Misreporting in Sensitive Health Behaviors and its Impact on Treatment Effects

An Application to Intimate Partner Violence

Jorge Agüero* Verónica Frisancho†

*University of Connecticut, Department of Economics and El Instituto (jorge.aguero@uconn.edu)
†Inter-American Development Bank, Research Department (vfrisancho@iadb.org)

August 11, 2017

Congreso de la Asociación Peruana de Economía

Motivation: Reporting Bias

- Much of the empirical work in economics relies on self-reported data
  - However, there could be reporting biases due to mistakes, limited attention, lack of recollection, behavioral biases, stigma, etc
  - More worrisome in the case of sensitive topics
- Non-classical measurement error in the dependent variable yields biased causal effects
  - Crucial in the case of risky behaviors since prevention and mitigation efforts are shaped by "risk factors"
- Several papers rely on administrative data to address misreporting
  - Bharadwaj, Pai, and Suziedelyte (2015), Bound, Brown, and Mathiowet (2001); Butler, Burkhauser, Mitchell, and Pincus (1987); Johnston, Propper, and Shields (2009); ONeill (2012)

Motivation: Intimate Partner Violence (IPV)

- Violence against women is a major public health problem
  - Growing number of studies trying to identify main drivers (e.g., Angelucci et al., 2008; Aizer, 2010; Hidrobo et al., 2016)
- Two features of IPV generate large potential for misreporting:
  - Invisible: takes place behind close doors
  - Perpetrator is known to the victim: this could increase the costs of exposing him (attachment, loss of economic support, retaliation, or stigmatization)
- WHO's gold standard ask direct questions about violent events
  - Main example: Demographic and Health Surveys (DHS/ENDES)
  - Despite great progress in protocols, risk of exposure persists
- Poor quality and coverage of IPV with administrative data.
  - Using 22 DHS, Palermo, Bleck, and Peterman (2014) find that only 7% of women who experienced violence made a formal report

This Paper

- Measures and characterizes the bias in direct reporting of IPV
  - We compare prevalence rates obtained from DHS and list experiments
  - This is done as a (cheaper) alternative in the absence of administrative data
  - We study the distortions introduced by misreporting in the estimation of causal effects

Literature Review

- Recent literature focusing on measurement error on sensitive questions:
  - Comparison to administrative records: earnings (Gottschalk and Huynh, 2010), body mass index (O'Neill, 2012), mental health (Bharadwaj et al., 2015), and plot size (Gourlay et al., 2017)
  - List experiments: Loan proceeds (Karlan and Zinman, 2012), illegal migration (McKenzie and Siegel, 2013), and LGBT population and anti-gay sentiment (Coffman et al., 2015)
  - Qualitative approaches: risky behavior (Blattman et al., 2016)

- In the case of IPV:
  - Administrative records are also biased due to exposure costs
  - Joseph et al. (2017) uses list experiments but with some limitations
  - Qualitative methods may not significantly provide more privacy to respondent

List Experiments: Design

- Control (C) and treatment (T) provided with a list of statements and requested to provide **number** of them that holds true
  - C gets S neutral statements; T gets S plus a *sensitive* statement
- Let $d_{is} = 1$ if statement s is true for individual i and 0 otherwise
- However, we only observe how many of those are true: $D_i = \sum_s d_{is}$
- Random assignment of the treatment at the individual level implies:

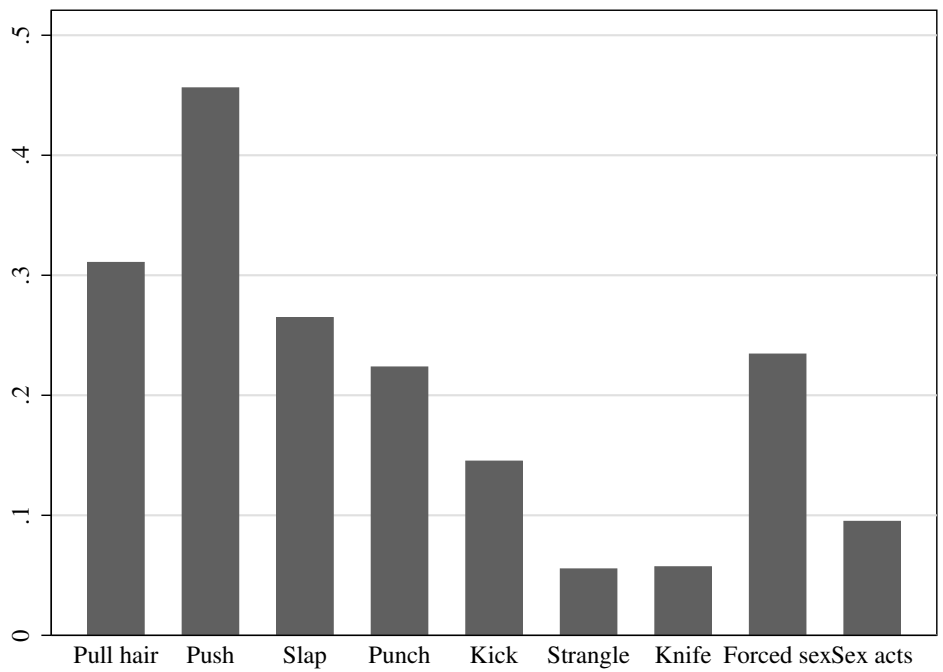
$$E_i\left(\sum_s^S d_{is}|T\right) = E_i\left(\sum_s^S d_{is}|C\right)$$

- Thus, prevalence rate of sensitive item is given by:

$$\rho = E_i\left[\left(\sum_s^{S+1} d_{is}|T\right) - \left(\sum_s^S d_{is}|C\right)\right]$$

Sample & Data

- Female clients of microfinance institution in poor districts in Lima
- 1223 interviews (1078 valid) between July 1st and August 25th, 2015
  - Large sample size allows us to have separate questionnaires for C and T
- High IPV prevalence rates as measured by DHS:
  - 62% ever experienced physical/sexual IVP



Implementation

- Questionnaire design and application considerations:
  1. Piloting of neutral statements
  2. Surveyors
     - Females with previous experience on gender/gender based violence topics
     - Sensitization and special training to conduct list experiments
     - Selection based on performance during training
  3. Visual aids Example
- Randomization at the individual level was successful Balance

Structure of the questionnaire

<table>
<tr><td colspan="2">Questionnaire</td></tr>
<tr><td>Control</td><td>Treatment</td></tr>
<tr><td colspan="2">Consent form and introduction
Demographics
Memory test
Direct questions about emotional violence</td></tr>
<tr><td>Direct questions about physical and sexual violence</td><td rowspan="2">List (5 statements) with indirect questions about physical and sexual violence</td></tr>
<tr><td>List (4 statements)</td></tr>
<tr><td colspan="2">Satisfaction with ADRA
End of questionnaire</td></tr>
</table>

Estimation

- Let prevalence rates reported under DHS methods be denoted by p
- Let $D_i = \sum_s^S d_{is}$ if $i \in C$ and $D_i = \sum_s^{S+1} d_{is}$ if $i \in T$
- If T_i denotes treatment assignment:

$$D_i = \alpha + \rho T_i + \xi_i$$

ρ measures prevalence under indirect methods and $(\rho - p)$ measures the bias

- We can also measure prevalence rates for different sub-samples:

$$D_i = \alpha + \rho T_i + \gamma x_i + \zeta T_i \cdot x_i + \xi_i$$

where $(\rho + \zeta)$ measures prevalence when $x_i = 1$

  - Comparison done with $(p|x_i)$ to capture bias

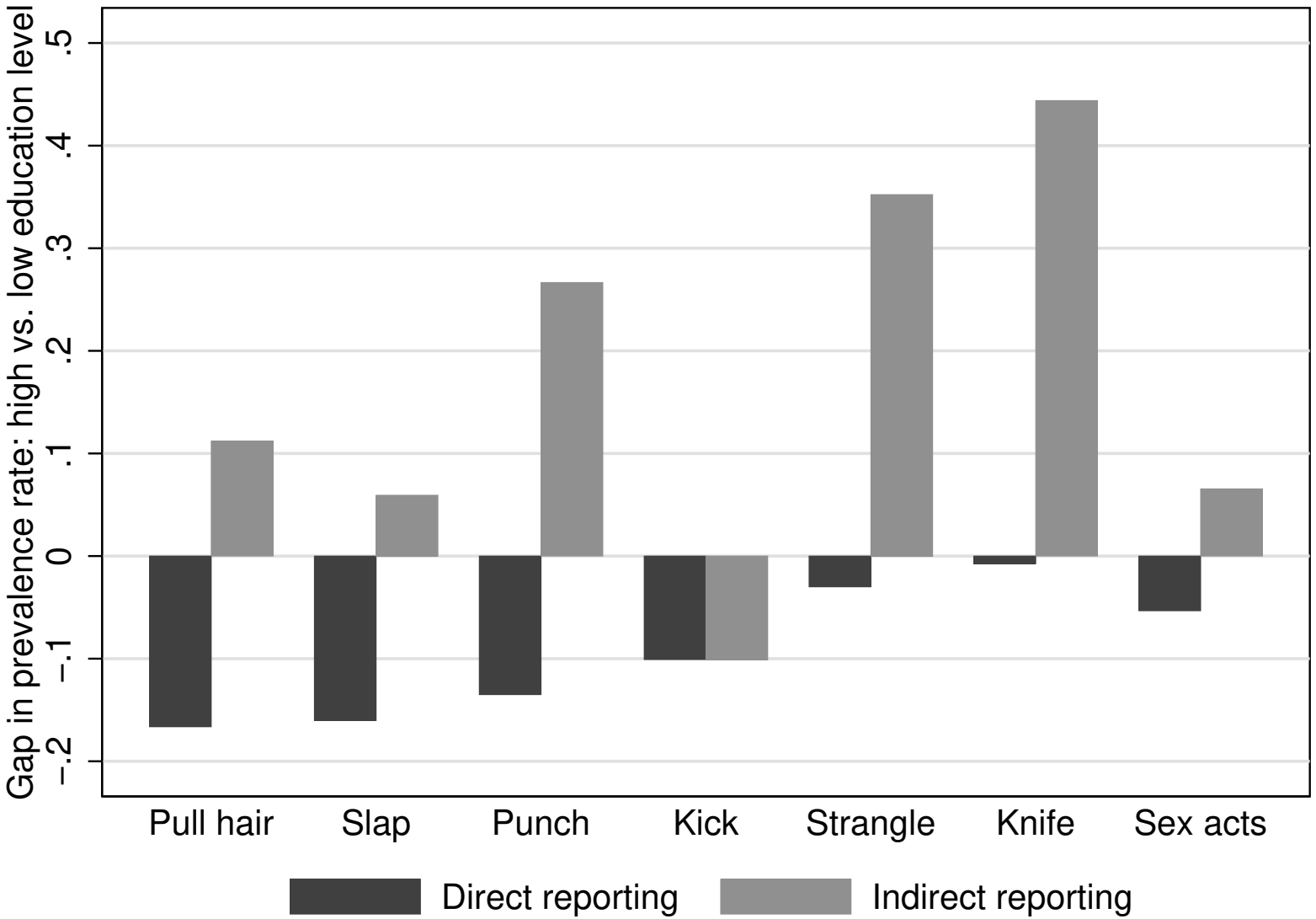
Results: Difference in prevalence rates of IPV

Violent act	List experiments (ρ)	Direct reporting (p)	$(\rho - p)$	
Pull hair	0.42	0.31	0.11	*
Slap	0.17	0.27	-0.09	
Punch	0.17	0.22	-0.05	
Kick	0.13	0.15	-0.02	
Strangle	-0.02	0.06	-0.08	
Knife	0.05	0.06	-0.01	
Sex acts	0.05	0.10	-0.04	
χ^2		8.12		
Prob $> \chi^2$		0.322		

Heterogenous effects: By education Level

	Less than college				College			
Violent act	List (ρ)	Direct (p)	($\rho - p$)		List (ρ)	Direct (p)	($\rho - p$)	
Pull hair	0.40	0.34	0.06		0.51	0.17	0.34	**
Slap	0.16	0.30	-0.13	*	0.22	0.13	0.09	
Punch	0.13	0.25	-0.12		0.40	0.11	0.28	*
Kick	0.14	0.16	-0.02		0.04	0.06	-0.02	
Strangle	-0.09	0.06	-0.15	**	0.27	0.03	0.24	*
Knife	-0.03	0.06	-0.09		0.41	0.05	0.36	***
Sex acts	0.04	0.10	-0.07		0.11	0.05	0.05	
χ^2		10.62				22.02		
Prob $> \chi^2$		0.156				0.003		

