EFFECTS OF PEER COUNSELING TO SUPPORT BREASTFEEDING:
ASSESSING THE EXTERNAL VALIDITY OF A RANDOMIZED FIELD EXPERIMENT

Onur Altindag
Theodore J. Joyce
Julie A. Reeder

Effects of Peer Counseling to Support Breastfeeding: Assessing the External Validity of a
Randomized Field Experiment
Onur Altindag, Theodore J. Joyce, and Julie A. Reeder
NBER Working Paper No. 21013
March 2015
JEL No. I12,I18

ABSTRACT

In an effort to improve breastfeeding, the Oregon WIC Program tested whether a relatively low-cost
telephone peer counseling initiative to support breastfeeding could increase the initiation and duration
of exclusive breastfeeding among its participants. They conducted a large randomized field experiment
(RFE) with over 1900 women from four WIC agencies in the state. In this study we use data from
the RFE along with administrative data from the rest of the state to assess whether the results from
the RFE can be extended to other agencies in the state. We find small or non-existent effects of peer
counseling in the non-experimental settings, which suggest that the experimental estimates may reflect
Hawthorne effects. We present evidence of selection into RFE in that exclusive breastfeeding among
the controls is significantly greater than among women who were offered but declined to participate
in the RFE as well as from women in the rest of the state who had no access to peer counseling. We
conclude that despite the strong internal validity of the RFE, extending the program to other agencies
in the state would have a limited impact at best on exclusive breastfeeding.

Onur Altindag
NBER
5 Hanover Square, 1602
NY, NY 10004
raltindag@gradcenter.cuny.edu

Theodore J. Joyce
NBER
5 Hanover Square, 16th Floor
Suite 1602
New York, NY 10004-2630
theodore.joyce@baruch.cuny.edu

Julie A. Reeder
Oregon WIC Program
800 NE Oregon Street
Suite 865
Portland, OR 97232-2187
julie.a.reeder@state.or.us

Abstract

A long-standing tension within the Special Supplemental Nutrition Program for Women, Infants and Children or WIC is promotion of breastfeeding while making available free infant formula. The WIC program purchases over half of all infant formula sold in the US and unsurprisingly, breastfeeding initiation and duration is substantially lower among WIC participants than their eligible non-participants. In an effort to improve breastfeeding, the Oregon WIC Program tested whether a relatively low-cost telephone peer counseling initiative to support breastfeeding could increase the initiation and duration of exclusive breastfeeding among its participants. They conducted a large randomized field experiment (RFE) with over 1900 women from four WIC agencies in the state. They found significant increases in exclusive breastfeeding among Spanish- but not English-speaking clients. In this study we use data from the RFE along with administrative data from the rest of the state to assess whether the results from the RFE can be extended to other agencies in the state. We use randomization as an instrument to estimate the effect of treatment on the treated and we compare these to non-experimental estimates of the effect of peer counseling from the same or similar WIC agencies as the RFE. We find small or non-existent effects of peer counseling in the non-experimental settings, which suggest that the experimental estimates may reflect Hawthorne effects. We present evidence of selection into RFE in that exclusive breastfeeding among the controls is significantly greater than among women who were offered but declined to participate in the RFE as well as from women in the rest of the state who had no access to peer counseling. We conclude that despite the strong internal validity of the RFE, extending the program to other agencies in the state would have a limited impact at best on exclusive breastfeeding.

1 Introduction

The enormous growth in randomized field experiments (RFEs) in the social sciences over the past 10 years has reignited a vigorous debate as to their utility to inform policy.[1] A primary criticism is that the generalizability of even well-designed RFEs with strong internal validity is limited by the narrowness of the questions and the specific circumstances of the studies. This concern is longstanding, but the renewed debate has pushed researchers involved in RFEs to discuss in more detail the generalizability of their results.[2]

In this study, we explore the external validity of the results from a large RFE of peer counseling to promote breastfeeding among women in Oregon enrolled in the Supplemental Nutrition Program for Women, Infants and Children (WIC). Specifically, we ask whether the results of a peer counseling RFE conducted at 4 WIC agencies would obtain in the other 30 agencies within the state. The question has meaningful policy implications as the U.S. Surgeon General has proposed making peer counseling to support breastfeeding a core WIC service. The motivation for expanding peer counseling emanates from a long-standing tension within WIC, which promotes breastfeeding while making available free infant formula. WIC, for example, purchases over 57 percent of all infant formula sold in the US (Oliveira et al., 2010). Not surprisingly, breastfeeding initiation and duration is substantially lower among WIC participants than their eligible non-participants (Jacknowitz et al., 2007). Closing this gap is core to WIC's mission given numerous reports that exclusive breastfeeding improves infant and maternal health.[3]

To promote breastfeeding among WIC participants, the United States Department of Agri-

---

[1]See for example Campbell et al. (1963); Rodrik (2008); Banerjee and Duflo (2008); Angrist and Pischke (2008); Deaton (2010); Imbens (2010); Ravallion (2012).

[2]As Campbell et al. (1963) wrote over 60 years ago, "While internal validity is the *sine qua non*, and while the question of external validity, like the question of inductive inference, is never completely answerable, the selection of designs strong in both types of validity is obviously ideal." (page 5). See also Kramer and Shapiro (1984) and Rothwell (2005) for discussion in the medical literature and Heckman and Smith (1995) and Teele (2014) for discussions in the social sciences.

[3]See for example Kramer et al. (2001); Kramer and Kakuma (2004); Bachrach et al. (2003); Ip et al. (2009); Eidelman et al. (2012).

cultures Food and Nutrition Service (FNS) initiated the promotion of peer counseling in 2004 (McLaughlin et al., 2004). The essential idea is that peers, in this case women who have been on WIC and have successfully breastfed, provide support and guidance for women who intend to or are trying to breastfeed. However, evidence on the effectiveness of peer counseling is mixed. Numerous observational studies have reported increases in breastfeeding initiation and duration associated with peer counseling programs for WIC clients.[4] However, a systematic literature review of peer counseling initiatives has characterized many of these observational studies as of moderate to poor quality (Ingram et al., 2010). The most convincing evidence as to the effectiveness of peer counseling among low-income women has come from three well-executed randomized control trials (RCTs) in the United States (Chapman et al., 2004; Anderson et al., 2005; Bonuck et al., 2005). Yet the intensity of the intervention in each of the three RCTs far exceeds what is currently offered in most local WIC agencies, which raises doubts as to the scalability of such support (Reeder et al., 2014).

To test whether peer counseling could prove effective if delivered in a less resource intensive manner, researchers from the Oregon WIC program undertook the largest randomized field experiment (RFE) to date in the US. The objective was to assess whether a telephone peer counseling program among WIC clients could increase the initiation and duration of exclusive breastfeeding (Reeder et al., 2014). One thousand, nine hundred and forty-eight WIC clients in four WIC agencies were randomly assigned among three study arms and stratified by whether they spoke English or Spanish. Researchers found that non-exclusive breastfeeding for at least 3 months increased by 8 percentage points among English-speaking women and 17 percentage points among Spanish-speaking participants while exclusive breastfeeding of the same duration increased by 8 percentage points, but only among the Spanish-speakers.

In this study, we extend Oregons RFE of peer counseling in several ways. First, Reeder et al. (2014) estimated the effect of being assigned to the treatment group or the intention-

---

[4]See for example Grummer-Strawn and Mei (2004); Gross et al. (2009); Yun et al. (2010); Schafer et al. (1998); Shaw and Kaczorowski (1999); Bolton et al. (2009); Gill et al. (2007); Olson et al. (2010).

to-treat (ITT). However, 6 percent of women in the treatment group never respond to any calls by the peer counselors and 29 percent of English-speaking clients never interacted with a peer counselor in the postpartum period. Thus, we use the randomization indicator as an instrument to estimate the effect of treatment on the treated (TOT). The TOT results provide the most comparable estimates to those from non-experimental settings that compare women who received peer counseling relative to those who declined. A second extension of the analysis by Reeder et al. (2014) is a test for heterogeneous treatment effects based on a recent approach to endogeneous stratification (Abadie et al., 2013).

As a third extension, we explore the external validity of the results from Oregons peer counseling RFE aided by a unique set of circumstances. Specifically, we have breastfeeding data on those who chose not to participate in the RFE as well as data on WIC clients in counties without peer counseling. Moreover, the same peer counseling program implemented in the experimental agencies was also offered in two other agencies over the same time period as the RFE but without randomization. Finally, three of the four experimental agencies continued to offer peer counseling to WIC clients without randomization for three years after the RFE was over. Thus, we estimate the effect of peer counseling on breastfeeding with observational data in the same and similar sites as the RFE and compare it to the instrumented estimates of the treatment on the treated (TOT) from the experimental sample. Importantly, data for the RFE were collected by the state-wide administrative system in same manner as data from the non-experimental settings, an important feature of a within-study design (Heckman and Smith, 1995).

To preview key findings, we uncover numerous sources of selection and possible Hawthorne effects that limit generalizability of the RFE. For example, English-speaking participants randomized into the control group had significantly higher rates of exclusive breastfeeding than women from the same counties who chose not to participate—evidence of positive selection into the study. Moreover, both English- and Spanish-speaking women in the control group had higher rates of exclusive breastfeeding than women in the rest of the state who had

no access to peer counseling. Second, we show that non-experimental estimates of treatment effects are statistically zero or substantially weaker than treatment effects obtained from the RFE. The diminished effect appears unrelated to self-selection into peer counseling, which suggests that the supervision of peer counselors and the attention paid to WIC participants in an experimental setting may be difficult to sustain in the routine provision of peer counseling services. We conclude that a relatively low-cost telephone peer counseling program broadly applied would have a limited impact at best if offered as a standard service of WIC.

2 Background

2.1 Peer Counseling

In the typical model a peer counselor meets with the expectant mother during pregnancy, and then up to six months or a year after birth. The goal is to promote exclusive breast-feeding for at least six months, which the literature suggests is necessary to reap the full benefits of breastfeeding (American Academy of Pediatrics, 2012). Several observational studies evaluated the effect of peer counseling among WIC clients in agencies with and without a peer counseling program. Each reported gains in breastfeeding initiation in local WIC agencies with peer counseling but sample sizes were small and research designs weak (Grummer-Strawn and Mei, 2004; Schafer et al., 1998; Shaw and Kaczorowski, 1999; Gill et al., 2007; Bolton et al., 2009). Larger studies of WIC clients—18,789 in Maryland and 29,881 from Missouri—were able to adjust estimated program effects with a sizeable number of covariates (Gross et al., 2009; Yun et al., 2010). However, in both studies there were important differences by race and ethnicity between those exposed and unexposed to peer counseling, and estimates of program effects were sensitive to adjustment. In the strongest observational study, researchers used as the comparison group WIC clients who requested peer counseling but were denied because of a lack of counselors (Olson et al., 2010). Peer counseling was associated with a 7.4 percentage point increase in non-exclusive breastfeeding for at least 6 months relative to the mean of 10.4 percent among those in the comparison

group.

