Can Nonexperimental Methods Provide Unbiased Estimates of a Breastfeeding Intervention? A Within-Study Comparison of Peer Counseling in Oregon

Onur Altindag[1], Theodore J. Joyce[2],
and Julie A. Reeder[3]

Abstract

Between July 2005 and July 2007, the Oregon Supplemental Nutrition Program for Women, Infants and Children program conducted the largest randomized field experiment (RFE) ever in the United States to assess the effectiveness of a low-cost peer counseling intervention to promote exclusive breastfeeding. We undertook a within-study comparison of the intervention using unique administrative data between July 2005 and July 2010. We found no difference between experimental and nonexperimental estimates but failed to determine correspondence based on more stringent criteria. We show that tests for nonconsent bias in the benchmark RFE

[1] Bentley University, Waltham, MA, USA
[2] Department of Economics and Finance, Baruch College, New York, NY, USA
[3] Oregon WIC Program, Portland, OR, USA

Corresponding Author:
Onur Altindag, Bentley University, 175 Forest St., Waltham, MA 02452, USA.
Email: oaltindag@bentley.edu

might provide an important signal as to confounding in the nonexperimental estimates.

Program evaluation based on a randomized controlled trial provides the most credible evidence of an intervention's effectiveness. Randomized controlled trials are considered the gold standard of evaluation in the clinical literature that explains, in part, the recent growth in randomized field experiments (RFEs) in the social sciences. Yet, RFEs are often costly to implement and in many cases unethical. Thus, nonexperimental methods are still widely used to evaluate programs in real-world settings. In an effort to improve the internal validity of nonexperimental evaluations, researchers have used within-study comparisons (WSCs) to assess how, when, and where observational studies are most likely to generate unbiased estimates of program effectiveness. Since LaLonde's (1986) seminal study, researchers have articulated the theoretical underpinnings of WSC, detailed the practical requirements for a successful WSC, and developed a rigorous measure of comparability (Cook et al., 2008; Steiner & Wong, 2018; Wong & Steiner, 2018).

In this study, we apply these proposed WSC methods to a large-scale peer counseling intervention to support breastfeeding among women in the Supplemental Nutrition Program for Women, Infants and Children (WIC) in Oregon. Between July 2005 and July 2007, the Oregon WIC program conducted the largest RFE ever in the United States to assess the effectiveness of a low-cost peer counseling intervention (Reeder, Joyce, Sibley, Arnold, & Altindag, 2014). The RFE was conducted in 4 of the Oregon's 34 local WIC agencies and included over 1,800 WIC participants. We use WIC clients from 27 agencies that were not part of the RFE to replace the control group and test for equivalence between the experimental and nonexperimental estimates of peer counseling on exclusive breastfeeding.

A breastfeeding intervention is a particularly suitable application of a WSC. Studies as to the health effects of breastfeeding rely principally on nonexperimental designs as it would be unethical to experimentally manipulate breastfeeding. The relatively modest number of randomized studies of breastfeeding interventions typically augment existing support (Anderson, Damio, Young, Chapman, & Pérez-Escamilla, 2005; Bonuck, Trombley,

Freeman, & McKee, 2005; Chapman, Damio, Young, & Pérez-Escamilla, 2004; Kramer et al. 2001). Second, 79% of women in the United States have ever breastfed, but only 18% have exclusively breastfed for 6 months.[1,2] The minority women who exclusively breastfeed likely depends on a myriad of factors suggesting that selection bias may be particularly profound in this context. This study is the first WSC of a breastfeeding intervention and thus has the potential to improve the methodological rigor of nonexperimental evaluations.

Our primary goal is to replicate the results from the original RFE with respect to exclusive breastfeeding using WIC administrative data on client characteristics and their breastfeeding behaviors. A general consensus from the WSC literature is that nonexperimental and experimental estimates are more likely to be similar when (1) the set of covariates is large, (2) subjects in the control and treatment groups are geographically and temporally near, (3) pretreatment outcomes are available, and (4) the data come from the same source and are measured in the same manner (Cook et al., 2008). We meet two of these criteria: (i) the study subjects all participated in WIC in same state over a fixed period and (ii) all breastfeeding outcomes and covariates in the RFE and comparison groups were measured similarly, collected by WIC staff and stored in TWIST, Oregon's administrative system for WIC. We lack a rich set of covariates beyond basic demographics and must rely on general similarity of WIC participants in a single state. We also lack a "pretest" measure such as previous breastfeeding experience, although the later would only be relevant for women with a previous live birth.

We compare the experimental and nonexperimental estimates in three ways. First, we replace the control observations in the RFE with WIC clients in the state who were unexposed to peer counseling during the study period and did not have an opportunity to participate in the RFE. Second, three of the four WIC agencies that participated in the RFE continued to offer peer counseling nonexperimentally to women enrolling in WIC at their new pregnancy visit from the end of the RFE in July 2007 through July 2010. We estimate the effect of peer counseling on exclusive breastfeeding in the post-RFE period and test its equivalence to its estimated impact from the RFE. The latter is a nonrandomized version of an independent WSC (Wong & Steiner, 2018). In an independent WSC, participants are randomized between the experimental and nonexperimental arms of the study, and researchers test for correspondence of the treatment effects between the two arms. Randomization ensures balance in expectation between the two arms. In our version, WIC clients are not randomized

between "arms." Nevertheless, the continuous enrollment of women with similar characteristics who live in the same residential areas and who receive services from many of the same staff is likely to mitigate bias resulting from unmeasured factors.

A third approach by which to compare experimental and nonexperimental estimates of program effectiveness uses a difference-in-differences (DD) design by comparing the breastfeeding outcomes of WIC participants in three counties before and after peer counseling services became available on a voluntary basis to their peers in counties in which peer counseling was never or always offered.

To preview our findings, we consistently report no statistical difference between the experimental and nonexperimental estimates in both the dependent and quasi-independent WSC. And yet, we failed to demonstrate correspondence between the experimental and nonexperimental estimates despite a well-powered RFE and large numbers of comparison subjects. We also found evidence of nonconsent bias in the RFE, suggesting that our set of covariates was too limited to mitigate an omitted-variable problem. One solution is to augment administrative data with surveys of subjects prior to the intervention, but this is often infeasible in many settings (Shadish, Clark, Steiner, & Hill, 2008). Tests for nonconsent bias in the benchmark RFE, however, can provide an important signal as to potential confounding in the nonexperimental estimates.

Background and Literature

Peer Counseling

Peer counselors are women who have breastfed successfully in the past and who provide encouragement and advice to new mothers during pregnancy and after birth. The ultimate goal is to support exclusive breastfeeding for at least 6 months, which the literature suggests is necessary to reap the full benefits of breastfeeding (American Academy of Pediatrics, 2012). In 2011, the U.S. Surgeon General recommended that peer counseling become a core service of the WIC program. By 2014, 69% of all WIC agencies in the United States offered some type of peer counseling support (Epstein & Collins, 2015).

The effectiveness of peer counseling remains ambiguous. Several observational studies have evaluated the effect of peer counseling among WIC clients in agencies with and without a peer counseling program. Each of these studies reported increases in breastfeeding initiation in local WIC

agencies with peer counseling, but sample sizes were small and research designs were weak (Bolton, Chow, Benton, & Olson, 2009; Gill, Reifsnider, & Lucke, 2007; Grummer-Strawn & Mei, 2004; Grummer-Strawn et al., 1997; Schafer, Vogel, Viegas, & Hausafus, 1998; Shaw & Kaczorowski, 1999). Larger studies of WIC clients—18,789 in Maryland and 29,881 from Missouri—were able to adjust estimated program effects with a sizable 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 counseling capacity (Olson, Haider, Vangjel, Bolton, & Gold, 2008). Peer counseling was associated with a 7.4% points increase in any breastfeeding for at least 6 months relative to the mean of 10.4% among those in the comparison group. A follow-up study of the same program reached similar conclusions. Women who received peer counseling were 8.7% points more likely to be breast-feeding at 6 months relative to the comparison group (Haider, Chang, Bolton, Gold, & Olson, 2014).

Three high-quality RCTs of peer counseling among low-income women in the United States reported significant increases in breastfeeding initiation and duration (Anderson et al., 2005; Bonuck et al., 2005; Chapman et al., 2004). Each study 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). The researchers reported significant differences in nonexclusive breastfeeding up to 6 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 at least three home visits postpartum (Chapman et al., 2004). The study found no differences in exclusive breastfeeding at any point postpartum, but women in the treatment group were less likely not to breastfeed at 1 and 3 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 3 months, the risk of nonexclusive breastfeeding was higher among the controls than among those in the intervention group (Risk ratio = 1.30, $p < .05$).[3]

Common characteristics of the three RCTs are their sample size, between 50 and 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 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, Damio, Young, Chapman, and Pérez-Escamilla (2005) reports increases in exclusive breastfeeding as a result of three prenatal and daily hospital visits by the peer counselor, followed by nine postpartum, in-person visits. The support provided by the current peer counseling programs offered by WIC do not come close to the level of service provided in the three RCTs. For WIC to offer the scale of peer support provided in the RCTs nationally would require very ambitious funding.

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 entirely by telephone could achieve substantial gains in exclusive breastfeeding.[4] The health benefits of breastfeeding are largely limited to exclusive breastfeeding for at least 6 months (American Academy of Pediatrics, 2012). The Oregon WIC program focused on exclusive breastfeeding because 90% of WIC participants in Oregon ever breastfeed, the highest rate of any breastfeeding in the nation among WIC recipients. However, less than 40% of Oregon WIC clients exclusively breastfeed for 6 months (see https://www.oregon.gov/oha/PH/HEALTHYPEOPLEFAMI LIES/WIC/Documents/hcpt-bf-support-education.pdf [last accessed on August 13, 2018]).

