Appendix 1. Propensity Score Matching for Maternal and Newborn Outcomes

Propensity models compared to standard models

	Nulliparas		Multiparas	
	aRR (95% CI) *Manuscript Method, full sample*	RR (95%CI) **PS-matched sample** 2:1 Matching: *Midwife n=1677; Obstetrician n=3138*	RR (95% CI) *Manuscript Method, full sample*	RR (95%CI) **PS-matched sample** 2:1 Matching *Midwife n= 2096 Obstetrician n=4086*
Type of Birth				
SVD	1.12 (1.09-1.16)	1.13 (1.10, 1.16)	1.02 (1.02-1.03)	1.03 (1.02, 1.03)
OVD	0.73 (0.57-0.93)	0.68 (0.55, 0.85)	0.30 (0.14-0.63)	0.27 (0.15, 0.49)
CS Delivery	0.68 (0.57-0.82)	0.62 (0.53, 0.72)	0.57 (0.36-0.89)	0.50 (0.31, 0.79)
Complications				
3rd or 4th degree laceration[†]	0.79 (0.58-1.07)	0.78 (0.59, 1.01)	* underpowered	
Neonatal complications				
Shoulder dystocia	* underpowered		1.51 (1.20, 1.91)	1.47 (1.15, 1.88)

Propensity scores were used as an alternate approach to control for confounding between the exposed (midwife) and control (obstetrician) groups. We created the propensity scores (PS) using the same covariates as in the main analyses (BMI, age, race, height, epidural and induction). These were selected a priori based on subject area knowledge and were associated with the exposure (midwife v physician) and outcomes of interest in bivariate models. The PS for this study was the predicted probability of being a midwife client based on the covariates in main analysis. Propensity scores were used to match controls (obstetrician cases) to treated cases (midwife group) with a greedy 2:1 matching (without replacement) and a caliper size of 0.01. We assessed for improvement in balance on covariates between the midwife and obstetrician groups after PS matching by assessing for a reduction in the standardized mean differences comparing the unmatched to PS-matched samples. After matching, all SMDs were <0.1 indicating samples were now balanced on these covariates. We estimated relative risks for the propensity matched sample using generalized models with robust standard errors (modified Poisson regression as with the main analyses). All models were run in SAS 9.4 after using the PSMATCH function to generate the matched sample.