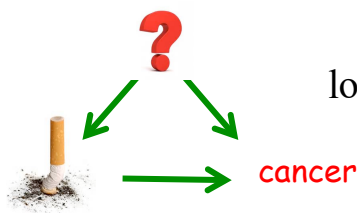


Propensity Score and Instrumental Variable

Guilherme J. M. Rosa
University of Wisconsin-Madison

Propensity Score

- **Propensity Score (PS):** Conditional probability of assignment to a particular category of the causal variable given the values of the confounder set (Rosenbaum and Rubin 1983)
- **Three different techniques:** Matched Samples, Stratification, and Regression



$$PS_i = \Pr(\text{smoke} \mid x_i) = p_i$$

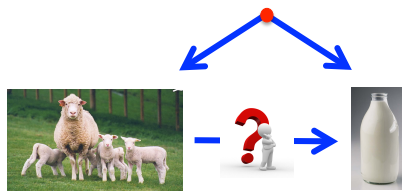
$$\text{logit}(p_i) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_m x_{mi}$$

⏟
confounders

Example

Inferring the Causal Effect of Number of Lambs Born on Milk Yield in Dairy Sheep

- Association between litter size (prolificacy) and milk yield (MY) has been shown in several species: mice (Skjervold 1976, Knight et al. 1986), rats (Yagil et al. 1976), pigs (Auldust 1998), goats (Heyden et al. 1978)



- Potential Confounders:
Age (parity)
Genetics
Year, Season, etc.

Data Description

- 4,319 lactation records collected (from 1997 to 2013)
- 1,534 crossbred dairy ewes from the Spooner Agricultural Research Station of the UW-Madison
- Average breed composition: mostly of dairy breeds (East Friesian = 45.2% and Lacaune = 27.6%), with smaller percentages for other breeds (Dorset, Polypay, Targhee, Romanov, Rideau, Kathadin, Rambouillet, Finnsheep, Texel, and Hampshire)
- Individual breed compositions were assigned into two explanatory covariates with three categories each: <50%, 50-75% and >75% EF or L
- 1st through the 6th lactation (1st, 2nd, 3rd+))
- Prolificacy (number of lambs born per parturition): single or multiple births

Statistical Analysis

- **Step 1:** Determine the "confounder set" of variables; i.e. those displaying simultaneous effect on both prolificacy and MY → lactation number, dairy breed proportions
- **Step 2:** Compute the PS for each observational unit. The distribution of PS can be compared between levels of the causal variable to check whether there are systematic differences between levels.
- **Step 3:** Apply a PS technique, such as Matched Samples (MS), i.e. ewes with "similar" background confounding effects (similar PS value) but under a different level of the causal variable. The criterion for similarity was defined by a caliper size (maximum distance between the PS values of each ewe) equal to 20% of one standard deviation of the logit of the PS (Austin and Mamdani 2006, Austin 2009)

Statistical Analysis

- **Step 4:** Perform a "balance check", i.e. check whether matching on the PS led to similar distribution of prolificacy across different levels of each of the k confounding variables (Ho et al., 2007)
- **Step 5:** Estimate target causal effect (e.g. using mean difference between outcome values within pairs)

Alternative: Bias-corrected matching estimator (Abadie and Imbens 2002, Guo and Fraser 2010), which uses LS regression to adjust for any remaining differences in PS within a matched pair. The correction is performed by modifying the value of one member in the pair as a function of the $E(Y_t | x_k)$, in which Y_t is the outcome given the causal variable t , and x_k are the confounding variables.

All five steps of the analysis performed using the R package "nonrandom" (Stampf 2014)