As detailed in Reeder, Joyce, Sibley, Arnold, and Altindag (2014), the intent-to-treat (ITT) results were mixed. The probability of any breastfeeding for at least 3 months among women assigned to the treatment group was 22% greater than women in the control group, but gains in exclusive breastfeeding were limited to Spanish-speaking clients only. The authors could only speculate as to why the treatment effects with respect to exclusive breastfeeding were limited to Spanish-speaking clients. They noted that women who self-identified as Hispanic but who conducted their interviews in English were much less likely to exclusively breastfeed than Spanish-speaking clients. A similar pattern was evident in national data on breastfeeding as collected by the U.S. Centers for Disease Control and Prevention. Understanding why Spanish-speaking immigrants may be more receptive to peer counseling provided by counselors who spoke their language needed further study (Reeder et al., 2014).

Combining Experimental and Observational Data

One of the earliest and most famous examples of combining observational with experimental data is the 1954 polio vaccine trial. Researchers conducted a double-blind placebo-controlled experiment with over 400,000 participants, which ran simultaneously with an observational study of over 900,000 children whose parents had volunteered to have their children vaccinated. Results in both studies were consistently supportive of vaccine efficacy. There were, however, clear patterns of positive selection bias in the observational arm as the incidence of polio was lower among both the vaccinated and unvaccinated children in the observational arm compared to the vaccinated and unvaccinated in the experimental arm (Meier, 2006). In a more recent and well-known example, LaLonde (1986) tested whether econometric approaches to correct for selection bias in observational studies could mimic the findings from a field experiment of a training and employment program. Lalonde replaced the controls from the experiment with a comparison group drawn from extant sources and reestimated the training program effects. He concluded that econometric techniques were inadequate at replicating the results from the randomized trial. Subsequent studies using LaLonde's data were less pessimistic (Dehejia, 2005; Dehejia & Wahba, 1999; Heckman, Ichimura, & Todd, 1997; Smith & Todd, 2005) This and subsequent work were termed "within-study" designs or within-study comparisons (WSCs).

The use of WSCs has grown with the increase in RFEs, and design elements of the most successful WSCs have been refined (Cook et al., 2008; Wong, Steiner, & Anglin, 2018). There are two broad type of WSCs (Wong & Steiner, 2018). In an independent WSC, subjects are recruited and randomly assigned to an experimental arm and a choice arm. Those in the experimental arm are randomized again between treatments. Those in the choice arm select their treatment. This design is referred to as doubly randomized preference trial.[5] Shadish, Clark, Steiner, and Hill (2008) were the first to use an independent WSC to test for bias in nonexperimental settings by comparing treatment effects obtained from the experimental arm to those in choice arm after applying various forms of adjustment. The great advantage of the independent WSC is that randomization between arms insures that participant characteristics are balanced in expectation. Implementing an independent WSC in a field setting, however, is challenging as researchers must recruit sufficient participants to power both arms.

The more common form of a WSC generates nonexperimental estimates of treatment effects by pairing the treatment group from the experiment with a comparison group drawn from other sources. Wong and Steiner (2018) term this a dependent WSC because of the dual use of the treatment group in generating both experimental and nonexperimental estimates. The quality of the WSC depends on the similarity of the comparison group to those in the treatment group not only in participant characteristics but how, when, and where data are collected. The advantage of the dependent WSC is the potential availability of large numbers of subjects for the comparison group. The disadvantage is the difficulty of identifying the source of non-experimental bias when there are significant differences between the estimated experimental and nonexperimental treatment effects. The main challenge is to assess whether these differences are due to how the outcomes are measured, the limited set of covariates used for adjustment, or whether the experimental estimates are exceptional.

Testing for Correspondence

Another challenge in WSCs is determining whether the difference between the experimental and nonexperimental estimates is sufficiently small to imply correspondence between the two. An obvious start is to compare the absolute difference between the nonexperimental and experimental estimates ($\tau_{NE} - \tau_{RFE}$). Failure to reject the null hypothesis of no difference would suggest that the two estimates correspond. As Steiner and Wong (2018) note, however, researchers should rule out differences due to sampling error. They recommend using a "tolerance threshold" (δ) that analysts would consider an inconsequential difference given the context. What constitutes a negligible difference is clearly subjective, so Steiner and Wong (2018) recommend 0.1 of a standard deviation (SD) of the outcome. The formal test of equivalence uses a composite null of the form H_{01}: $\tau_{NE} - \tau_{RFE} \geq \delta$ and H_{02}: $\tau_{NE} - \tau_{RFE} \leq -\delta$, whereas the rejection of both one-sided nulls is evidence of equivalence. In an effort to jointly interpret the difference and equivalence tests, we follow Tryon and Lewis (2008) and Steiner and Wong (2018) and use a two-by-two matrix that combines tests of statistical difference with tests of equivalence. The approach allows for indeterminate inferences when the test of equivalence fails to reject the composite null, but the difference test finds evidence of correspondence.

Statistical tests of significance and equivalence necessitate a standard error for differences between experimental and nonexperimental estimates. In an independent WSC, randomization insures the covariance of the two

estimates is zero in expectation. In the dependent WSC, however, the treatment group is used in both experimental and nonexperimental estimates insuring a nonzero covariance. Following Steiner and Wong (2018), we bootstrap the standard error of the difference, which allows the treatment arm to contribute to the variance–covariance matrix in both experimental and nonexperimental estimates.

Empirical Framework

Data

The Oregon RFE. The Oregon RFE involved 1,885 women in four WIC agencies who registered for WIC during their prenatal visit and who agreed to participate. Participants were stratified between English and Spanish speakers and then randomized into 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 2 weeks before the expected due date, the third within 1-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 and 4 postpartum. There were no meaningful differences in the breastfeeding outcomes between women in the low- and high-intensity groups, so researchers combined the two treatment groups. The ITT results with respect to exclusive breastfeeding were mixed. Spanish-speaking clients were between 6% and 8% points more likely to exclusively breastfeed at 1, 3, and 6 months postpartum relative to the controls, but there was no effect of peer counseling among English-speaking clients (see Reeder et al., 2014, for a detailed description of the RFE).

The dependent WSC. To implement the dependent WSC, we combine data from Oregon's RFE on peer counseling with administrative data on WIC participants in 27 agencies with no peer counseling services during the period of the RFE: July 2005 through July 2007.[6] A map of Oregon indicating the treatment and comparison agencies is displayed in Figure 1. The flowchart in Figure 2 shows the data used in the dependent WSC. The 560 WIC clients in control group of the RFE are replaced by 24,857 WIC clients
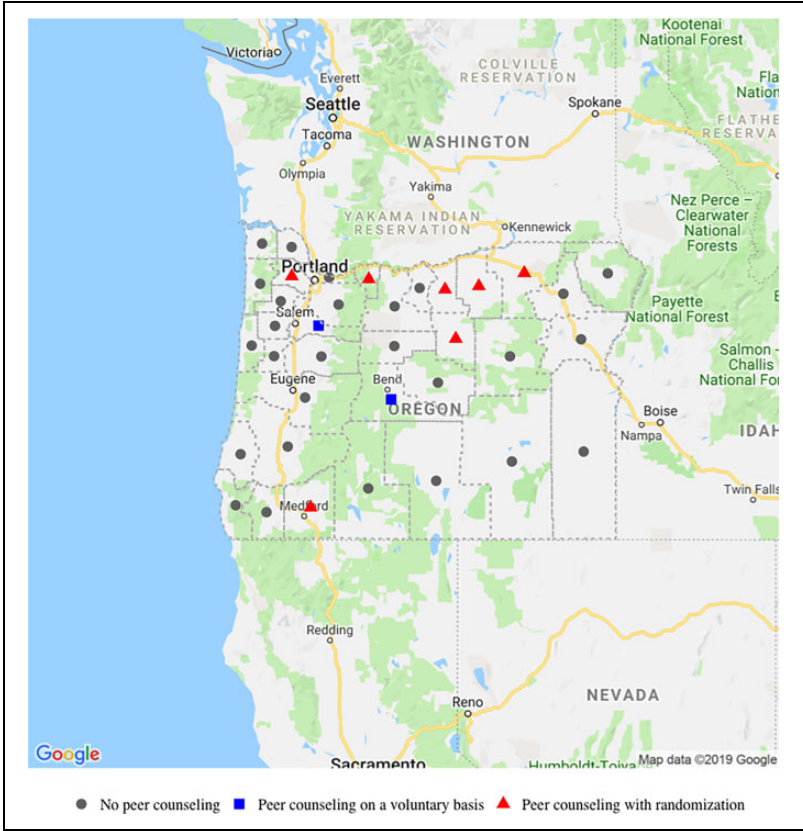
Figure 1. Counties in Oregon with and without peer counseling for breastfeeding, 2005–2007.

from the 27 WIC agencies that did not provide peer counseling during the study period.