Three high-quality RCTs of peer counseling among low-income women in the U.S. reported significant gains in breastfeeding initiation and duration (Chapman et al., 2004; Anderson et al., 2005; Bonuck et al., 2005). Each involved prenatal and postpartum home and hospital visitation as well as telephone follow-up as needed. In one study, professional lactation consultants were used instead of peer counselors (Bonuck et al., 2005). They reported significant differences in non-exclusive breastfeeding up to six months postpartum, but no difference in exclusive breastfeeding of any duration. In another RCT, new mothers received at least one daily visit by a peer counselor while in the hospital and at least three home visits postpartum (Chapman et al., 2004). They found no differences in exclusive breastfeeding at any point postpartum but women in the treatment group were less likely not to breastfeed at one and three months relative to the controls. In a follow-up study to test whether more intensive counseling might improve exclusive breastfeeding, women in the treatment groups were offered three prenatal and nine postpartum visits in addition to daily hospital visits by a peer counselor (Anderson et al., 2005). After three months, the risk of non-exclusive breastfeeding was higher among the controls than among those in the intervention group (RR=1.30, $p > .05$).[5]

Common characteristics of the three RCTs are sample size, between 50-200 women in each experimental arm of the study, and their large Hispanic populations. Although most appeared powered to detect differences in breastfeeding initiation and duration, only one

---

[5]There have been numerous high-quality RCTs of peer counseling (PC) for breastfeeding conducted outside the US. Researchers in England used telephone or hospital visitation to reach clients but reported no impact of PC on initiation or duration of breastfeeding (Graffy et al., 2004). A peer counseling intervention similar to the RCT in our study was conducted in Canada in the late 1990s (MacArthur et al., 2009). New and expectant mothers were contacted by telephone within 48 hours of delivery and for as many times in the next three months as seemed necessary. The authors reported impressive gains in exclusive breastfeeding at 12 weeks: 56% of women in the treatment group exclusively breastfed as compared to 40.3% among controls. However the women in the study included upper and middle-income women with substantial education. The generalizability to a low-income population is not clear. More generally, a recent review of prenatal peer support for breastfeeding concluded that universal PC support was not effective, but that PC targeted at women who are considering breastfeeding did increase breastfeeding initiation rates (Ingram et al., 2010).

had sufficient power to detect less than large differences in exclusive breastfeeding (Anderson et al., 2005). A second characteristic of almost all RCTs is the provision of in-home and in-hospital visits in both the prenatal and postpartum periods. For example, Anderson et al. (2005) reports increases in exclusive breastfeeding as a result of three prenatal and daily hospital visits, followed by nine post-partum visits. The support provided by the current peer counseling programs funded by the USDA does not come close to the level of service provided in the three RCTs. Scaling up the peer support provided in the RCTs to the national level would appear unrealistic in the current fiscal environment.

2.2   Oregon's Experiment

An important motivation for the RFE conducted in Oregon was to assess whether a relatively low-cost peer counseling program in which support was provided almost exclusively by telephone could achieve substantial gains in breastfeeding. As detailed in Reeder et al. (2014), the intention-to-treat (ITT) results were mixed. The probability that women assigned to the treatment group breastfed non-exclusively for at least 3 months was 22 percent greater than women in the control group and 30 percent greater among Spanish speakers. Non-exclusive breastfeeding for at least 6 months increased by 14 percentage points among Spanish-speaking women in the treatment group relative to an overall non-exclusive breastfeeding prevalence of 45% among Spanish-speaking controls.

Peer counseling was also associated with decreases in any and exclusive breastfeeding cessation, but these gains also were limited to Spanish-speakers only. The findings for Spanish-speaking WIC participants are broadly consistent with the three RCTs in the US that also evaluated peer counseling interventions. In two of the RCTs 80 percent of participants were Hispanic with approximately half designating Spanish as their preferred language and at least 70 percent participated in WIC (Chapman et al., 2004; Bonuck et al., 2005). The third RCT had a smaller proportion of Hispanics (Anderson et al., 2005). In the Oregon RFE as well as the other three RCTs not all women in the treatment group take-up the peer

counseling or use peer counselors to the same degree. Thus, the effect of peer counseling on women who partake of treatment is unclear due to the obvious selection involved with compliance. In this study we extend the results from the Oregon RFE by using randomization as an instrument to estimate the effect of peer counseling among women who interact with the counselors. To the extent that randomization satisfies the exclusion restriction and given one-sided compliance, we can interpret the instrumented estimates as the effect of treatment on the treated (TOT). [6] We then compare the TOT estimate from the RFE to estimates of average treatment effects obtained in the non-experimental setting.

3   Empirical Framework

3.1   Data

The study is based on two samples of WIC clients in Oregon during the period of July 2005 through June 2010. The analysis is anchored by the results from the RFE conducted at four WIC agencies between July 2005-June 2007. There are 34 local WIC agencies in Oregon that, with few exceptions, are organized at the county level.[7] A description of the RFE and the intention-to-treat (ITT) results are reported in Reeder et al. (2014). The second sample, henceforth the non-experimental sample, consists of all WIC clients not in the RFE from June 2005 to July of 2010. All data pertaining to the characteristics of women on WIC and their breastfeeding outcomes are entered by staff at WIC agencies to the State's centralized Information System Tracker database, TWIST.

3.1.1   Sample 1: The Randomized Field Experiment

Nineteen hundred and forty-eight English or Spanish speaking women attending a new pregnant appointment for WIC between July 2005 and July 2007 at one of the four WIC agencies

---

[6] With one-sided compliance there are no "always takers" (Angrist et al., 1996; Angrist and Pischke, 2008). There were no always takers in the Oregon RFE as peer counselors were only in contact with those in the treatment group.

[7] Oregon has 36 counties. Several rural counties are serviced by one local WIC agency.

in the State consented to be in the study. The four agencies were the counties of Hood River, Jackson, Umatilla and Washington. The counties served by each agency are show in blue in Figure 1. Sixty-four women miscarried or moved out of the state leaving 1,884 women that were assigned to one of the three study arms. Women were stratified between English and Spanish speakers and then randomized in the three treatment arms. The control group received standard WIC Program breastfeeding promotion and support but did not have any contact with a peer counselor. The low-frequency treatment group was eligible to receive four planned, peer-initiated contacts: the first after the initial prenatal assignment; the second two weeks before the expected due date; the third within one-week postpartum; and the fourth approximately 2 weeks postpartum. The high-frequency treatment group was eligible to receive eight planned peer-initiated contacts. The first four contacts were the same as the low-frequency group with the additional four occurring at months 1-4 postpartum. There were no meaningful differences in the breastfeeding outcomes between women in the low and high intensity groups so following Reeder et al. (2014), we have combined them in the analyses that follow.

The peer counselors were current or former WIC clients within the past five years. They had to have breastfed an infant for at least 6 months, have sufficient literacy in English or Spanish to complete the paper work, and be able to commit at least 10 hours a week to counseling. All peer counselors received a state-provided three-day training grounded in the USDA's *Loving Support* curriculum at the central administrative office in Portland. Afterwards, the peer counselors returned to their respective WIC agencies for further orientation and training.[8].

3.1.2   Sample 2: Non-Experimental Data on Peer Counseling

The same peer counseling program as in the RFE was offered in two WIC agencies located in Marion and Deschutes counties from 2005-2010. Marion and Deschutes counties are the

---

[8]See Reeder et al. (2014) for more details regarding the RFE.

5th and 7th most populous counties in the state (Figure 1). Jackson and Washington counties in the RFE are ranked 6th and 2nd, respectively. Women attending their new pregnant appointment at the WIC agencies in Marion and Deschutes were offered the opportunity to receive the same high-frequency peer counseling support as offered to women in the RFE. The important difference from the RFE was that any WIC client who desired peer counseling was enrolled. Of the 12,348 women enrolled in WIC in Marion and Deschutes between July 2005 and July 2010, 1,774 (14.2%) requested peer counseling support.

Another group of WIC clients were also offered peer counseling in a non-experimental setting. Three of the four agencies that participated in the RFE, Jackson, Umatilla and Washington continued to offer peer counseling to women enrolling in WIC at their new pregnancy visit from the end of the RFE in July of 2007 through July, 2010. Thus, 3,577 WIC clients out of 13,094 eligible women (27.3%) were enrolled in a voluntary peer counseling program between July 2007 and June 2010. Lastly, we have data on 24,857 WIC clients from the agencies in Oregon that did not offer or provide peer counseling for breastfeeding between July 2007 and July 2010.

3.1.3 Outcomes and Covariates

We focus on exclusive breastfeeding at one, three and six months postpartum. At each certification visit up to two years, mothers were asked how they were feeding their baby. Duration of exclusive breastfeeding was derived from the first time that the mother reported to WIC that she had stopped breastfeeding or introduced formula and the timing of each. Exclusive breastfeeding duration was recorded in weekly intervals for the first month and then at intervals of 5, 9, 13, 18, 22, 26, 31, 35, 39, 43, 47, 52 weeks and more than 52 weeks.

We focus on exclusive breastfeeding because the explicit goal of the peer counseling initiative was to increase the prevalence of exclusive breastfeeding for at least six months. The

health benefits of breastfeeding have been related to exclusive breastfeeding and not just any breastfeeding (American Academy of Pediatrics, 2012). In addition, exclusive breastfeeding is reported more completely than non-exclusive breastfeeding. The duration of non-exclusive breastfeeding cannot be determined until a woman stops breastfeeding completely. As a result, non-exclusive breastfeeding duration was missing for women who reported breastfeeding at their last WIC re-certification visit but then left WIC before their next scheduled re-certification appointment. Exclusive breastfeeding duration tends to end earlier in the postpartum period and thus, is measured more completely. In the Oregon RFE, non-exclusive breastfeeding was missing in 19 percent of cases whereas exclusive breastfeeding was missing for only 8 percent (Reeder et al., 2014).

The administrative data system, TWIST, also has information on a WIC client's age, educational attainment, family income, marital status, race/ethnicity, spoken language as well as month of enrollment in WIC.