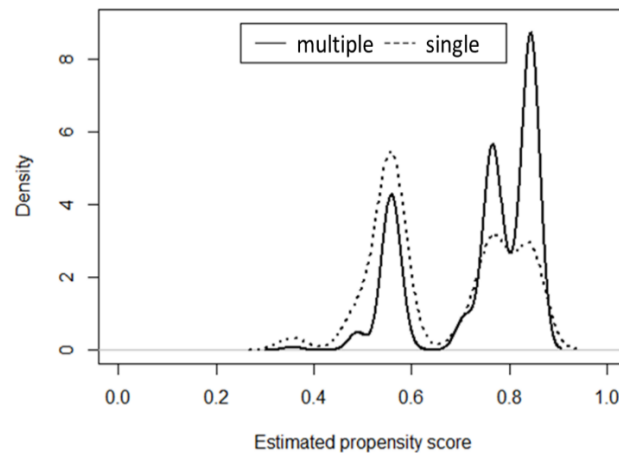
Results

Standardized differences for each confounder variable (Lactation, East Friesian proportion and Lacaune breed proportion) before and after matching, in percentage (%)

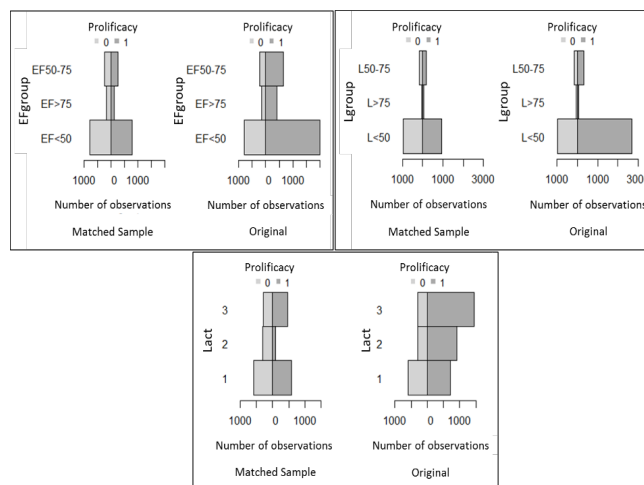
	Standardized differences – cutoff point 20	
	Before Matching	After Matching
Lactation	61.94	16.86 ^b
East Friesian proportion	4.51 ^b	8.54 ^b
<u>Lacaune</u> Breed proportion	15.12 ^b	5.10 ^b

^b balanced according to the cutoff of 20%

Distribution of estimated probabilities of assignment (propensity score) in each prolificacy level (single and multiple lamb birth) conditionally on the confounders (lactation number, and breed composition of EF and L)



Distribution of prolificacy for each confounder variable, EF and L breed composition (Efgroup and Lgroup), and Lactation (Lact), before and after matching. Prolificacy = 0 and 1 for single and multiple birth ewes, respectively.



Results

Estimated causal effect of prolificacy on MY using Propensity Scores with Matched Samples, as well as using marginal and partial regression of prolificacy on MY.

	Effect (L/lamb)	SE	95% CI
Simple Matching	20.52*	3.77	[13.13, 27.91]
Bias-corrected Matching	12.62*	3.63	[5.50, 19.74]
Marginal regression	43.93*	3.87	[36.34, 51.52]
Partial Linear regression	3.25	3.21	[-3.04, 9.56]

Ferreira VC, Valente BD, Thomas DL and Rosa GJM. Causal effect of prolificacy on milk yield in dairy sheep using propensity score. *Journal of Dairy Science* 100: 8443-8450, 2017.

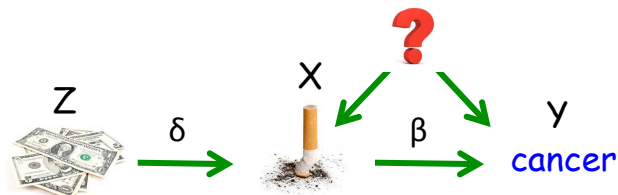
Instrumental Variable (IV)



$$\hat{\beta}_{\text{OLS}} = (X^T X)^{-1} X^T Y$$

$$\hat{\beta}_{\text{IV}} = (Z^T X)^{-1} Z^T Y$$

Instrumental Variable (IV)



Two-stage estimation:

1. Regress X on Z: $\hat{\delta} = (Z^T Z)^{-1} Z^T X$ and save predicted values $\hat{X} = Z\hat{\delta} = Z(Z^T Z)^{-1} Z^T X = P_Z X$
2. Regress Y on \hat{X} : $\hat{\beta}_{\text{2SLS}} = (X^T P_Z X)^{-1} X^T P_Z Y$