The quasi-independent WSC. In the quasi-independent WSC, we compare the effects of peer counseling on exclusive breastfeeding from the RFE with nonexperimental estimates obtained from WIC clients who enrolled in peer counseling in three of the four RFE agencies in the 3 years immediately following the RFE (July 2007 and July 2010). The bottom row of Figure 3 shows the experimental and nonexperimental samples. Similar to an independent WSC, the treatment group is not used to obtain the
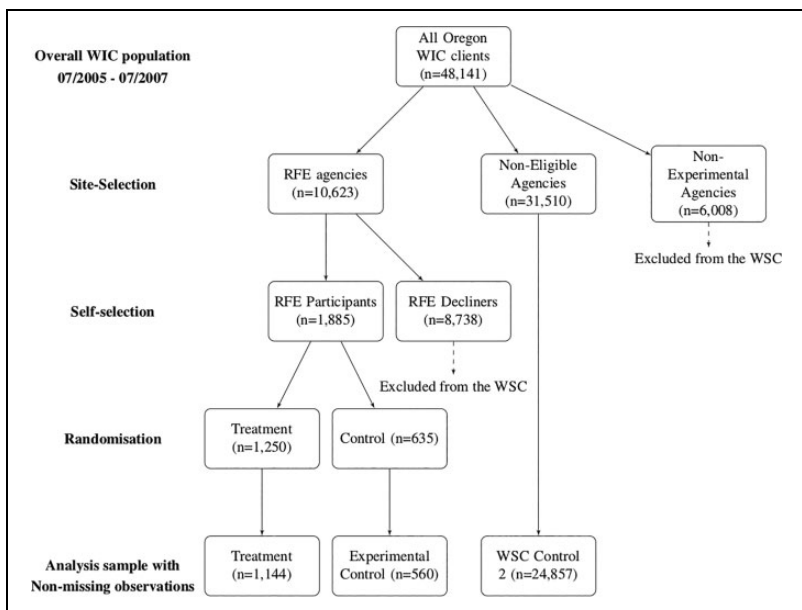
Figure 2. Within-study comparison design: dependent arm approach. *Adapted from* Wong and Steiner (2018).

nonexperimental estimates. In a truly independent WSC, the experimental and nonexperimental arms of the study would have been randomly drawn from the 1,885 participants of the RFE (Figure 3). In lieu of random assignment to the choice arm, we use the 14,716 women from three WIC agencies in the RFE who enrolled in WIC in the 3 years immediately following the RFE as the nonexperimental group. The assumption is that the nonexperimental sample is "as-if" randomly assigned given the same clinics, many of the same staff, and WIC clients enrolled from the same counties.

Difference-in-differences (DD) Analysis. The DD analysis includes all women who enrolled in WIC between July 2005 and July 2010. We exclude WIC clients from the treatment group of the RFE. To account for this loss, we weight the controls by the inverse of the proportion of their sample size in the RFE. Randomization insures that weighting provides the population prevalence of breastfeeding in the experimental counties in the absence of peer counseling services. We then estimate the probability of exclusive breastfeeding before and after the availability of peer counseling services in
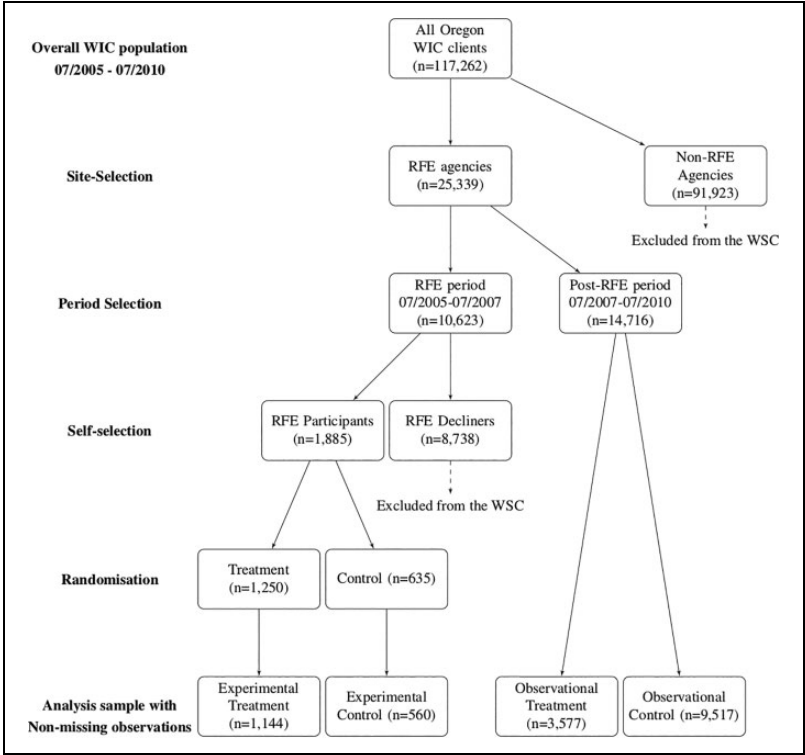
Figure 3. Within-study comparison design: quasi-independent arm approach. *Adapted from* Wong and Steiner (2018).

July 2007 among WIC clients in 3 counties that offered the services relative to the 2 counties that always offered peer counseling and the remaining 30 counties (27 agencies) that never provided peer counseling services.

Outcomes and covariates. We assess the impact of the peer counseling service on exclusive breastfeeding at 1, 3, and 6 months postpartum. At each certification visit up to 2 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. Covariates include the WIC client's age, educational attainment, family income, marital status, race/ethnicity, spoken language, and month of enrollment in WIC.

We focus on exclusive breastfeeding because the explicit goal of Oregon's peer counseling initiative was to increase the prevalence of exclusive breastfeeding for at least 6 months. The health benefits of breastfeeding have been associated with exclusive breastfeeding rather than any breastfeeding (Kramer et al., 2001). In addition, exclusive breastfeeding is reported more completely than any breastfeeding.[7] In the Oregon RFE, the duration of any breastfeeding was missing in 19% of cases, whereas exclusive breastfeeding was missing for only 8% (Reeder et al., 2014).

A key feature of our study is that we have data not only on the characteristics of WIC clients statewide but their breastfeeding outcomes as well. In addition, all data are standardized across WIC agencies. Program staff at each local agency enter all information into the State's centralized Information System Tracker database. This provides consistency in the collection and measurement of data, which is an important requirement of a WSC.

Results

We present the results in three parts. The first section contains the results from the dependent WSC. In the second section, we compare estimates of peer counseling on exclusive breastfeeding from the RFE to nonexperimental estimates in three of the four RFE agencies after the experiment ended, which we term as a quasi-independent WSC. In the last section, we present results from the statewide DD analysis of peer counseling.

Dependent WSC

Table 1 shows the mean characteristics or participants in the RFE (column 1), WIC clients in rest of the state without peer counseling services (column 2), and the normalized differences between the two (column 3). The top panel pertains to English-speaking clients and the middle panel to Spanish-speaking clients. The bottom panel shows three characteristics of the counties in which the WIC agencies are located: the percentage of the population that is poor, the percentage on food stamps, and the percentage of population who identify as Hispanic. The normalized differences suggest no major imbalances among the individual and county characteristics between the two groups. We also tested for differences in the ratio of variances between groups, and none showed evidence of imbalance.[8]

Table 1. Characteristics of Oregon WIC Participants in Experimental and Nonexperimental Samples, 2005–2010.

English-Speaking WIC Clients

	RFE		Rest of State		Post-RFE	
	Mean		Mean	Norm. Diff.	Mean	Norm. Diff.
Age	25.9		24.8	0.14	25.0	0.12
Education						
<High school	0.27		0.21	0.09	0.25	0.03
≥High school	0.71		0.74	−0.04	0.73	−0.03
Unknown	0.03		0.05	−0.09	0.02	0.01
Marital status						
Single	0.44		0.46	−0.03	0.45	−0.02
Married/partner	0.53		0.45	0.11	0.50	0.04
Unknown	0.03		0.09	−0.18	0.05	−0.06
Family income	1,323.3		1,133.4	0.13	1,226.5	0.07
N	889		19,351		9,303	

Spanish-Speaking WIC Clients

	RFE		Rest of State		Post-RFE	
	Mean		Mean	Norm. Diff.	Mean	Norm. Diff.
Age	28.9		27.0	0.22	28.1	0.09
Education						
<High school	0.57		0.59	−0.03	0.53	0.06
≥High school	0.38		0.33	0.07	0.42	−0.06
Unknown	0.05		0.08	−0.08	0.05	0.01

(continued)

165

Table 1. (continued)

Spanish-Speaking WIC Clients

| | RFE | Rest of State | | Post-RFE | |
	Mean	Mean	Norm. Diff.	Mean	Norm. Diff.
Marital status					
Single	0.20	0.20	−0.01	0.24	−0.07
Married/partner	0.77	0.71	0.10	0.70	0.12
Unknown	0.03	0.09	−0.17	0.07	−0.11
Family income	1,564.5	1,378.2	0.16	1,501.0	0.05
N	815	5,506		3,791	

Population Characteristics of Counties as of 2005

| | RFE | Rest of State | | Post-RFE | |
	Mean	Mean	Diff.	Mean	
% poor	0.11	0.14	−0.03	0.11	0.00
% food stamps	0.09	0.12	−0.03	0.09	0.00
% Hispanic	0.13	0.09	0.04	0.13	0.00

Note. This table reports the average background characteristics of WIC clients in the randomized field experiment (RFE) and contrasts them with WIC clients from different samples. There are 34 local WIC agencies in Oregon. Four agencies participated in the RFE from July 2005 to July 2007, although one agency dropped out of the study within the first year due to staffing problems and weak enrollment. Twenty-eight agencies had no peer counseling services (PC) throughout the study period, and we refer to WIC clients in these agencies as the Rest of State. Two agencies offered PC services but were not part of the REF. The Post-RFE sample includes WIC clients from July 2007 to July 2010 from the three agencies that participated in the REF for the complete study period and that offered PC services on a nonexperimental basis after the RFE was completed. Population characteristics of counties that the agencies serve come from Area Resource Files, 2005. WIC = Supplemental Nutrition Program for Women, Infants and Children.