3.2 Analysis

3.2.1 Analysis of the RFE

We first present ITT estimates of offers of peer counseling on dichotomous indicators of exclusive breastfeeding at least one, three and six months in the RFE.

$$BF_i = \beta_0 + \beta_1 Z_i + X_i'\theta + \varepsilon_i \tag{1}$$

where BF_i a breastfeeding outcome for person i, Z is a dummy variable and equals 1 if the woman is assigned to the peer counseling treatment group, X is a vector of baseline controls (age, age squared, race/ethnicity, education, marital status, the month of pregnancy when certified for WIC, natural logarithm of monthly income and WIC agency indicators), and ε denotes the random component. Based on a linear probability model, the estimated coefficient, β_1, provides an unbiased estimate of the intention-to-treat (ITT). All analyses are stratified by whether the counseling was conducted in English or Spanish.

One way in which we extend the work of Reeder et al. (2014) is by estimating the effect of peer counseling among women who actually received support. We define treatment as women who received at least one peer counseling call in the postpartum period. As such, 71 percent of English-speaking women and 89 percent of Spanish-speaking women assigned to the peer counseling group actually received treatment. Because interactions with a peer counselor is a choice, we instrument the receipt of peer counseling with the randomization indicator. The first stage is as follows:

$$T_i = \pi_0 + \pi_1 Z_i + X_i'\Gamma + v_i \tag{2}$$

where T equals one if the woman received any postpartum PC service and zero otherwise where π_1 captures the effect of assignment to treatment group on receiving peer counseling service. We use the predicted value from equation (2), $\widehat{T}_i$ to estimate the following second stage:

$$BF_i = \gamma_0 + \gamma_1 \widehat{T}_i + X_i'\xi + \eta_i \tag{3}$$

where the coefficient, γ_1, is the estimate of TOT. Given one-sided compliance, this represents the effect of peer counseling among women who interacted with counselors (Bloom, 1984; Angrist and Pischke, 2008).[9]

Another extension of Reeder et al. (2014) is a test of heterogeneous treatment effects. A common approach is to interact the assignment to treatment indicator Z with baseline characteristics. Such a tactic is somewhat *ad hoc* and can lead to specification searches. A recent alternative uses the entire set of pre-treatment covariates to predict outcomes among the control group (Abadie et al., 2013). The estimated parameters from this regression are applied to the entire sample in the RFE and treatment effects are estimated within different quantiles of the predicted outcome. To avoid over-fitting we use the repeated split sample

---

[9]One-sided compliance means women in the control group had no access to peer counseling. In lexicon of Angrist et al. (1996), there were no always takers.

(RSS) algorithm suggested by the authors.[10]

3.2.2 External Validity

We begin the assessment of external validity[11] by generating bounds on the treatment effects following Manski (1999, 2013). We first assume no knowledge of the treatment effects for the unobserved counterfactuals (Manski, 1999).[12]. We then narrow the bounds by imposing various restrictions (Manski, 2013). Although the range of treatment effects remains relatively wide, they provide a transparent set of goal posts with which to compare estimates of treatment effects from subsequent exercises.

In the next part of the study we use WIC administrative data for women not-involved in the RFE. Following the within-study literature, we compare the breastfeeding outcomes of the randomized out controls to women who were offered but chose not to participate in the RFE (LaLonde, 1986; Smith and Todd, 2005). Specifically,

$$BF_i = \rho_0 + \rho_1 C_i + X_i'\Psi + \eta_i \tag{4}$$

where C_i equals one if the woman was in the control group of the RFE and zero if she did not participate. The coefficient, ρ_1, captures differences in breastfeeding outcomes between

---

[10] A Stata version of the algorithm by Jeremy Ferwerds is available at `https://ideas.repec.org/c/boc/bocode/s457801.html` (last accessed December 31, 2014).

[11] External validity concerns inferences about the extent to which a causal relationship holds over variation in persons, settings, treatments and outcomes (Shadish et al., 2002). The broadness of the definition explains why convincing demonstrations of external validity have proven elusive. One approach to improving the generalizability of results from RFEs is more experiments addressing similar questions but in different settings (Banerjee and Duflo, 2008; Imbens, 2010). Even with multiple experiments, however, selective participation by subjects, heterogeneous treatment effects, randomization bias and general equilibrium effects remain challenges to external validity (Heckman and Smith, 1995; Rodrik, 2008; Deaton, 2010). The recent focus on external validity has motivated approaches to testing for heterogeneous treatment effects across multiple sites (Crump et al., 2008), but the generalizability of a single RFE remains more art than science (Woolcock, 2013).

[12] Let $E[Y_1|D = 0]$ be the expected exclusive breastfeeding outcome of being offered peer counseling on those in the treatment group had they been assigned to the controls and let $E[Y_0|D = 1]$ be the expected outcome of the controls had they been assigned to the treatment group. The no information bounds assume $E[Y_1|D = 0] = 1$ and $E[Y_0|D = 1] = 0$ for the upper bound and the reverse for the lower bound.

the controls and non-participants in the absence of treatment. The sign and statistical significance of, ρ_1, provides evidence on selection into the RFE. We extend this exercise to the other WIC agencies that were not part of the RFE. We compare the probability of exclusive breastfeeding between the controls from the RFE and women who either had no access to peer counseling or who chose not to use the service when available.

In last section we use equation (1) to estimate the effects of peer counseling on breastfeeding outcomes in a non-experimental setting. We rely on statistical controls rather than the randomization to mitigate selection bias. Comparing these to the experimental estimates provides several possible insights. For instance, by estimating non-experimental effects in the same agencies as the RFE, we largely hold differences in staff or the population of WIC clients constant. Therefore, if we find that the non-experimental estimates are smaller than those from the RFE, it would point to adverse selection among those who choose the peer counseling, but it would also be consistent with Hawthorne effects as the effort and scrutiny applied during the RFE was not sustainable afterwards. Similarly, if the non-experimental effects are larger than those from the RFE, then women choosing to work with peer counselors are favorably selected in terms of breastfeeding and this dominates any Hawthorne effects. Moreover, comparison of experimental and non-experimental results across different agencies provides additional insight. If as we show below, the non-experimental treatment effect estimated in all the agencies are less than the experimental TOT effects, then the omitted variable bias may be working in one direction (Pritchett and Sandefur, 2014). This would be an important consideration when deciding to scale up the program.

We estimate all equations by ordinary least squares and we adjust the standard errors for general forms of heteroscedasticity. For the models in which women were not randomized, we weight regressions by propensity scores.[13] Our results are robust to the use of logistic

---

[13]We weight the treated group—those receiving peer counseling—by 1 and the comparison group by $\frac{p_i}{1-p_i}$ where p_i is the estimated probability of receiving peer counseling for individual i. This provides an estimate of the TOT (Imbens and Wooldridge, 2008).

regression instead of OLS and weighted as compared to unweighted regressions.

4   Results

We present the results in two parts. In the first part, we analyze effects of peer counseling on exclusive breastfeeding based on the RFE in Oregon. In the second part we explore the potential generalizability of the findings.

4.1   Results from the RFE

The estimates from the RFE for English- and Spanish-speaking clients are shown in Table 1. We show coefficients for exclusive breastfeeding at one, three and six months postpartum adjusted for covariates.[14] The first column shows the ITT estimates. These reveal no effect of the offer of peer counseling on exclusive breastfeeding among English speakers at any point postpartum (Table 1, column 1). For instance, women assigned to the treatment group were three percentage points more likely to exclusive breastfeeding at one month postpartum but the coefficient is statistically insignificant. There is no effect at three of six months postpartum. This is not true for Spanish speakers (Table 1, column 4). The offer of peer counseling increases exclusive breastfeeding at one, three and six months postpartum by between seven and 10 percentage points. These are meaningful increases relative to the prevalence of breastfeeding among the controls.

Weighted least squares estimates (WLS) of the effect of actually receiving peer counseling on exclusive breastfeeding is shown for English and Spanish speakers in columns (2) and (5), respectively. Estimates for English speakers differ markedly from the ITT estimates at one and three months postpartum. Exclusive breastfeeding is nine to 10 percentage points greater among the English speakers who actually received peer counseling support relative to both the treated and controls who did not receive counseling. The first stage in column (3)

---

[14]The unadjusted coefficients differ little. See Reeder et al. (2014) for a comparison of the adjusted and unadjusted ITT effects.

indicates that after adjusting for the covariates, the probability of receiving peer counseling in the treatment group is 0.76. The differences between the ITT and WLS estimates therefore suggests that English-speaking women in the treatment group who received peer counseling were positively selected toward breastfeeding. However, when we correct for selection by using the randomization indicator as an instrument, we find no effect of peer counseling on exclusive breastfeeding (Table 1, column 3).

As with the ITT results, the pattern for Spanish-speaking clients differs importantly from those for English-speaking clients. In brief, the ITT, WLS and TOT estimates of peer counseling are essentially the same at each point postpartum. For example, at three months postpartum, the ITT, OLS and TOT estimated effects are between 10 and 11 percentages points, an increase of roughly 20 percent relative to the control mean of 51 percent. The consistency across estimates is perhaps unsurprising given that 90 percent of Spanish-speaking women in the treatment group actually received peer counseling in the prenatal and postpartum period as shown in the first stage in column (6).[15]

4.2 Heterogeneous Treatment Effects

In Table 2 we show results using endogenous stratification to describe differences in the estimated treatment effects across terciles of predicted breastfeeding within each language group. There is no effect of being assigned to the treatment group among English-speaking clients for any tercile, which is consistent with the ITT and TOT results in Table 1. However, among Spanish-speaking clients the largest effects of peer counseling occur among women from the lower and middle terciles; these women are 11 to 12 percentage points more likely to exclusively breastfeed for at least one month if assigned to the treatment group than their counterparts among the controls.

---

[15]This differs slightly from the 89 percent of Spanish-speaking women in the treatment group who received treatment as reported previously because the 90 percent pertains only to women with non-missing data on exclusive breastfeeding.

In summary, our findings extend the ITT estimates in Reeder et al. (2014). There is evidence of selection bias in the WLS estimates of the TOT among English-speaking clients. This is potentially important because implementation of a peer counseling program would involve self-selection. The consistent results among Spanish-speaking women across different estimates are also of interest for they suggest a strong willingness to participate with impressive gains. Moreover, Spanish-speaking clients with a lower predicted propensity to exclusively breastfeed appear to benefit the most from peer counseling.

