of significance of the estimates: *** p<0.01, ** p<0.05, * p<0.10
Source: Authors' computations

Hypothesis 11: PP increases household consumption and income

Table 29. Household expenditures: Share to total expenditures

Outcomes		Sharp RD			Fuzzy RD		
		CER Optimal	MSE Optimal	Sample	CER Optimal	MSE Optimal	Sample
Share of food to total expenditures	Impact	1.36	1.31	0.35	1.70	1.65	0.43
	Robust p-value	0.18	0.15	0.21	0.21	0.20	0.21
	Non-Pantawid	63.13	63.13	63.42	63.13	63.06	63.41
	Number of obs.	2,699	3,357	5,523	2,277	2,848	5,523
Share of nonfood to total expenditures	Impact	-1.36	-1.31	-0.35	-1.70	-1.65	-0.43
	Robust p-value	0.18	0.15	0.21	0.21	0.20	0.21
	Non-Pantawid	36.87	36.87	36.58	36.87	36.94	36.59
	Number of obs.	2,699	3,357	5,523	2,281	2,848	5,523
Share of education to total expenditures	Impact	-0.03	0.03	0.14	-0.19	-0.08	0.17
	Robust p-value	0.85	0.99	0.48	0.51	0.67	0.48
	Non-Pantawid	2.36	2.35	2.29	2.42	2.38	2.29
	Number of obs.	3,026	3,670	5,523	2,252	2,819	5,523
Share of clothing and footwear to total expenditures	Impact	0.27***	0.25***	0.22***	0.32***	0.31***	0.26***
	Robust p-value	0.01	0.01	0.00	0.01	0.01	0.00
	Non-Pantawid	1.06	1.09	1.17	1.05	1.08	1.17
	Number of obs.	2,384	2,987	5,523	2,375	2,975	5,523

Hypothesis 12: PP does not encourage dependency

Table 33. Employment

Outcomes		Sharp RD			Fuzzy RD		
		CER Optimal	MSE Optimal	Sample	CER Optimal	MSE Optimal	Sample
Labor force participation	Impact	0.10	0.79	0.69	-0.21	0.12	0.85
	Robust p-value	0.99	0.58	0.12	0.88	0.99	0.12
	Non-Pantawid	57.97	57.62	57.50	58.18	57.97	57.48
	Number of obs.	12,564	15,143	22,315	10,172	12,549	22,315
Employment	Impact	-2.69**	-2.59***	0.08	-3.24**	-3.25***	0.10
	Robust p-value	0.01	0.01	0.75	0.02	0.01	0.75
	Non-Pantawid	92.99	92.80	91.83	93.09	92.93	91.82
	Number of obs.	4,784	6,077	12,860	4,655	5,918	12,860
Usual work hours per week in primary occupation	Impact	1.78	1.11	0.28	2.20	1.37	0.35
	Robust p-value	0.12	0.25	0.51	0.13	0.28	0.51
	Non-Pantawid	39.47	39.71	39.81	39.41	39.67	39.80
	Number of obs.	5,710	7,021	11,731	5,718	7,029	11,731
Other job or business besides primary occupation	Impact	2.81*	2.30*	0.18	3.21*	2.64	0.22
	Robust p-value	0.07	0.09	0.71	0.09	0.11	0.71
	Non-Pantawid	5.30	5.41	6.92	5.27	5.37	6.92
	Number of obs.	4,005	5,132	11,675	4,297	5,471	11,675

Table 33 *(continued)*

Outcomes		Sharp RD			Fuzzy RD		
		CER Optimal	MSE Optimal	Sample	CER Optimal	MSE Optimal	Sample
Usual work hours per week in other jobs	Impact	1.05	1.62	2.27	1.12	1.76	2.84
	Robust p-value	0.84	0.77	0.42	0.79	0.69	0.42
	Non-Pantawid	17.04	16.66	16.94	17.06	16.68	16.83
	Number of obs.	315	387	816	310	377	816
Total usual work hours per week	Impact	2.62**	1.92*	0.49	2.36*	1.58	0.60
	Robust p-value	0.03	0.07	0.38	0.08	0.18	0.38
	Non-Pantawid	40.20	40.44	40.94	40.37	40.61	40.93
	Number of obs.	5,108	6,360	11,732	6,388	7,721	11,732
Looking for additional work if employed	Impact	-0.73	-0.70	0.27	-0.85	-1.04	0.33
	Robust p-value	0.59	0.51	0.91	0.65	0.52	0.90
	Non-Pantawid	9.14	8.90	8.32	9.46	9.17	8.31
	Number of obs.	5,439	6,726	11,871	4,483	5,684	11,871
Unemployed and looking for work	Impact	-14.13	-14.02	-4.22	-16.35	-18.13	-5.10
	Robust p-value	0.19	0.14	0.19	0.24	0.14	0.20
	Non-Pantawid	36.76	36.69	30.29	37.03	37.42	30.38
	Number of obs.	502	607	1,043	469	566	1,043