Table 2 displays results from the dependent WSC. We display the ITT estimates of peer counseling on exclusive breastfeeding in the column (1) and the nonexperimental estimates in column (3). We adjust all estimates for individual and county characteristics by ordinary least squares with robust standard errors are in parentheses. The top panel is for English-speaking clients, and the bottom panel is for Spanish-speaking clients. As reported in Reeder et al. (2014), we find no effect of peer counseling among English-speaking clients in the RFE but positive and statistically significant effects of peer counseling on exclusive breastfeeding in the nonexperimental estimates at 1 and 3 months. We fail to reject the null hypothesis of no difference (H_0: $\Delta = \tau_{NE} - \tau_{RFE} = 0$) at 1 and 3 months. We also fail, however, to reject the null of nonequivalence (H_0: $|\Delta| \geq 0.1$ SD) between the nonexperimental and RFE estimates at all three points postpartum among English-speaking clients. Correspondence between the two estimates, thus, is indeterminate.

The WSC results for Spanish-speaking clients are similar. Peer counseling increases exclusive breastfeeding among participants of the RFE at each point postpartum, while the nonexperimental estimates are even larger. Specifically, peer counseling increases exclusive breastfeeding at 6 months by 6.8% points in the RFE and by a much larger 14.7% points based on the nonexperimental estimates, a statistically significant difference of 7.9% points. Similarly, we cannot reject the composite null of nonequivalence. Following Steiner and Wong (2018), we conclude the experimental and nonexperimental estimates at 6 months postpartum among Spanish-speaking clients are different or fail the test of correspondence (Table 2).

Quasi-Independent WSC

We again display the ITT estimates of peer counseling on exclusive breastfeeding from the RFE in Table 3 (column 1). We contrast them with the nonexperimental estimates obtained from WIC clients in the post-RFE period. The advantage of this comparison is that the within-agency comparison may better control for hard-to-measure differences in agency personal, agency culture, and the WIC clients specific to these counties. The results with respect to the correspondence between the experimental and nonexperimental estimates in Table 3 are the same as in Table 2: statistical indeterminacy. Specifically, we fail to reject the null of no difference (H_0: $\Delta = 0$) suggesting correspondence, but then we fail to reject the null of nonequivalence (H_0: $|\Delta| \geq 0.1$ SD).

Table 2. Within-Study Comparison of Intent-to-Treat Effects—Dependent Simultaneous Design.

English-Speaking WIC Clients

Exclusive Breastfeeding Duration	RFE		Rest of State		$H_0: \Delta = 0$ $H_0: |\Delta| \geq 0.1$ SD	$H_A: \Delta \neq 0$ $H_A: |\Delta| < 0.1$ SD	Overall Result
	τ_{RFE}	$\bar{y}$	τ_{NE}	$\Delta = \tau_{NE} - \tau_{RFE}$			
One month	0.028 (0.035)	0.55 [0.50]	0.067 (0.021)**	0.040 (0.030)	Fail to reject	Fail to reject	Indeterminacy
Three months	0.021 (0.033)	0.36 [0.48]	0.049 (0.020)*	0.027 (0.028)	Fail to reject	Fail to reject	Indeterminacy
Six months	−0.022 (0.029)	0.21 [0.41]	0.026 (0.017)	0.048 (0.023)†	Fail to reject	Fail to reject	Indeterminacy
N	889		19,931				

Spanish-Speaking WIC Clients

Exclusive Breastfeeding Duration	RFE		Rest of State		$H_0: \Delta = 0$ $H_0: |\Delta| \geq 0.1$ SD	$H_A: \Delta \neq 0$ $H_A: |\Delta| < 0.1$ SD	Overall Result
	τ_{RFE}	$\bar{y}$	τ_{NE}	$\Delta = \tau_{NE} - \tau_{RFE}$			
One month	0.087 (0.038)*	0.56 [0.50]	0.125 (0.022)**	0.038 (0.033)	Fail to reject	Fail to reject	Indeterminacy
Three months	0.095 (0.037)*	0.46 [0.50]	0.146 (0.022)**	0.051 (0.031)	Fail to reject	Fail to reject	Indeterminacy
Six months	0.068 (0.036)†	0.35 [0.48]	0.147 (0.021)**	−0.079 (0.030)**	Reject	Fail to reject	Difference
N	815		6,070				

Note. This table reports the intent-to-treat effects of peer counseling on exclusive breastfeeding for 1 month, 3 months, and 6 months from different samples. Outcome mean for the control group in randomized field experiment is indicated with $\bar{y}$ sample standard errors in brackets. The regressions control for age, age squared, natural logarithm of family income, high school graduation, marital status, and agency fixed effects plus indicators for missing observations for each of the independent variables. The columns "diff." show the difference between the nonexperimental and experimental estimate with their bootstrapped standard errors in parentheses. The standard errors in individual intent-to-treat effects are robust to general form of heteroskedasticity. The test of equivalence is based on rejection of the composite null hypothesis ($H_0: |\Delta| \geq 0.1$ SD). The overall test of equivalence combines results from the absolute difference ($H_0: \Delta = 0$) and the composite null ($H_0: |\Delta| \geq 0.1$ SD) following Steiner and Wong (2018). The samples are described in the notes to Table 1. Significance levels are indicated by †10%, *5%, and **1%.

Table 3. Within-Study Comparison of Intent-to-Treat Effects—Quasi-Independent Simultaneous Design.

English-Speaking WIC Clients

Exclusive Breastfeeding Duration	RFE		post-RFE		$H_0: \Delta = 0$ $H_0: |\Delta| \geq 0.1$ SD		Overall Result
	τ_{RFE}	$\bar{y}$	τ_{NE}	$\Delta = \tau_{NE} - \tau_{RFE}$	$H_A: \Delta \neq 0$	$H_A: |\Delta| < 0.1$ SD	
One month	0.028 (0.035)	0.55 [0.50]	0.029 (0.012)*	0.002 (0.036)	Fail to reject	Fail to reject	Indeterminacy
Three months	0.021 (0.033)	0.36 [0.48]	0.030 (0.011)**	0.008 (0.034)	Fail to reject	Fail to reject	Indeterminacy
Six months	−0.022 (0.029)	0.21 [0.41]	0.020 (0.009)*	0.042 (0.029)	Fail to reject	Fail to reject	Indeterminacy
N	889		9,303				

Spanish-Speaking WIC Clients

Exclusive Breastfeeding Duration	RFE		post-RFE		$H_0: \Delta = 0$ $H_0: |\Delta| \geq 0.1$ SD		Overall Result
	τ_{RFE}	$\bar{y}$	τ_{NE}	$\Delta = \tau_{NE} - \tau_{RFE}$	$H_A: \Delta \neq 0$	$H_A: |\Delta| < 0.1$ SD	
One month	0.087 (0.038)*	0.56 [0.50]	0.031 (0.017)†	−0.056 (0.041)	Fail to reject	Fail to reject	Indeterminacy
Three months	0.095 (0.037)*	0.46 [0.50]	0.042 (0.016)*	−0.052 (0.042)	Fail to reject	Fail to reject	Indeterminacy
Six months	0.068 (0.036)†	0.35 [0.48]	0.038 (0.015)*	−0.030 (0.038)	Fail to reject	Fail to reject	Indeterminacy
N	815		3,791				

Note. This table reports the intent-to-treat effects of peer counseling on exclusive breastfeeding for 1 month, 3 months, and 6 months from different samples. Outcome mean for the control group in randomized field experiment is indicated with $\bar{y}$ sample standard errors in brackets. The regressions control for age, age squared, natural logarithm of family income, high school graduation, marital status, and agency fixed effects plus indicators for missing observations for each of the independent variables. The columns "diff." shows the difference between the nonexperimental and experimental estimate with their bootstrapped standard errors in parentheses. The standard errors in individual intent-to-treat effects are robust to general form of heteroskedasticity. The test of equivalence is based on rejection of the composite null hypothesis ($H_0: |\Delta| \geq 0.1$ SD). The overall test of equivalence combines results from the absolute difference ($H_0: \Delta = 0$) and the composite null ($H_0: |\Delta| \geq 0.1$ SD) following Steiner and Wong (2018). The samples are described in the notes to Table 1. Significance levels are indicated by †10%, *5%, and **1%.

Despite the statistical indeterminacy as to the correspondence between the experimental and nonexperimental estimates in both the dependent and quasi-independent WSC, there are noticeable differences between the estimates in Tables 2 and 3. Among the English-speaking clients in the quasi-independent WSC (Table 3), the nonexperimental estimates are half the magnitude of the nonexperimental estimates obtained using WIC clients from the rest of the state as the comparison group (Table 2). The nonexperimental estimates of peer counseling among Spanish-speaking clients in the bottom half of Table 3 are also substantially smaller than the nonexperimental estimates among Spanish-speaking clients in the dependent WSC (Table 2). And yet, such small differences between the experimental and nonexperimental estimates do not have the statistical power to reject nonequivalence. This underscores the statistical power needed to achieve equivalence with a tolerance threshold of 0.1 *SD*s in this setting. For instance, we would need a tolerance threshold to be between 0.12 and 0.22 *SD*s among English speakers and between 0.19 and 0.25 *SD*s among Spanish speakers to achieve full correspondence between experimental and nonexperimental estimates holding all else constant.