4.3 External Validity

4.3.1 Partial Identification

Manski (2013) argues that analysts fail to appreciate the incredible uncertainty inherent in point estimates of treatment effects. Even with randomized field experiments and with the assumption of homogenous treatment effects, generalizability may be limited. The non-random selection of sites in which an experiment is conducted as well as non-random selection by participants into the study can render external validity questionable. In this section, we focus on the selection issue to show a range of possible treatment effects that may hold statewide under various restrictions. Table 3 documents the average treatment effect bounds under different assumptions about the range of unobserved counterfactual treatment effects of exclusive breastfeeding. For example, panel (1) shows treatment effect bounds assuming no knowledge of treatment effects for the unobserved counterfactuals. The bounds have a length of one and are clearly uninformative. We narrow the bounds by using the minimum and maximum rate of exclusive breastfeeding in the 28 WIC agencies in Oregon that had no peer counseling program as the two counterfactuals. As shown in panel (2), bounds contract by more than a half but are still quite wide. For instance, the lowest rate of exclusive breastfeeding at three months among Spanish-speaking WIC participants in agencies without peer counseling is 0.12 and the highest rate is 0.55. Using these instead of rates of zero and one yield treatment effect bounds of -0.15 and 0.30 (panel 2). Lastly we assume that peer

counseling has no negative effects on exclusive breastfeeding, which creates a lower bound of zero. We combine that restriction with the min-max counterfactuals to generate the bounds in panel (3). For Spanish-speakers, the bounds have a length of approximately 30 percentage points and are wide enough to contain the 95 percent confidence intervals around the point estimates of the treatment effects for Spanish speakers from the RFE at one and three months postpartum in Table 1. The bounds are considerably narrower for English-speaking clients because the point estimates of the treatment effects are much less as is the spread between the high and low rates of exclusive breastfeeding among WIC agencies.

4.3.2 Participation Selection

In Oregons RFE, all eligible WIC participants who presented at a "new pregnant" appointment were asked if they would be willing to participate in a study to determine the effectiveness of a telephone peer counseling program to support exclusive breastfeeding.[16] Under the assumption of homogeneous treatment effects, results from the RFE would generalize to non-participants even with non-random selection into the study. However, homogeneous treatment effects is a strong assumption. If we relax that assumption, and if women who chose to participate differ from those who declined, then results may not apply to the larger population of non-participants. As a first step in understanding potential differences between participants and non-participants of the RFE, we compare baseline characteristics of RFE participants to four groups[17] of WIC clients as shown in Table 4. For each sample, we report the differences in means with respect to the RFE participants and report the result of a t-test based on the difference for each covariate. Sample sizes, however, greatly vary across samples thus we also report the standardized differences. Unlike the t-test, the standardized

---

[16]Eligibility was extremely broad. Any WIC participant who indicated that she was interested in breast-feeding or undecided was considered eligible. Oregon has the highest rate of breastfeeding initiation in the country among women on WIC at over 91 percent (Polhamus et al., 2011). Essentially, all women attending a new pregnant visit were eligible.

[17]Women on WIC who were offered participation in the RFE but declined; WIC clients from agencies in the rest of the state in which there was no peer counseling available; women from agencies that participated in the RFE but for the period after the RFE (2007-2010) in which peer counseling was available to anyone without randomization, and women from the WIC agencies in Marion and Deschutes counties in which peer counseling was available over the entire period, 2005-2010.

differences are insensitive to sample size (Imbens and Wooldridge, 2008).

There are no glaring differences between the RFE participants and non-participants in age, maternal education, family income, and marital status. Standardized differences are less than 0.25 in all cases, a threshold above which estimates may become sensitive to the specification. The lack of difference in these covariates is not surprising given that all women are pregnant and meet the eligibility criteria for WIC.

As a further check on balance across the four samples, we plot the distribution of propensity scores for each group relative to those in the RFE. As a point of comparison, we show propensity scores for those assigned to the treatment and control group in the RFE (Figures 2a and 2b). As expected, overlap is impressive. However, the overlap is also notable between participants and non-participants in the RFE (Figures 2c and 2d) as well as between women in the other three groups (Figures 2e-2j).

Despite the apparent balance along observables, our set of covariates is limited and non-random sorting into the RFE as well as difference across WIC agencies remains a concern. As a further examination, we estimate equation (4) comparing the breastfeeding outcomes of the RFE controls to those of women who were offered participation and declined as well as to women from agencies with no peer counseling services. The results from this exercise reveal a clear pattern of selection into the RFE among English speakers (Table 5, Panel A). Women in the control group are between 5 and 7 percentage points more likely to exclusively breastfeed relative to women who declined participation in the RFE. A similar pattern appears when we contrast controls with women on WIC in the rest of the state who had no access to peer counseling, although the magnitude is more muted (Table 5, panel B). There are also large differences in exclusive breastfeeding between English-speaking controls and women in the post-RFE agencies who received no peer counseling (Panel C) and women form the Marion and Deschutes agencies who also declined peer counseling (Table 5, Panel D).

The pattern of selection into the RFE is less consistent among Spanish-speaking women in WIC. Controls in the RFE are no more likely to exclusively breastfeed at one and three months postpartum than are non-participants but they are 6-7 percentage more likely to exclusively breastfeed when compared to their counterparts in the rest of the state (Table 5, Panels A and B). Similarly, there are no differences in exclusive breastfeeding between the RFE controls and their counterparts in the post-RFE period, but large differences when compared to Spanish-speaking clients in Marion and Deschutes (Table 5, Panels C and D).

4.3.3   Non-experimental estimates of peer counseling

In this last section we examine the association between the use of peer counseling and exclusive breastfeeding in the two non-experimental settings. We present adjusted and unadjusted estimates of the TOT for each breastfeeding outcome. We use weighted least squares to mitigate potential selection bias. Consider results from the three agencies involved in the RFE in the post-RFE period (Table 6, Panel A). The adjusted estimates of peer counseling for the English speaking participants are all less than three percentage points for each breastfeeding outcome. Although statistically significant they are clinically small and similar in magnitude to the TOT estimates for English-speaking women in RFE (Table 1). The adjusted estimates for Spanish-speaking participants are between three and four percentage points, roughly 40 percent smaller than the unadjusted estimates and approximately 60 percent smaller than the TOT estimates from the RFE (Table 6, Panel A).

Estimates from the two agencies not involved in the RFE are even smaller than those in RFE agencies and are not statistically significant (Table 6, Panel B). Based on the non-experimental estimates, one would conclude that effects of peer counseling on exclusive breastfeeding ranged from small and positive to nonexistent among both English and Spanish-speaking women.

4.4   Discussion

The pattern of results from Tables 1-6 suggests the following. First, there is little experimental or non-experimental evidence that a primarily telephone peer counseling program among English-speaking women on WIC increased exclusive breastfeeding in a clinically meaningful manner. Moreover, there is a clear pattern of selection bias among English-speaking women on WIC who volunteered for the RFE. The RFE controls were more likely to exclusively breastfeed relative to women who chose not to participate in the RFE. But there was also selection bias among women assigned to the treatment group. English-speaking women in the RFE who interacted with peer counselors were more likely to exclusively breastfeed than those in the treatment and control groups who did not receive counseling. In other words, compliance was not random even among those assigned to the treatment arm.

The pattern among Spanish-speaking women on WIC is different. First, estimates from the RFE indicate that the telephone peer counseling program increased exclusive breastfeeding at one, three and six months postpartum. In addition, almost 90 percent of Spanish-speaking women in the treatment group received counseling, which rendered the ITT and TOT estimates indistinguishable. There was also no pattern of selection bias into the RFE as the Spanish-speaking controls were not more likely to exclusively breastfeed than were non-participants. Lastly, the estimated effects of peer counseling in post-RFE were qualitatively similar although approximately 60 percent less in magnitude than those in the RFE. There was no obvious evidence of selection bias as exclusive breastfeeding among women who chose not to use counseling in the post-RFE period did not differ from that of the RFE controls. If we assume that selection bias is not a good explanation for the diminished impact of peer counseling in the non-experimental period, then how best to explain it? One consideration is Hawthorne effects. Peer counselors in the RFE were under much more scrutiny than those in the post-RFE period. Counselors were required to maintain logs of each interaction with clients and reports were sent twice monthly to researchers at the local and state WIC offices (Reeder et al., 2008). After the RFE, reporting by peer counselors was

22

less formal in these same agencies as was also the case in the Marion and Deschutes agencies.

Why might Spanish-speaking clients in the RFE have been more receptive to peer counseling that their English-speaking counterparts? We can only speculate, but if Spanish-speaking is a proxy for relatively recent immigration and perhaps social isolation, then support by a native-speaking counselor may have been more valued by these clients relative to English-speaking women with greater access to networks of support within their community. As Reeder et al. (2014) note, Spanish-speaking clients responded more readily to peer counselors' calls than did English-speaking clients. Moreover, Spanish-speaking women on WIC were more likely to exclusively breastfeed than were their English-speaking peers who self-identify as Hispanic.[18]

5 Conclusion

Oregon's RFE of a telephone peer counseling program to support exclusive breastfeeding among women on WIC was the largest such intervention to date in the US. The study came at an important junction as the U.S. Surgeon General has called for making peer counseling to support breastfeeding available to all women on WIC. Our analysis of Oregons RFE and its generalizability suggest caution before scaling up the program. Internally valid and clinically meaningful effects were limited to Spanish-speaking clients. However, we found only a weak association between peer counseling and increased rates of exclusive breastfeeding in the non-experimental settings even when provided in the same agencies that had successfully implemented the RFE. We suspect that the attenuation reflects the heightened attention to the work of peer counselors in the RFE that is difficult to maintain outside the experimental context. Oregon is not representative of the US demographically and it is a national leader

---

[18]The Oregon WIC program discontinued its telephone-based peer counseling program based on findings from the RFE. In its place Oregon has created a new model centered on in-person, group prenatal sessions where women explore not just the benefits of breastfeeding but strengthen their own intrinsic motivators, support networks, and self-advocacy skills. Enrollment in the program remains a voluntary option for new pregnant participants. In order to provide the intensity of services needed to increase long term exclusive breastfeeding, local agency caseload requirements were lowered from 30% of pregnant participants required during the RFE to 17% of pregnant participants.

in breastfeeding initiation. This limits generalizability of the RFE. But even within State, the external validity of the results from the RFE appears limited.

References

ABADIE, A., M. M. CHINGOS, AND M. R. WEST (2013): "Endogenous stratification in randomized experiments," *NBER Working Paper No. 19742.*

AMERICAN ACADEMY OF PEDIATRICS, A. (2012): "Breastfeeding and the use of human milk," *Policy statement, Pediatrics*, 129, e827–41.