RD = regression discontinuity; CER = coverage error rate; MSE = mean square error; obs. = observation
Notes: Treatment and control means are calculated using predicted values from replicating the rdrobust routine using least-squares regression. The p-value presented is from the robust version of the estimation that corrects for bias. P-values are rounded off to two decimal places. Asterisks reflect level of significance of the estimates: *** p<0.01, ** p<0.05, * p<0.10
Source: Authors' computations

Table 38. Grit

Outcomes		Sharp RD			Fuzzy RD		
		CER Optimal	MSE Optimal	Sample	CER Optimal	MSE Optimal	Sample
Ask for help when lesson is difficult	Impact	3.03**	1.86	0.75	4.40**	3.88**	0.86
	Robust p-value	0.03	0.11	0.43	0.02	0.01	0.43
	Non-Pantawid	88.19	88.81	89.72	87.42	87.88	89.70
	Number of obs.	5,373	6,373	8,763	4,096	5,058	8,763
Strive to get higher grades	Impact	2.98**	2.68**	2.15**	3.53*	3.41*	2.48**
	Robust p-value	0.05	0.05	0.04	0.07	0.05	0.04
	Non-Pantawid	88.71	88.92	89.05	88.46	88.68	88.99
	Number of obs.	4,937	5,933	8,756	4,196	5,193	8,756
Finish school work before playing or resting	Impact	5.79**	4.38**	2.39	5.58**	4.25**	2.76
	Robust p-value	0.01	0.03	0.12	0.02	0.04	0.12
	Non-Pantawid	71.62	72.17	73.60	71.83	72.38	73.52
	Number of obs.	3,650	4,537	8,748	4,152	5,110	8,748
Finish school work despite lack of time and resources	Impact	2.23	2.95	2.99*	2.60	3.44	3.45*
	Robust p-value	0.31	0.16	0.07	0.29	0.14	0.07
	Non-Pantawid	82.01	81.76	82.09	81.90	81.65	82.00
	Number of obs.	4,146	5,111	8,762	4,182	5,178	8,762
Grit index	Impact	0.13**	0.12**	0.08**	0.16**	0.15**	0.10**
	Robust p-value	0.01	0.01	0.04	0.02	0.01	0.04
	Non-Pantawid	3.30	3.31	3.34	3.29	3.30	3.34
	Number of obs.	4,592	5,605	8,776	4,163	5,132	8,776

RD = regression discontinuity; CER = coverage error rate; MSE = mean square error; obs. = observation
Notes: Treatment and control means are calculated using predicted values from replicating the rdrobust routine using least-squares regression. The p-value presented is from the robust version of the estimation that corrects for bias. P-values are rounded off to two decimal places. Asterisks reflect level of significance of the estimates: *** $p<0.01$, ** $p<0.05$, * $p<0.10$
Source: Authors' computations

Müller T, Shaikh M (2018). "Your retirement and my health behavior: Evidence on retirement externalities from a fuzzy regression discontinuity design" *Journal of Health Economics.* 57: 45-59.

CHAPTER BREAK

Subtitle

Background

- Retirement age reforms to secure viability of social security funds
- Reforms should account for adverse social and economic effects due to individuals retiring earlier or later
- Retirement → events that have interaction effects in the HH and thus affect behaviour of others
- Intra-household externalities - the behavior or characteristics of individuals directly affect the behaviour of others within the household

Background

- Aim of paper: Identify the causal effect of being retired (a characteristic of the individual) on the health behavior of the other individual (partner) in the household.
- Use a fuzzy regression discontinuity design - accounts for the endogeneity of the retirement decision by exploiting the legislation on retirement eligibility, which makes the probability of being in retirement a discontinuous function of age
- Contribution - assesses the different determinants of specific risky behaviors such as smoking, alcohol consumption and physical activity. Extend literature by analyzing the impact of spousal retirement on health status and risky behaviors