Nonconsent Bias in the RFE

A limitation of our WSC is the restricted set of covariates with which to adjust the nonexperimental estimates. The lack of correspondence could be due to omitted variables, despite the lack of imbalance among observable characteristics between WIC clients in the treatment and comparison groups. Fortunately, we know the breastfeeding outcomes of women who declined to participate in the RFE. In an effort to gauge the importance of nonconsent bias net of observable characteristics, we compare the breastfeeding outcomes of women in the control group of the RFE with those who declined to participate when offered. Neither group had access to peer counseling and thus any difference in exclusive breastfeeding adjusted for the observables would point to nonconsent bias (Kramer, 1984; Marcus, 1997).[9] In Table 4, we show the estimates from the following regression:

$$BF_i = \rho_0 + \rho_1 C_i + X_i' \Psi + \eta_i, \tag{1}$$

where C_i equals 1 if the woman was in the control group and 0 if she declined to participate. Let X_i be the set of observable characteristics listed in Table 1. If women who agree to participate in the RFE are more inclined

Table 4. Exclusive Breastfeeding Duration Differences in RFE Participants and Nonparticipicants, 2005–2007.

Exclusive breastfeeding for	English-Speaking WIC Clients			Spanish-Speaking WIC Clients		
	1 Month	3 Months	6 Months	1 Month	3 Months	6 Months
Difference	0.07 (0.03)†	0.05 (0.03)†	0.06 (0.02)*	0.00 (0.03)	0.02 (0.03)	0.05 (0.03)†
Mean outcome	0.47	0.30	0.17	0.52	0.38	0.25
N		4,831			2,523	

Note. This table reports the differences in exclusive breastfeeding between the control group from Oregon's randomized field experiment (RFE) and WIC participants in the same agencies who were offered participation in the RFE but declined. The reported differences are estimated with ordinary least squares regression. The regressions control for age, age squared, natural logarithm of family income, high school graduation, marital status, and agency fixed effects plus indicators for missing observations for each of the independent variables. Heteroskedasticity-consistent standard errors are in parentheses. Significance levels are indicated by †10%, *5%, and **1%.

to exclusively breastfeed than those who chose not to participate, then ρ_1 should be positive. Among the English-speaking patients, for instance, there is clear evidence of selection into the RFE (Table 4). Women assigned to the control group and who receive no peer counseling are between 5% and 7% points more likely to exclusively breastfeed at 1, 3, and 6 months than women who chose not to participate in the RFE and who had no access to peer counseling. The 6% point difference in exclusive breastfeeding at 6 months is large given a mean of 17%. There is less evidence of nonconsent bias among the Spanish-speaking clients except at 6 months postpartum.

Nonconsent bias has important implications for WSC. Eighty-three percentage of WIC clients eligible to participate in the RFE declined to join. These women appeared to differ in the baseline propensity to exclusive breastfeed relative to the RFE participants. The proportion of WIC clients from agencies not involved in the RFE who would not participate in an RFE if offered is likely to have similar propensities toward exclusive breastfeeding. Thus, the presence of nonconsent bias among the RFE agencies suggests that differences between experimental and nonexperimental estimates of the effect of peer counseling on exclusive breastfeeding based on comparison groups drawn from WIC agencies in the rest of the state are likely to be similarly confounded. This might explain the lack of correspondence in the WSCs despite no statistical difference between the experimental and nonexperimental treatment effect estimates.

DD

In this section, we present estimates of a reduced-form DD research design to compare exclusive breastfeeding among WIC clients in counties in which peer counseling services were available on a voluntary basis relative to counties that never or always offered these services.[10] Specifically, we took all women in three experimental counties who did not participate in the RFE and combined them with the controls from the RFE and weighted the controls by the inverse of the proportion of RFE participants in the treatment group. This provided the correct estimate of breastfeeding among all women WIC participants in the three counties who were unexposed to peer counseling. As noted previously, these three counties offered peer counseling in a nonexperimental setting after the RFE was completed. To estimate the DD, we compared the changes in exclusive breastfeeding in these counties before and after voluntary peer counseling services were offered relative to the change in breastfeeding among counties that never or always had peer counseling services.

The DD estimate is directly comparable to the ITT estimates from the RFE under two assumptions: first, that the availability of peer counseling services was quasi-randomly assigned and second, that trends in exclusive breastfeeding prior to the availability of peer counseling services were the same in the two groups of counties. Given that the DD is estimated at the aggregate level, we do not apply a direct test of correspondence with estimates from the RFE. Nevertheless, with the availability of administrative data, the DD provides a readily accessible approach to program evaluation. We estimate the DD with following regression:

$$BF_{ijt} = \theta_0 + \theta_1 PC_{jt} + X_i \pi + \tau_t + \lambda_j + \eta_{ijt}. \qquad (2)$$

As before, BF_{ijt} is a dichotomous indicator of whether WIC client i, in county j, who delivered in quarter t, exclusively breastfed; PC_{jt} is one if peer counseling was available in county j, and period t and zero if not; X_i is a vector of client characteristics and the terms τ_t and λ_j control for quarter of birth and agency fixed effects. The coefficient, θ_1, is the estimated difference in exclusive breastfeeding among WIC participants in the three agencies from before to after peer counseling services became available on a voluntary basis relative to the change in exclusive breastfeeding in agencies that never offered or always offered peer counseling services. The standard errors are clustered at agency level.

We show quarterly rates of exclusive breastfeeding at 1, 3, and 6 months postpartum by language and whether the WIC clients lived in a county with peer counseling (Figure 4). The shaded bar marks the beginning of peer counseling services offered on a voluntary basis in three WIC agencies that participated in the RFE. Among the English-speaking clients, there is no difference in exclusive breastfeeding at any point postpartum either before or after peer counseling becomes available in the three WIC agencies (Figure 4, top panel, dark line) relative to the rest of the state. The preintervention trends visually support the parallel trend assumption and the lack of any effect of peer counseling availability on exclusive breastfeeding is consistent with the ITT results for English-speaking clients from the RFE. Among the Spanish-speaking clients, there is a modest separation in rates of exclusive breastfeeding between the two groups of counties prior to the availability of peer counseling and that difference becomes larger after peer counseling is available (Figure 4, bottom panel). Again, the parallel trend assumption appears to hold, but the data are noisier due to the smaller number of Spanish speakers in our sample. There is some visual evidence of an increase in exclusive breastfeeding among Spanish-speaking clients associated with the availability of peer counseling services.[11]
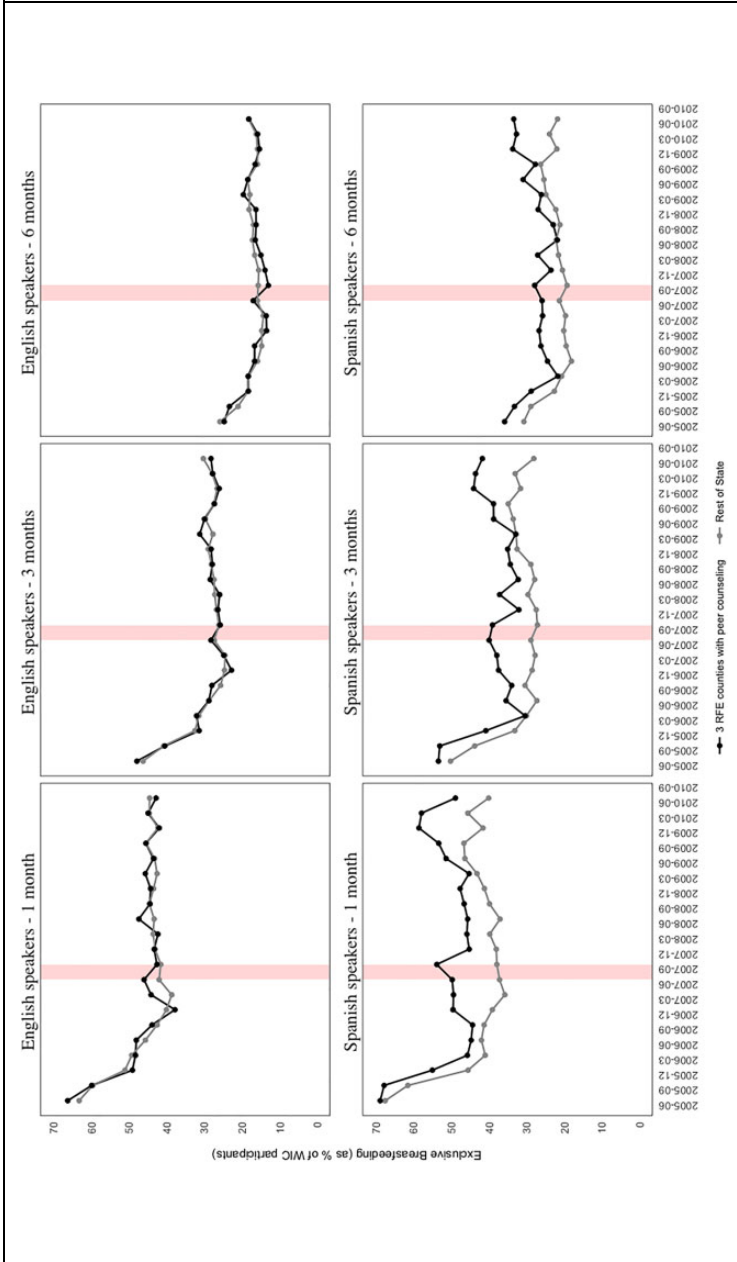
Figure 4. Trends in exclusive breastfeeding by the availability of peer counseling services, 2005–2010. The pink bar indicates the third quarter of 2007, after which the peer counseling was available in three randomized field experiment counties. No peer counseling was available at any time in the rest of State.