ANDERSON, A. K., G. DAMIO, S. YOUNG, D. J. CHAPMAN, AND R. PÉREZ-ESCAMILLA (2005): "A randomized trial assessing the efficacy of peer counseling on exclusive breast-feeding in a predominantly Latina low-income community," *Archives of Pediatrics & Adolescent Medicine*, 159, 836–841.

ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): "Identification of causal effects using instrumental variables," *Journal of the American Statistical Association*, 91, 444–455.

ANGRIST, J. D. AND J. S. PISCHKE (2008): *Mostly harmless econometrics: an empiricist's companion*, Princeton University Press.

BACHRACH, V. R. G., E. SCHWARZ, AND L. R. BACHRACH (2003): "Breastfeeding and the risk of hospitalization for respiratory disease in infancy: a meta-analysis," *Archives of Pediatrics & Adolescent Medicine*, 157, 237–243.

BANERJEE, A. V. AND E. DUFLO (2008): "The experimental approach to development economics," *NBER Working Paper No. 14467.*

BLOOM, H. S. (1984): "Accounting for no-shows in experimental evaluation designs," *Evaluation Review*, 8, 225–246.

BOLTON, T. A., T. CHOW, P. A. BENTON, AND B. H. OLSON (2009): "Characteristics associated with longer breastfeeding duration: an analysis of a peer counseling support program," *Journal of Human Lactation*, 25, 18–27.

BONUCK, K. A., M. TROMBLEY, K. FREEMAN, AND D. MCKEE (2005): "Randomized, controlled trial of a prenatal and postnatal lactation consultant intervention on duration and intensity of breastfeeding up to 12 months," *Pediatrics*, 116, 1413–1426.

CAMPBELL, D. T., J. C. STANLEY, AND N. L. GAGE (1963): *Experimental and quasi-experimental designs for research*, Houghton Mifflin Boston.

CHAPMAN, D. J., G. DAMIO, S. YOUNG, AND R. PEREZ-ESCAMILLA (2004): "Effectiveness of breastfeeding peer counseling in a low-income, predominantly Latina population: a randomized controlled trial," *Archives of Pediatrics & Adolescent Medicine*, 158, 897–902.

CRUMP, R. K., V. J. HOTZ, G. W. IMBENS, AND O. A. MITNIK (2008): "Nonparametric tests for treatment effect heterogeneity," *The Review of Economics and Statistics*, 90, 389–405.

DEATON, A. (2010): "Instruments, randomization, and learning about development," *Journal of Economic Literature*, 424–455.

EIDELMAN, A. I., R. J. SCHANLER, M. JOHNSTON, S. LANDERS, L. NOBLE, K. SZUCS, L. VIEHMANN, ET AL. (2012): "Breastfeeding and the use of human milk," *Pediatrics*, 129, e827–e841.

GILL, S. L., E. REIFSNIDER, AND J. F. LUCKE (2007): "Effects of support on the initiation and duration of breastfeeding," *Western Journal of Nursing Research*, 29, 708–723.

GRAFFY, J., J. TAYLOR, A. WILLIAMS, AND S. ELDRIDGE (2004): "Randomised controlled trial of support from volunteer counsellors for mothers considering breast feeding," *BMJ*, 328, 26.

GROSS, S. M., A. K. RESNIK, C. CROSS-BARNET, J. P. NANDA, M. AUGUSTYN, AND D. M. PAIGE (2009): "The differential impact of WIC peer counseling programs on breastfeeding initiation across the state of Maryland," *Journal of Human Lactation*, 25, 435–443.

Grummer-Strawn, L. M. and Z. Mei (2004): "Does breastfeeding protect against pediatric overweight? Analysis of longitudinal data from the Centers for Disease Control and Prevention Pediatric Nutrition Surveillance System," *Pediatrics*, 113, e81–e86.

Heckman, J. J. and J. A. Smith (1995): "Assessing the case for social experiments," *The Journal of Economic Perspectives*, 85–110.

Imbens, G. (2010): "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48, 399–423.

Imbens, G. M. and J. M. Wooldridge (2008): "Recent developments in the econometrics of program evaluation," *NBER Working Paper No. 14251*.

Ingram, L., C. MacArthur, K. Khan, J. J. Deeks, and K. Jolly (2010): "Effect of antenatal peer support on breastfeeding initiation: a systematic review," *Canadian Medical Association Journal*, 182, 1739–1746.

Ip, S., M. Chung, G. Raman, T. A. Trikalinos, and J. Lau (2009): "A summary of the agency for healthcare research and quality's evidence report on breastfeeding in developed countries," *Breastfeeding Medicine*, 4, S–17.

Jacknowitz, A., D. Novillo, and L. Tiehen (2007): "Special supplemental nutrition program for women, infants, and children and infant feeding practices," *Pediatrics*, 119, 281–289.

Kramer, M. S., B. Chalmers, E. D. Hodnett, Z. Sevkovskaya, I. Dzikovich, S. Shapiro, J.-P. Collet, I. Vanilovich, I. Mezen, T. Ducruet, et al. (2001): "Promotion of Breastfeeding Intervention Trial (PROBIT): a randomized trial in the Republic of Belarus," *JAMA*, 285, 413–420.

Kramer, M. S. and R. Kakuma (2004): "The optimal duration of exclusive breastfeeding," in *Protecting Infants Through Human Milk*, Springer, 63–77.

KRAMER, M. S. AND S. H. SHAPIRO (1984): "Scientific challenges in the application of randomized trials," *JAMA*, 252, 2739–2745.

LALONDE, R. J. (1986): "Evaluating the econometric evaluations of training programs with experimental data," *The American Economic Review*, 604–620.

MACARTHUR, C., K. JOLLY, L. INGRAM, N. FREEMANTLE, C.-L. DENNIS, R. HAMBURGER, J. BROWN, J. CHAMBERS, AND K. KHAN (2009): "Antenatal peer support workers and initiation of breast feeding: cluster randomised controlled trial," *BMJ*, 338.

MANSKI, C. F. (1999): *Identification problems in the social sciences*, Harvard University Press.

——— (2013): *Public policy in an uncertain world: analysis and decisions*, Harvard University Press.

MCLAUGHLIN, J. E., N. R. BURSTEIN, F. TAO, AND F. M. K (2004): "Breastfeeding intervention design study-final evaluation design and analysis plan," *Report WIC-04-BFDSN, US Department of Agriculture, Food and Nutrition Services, Office of Analysis, Nutrition and Evaluation.*

OLIVEIRA, V., E. FRAZÃO, AND D. M. SMALLWOOD (2010): "Rising infant formula costs to the WIC program: recent trends in rebates and wholesale prices," *Report No. 59384, United States Department of Agriculture, Economic Research Service.*

OLSON, B. H., S. J. HAIDER, L. VANGJEL, T. A. BOLTON, AND J. G. GOLD (2010): "A quasi-experimental evaluation of a breastfeeding support program for low income women in Michigan," *Maternal and Child Health Journal*, 14, 86–93.

POLHAMUS, B., K. DALENIUS, H. MACKINTOSH, B. SMITH, AND G.-S. L (2011): "Pediatric nutrition surveillance report," *U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.*

PRITCHETT, L. AND J. SANDEFUR (2014): "Context matters for size: why external validity claims and development practice do not mix," *Journal of Globalization and Development*, 4, 161–197.

RAVALLION, M. (2012): "Fighting poverty one experiment at a time: a review of Abhijit Banerjee and Esther Duflo's poor economics: a radical rethinking of the way to fight global poverty," *Journal of Economic Literature*, 50, 103–114.

REEDER, J., K. SIBLEY, D. ARNOLD, AND S. EUTENEIER (2008): "Effect of a telephone-based peer counseling program on duration and exclusivity of breastfeeding among Oregon WIC clients," *Oregon WIC Program Final Report: Portland, Oregon.*

REEDER, J. A., T. JOYCE, K. SIBLEY, D. ARNOLD, AND O. ALTINDAG (2014): "Telephone peer counseling of breastfeeding among WIC participants: a randomized controlled trial," *Pediatrics*, 134, e700–e709.

RODRIK, D. (2008): "The new development economics: we shall experiment, but how shall we learn?" *HKS Working Paper No. RWP08-055.*

ROTHWELL, P. M. (2005): "External validity of randomised controlled trials:to whom do the results of this trial apply?," *The Lancet*, 365, 82–93.

SCHAFER, E., M. K. VOGEL, S. VIEGAS, AND C. HAUSAFUS (1998): "Volunteer peer counselors increase breastfeeding duration among rural low-income women," *Birth*, 25, 101–106.

SHADISH, W. R., T. D. COOK, AND D. T. CAMPBELL (2002): *Experimental and quasi-experimental designs for generalized causal inference*, Houghton Mifflin Company, Boston.

SHAW, E. AND J. KACZOROWSKI (1999): "The effect of a peer counseling program on breastfeeding initiation and longevity in a low-income rural population," *Journal of Human Lactation*, 15, 19–25.

SMITH, J. A. AND P. E. TODD (2005): "Does matching overcome LaLonde's critique of nonexperimental estimators?" *Journal of Econometrics*, 125, 305–353.

TEELE, D. L. (2014): *Field experiments and their critics: essays on the uses and abuses of experimentation in the social sciences*, Yale University Press.

WOOLCOCK, M. (2013): "Using case studies to explore the external validity of complexdevelopment interventions," *Evaluation*, 19, 229–248.

YUN, S., Q. LIU, K. MERTZLUFFT, C. KRUSE, M. WHITE, P. FULLER, AND B.-P. ZHU (2010): "Evaluation of the Missouri WIC (Special Supplemental Nutrition Program for Women, Infants, and Children) breast-feeding peer counselling programme," *Public Health Nutrition*, 13, 229–237.