Data

- 4 waves (2004, 2006/2007, 2011/2012, 2013) of Survey of Health Ageing and Retirement in Europe (SHARE)
- Retirement – self-reported indicator about current job situation, equal to one if "retired" is selected as current job situation
- Inclusion - Individuals within a +-3 window of the official retirement age

Data

- Outcome variables – health behaviors
  - Physical activity – two types of physical activity, moderate or vigorous
    - Binary variables that take on a value of one if respondent reports “more than once a week” and zero if they report once a week or less
  - Alcohol consumption
  - Smoking
    - Currently smoking
    - Cigarettes per day
  - Subjective health status
    - Self-reported health status measured on a 5 point scale

Data

- Treatment variable
  - Forcing variable – age of individual (continuous) at the time of interview
  - Cutoff - Retirement eligibility threshold (eligibility for pensions)
  - Countries have different retirement eligibility thresholds
  - Some individuals retire earlier → a fuzzy regression discontinuity design is used which allows for a discrete increase in the retirement probability

Identification strategy: Fuzzy RDD

- Endogeneity in the partner's retirement status.
    - Omitted variable bias –grandparenting increase likelihood to retire, reduce smoking
    - Reverse causality –decision to retire may depend on health and subsequent health behavior.
    - Standard regression techniques lead to biased and inconsistent estimates.
- Imperfect compliance - Retirement eligibility does not necessarily imply that individuals are actually retired, retirement not mandatory
    - Discontinuity in the probability of retirement is smaller than 100% at the official retirement age

Design

- Partner's age is the forcing variable (X_i^p) that partially determines spousal retirement
- Use the discontinuity in the retirement probability as an instrumental variable for partner's retirement status.
- Apply 2SLS to estimate the parametric equations of the form:

$$Y_i = \alpha + \tau_1 D_i^p + \beta_1 \widetilde{X_i^P} + \beta_2 \widetilde{X_i^P} D_i^p + \tau_2 D_i + \beta_3 \widetilde{X}_i + \beta_4 \widetilde{X}_i D_i + \lambda_t + \Lambda_i + \epsilon_i$$

$$D_i^p = \gamma + \gamma_1 \tilde{X}_i^p + \delta_1 T_i^p + \delta_2 \tilde{X}_i^p T_i^p + \lambda_t + \Lambda_i + \nu_i$$

$$D_i = \gamma + \gamma_1 \tilde{X}_i + \delta_1 T_i + \delta_2 \tilde{X}_i T_i + \lambda_t + \Lambda_i + \mu_i$$

- Where:
  - Y is the indicator of the individual's health behavior
  - D_i^p is the indicator for partner's retirement status of individual i
  - D_i is the indicator of own retirement status
  - $\tilde{X}_i^P = (X_i^p - c)$ is individual i's partner age centered at the country specific retirement age
  - $\tilde{X}_i = (X_i - c)$ is the own age centered at the official retirement age.

Design

- Instrument for spousal and own retirement - respective retirement age threshold crossing indicators:

$$T_i^P = 1[\tilde{X}_i^p \geq 0] \text{ and } T_i = 1[\tilde{X}_i \geq 0]$$

  - which equal one if (partner's) age exceeds the official country specific retirement age.
- Effects of spousal and own retirement are captured by the parameters $\tau_1 and\ \tau_2$
- Endogenous treatment variables D_i^p and D_i linked to set of exogenous variables and the instruments.
- Estimated treatment effects are local average treatment effects, i.e., effect for those individuals who exit the labor market into retirement due to being eligible for old-age pensions, "retirement age compliers"