Estimates of θ_1 from Equation 2 are shown in Table 5. These DD estimates seem to confirm what appears visually apparent in Figure 4: no association between the availability of peer counseling services and exclusive breastfeeding among English speakers. Even estimates at the tail of the 95% confident interval offer no substantive clinical impact. By contrast, the DD estimates for Spanish speakers indicate that the availability of peer counseling increases exclusive breastfeeding by 4.2% and 2.5% points at 1 and 3 months, respectively. These estimates are substantially smaller than the ITT estimates from the RFE. However, because 90% of Spanish-speaking clients assigned to the treatment group in the RFE used a peer counselor, the ITT estimates from the RFE are approximately equal to the effect of treatment on the treated. To make the DD results more comparable to those from the RFE, we inflate the DD estimates by the take-up rate among Spanish-speaking clients (31%). These estimates suggest that those who receive peer counseling increase exclusive breastfeeding by 13.5% (4.2/0.31) points at 1 month and 8.1% (2.5/0.31) points at 3 months. These point estimates are larger than those from the RFE and not surprisingly, similar to the nonexperimental estimates that used the WIC clients from the rest of the state as the comparison group (Table 2). The DD estimates in Table 5 inflated by the peer counseling take-up rate are also much larger than the nonexperimental estimates from the quasi-independent WSC in Table 3.

The interpretation of the DD as unbiased estimate of ITT rests on the assumption that trends in exclusive breastfeeding prior to the availability of services between the two groups of counties are parallel. As further test of the latter, we allow the DD estimates to vary by quarter before and after peer counseling services became available. We then test whether the leads and lags of the DD are jointly different from zero in both the pre- and postperiods. The results are displayed in Figure 5.[12] There is no evidence of an association between peer counseling services and exclusive breastfeeding among English-speaking clients. The pattern is more complex among Spanish speakers. Differences in exclusive breastfeeding at both 1 and 3 months postpartum are different from zero in the preperiod as well as in the postperiods. This points to a violation of the parallel-trends assumption suggesting that there are differences in exclusive breastfeeding trends between the three treatment counties and the rest of the state that are not explained by the availability of peer counseling services.

Table 5. Observational Estimates of Peer Counseling Availability on Exclusive Breastfeeding by Duration and Language, 2005–2010.

Exclusive breastfeeding for	English-Speaking WIC Clients			Spanish-Speaking WIC Clients		
	1 Month	3 Months	6 Months	1 Month	3 Months	6 Months
DD estimate	−0.001 (0.01)	−0.002 (0.01)	−0.011 (0.01)	0.042 (0.01)**	0.025 (0.01)**	0.008 (0.01)
Mean outcome	0.449	0.287	0.168	0.450	0.332	0.234
N		76,213			21,986	
PC take-up		0.218** (0.04)			0.307** (0.03)	
N		91,883			24,064	

Note. This table reports the difference-in-differences estimates of peer counseling service availability on exclusive breastfeeding. The reported differences are estimated with ordinary least squares regression. The regressions control for age, age squared, natural logarithm of family income, child's birth quarter-year fixed effects, and agency fixed effects plus indicators for missing observations for each of the independent variables. Standard errors are clustered at agency level and are in parentheses. Significance levels are indicated by *5%, and **1%. WIC = Supplemental Nutrition Program for Women, Infants and Children.
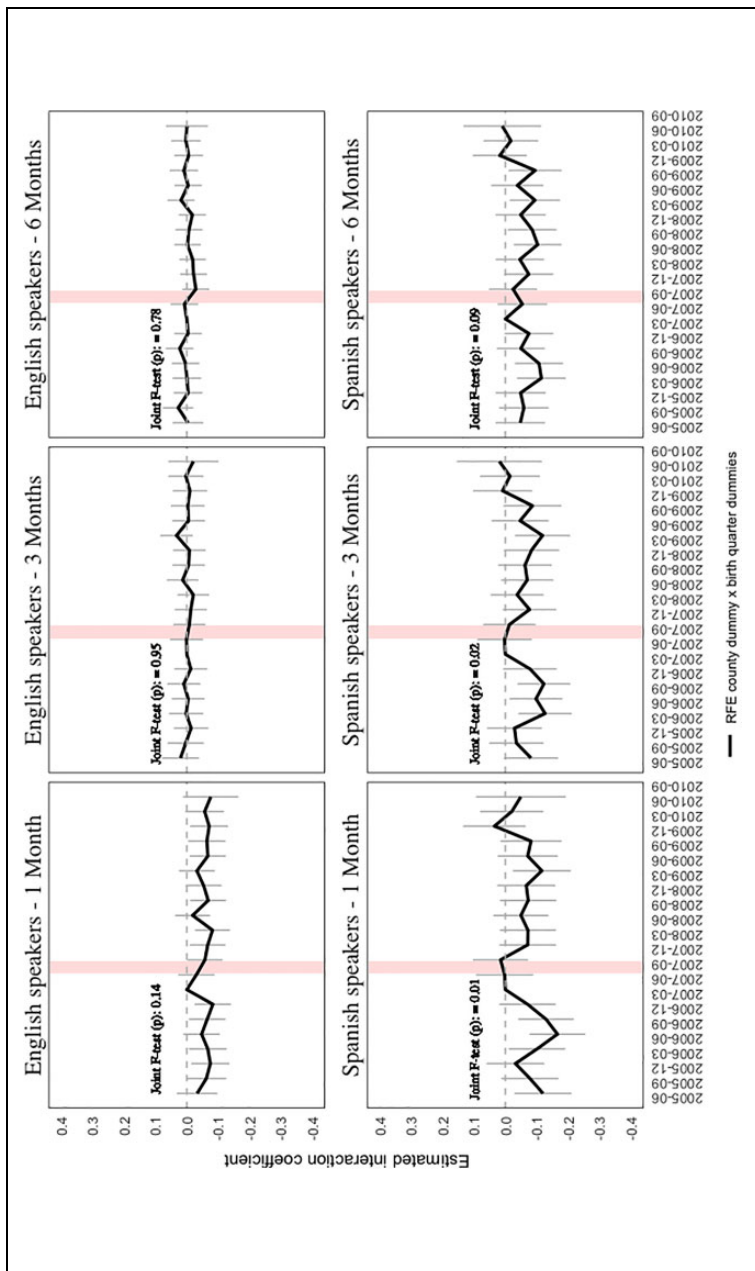
Figure 5. Dynamic difference-in-differences estimates of the availability of peer counseling services on exclusive breastfeeding duration, 2005–2010. Each panel shows the regression coefficients on the interaction between the quarter of birth and a dummy variable indicating the availability of peer counseling services (see footnote 12 in the text). The pink bar indicates the third quarter of 2007, after which peer counseling was available in three randomized field experiment counties. No peer counseling was available at any time in the rest of State.
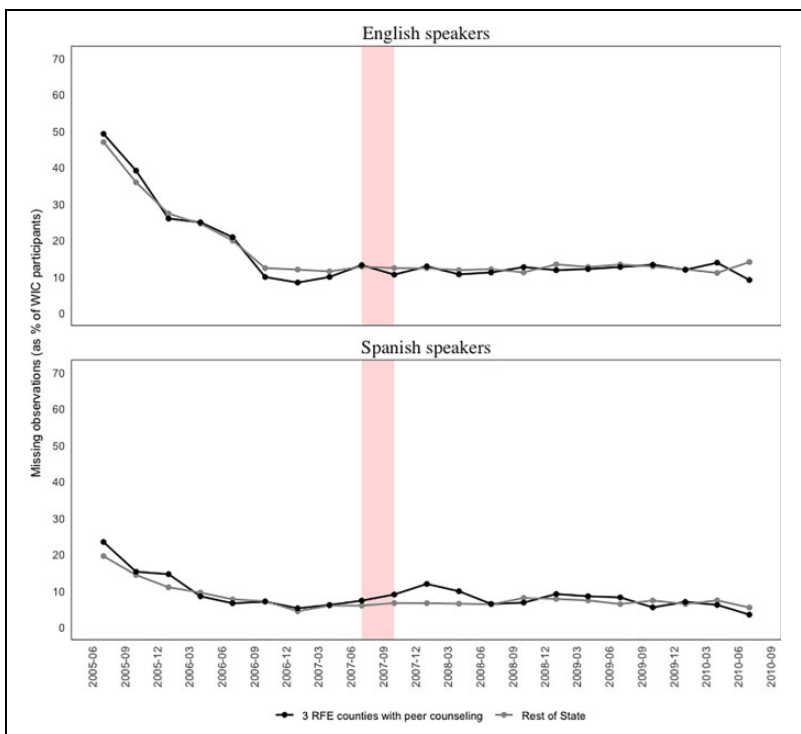
Figure 6. Trends in the missing observations for exclusive breastfeeding by the availability of peer counseling services, 2005–2010. The pink bar indicates the third quarter of 2007, after which the peer counseling was available in three randomized field experiment counties. No peer counseling was available at any time in the rest of State.

Discussion

RFEs provide strong internal validity. The results, however, are specific to places, people, and time periods, which can make them difficult to replicate using nonexperimental methods. In this study, we undertook a WSC of peer counseling programs to promote exclusive breastfeeding among WIC clients in Oregon. The benchmark RFE was the largest RFE of peer counseling in the United States undertaken to date. Given the availability of statewide administrative data on breastfeeding among WIC clients, we obtained nonexperimental estimates in three ways.

First, we undertook a simultaneous, dependent WSC using WIC clients in the rest of the state that had no exposure to peer counseling as the comparison group. There was no statistical difference between the experimental and nonexperimental in all cases but one, Spanish-speaking clients at 6 months postpartum. Tests of correspondence were indeterminate except for this one group of Spanish-speaking clients that lacked correspondence.