Figure 1: Counties in Oregon with and without Peer Counseling for Breastfeeding, 2005-2007

Legend

- No peer counseling
- RCT of peer counseling
- Non-RCT peer counseling
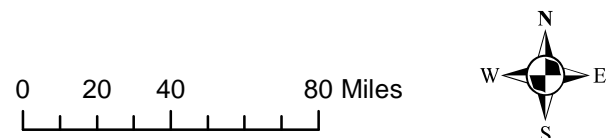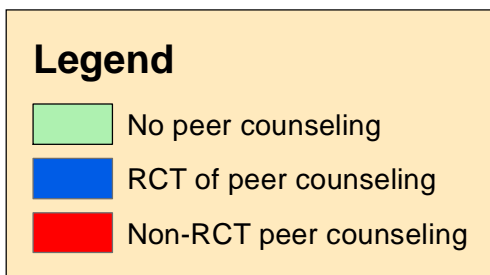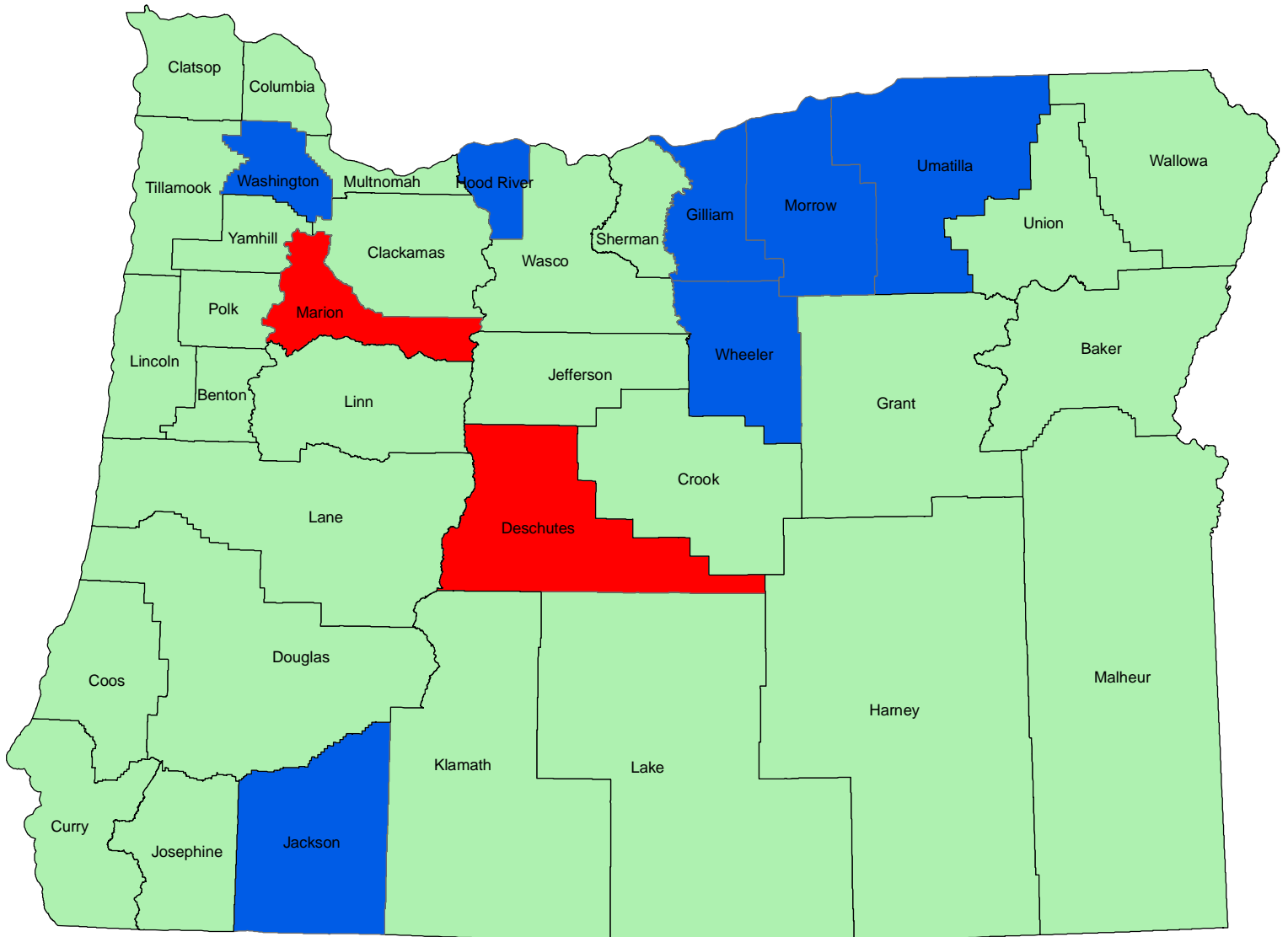
Figures 2a-2d. Propensity Score Distribution by Sample
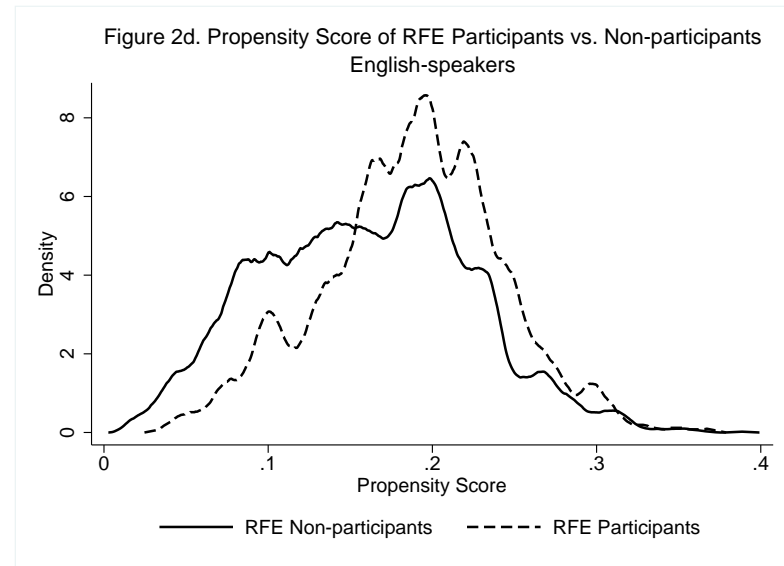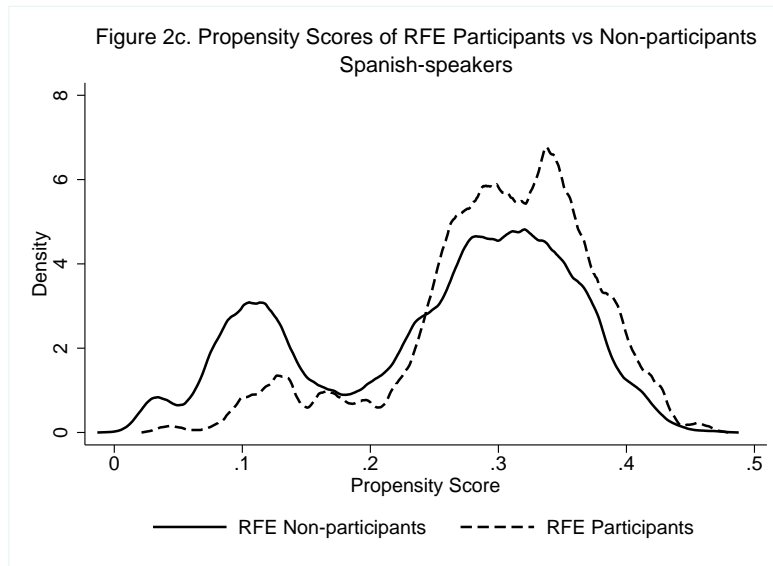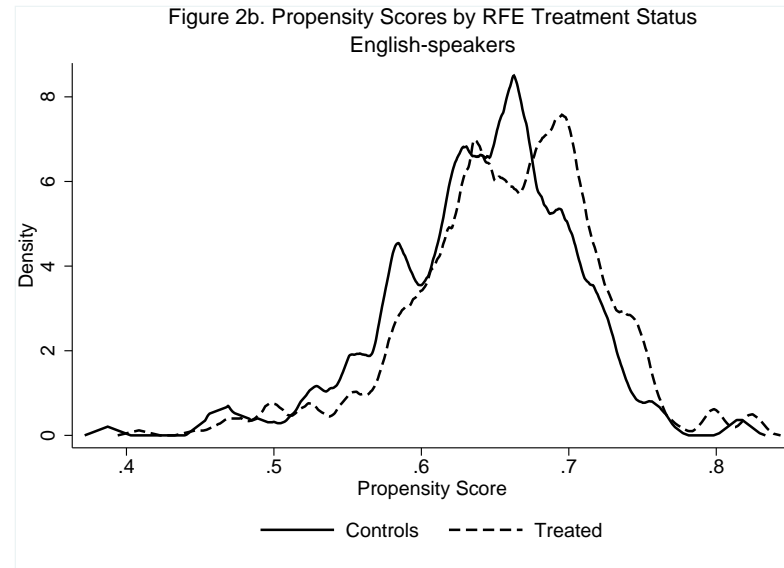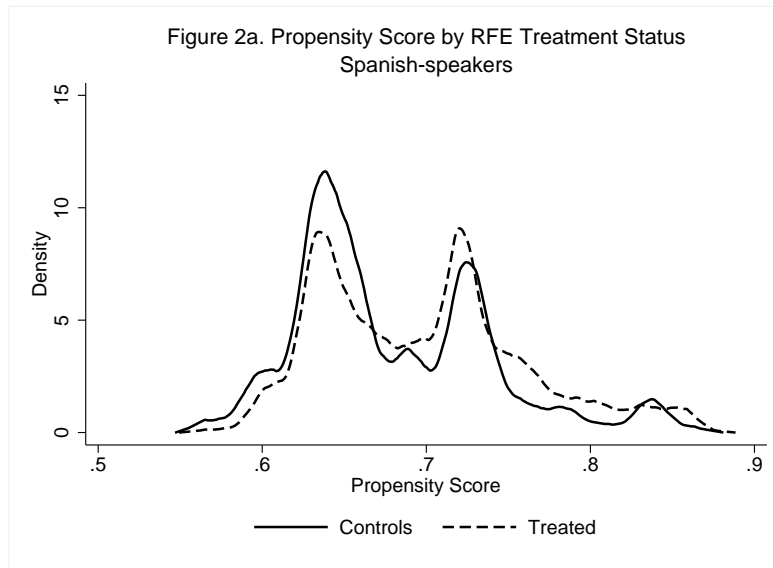
Figures 2e-2h. Propensity Score Distributions by Sample

Figure 2e. Propensity Scores of RFE Participants vs. Rest of State Spanish-speakers