Checking validity of RDD: Discontinuity in the retirement rate

T. Müller, M. Shaikh / Journal of Health Economics 57 (2018) 45–59 51

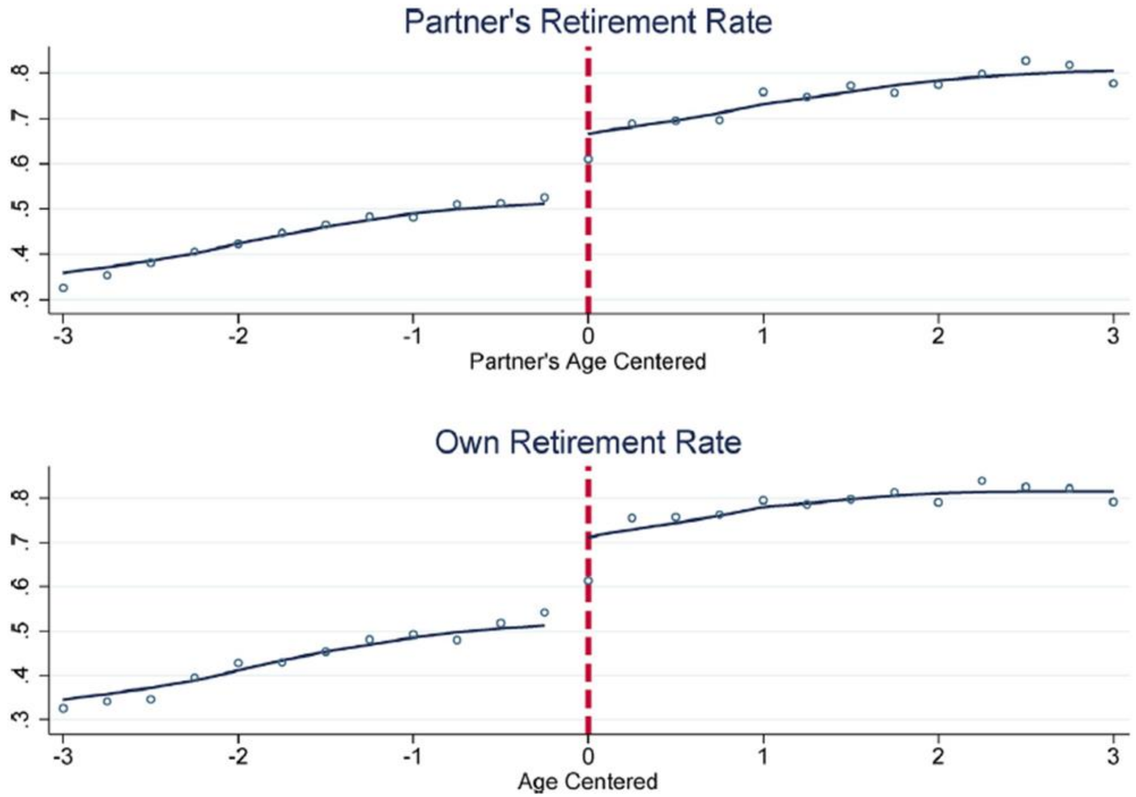
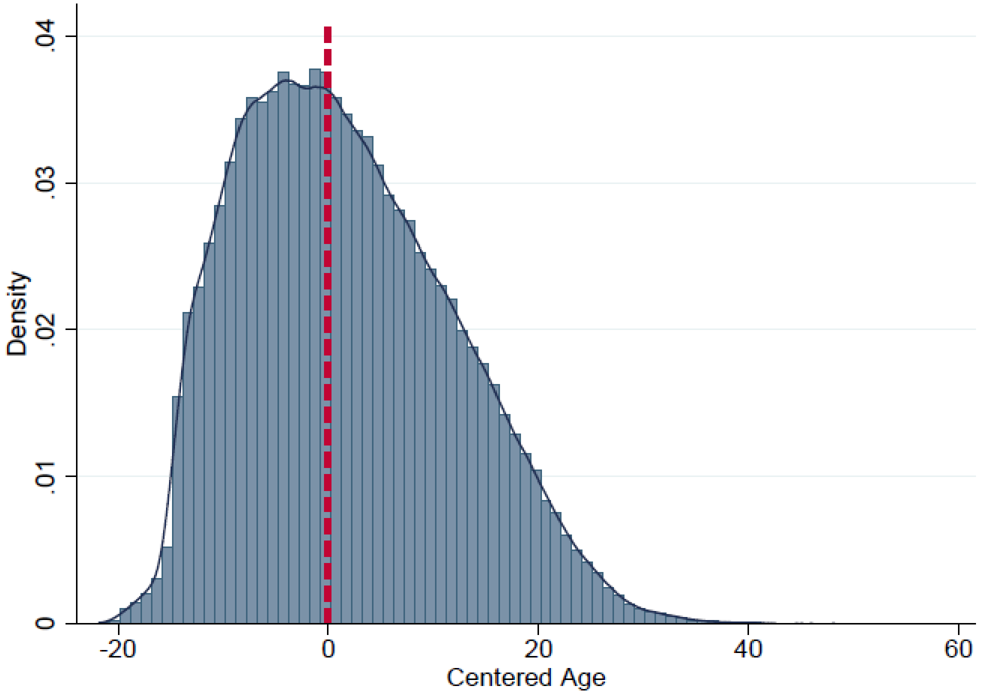


Fig. 1. First-stage effects: discontinuities in the retirement rates. *Note:* The figure shows the discontinuities in the retirement rates at the country-specific retirement ages. Partner's (own) age is centered at the retirement cutoffs. The scatters are overlaid with local polynomial smooths.