IPV and education gradient: by method



Robustness checks

- We do not find differences by other characteristics (e.g., age, marital status, employment, loan size and saving balance)
- Results are not driven by memory (balanced on memory test)
- Cannot be driven by "complexity" of list experiments: within education category, we compare 4 vs. 4+1 statements
- Asking directly about violent episodes to the control group did not affect results: balanced across satisfaction with ADRA (last questions of the survey)

How do our findings affect current literature?

- Consider a model like the following:

$$y_i = \beta x_i + \epsilon_i \qquad i = 1, \ldots, N, \text{where } \epsilon_i \sim N(0, 1)$$

...where y_i is measured with some noise:

$$\tilde{y}_i = y_i + \omega_i$$

- Let $x_i = \gamma \epsilon_i + \tau_i$
  - Endogeneity is present whenever $\gamma \neq 0$. Let $\tau_i \sim N(0, \kappa)$
- Measurement error includes a classical and a non-classical component:

$$\omega_i = \phi x_i + \nu_i$$

Causal Estimation: Endogeneity and Measurement Error

$$\begin{aligned}\hat{\beta}_{\text{OLS}} &= \beta + \frac{\text{cov}(\epsilon_i, x_i)}{\text{var}(x_i)} + \frac{\text{cov}(\omega_i, x_i)}{\text{var}(x_i)} \\ &= \beta + \gamma \frac{\text{var}(\epsilon_i)}{\text{var}(x_i)} + \phi\end{aligned}$$

- RCT and IV methods set $\gamma \frac{\text{var}(\epsilon_i)}{\text{var}(x_i)}$ to zero... but do not get rid of ϕ!

Measuring Bias due to Misreporting

- Remember that measurement error: $\omega_i = \phi x_i + \nu_i$
- List experiments allow us to directly measure ϕ
  - We can directly correct estimates of β obtained from RCTs or IVs
- Example:

	$\hat{\beta}$	Bias due to Misreporting
Pulled hair	-0.143	0.059
Attacked with a knife	0.009	0.063

Conclusions (I)

- We use experimental methods to measure reporting biases in IPV
- We are the first to measure the reporting bias relative to gold-standard (DHS)
  - On average, there are no significant differences in direct versus indirect reporting
  - Underreporting is concentrated among college educated women in our sample
  - Big enough to *reverse* the education gradient: more education $\rightarrow$ more violence under list experiments!

Conclusions (II)

- We also contribute to the literature on measurement error:
  - Even with random variation in x_i, non-classical measurement biases treatment effects
  - We propose list experiments as an inexpensive way to measure ϕ:
    - US$ 8 per survey, less if add-on module to instrument
    - Extremely useful for RCTs in the making
- Cheap(er) alternative when administrative data are not available

Misreporting in Sensitive Health Behaviors and its Impact on Treatment Effects

An Application to Intimate Partner Violence

Jorge Agüero* Verónica Frisancho†

*University of Connecticut, Department of Economics and El Instituto (jorge.aguero@uconn.edu)
†Inter-American Development Bank, Research Department (vfrisancho@iadb.org)

August 11, 2017

Congreso de la Asociación Peruana de Economía

Balance: Demographic Characteristics

Variable	Control	(T-C)	N
Age	43.825	0.903	1078
	[11.604]	[0.693]	
Married	0.798	-0.007	1078
	[0.402]	[0.025]	
Literate	1.959	0.002	1078
	[0.199]	[0.012]	
Spanish is not mother tongue	0.114	0.019	1078
	[0.318]	[0.020]	
Household head	0.313	0.07	1078
	[0.464]	[0.029]**	
Works	0.73	0.005	1078
	[0.444]	[0.027]	
Less than complete primary	0.109	0.017	1078
	[0.312]	[0.020]	
Primary education	0.266	-0.036	1078
	[0.442]	[0.026]	
Secondary education	0.45	-0.019	1078
	[0.498]	[0.030]	
Higher education	0.175	0.039	1078
	[0.380]	[0.024]	
Number of children	2.987	-0.013	1076
	[1.891]	[0.102]	
Number of children under 12 under her care	0.897	-0.025	1060
	[1.641]	[0.083]	
Memory test: % words remembered right after	0.85	0.026	1078
	[0.357]	[0.021]	
Memory test: % words remembered at the end	0.489	0.038	1078
	[0.500]	[0.030]	
Always lived in current locality	0.632	-0.028	1078
	[0.483]	[0.030]	