Figure 2f. Propensity Score of RFE Participants v Rest of State English-speakers

Figure 2g. Propensity Scores of RFE Participants vs. Post-RFE Spanish-speakers

Figure 2h. Propensity Scores of RFE Participants vs Post-RFE English-speakers
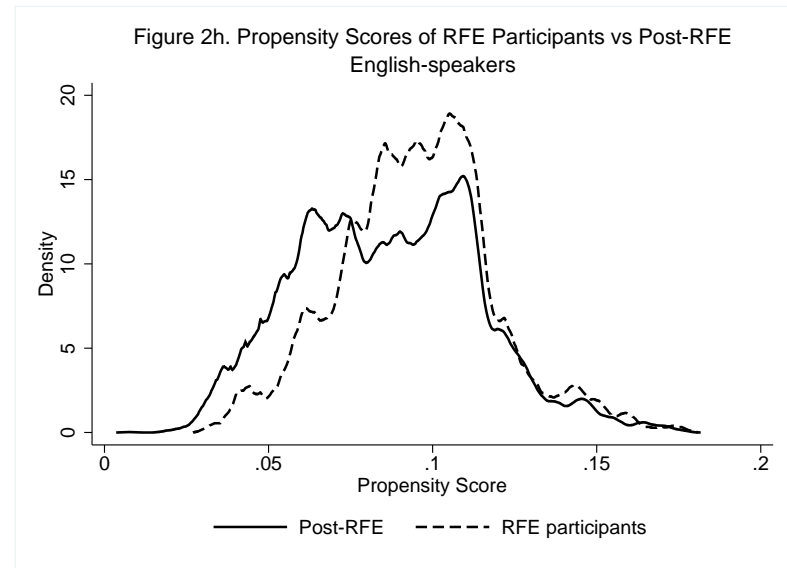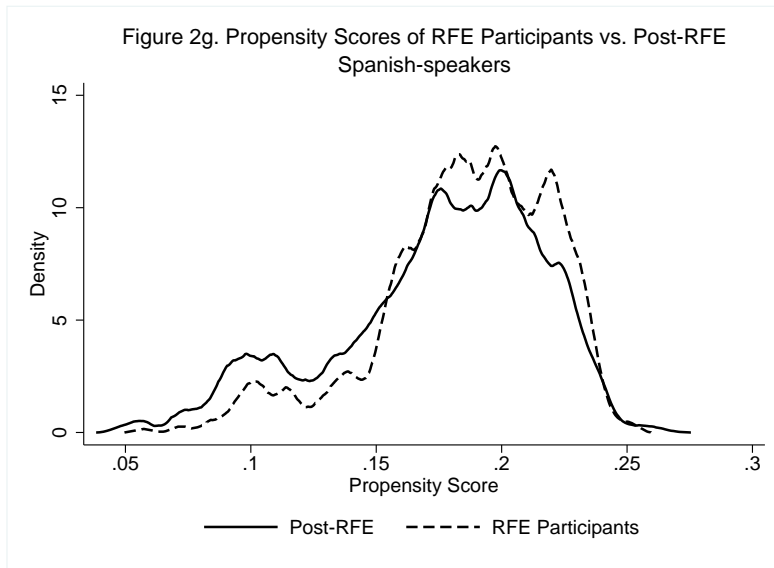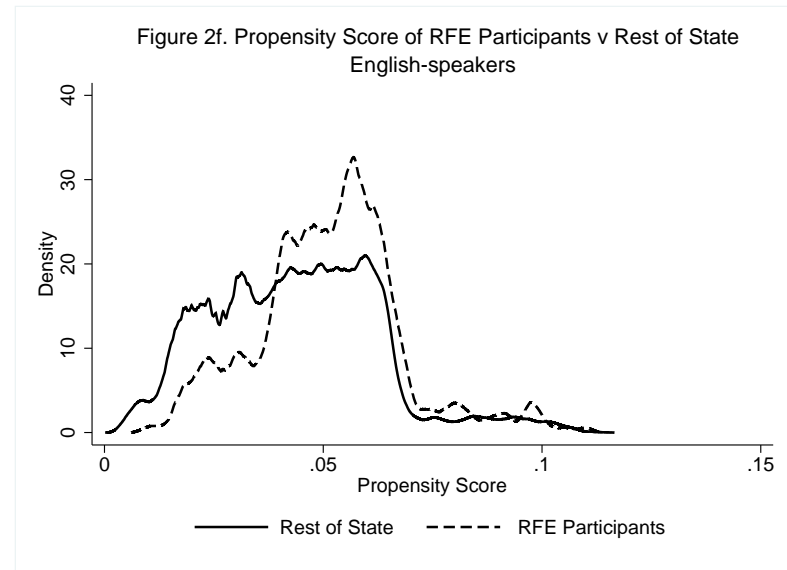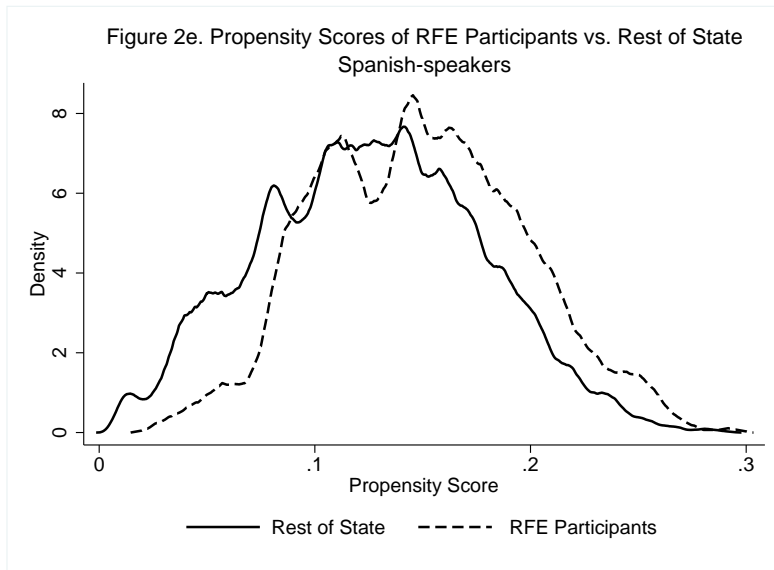
Figures 2i-2j. Propensity Score Distributions by Sample

Figure 2i. Propensity Scores of RFE Participants vs. Marion-Deschutes
Spanish-speakers

Figure 2j. Propensity Scores of RFE Participants vs. Marion-Deschutes
English-speakers

Table 1. Effect of Peer Counseling on Exclusive Breastfeeding from Oregon RFE by Duration and Language, 2005 - 2007

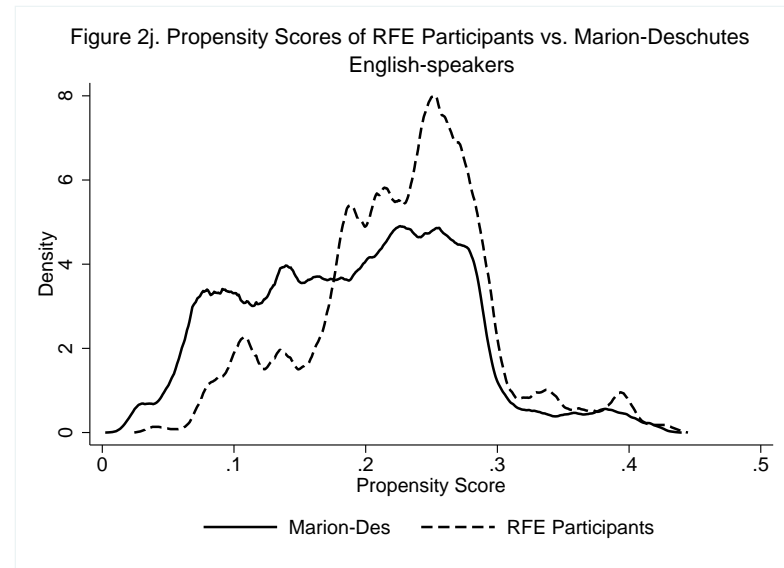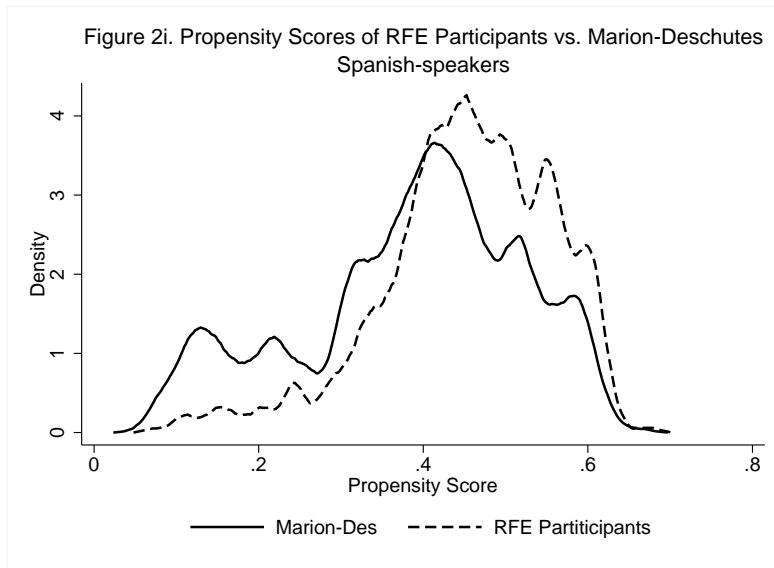| Exclusive breastfeeding for at least	English Speaking WIC Clients			Spanish Speaking WIC Clients		
	ITT (1)	OLS (2)	TOT (3)	ITT (4)	OLS (5)	TOT (6)
One month	0.029	0.115**	0.039	0.091*	0.103**	0.100*
	(0.04)	(0.03)	(0.05)	(0.04)	(0.04)	(0.04)
R^2	0.067	0.079	0.073	0.060	0.063	0.063
Control group mean	0.534	0.495	0.495	0.518	0.511	0.511
Three months	0.011	0.095**	0.014	0.104**	0.114**	0.115**
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
R^2	0.066	0.075	0.069	0.071	0.074	0.074
Control group mean	0.346	0.312	0.312	0.402	0.390	0.390
Six months	-0.026	0.019	-0.035	0.073*	0.089*	0.081*
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)
R^2	0.043	0.043	0.038	0.057	0.060	0.060
Control group mean	0.227	0.204	0.204	0.315	0.298	0.298
First Stage			0.763**			0.903**
			(0.02)			(0.01)
Number of observations		889			815	

Note: This table reports the intent-to-treat (ITT), ordinary least squares (OLS) and treatment on the treated (TOT) estimates of the effects of peer counseling on exclusive breastfeeding for at least one, three and six months. The endogenous variable in columns (3) and (6) is the receipt of peer counseling, and the instrument is a randomization indicator for the treatment group. All regressions control for mother's age, age-squared, race/ethnicity, education, marital status, the month pregnancy when certified for WIC, the natural logarithm of family income, WIC agency fixed effects plus indicator variables for missing month pregnancy and family income. Heteroskedasticity-consistent standard errors are in parentheses. Significance levels are indicated by $* < .05$, $** < .01$.