Validity checks

No manipulation assignment variable

Figure 2: Density of forcing variable

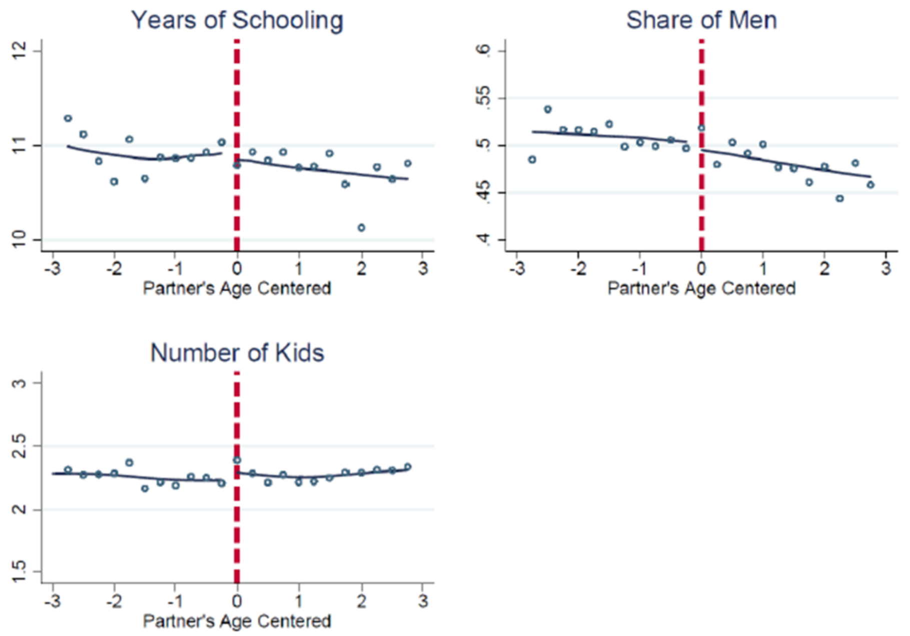


Notes: The figure shows the distribution of age around the official retirement cutoff. Age is centered at the country-specific retirement age. The dashed line indicates the retirement cutoff.

Validity checks

Balance of observed and unobserved characteristics around the the cutoff

Figure 3: Baseline covariates



Notes: The figure shows the reduced-form effects for the predetermined covariates education, gender and number of kids around the retirement cutoff (dashed lines). The scatters are overlaid with local polynomial smooths (dark blue lines). Partner's age is centered at the country-specific retirement age.

Table 3
First-stage effects.

First-stage effects		
Dependent variable: partner's retirement		
$I(\text{age}^p > \text{retirement age})$	0.24***	0.24***
	(0.01)	(0.01)
$I(\text{age}^p > \text{retirement age}) \times \tilde{age}^p$	–	–0.04***
		(0.00)
Dependent variable: own retirement		
$I(\text{age} > \text{retirement age})$	0.35***	0.34***
	(0.01)	(0.01)
$I(\text{age} > \text{retirement age}) \times \tilde{age}$	–	–0.03***
		(0.00)
Country and year FE	Yes	Yes
Flexible age function	No	Yes
Number of observations	23,598	23,598
Cragg–Donald Wald F-stats	437.03	201.65

Note: First-stage regression results using spousal and own retirement as dependent variables. The instrument used in specification 1 is a binary indicator for (partner's) age crossing the official retirement age; specification 2 then adds an interaction of that indicator with (partner's) age. An analogous first-stage is constructed for the interactions of (partner's) retirement with (partner's) age as in the second specification. Standard errors clustered at the individual level.
$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4
Retirement effects: physical activity and smoking.