Significance levels (* 10%; ** 5%; *** 1%) captured through OLS estimation accounting for clustered (school) standard errors and cell fixed effects)

Balance: Financial Situation and Partner Characteristics

Variable	Control	(T-C)	N
Average loan size in past 4 cycles	1552.664	8.921	1025
	[1178.413]	[72.065]	
Average savings balance in past 4 cycles	791.688	77.259	1025
	[861.449]	[63.958]	
High loan size and high savings balance	0.284	0.038	1078
	[0.451]	[0.028]	
Jealous when speaking to other men	0.979	0.195	1077
	[7.224]	[0.488]	
Accuses her of being unfaithful	0.452	0.521	1078
	[4.196]	[0.420]	
Prevents her from visiting or being visited by friends	0.801	-0.203	1077
	[7.233]	[0.408]	
Limits contact with family	1.096	-0.511	1078
	[9.310]	[0.477]	
Wants to know where she is at all times	0.828	-0.34	1077
	[5.909]	[0.251]	
Does not trust her with money	0.428	0.374	1077
	[4.199]	[0.375]	
Humiliates her in public	0.555	0.018	1078
	[4.196]	[0.261]	
Calls her ignorant or idiot	0.538	0.37	1078
	[4.196]	[0.375]	
Calls her lazy, useless, or sleepy	0.45	0.006	1078
	[4.196]	[0.261]	
Threatened to harm her or someone close to her	0.512	-0.368	1078
	[5.913]	[0.250]	
Threatened to leave, take children, or cut off financial support	0.68	-0.362	1078
	[5.910]	[0.251]	

Significance levels (* 10%; ** 5%; *** 1%) captured through OLS estimation accounting for clustered (school) standard errors and cell fixed effects)

Balance: Survey Application

Variable	Control	(T-C)	N
Interruption by men	0.045	0	1078
	[0.207]	[0.013]	
Interruption by partner	0.007	-0.003	1078
	[0.084]	[0.004]	
Presence partner	0.018	-0.006	1078
	[0.133]	[0.007]	

Significance levels (* 10%; ** 5%; *** 1%) captured through OLS estimation accounting for clustered (school) standard errors and cell fixed effects)

back

Visual aids

35) ¿Cuántas **de las siguientes afirmaciones son correctas?**

Alguna vez….

- Ha comprado una TV con HD.
- Ha salido a pasear con sus hijos.
- Ha ayudado a sus hijos con sus tareas.
- Se ha comprado ropa muy cara.
- Su pareja le ha jalado el pelo.

( ) 1 ( ) 2 ( ) 3 ( ) 4 ( ) 5 ( ) NS/NR

back

Joint Significance of $(\rho - p)$

	χ^2	Prob $> \chi^2$
Age		
$<$50	4.124	0.765
50+	8.219	0.314
Civil status		
Single	13.436	0.062
Married	4.318	0.742
Education level		
Less than tertiary	10.617	0.156
Completed tertiary	22.018	0.003
Mother tongue		
Spanish	10.934	0.142
Other language	7.306	0.398
Memory test		
Low score	3.993	0.781
High score	6.598	0.472
Household head		
Not the head	8.781	0.269
Head	4.729	0.693
Employment		
Does not work	6.218	0.515
Works	6.481	0.485
Loan size in ADRA		
Low	16.087	0.024
High (p75+)	9.319	0.231
Savings balance in ADRA		
Low	12.842	0.076
High (p75+)	4.810	0.683

Simulated Bias in OLS estimates ($\phi = -1$)

- OLS may yield *less* biased estimates whenever the correlation between x_i and ϵ_i (γ) has opposite sign than the correlation between x_i and ω_i (ϕ)