Table 2. Intention-to-Treat Effects of Peer Counseling by Terciles of Predicted Exclusive Breastfeeding

Exclusive breastfeeding for at least	English Speaking WIC Clients			Spanish Speaking WIC Clients		
	1 Month (1)	3 Months (2)	6 Months (3)	1 Month (4)	3 Months (5)	6 Months (6)
Lower tercile	0.001	-0.025	-0.047	0.119[+]	0.089	0.074
	(0.06)	(0.05)	(0.04)	(0.07)	(0.06)	(0.06)
Middle tercile	0.047	0.041	0.008	0.113*	0.104*	0.060
	(0.05)	(0.05)	(0.04)	(0.06)	(0.05)	(0.05)
Upper tercile	0.025	0.053	-0.015	-0.022	0.040	0.023
	(0.06)	(0.05)	(0.05)	(0.06)	(0.07)	(0.06)
Number of observations		889			815	

Note: This table reports the intent-to-treat (ITT) estimates of the effects of peer counseling by terciles of predicted exclusive breastfeeding. Estimated with repeated split sample (RSS) estimator as shown in Abadie et al. (2014). Variables that are used to predict the terciles are mother's age, education, marital status and the natural logarithm of family income. The number of repeated split sample repetitions is 500. Standard errors are in parentheses and based on 500 bootstrap samples. Significance levels are indicated by $+ < .1$, $* < .05$.

Table 3. Bounds of Treatment Effects of Peer Counseling by Identifying Restrictions

Exclusive breastfeeding for at least	1 Month		3 Months		6 Months	
	Lower	Upper	Lower	Upper	Lower	Upper
No information bounds						
(1) *English Speakers*	-0.45	0.55	-0.44	0.56	-0.47	0.53
Spanish Speakers	-0.45	0.55	-0.46	0.54	-0.49	0.51
Min, Max						
(2) *English Speakers*	-0.10	0.20	-0.10	0.19	-0.08	0.10
Spanish Speakers	-0.17	0.31	-0.15	0.30	-0.09	0.22
No Negative Effects & Min, Max						
(3) *English Speakers*	0.00	0.20	0.00	0.19	0.00	0.10
Spanish Speakers	0.00	0.31	0.00	0.30	0.00	0.22

Note: This table reports the average treatment effect bounds under different assumptions about the range of unobserved counterfactual treatment effects of exclusive breastfeeding. Panel (1) impose no restrictions on the possible treatment effects. In panel (2), the lower and upper bounds of the counterfactual treatment levels are restricted to the minimum and maximum rates of exclusive breastfeeding in the 28 WIC agencies that did not provide peer counseling services. Panel (3) imposes the same bound restrictions as panel (2) and additionally assumes that peer counseling could not result in negative treatment effects.

Table 4. Characteristics of Oregon WIC participants in Experimental and Non-Experimental Samples, 2005-2010

English Speaking WIC Clients

	RFE	Non Participants		Rest of State		Post RFE		Marion & Deschutes	
	Mean	Diff.	Norm. Diff.	Diff.	Norm. Diff.	Diff.	Norm. Diff.	Diff.	Norm. Diff.
Age	25.9	0.86**	0.11	1.06**	0.14	0.91**	0.12	1.08**	0.14
Education									
< High school	0.27	0.04**	0.07	0.05**	0.09	0.02	0.03	0.04*	0.06
≥ High school	0.71	-0.02	-0.03	-0.03[+]	-0.04	-0.02	-0.03	-0.04*	-0.06
Unknown	0.03	-0.03**	-0.09	-0.03**	-0.09	<0.01	0.01	<0.01	<0.01
Marital Status									
Single	0.46	0.05*	0.06	<0.01	<0.01	0.01	0.01	-0.02	-0.03
Married / Partner	0.51	0.01	0.01	0.06**	0.08	0.01	0.01	0.06**	0.08
Unknown	0.03	-0.06**	-0.17	-0.06**	-0.18	-0.02*	-0.06	-0.03**	-0.11
Family Income	1319	167**	0.12	186**	0.13	92**	0.07	152**	0.11
N	889	4522		19351		9303		9771	

Spanish Speaking WIC Clients

	RFE	Non Participants		Rest of State		Post RFE		Marion & Deschutes	
	Mean	Diff.	Norm. Diff.	Diff.	Norm. Diff.	Diff.	Norm. Diff.	Diff.	Norm. Diff.
Age	28.9	1.20**	0.15	1.87**	0.22	0.79**	0.09	0.97**	0.12
Education									
< High school	0.57	0.01	0.02	-0.02	-0.03	0.04*	0.06	0.09**	0.12
≥ High school	0.38	<0.01	<0.01	0.05**	0.07	-0.04*	-0.06	-0.11**	-0.15
Unknown	0.05	-0.01	-0.04	-0.03**	-0.08	<0.01	0.01	0.02*	0.07
Marital Status									
Single	0.20	0.02	0.03	<0.01	<0.01	-0.04*	-0.06	-0.01	-0.02
Married / Partner	0.77	0.06**	0.10	0.06**	0.09	0.07**	0.11	0.04*	0.07
Unknown	0.03	-0.08**	-0.21	-0.06**	-0.17	-0.03**	-0.11	-0.03**	-0.10
Family Income	1559	178**	0.15	181**	0.15	59[+]	0.05	185**	0.15
N	815	2272		5506		3791		2577	

Note: This table reports the average background characteristics of WIC clients in randomized field experiment (RFE) and contrast them with the WIC clients from different samples. The column "Diff." shows the difference in means for each of the other samples relative to the RFE. The column "Norm. Diff" shows the normalized differences. Non-participants are women from the same WIC agencies as those in the RFE who were offered participation in the study but declined. Women in the rest of the state are from the 28 other WIC agenices in the state that did not provide peer counseling services. The Post-RFE sample are WIC clients from three of the four experimental agencies that continued to offer peer counseling services after the RFE was completed. Marion and Deschutes are two other WIC agencies that offered peer counseling services to interested women from 2005-2010. Any difference smaller than 0.01 in absolute value is indicated with <0.01. Significance levels are indicated by [+] < .1, * < .05 , * < .01.

Table 5. Differences in Exclusive Breastfeeding Between Experimental Controls and Non-participants by Duration, Language and WIC Agencies

Exclusive breastfeeding for	English Speaking WIC Clients			Spanish Speaking WIC Clients		
	1 Month	3 Months	6 Months	1 Month	3 Months	6 Months
RFE Non-Participants (A)						
Difference	0.07*	0.05$^+$	0.06*	<0.01	0.01	0.06$^+$
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Mean Outcome	0.47	0.30	0.17	0.52	0.39	0.26
N		4831			2523	
Rest of State (B)						
Difference	0.05	0.03	0.04$^+$	0.06$^+$	0.07*	0.10**
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Mean Outcome	0.49	0.32	0.19	0.46	0.33	0.22
N		19660			5757	
Post RFE Non-Participants (C)						
Difference	0.10**	0.07*	0.06*	0.01	0.02	0.04
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Mean Outcome	0.44	0.28	0.17	0.50	0.38	0.28
N		7295			2782	
Marion & Deschutes (D)						
Difference	0.10**	0.04	0.04	0.11**	0.08**	0.08**
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Mean Outcome	0.43	0.30	0.19	0.42	0.32	0.23
N		8637			2527	

Note: This table reports the differences in exclusive breastfeeding between the control group from Oregon's randomized field experiment (RFE) and peer counseling non-participants from different samples. The reported differences are estimated with linear regression weighted by the inverse probability score of being in the experimental control group. Weights are 1 for women in the experimental control group and $\frac{p_i}{1-p_i}$ for those in the comparison group where p_i detones the propensity score for WIC client i. The propensity scores are estimated with logistic regression using age, age squared, the natural logarithm of family income plus indicators for education and marital status as predictors. Comparison groups are as follows. Panel (A): women from the same WIC agencies as those in the RFE who were offered participation in the study but declined; panel (B): WIC clients in the rest of the state from the 28 other WIC agenices that did not provide peer counseling services; panel (C): Non-participants WIC clients from three of the four experimental agencies that continued to offer peer counseling services after the RFE was completed; panel (D): Non-participant WIC clients from Marion-Deschutes that offered peer counseling services to interested women from 2005-2010. Any difference smaller than 0.01 in absolute value is indicated with <0.01. Heteroskedasticity-consistent standard errors are in parentheses. Significance levels are indicated by $^+ < .1$, * $< .05$, * $< .01$.

Table 6. Non-experimental Estimates of Peer Counseling on Exclusive Breastfeeding by Duration, Language and WIC Agencies

Exclusive breastfeeding for	English Speaking WIC Clients			Spanish Speaking WIC Clients		
	1 Month	3 Months	6 Months	1 Month	3 Months	6 Months
Post RFE Agencies (A)						
Unadjusted	0.04**	0.03*	0.01	0.05**	0.07**	0.05**
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Adjusted	0.03*	0.03**	0.02*	0.03^+	0.04**	0.04*
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Mean Outcome	0.44	0.27	0.17	0.48	0.35	0.25
N		9303			3791	
Marion & Deschutes (B)						
Unadjusted	0.02	<0.01	-0.00	-0.01	0.02	0.03
	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)
Adjusted	0.03**	0.02	0.01	<0.01	0.03	0.04
	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)
Mean Outcome	0.42	0.28	0.17	0.41	0.32	0.24
N		9771			2577	

Note: This table reports the differences in exclusive breastfeeding between the clients who received peer counseling relative those who did not. The reported differences are estimated with linear regression weighted by the inverse probability receiving peer counseling. Weights are 1 for women who received peer counseling and $\frac{p_i}{1-p_i}$ for those who did not where p_i detones the propensity score for WIC client i. The propensity scores are estimated with logistic regression using age, age squared, the natural logarithm of family income plus indicators for education and marital status as predictors. Regression samples are as follows. Panel (A): WIC clients from three of the four experimental agencies that continued to offer peer counseling services after the RFE was completed; panel (B): Two WIC agencies that offered peer counseling services to interested women from 2005-2010. Any difference smaller than 0.01 in absolute value is indicated with <0.01. Heteroskedasticity-consistent standard errors are in parentheses. Significance levels are indicated by $^+ < .1$, $* < .05$, $* < .01$.