The second WSC we termed a quasi-independent WSC. After the RFE was concluded, the WIC program continued to offer the same peer counseling services in three of the four WIC agencies involved in the RFE. We compared the estimates from the RFE to the nonexperimental estimates of peer counseling in the post-RFE period. The advantage of this design is that we implicitly held constant the WIC agency staff, the agency culture, and the characteristics of WIC clients serviced by these agencies. This highlighted an important difference with the dependent WSC that drew a comparison group of WIC clients from other WIC agencies across the state. The nonexperimental estimates from the quasi-independent WSC were smaller than those of the RFE, and in numerous instances, the differences in exclusive breastfeeding were close to zero or clinically irrelevant. Nevertheless, we were unable to infer correspondence between the experimental and nonexperimental estimates.

The third attempt to replicate results from the RFE involved a DD design contrasting the availability of peer counseling services on exclusive breastfeeding in three experimental WIC agencies to changes in the prevalence of exclusive breastfeeding in WIC agencies in the rest of the state. The DD estimates were consistent with those from the RFE. The availability of peer counseling was unassociated with exclusive breastfeeding among English-speaking clients but positively associated with exclusive breastfeeding among Spanish-speaking clients. Despite this congruence, differential trends in exclusive breastfeeding among Spanish-speaking clients prior to the availability of peer counseling services in the treated agencies relative to the rest of the states suggested confounding from unobservable differences between agencies.

Conclusion

Our application of the correspondence test promoted by Steiner and Wong (2018) offers two important insights. First, tests of correspondence require substantial statistical power in both experimental and nonexperimental

arms of the study. Despite over 1,800 participants in the RFE and large comparison groups, we were still underpowered to detect differences between the experimental and nonexperimental estimates with a tolerance level of 0.1 *SD*s in exclusive breastfeeding. We needed thresholds between 0.12 and 0.25 *SD*s to conclude correspondence with the existing data and effect sizes. Our findings are in line with Steiner and Wong's (2018) conjecture that few WSC studies would have achieved correspondence with a tolerance threshold of 0.1 *SD*s. Nevertheless, assessing correspondence based on a standard framework will facilitate comparisons across WSC and improve their application.

The second observation is that selection into the RFE, what we termed non-consent bias, is another challenge for WSC that has not received much attention. A rich set of covariates may mitigate this source of bias as demonstrated by Shadish et al. (2008). Preintervention surveys of eligible participants are feasible in an independent WSC but less so when the data on the comparison group are from administrative or other extant sources. We had a limited set of covariates, which can occur in dependent WSCs that rely on administrative data or other extant sources. Fortuitously, we had data on exclusive breastfeeding among eligible nonparticipants of the RFE. We were able to compare the adjusted breastfeeding outcomes of the controls in the RFE to those of the eligible nonparticipants. We found substantial nonconsent bias among English-speaking participants of the RFE. WIC clients who agreed to participate in the RFE were more favorably disposed to exclusive breastfeeding that those who declined to participate. The finding of nonconsent bias is unsurprising, given that only 17% of WIC clients who were offered the opportunity for peer counseling in the RFE agreed to participate. The nonconsent bias was an important signal that omitted-variable bias would likely lessen the likelihood of achieving correspondence between the experimental and nonexperimental estimates in the dependent WSC. Where feasible, testing for nonconsent bias in an RFE provides a useful indicator as whether the available covariates are sufficient to eliminate one source of potential noncorrespondence in a dependent or quasi-independent WSC.

Appendix A

Table A1. Observational Estimates of Peer Counseling Availability on Exclusive Breastfeeding by Duration and Language, 2006 (Q2)–2010 (Q3).

Exclusive breastfeeding for	English-Speaking WIC Clients			Spanish-Speaking WIC Clients		
	1 Month	3 Months	6 Months	1 Month	3 Months	6 Months
DD estimate	−0.004 (0.01)	0.000 (0.01)	−0.012† (0.01)	0.050** (0.01)	0.029** (0.01)	0.005 (0.01)
Mean outcome	0.430	0.272	0.162	0.425	0.314	0.226
N		64,137			17,509	

Observational Estimates of Peer Counseling Availability on Missing Outcome by Duration and Language, 2005 (Q3)–2010 (Q3)

	English-Speaking WIC Clients	Spanish-Speaking WIC Clients
Missing	0.001 (0.01)	0.004 (0.01)
Mean outcome	0.171	0.086
N	91,883	24,064

Note. The difference-in-difference estimates in the top panel are from a sample that drops the first five quarters of data in which missing data on breastfeeding were substantial. The lower panel tests whether missing data on exclusive breastfeeding are correlated with observed characteristics of the mother. The reported differences are estimated with ordinary least squares regression. The regressions control for age, age squared, natural logarithm of family income, child's birth quarter-year fixed-effects, and agency fixed-effects plus indicators for missing observations for each of the independent variables. Standard errors are clustered at agency level and are in parentheses. Significance levels are indicated by †10%, and **1%.

Declaration of Conflicting Interests

Funding

Notes

1. See https://www.cdc.gov/breastfeeding/pdf/2014breastfeedingreportcard.pdf. The reported health benefits of breastfeeding are limited to exclusive breastfeeding, and 6 months is the recommended duration by American Academy of Pediatrics.

2. The Promotion of Breastfeeding Intervention Trial (PROBIT) of breastfeeding remains the most important study in the literature (Kramer et al., 2001, 2003, 2008, 2009). Forty-three percentage of women in the treatment group exclusively breastfed at 3 months as compared to 6% among the controls. Differences in any breastfeeding at 3 months were more muted (73% vs. 60%). The health benefits observed in the PROBIT are largely attributed to the increase in exclusive breastfeeding (Kramer et al., 2003, 2008).

3. There have been numerous high-quality Randomized Controlled Trials (RCTs) of peer counseling (peer counseling) for breastfeeding conducted outside the United States. Researchers in England used telephone or hospital visitation to reach clients but reported no impact of peer counseling on initiation or duration of breastfeeding (Graffy, Taylor, Williams, & Eldridge, 2004). A peer counseling intervention similar to the RCT in our study was conducted in Canada in the late 1990s (Dennis, Hodnett, Gallop, & Chalmers, 2002). New and expectant mothers were contacted by telephone within 48 hr of delivery and for as many times in the next 3 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 peer counseling support was not effective but that peer counseling targeted at women who are considering breastfeeding did increase breastfeeding initiation rates (Ingram, MacArthur, Khan, Deeks, & Jolly, 2010).

4. The three U.S. RCT discussed previously all provided both in-person visits by the peer counselor to both the hospital and home of participants. The Oregon intervention used telephone counseling almost exclusively.

5. Doubly randomized preference trials originated in the clinical literature. Patients recruited to RCTs may have preference for a particular treatment. These preferences can affect the validity of the estimates when it is not possible to blind participants to their treatment. Under certain assumption, researchers can identify the separate effect of patient preferences from treatment effects (Long, Little, & Lin, 2008; Wing & Clark 2017).

6. Oregon has 36 counties but 33 Supplemental Nutrition Program for Women, Infants and Children program agencies. The counties of Umatilla, Morrow, Wheeler, and Gilliam are served by Umatilla–Morrow Head Start. Four agencies were recruited for the randomized field experiment (RFE) that covered seven counties (Hood River, Jackson, Washington Umatilla, Morrow, Wheeler, and Gilliam). Two agencies, one each in Marion and Deschutes counties, offered peer counseling on a nonexperimental basis during the entire study period (July 2005–July 2010). Thus, 27 WIC agencies had no peer counseling services during the study period.

7. The duration of any breastfeeding cannot be determined until a woman stops breastfeeding completely. As a result, any breastfeeding duration was missing for women who reported breastfeeding at their last WIC recertification visit but then left WIC before their next scheduled recertification appointment. Exclusive breastfeeding duration tends to end earlier in the postpartum period and thus is measured more completely.

8. Results are not shown and available from the authors upon request.

9. Heckman and Smith (1995) refer to this as randomization bias. We also note that the take-up of peer counseling is greater in the nonexperimental setting of the post-RFE (23.6%) than in the RFE itself (17.7%). This difference is not evidence of nonconsent bias but reflects the challenges of recruitment for the RFEs and the potential for bias (Heckman & Smith 1995; Kramer, 1984).

10. Note we also include the two WIC agencies that offered peer counseling services during the entire study period. The lack of change in the availability of peer counseling services in these two counties means they contribute to the counterfactual. As noted previously, we had excluded WIC clients from these agencies in both WSC samples, as shown in Figures 2 and 3.

11. The other distinctive pattern in Figure 4 is the elevated rate of exclusive breastfeeding at the beginning of the study period that declines steeply before leveling off around the third quarter of 2006. The high rate of exclusive breastfeeding is an artifact of missing data. Many cases of exclusive breastfeeding were not recorded in 2005 due to issues with the administrative system, TWIST. We deal

with the missing data in several ways. Figure 6 shows the percentage of clients for whom data on exclusive breastfeeding are missing. The top panel displays the three breastfeeding durations for English speakers and the bottom panel for Spanish speakers. Fortunately, the pattern is same in both sets of counties and for both languages, and it mimics the movement of exclusive breastfeeding in Figure 2. Given the proportion of missing data appears "balanced" between the treated and comparison counties, they should not affect the difference-in-differences (DD) estimates. As a further test, we create a dummy variable that is one if information on exclusive breastfeeding is missing for an individual WIC participant. In Equation 2, we replace the outcome variable with the missing data indicator to see whether the likelihood of a missing observation is related to our DD estimator. Lastly, we reestimate the DD beginning in the third quarter of 2006, after which the missing data on exclusive breastfeeding stabilized (Figure 6). As shown in Table A1, we don't find any evidence that the missing data drive our results.