Outcome variable	Moderate PA		Vigorous PA		Smoking		Cigarettes (including 0's)		Cigarettes (excluding 0's)	
Specification	RD (1)	RD (2)	RD (1)	RD (2)	RD (1)	RD (2)	RD (1)	RD (2)	RD (1)	RD (2)
Partner retired	−0.05*	−0.05*	−0.01	−0.00	0.01	0.02	−0.26	−0.31	6.57***	6.94***
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.51)	(0.54)	(2.45)	(2.62)
Retired	0.06***	0.06***	0.05*	0.06**	−0.00	−0.00	0.01	0.05	2.36	2.18
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.33)	(0.34)	(1.97)	(1.91)
Country and year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Flexible age function	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of observations	23,598	23,598	23,598	23,598	17,229	17,229	14,096	14,096	2785	2785

Note: Fuzzy RD estimates of the effects of partner's and own retirement for the outcomes moderate physical activity, vigorous physical activity, reporting to be a smoker and cigarettes smoked per day with one specification including the non-smokers and the other excluding them. The instruments for partner's and own retirement in model 1 are indicators for crossing the official country-specific retirement age; in model 2 interactions of the indicators with polynomials of centered age are added to the instrument set. Standard errors clustered at the individual level in parentheses.

* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Table 5
Retirement effects: alcohol consumption.

Outcome variable	Alcohol freq.		Drinks		Binge drinking		Drinking problem	
Specification	RD (1)	RD (2)	RD (1)	RD (2)	RD (1)	RD (2)	RD (1)	RD (2)
Partner retired	1.82***	1.64***	1.61***	1.58***	0.10***	0.10***	0.10***	0.10***
	(0.17)	(0.17)	(0.54)	(0.53)	(0.03)	(0.03)	(0.02)	(0.02)
Retired	1.04***	1.21***	0.65	0.50	0.05**	0.03	0.00	−0.01
	(0.13)	(0.13)	(0.41)	(0.35)	(0.02)	(0.02)	(0.02)	(0.02)
Country and year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Flexible age function	No	Yes	No	Yes	No	Yes	No	Yes
Number of observations	23,587	23,587	14,927	14,927	14,927	14,927	7744	7744

Note: Fuzzy RD estimates of the effects of partner's and own retirement on the frequency of alcohol consumption, the number of drinks consumed per day, an indicator for binge drinking (at least 5 drinks per day) and reporting to have a drinking problem. The instruments for partner's and own retirement in model 1 are indicators for crossing the official country-specific retirement age; in model 2 interactions of the indicators with polynomials of centered age are added to the instrument set. Standard errors clustered at the individual level in parentheses.

* $p < 0.1$.

** $p < 0.05$.

*** $p < 0.01$.

Table 6
Retirement effects: subjective health.

Retirement effects		
Dependent variable: subjective health		
Partner retired	−0.36***	−0.37***
	(0.08)	(0.08)
Retired	0.23***	0.24***
	(0.06)	(0.06)
Country and year FE	Yes	Yes
Flexible age function	No	Yes
Number of observations	21,854	21,854

Note: Fuzzy RD estimates of the effects of partner's and own retirement on subjective health. The instruments for partner's and own retirement in model 1 are indicators for crossing the official country-specific retirement age; in model 2 interactions of the indicators with polynomials of centered age are added to the instrument set. Standard errors clustered at the individual level.

* $p < 0.1$

** $p < 0.05$.

*** $p < 0.01$.

References

- Gertler P, Martinez S, Premand P, Rawlings L, Vermeersch C (2016). *Impact Evaluation in Practice, 2nd Edition*. Washington, DC: Inter-American Development Bank and the World Bank

- Glewwe P, Todd P. (2022) *Impact evaluation in international development: theory, methods, and practice:* World Bank Publications; 2022

- Khandker, S., G. Koolwal, and H Samad. (2010). *Handbook on Impact Evaluation: Quantitative Methods and Practices.* Washington, DC, The World Bank.

- Müller T, Shaikh M (2018). "Your retirement and my health behavior: Evidence on retirement externalities from a fuzzy regression discontinuity design" *Journal of Health Economics.* 57: 45-59.

- Orbeta, Aniceto, Jr. , Kris Ann Melad and Nina Victoria Araos (2023), "Reassessing the Impact of the Pantawid Pamilyang Pilipino Program: Results of the Third Wave Impact Evaluation" *PIDS Research Paper Series* 2023-06