12. We modify Equation 4 as follows: $BF_{ijt} = \sum_{k=-8}^{k=-1} \theta_k (PC_j * \tau_k) + \sum_{k=1}^{k=14} \phi_k (PC_j * \tau_k) + X_i' \beta + \tau_t + \lambda_j + \epsilon_{ijt}$. We do an F test for θ_k in the pre ($k = -8, -7, -.1$) and postperiods ($k = 1, 2, .14$) where the reference period is the quarter just before services became available ($k = 0$).

References

American Academy of Pediatrics. (2012). Breastfeeding and the use of human milk. *Pediatrics*. doi:10.1542/peds.2011-3552

Anderson, A. K., Damio, G., Young, S., Chapman, D. J., & Pérez-Escamilla, R. (2005). A randomized trial assessing the efficacy of peer counseling on exclusive breastfeeding in a predominantly Latina low-income community. *Archives of Pediatrics & Adolescent Medicine*, *159*, 836. doi:10.1001/archpedi.159.9.836

Bolton, T. A., Chow, T., Benton, P. A., & Olson, B. H. (2009). Characteristics associated with longer breastfeeding duration: An analysis of a peer counseling support program. *Journal of Human Lactation*, *25*, 18–27. doi:10.1177/0890334408325985

Bonuck, K., Trombley, K., Freeman, K., & McKee, D. (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. doi:10.1542/peds.2005-0435

Chapman, D. J., Damio, G., Young, S., & Pérez-Escamilla, R. (2004). Effectiveness of breastfeeding peer counseling in a low-income, predominantly Latina population: A randomized controlled trial. *Archives of Pediatrics & Adolescent Medicine*, *158*, 897. doi:10.1001/archpedi.158.9.897

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, *27*, 724–750. doi:10.1002/pam.20375

Dehejia, R. (2005). Practical propensity score matching: a reply to Smith and Todd. *Journal of econometrics*, *125*, 355–364.

Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association*, *94*, 1053–1062.

Dennis, C. L., Hodnett, E., Gallop, R., & Chalmers, B. (2002). The effect of peer support on breast-feeding duration among primiparous women: A randomized controlled trial. *Canadian Medical Association Journal*, *166*, 21–28.

Epstein, C., & Collins, A. (2015). *WIC breastfeeding peer counseling study phase II: Follow-up implementation report* (No. AG-3198-D-06-0105). Abt Associates, Inc. Retrieved September 27, 2018, from https://fns-prod.azur eedge.net/sites/default/files/ops/WICPeerCounseling-PhaseII.pdf

Gill, S. L., Reifsnider, E., & Lucke, J. F. (2007). Effects of support on the initiation and duration of breastfeeding. *Western Journal of Nursing Research*, *29*, 708–723. doi:10.1177/0193945906297376

Graffy, J., Taylor, J., Williams, A., & Eldridge, S. (2004). Randomised controlled trial of support from volunteer counsellors for mothers considering breast feeding. *British Medical Journal*, *328*, 26. doi:10.1136/bmj.328.7430.26

Gross, S. M., Resnik, A. K., Cross-Barnet, C., Nanda, J. P., Augustyn, M., & Paige, D. M. (2009). The differential impact of WIC peer counseling programs on breastfeeding initiation across the state of Maryland. *Journal of Human Lactation*, *25*, 435–443. doi:10.1177/0890334409342070

Grummer-Strawn, L. M., & Mei, Z. (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. doi:10.1542/peds.113.2.e81

Grummer-Strawn, L. M., Rice, S. P., Dugas, K., Clark, L. D., & Benton-Davis, S. (1997). An evaluation of breastfeeding promotion through peer counseling in mississippi WIC clinics. *Maternal and Child Health Journal*, *1*, 35–42. doi:10.1023/A:1026224402712

Haider, S. J., Chang, L. V., Bolton, T. A., Gold, J. G., & Olson, B. H. (2014). An evaluation of the effects of a breastfeeding support program on health outcomes. *Health Services Research*, *49*, 2017–2034. doi:10.1111/1475-6773.12199

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, *64*, 605–654.

Heckman, J. J., & Smith, J. A. (1995). Assessing the case for social experiments. *The Journal of Economic Perspectives*, *9*, 85–110.

Ingram, L., MacArthur, C., Khan, K., Deeks, J. J., & Jolly, K. (2010). Effect of antenatal peer support on breastfeeding initiation: A systematic review. *Canadian Medical Association Journal*, *182*, 1739–1746. doi:10.1503/cmaj.091729

Kramer, M. S. (1984). Scientific challenges in the application of randomized trials. *The Journal of the American Medical Association*, *252*, 2739. doi:10.1001/jama.1984.03350190041017

Kramer, M. S., Aboud, F., Mironova, E., Vanilovich, I., Platt, R. W., & Matush, L., . . . Promotion of Breastfeeding Intervention Trial (PROBIT) Study Group. (2008). Breastfeeding and child cognitive development: New evidence from a large randomized trial. *Archives of General Psychiatry*, *65*, 578–584. doi:10.1001/archpsyc.65.5.578

Kramer, M. S., Chalmers, B., Hodnett, E. D., Sevkovskaya, Z., Dzikovich, I., & Shapiro, S., . . . for the PROBIT Study Group. (2001). Promotion of breastfeeding intervention trial (PROBIT): A randomized trial in the republic of Belarus. *The Journal of the American Medical Association*, *285*, 413. doi:10.1001/jama.285.4.413

Kramer, M. S., Guo, T., Platt, R. W., Sevkovskaya, Z., Dzikovich, I., Collet, J. P., . . . Bogdanovich, N. (2003). Infant growth and health outcomes associated with 3 compared with 6 mo of exclusive breastfeeding. *The American Journal of Clinical Nutrition*, *78*, 291–295. doi:10.1093/ajcn/78.2.291

Kramer, M. S., Matush, L., Bogdanovich, N., Aboud, F., Mazer, B., Fombonne, E., . . . Platt, R. W. (2009). Health and development outcomes in 6.5-y-old children breastfed exclusively for 3 or 6 mo. *The American Journal of Clinical Nutrition*, *90*, 1070–1074. doi:10.3945/ajcn.2009.28021

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review*, *76*, 604–620.

Long, Q., Little, R. J., & Lin, X. (2008). Causal inference in hybrid intervention trials involving treatment choice. *Journal of the American Statistical Association*, *103*, 474–484.

Marcus, S. M. (1997). Assessing non-consent bias with parallel randomized and nonrandomized clinical trials. *Journal of Clinical Epidemiology*, *50*, 823–828. doi:10.1016/S0895-4356(97)00068-1

Meier, P. (2006). Biggest public health experiment ever: The 1954 field trial of the Salk poliomyelitis vaccine. In R. Peck (Ed.), *Statistics: A guide to the unknown* (4th ed., pp. 2–13). Belmont, CA: Thomson, Brooks/Cole.

Olson, B. H., Haider, S. J., Vangjel, L., Bolton, T. A., & Gold, J. G. (2008). A quasi-experimental evaluation of a breastfeeding support program for low income

women in Michigan. *Maternal and Child Health Journal*, *14*, 86. doi:10.1007/s10995-008-0430-5

Reeder, J. A., Joyce, T., Sibley, K., Arnold, D., & Altindag, O. (2014). Telephone peer counseling of breastfeeding among WIC participants: A randomized controlled trial. *Pediatrics*, *134*, e700–e709. doi:10.1542/peds.2013-4146

Schafer, E., Vogel, M. K., Viegas, S., & Hausafus, C. (1998). Volunteer peer counselors increase breastfeeding duration among rural low-income women. *Birth*, *25*, 101–106. doi:10.1046/j.1523-536x.1998.00101.x

Shadish, W. R., Clark, M. H., Steiner, P. M., & Hill, J. (2008). Can nonrandomized experiments yield accurate answers? A Randomized experiment comparing random and nonrandom assignments [with comments, rejoinder]. *Journal of the American Statistical Association*, *103*, 1334–1350.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators?. *Journal of econometrics*, *125*, 305–353.

Shaw, E., & Kaczorowski, J. (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. doi:10.1177/089033449901500108

Steiner, P. M., & Wong, V. C. (2018). Assessing correspondence between experimental and nonexperimental estimates in within-study comparisons. *Evaluation Review*, *42*, 214–247. doi:10.1177/0193841X18773807

Tryon, W. W., & Lewis, C. (2008). An inferential confidence interval method of establishing statistical equivalence that corrects Tryon's (2001) reduction factor. *Psychological Methods*, *13*, 272–277. doi:10.1037/a0013158

Wing, C., & Clark, M. H. (2017). What can we learn from a doubly randomized preference trial?—An instrumental variables perspective. *Journal of Policy Analysis and Management*, *36*, 418–437.

Wong, V. C., & Steiner, P. M. (2018). Designs of empirical evaluations of nonexperimental methods in field settings. *Evaluation Review*. doi:10.1177/0193841X18778918

Wong, V. C., Steiner, P. M., & Anglin, K. L. (2018). What Can Be Learned From Empirical Evaluations of Nonexperimental Methods?. *Evaluation review*, *42*, 147–175.

Yun, S., Liu, Q., Mertzlufft, K., Kruse, C., White, M., Fuller, P., & Zhu, B.-P. (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. doi:10.1017/S1368980009990668

Author Biographies

Onur Altindag is an assistant professor of Economics at Bentley University. He is a research fellow at the Economic Research Forum and a visiting scientist at the Harvard Center for Population and Development Studies.

Theodore J. Joyce is a professor of Economics at Baruch College and The Graduate Center, City University of New York. He is also a research associate with the National Bureau of Economic Research.

Julie A. Reeder is the senior research analyst at the State of Oregon WIC Program. She is also the current Chair of the American Public Health Association Food and Nutrition Section.