Received: 29 April 2022 | Revised: 15 July 2022 | Accepted: 1 October 2022

DOI: 10.1002/hec.4617

RESEARCH ARTICLE

Health Economics WILEY

Variation in the infant health effects of the women, infants, and children program by predicted risk using novel machine learning methods

Evan D. Peet[1] | **Dana Schultz**[1] | **Susan Lovejoy**[1] | **Fuchiang (Rich) Tsui**[2]

[1]RAND Corporation, Pittsburgh, Pennsylvania, USA

[2]Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correspondence
Evan D. Peet, RAND Corporation, 4570 Fifth Avenue, Ste 600, Pittsburgh, PA 15213, USA.
Email: epeet@rand.org

Funding information
Richard King Mellon Foundation

Abstract

The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) has an extensive literature documenting positive effects on infant health outcomes, specifically preterm birth, low birthweight, small size for gestational age, and infant mortality. However, existing studies focus on average effects for these relatively infrequent outcomes, thus providing no evidence for how WIC affects those at greatest risk of negative infant health outcomes. Our study focuses on documenting how WIC's infant health effects vary by level of risk. In doing so, we leverage a uniquely rich database describing maternal and infant outcomes and risk factors. Additionally, we use high dimensional data to generate predictions of risk and combine these predictions with the novel double machine learning method to stratify the effects of WIC by predicted risk. Our estimates of WIC's average treatment effects align with those in the existing literature. More importantly, we document significant variation in the effects of WIC on infant health by predicted risk level. Our results show that WIC is most beneficial among those at greatest risk of poor outcomes.

KEYWORDS
health policy, machine learning, maternal and child health, social services, treatment effect heterogeneity

1 | INTRODUCTION

The poor socioeconomic conditions that many women face are transmitted to the next generation to affect the health of their children (Aizer & Currie, 2014), and health influences the child's trajectory and long-term outcomes (Peet, 2021). One federal program that aims to support vulnerable populations and improve child outcomes and their trajectories, the Special Supplemental Nutrition Program for Woman, Infants, and Children (WIC) was established in 1972 to provide supplemental foods, nutrition education, and health-care referrals to low income pregnant and postpartum WIC under the age of five. Millions of mothers and their infants born in the U.S. each year receive monthly WIC checks or vouchers redeemable for specific packages of food at participating retailers (Hoynes & Schanzenbach, 2009). The food packages are tailored for each group—the main component of the infant package is infant formula, while the packages for breastfeeding women include a large variety and quantity of

foods. Studies generally show large, positive benefits of WIC participation, such as reducing low birthweight (LBW) (Bitler & Currie, 2005), prematurity, and neonatal intensive care unit admission (Sonchak, 2016). Consequently, studies show that the averted infant health costs exceed WIC's annual cost (Hoynes et al., 2009).

However, existing evaluations of WIC's effects on infant health do not consider how the effects of the program may vary, particularly, how WIC affects those at greatest risk of negative infant health outcomes. Understanding how WIC affects those at greatest risk is important for policy aiming to distribute limited resources to produce the greatest social benefits. Moreover, because negative infant health outcomes are relatively infrequent, average treatment effects may mask important benefits to certain groups. However, the question of WIC's estimated effects centers on selective participation. Because WIC participation is non-random, studies may overestimate WIC's benefits if participation is systematically correlated with unobserved positive influences on infant health. The converse may also be true—if WIC participants are more disadvantaged than non-participants, WIC's effects on infant health may be underestimated. Previous studies have attempted to account for selective participation by (1) narrowly defining the treatment and control groups (Bitler & Currie, 2005; Figlio et al., 2009; Joyce et al., 2005, 2008); (2) comparing children born to the same mother who both did and did not participate in WIC (Brien & Swann, 1997; Chatterji et al., 2002; Currie & Rajani, 2015; Kowaleski-Jones & Duncan, 2002); (3) using monotonicity restrictions to achieve partial identification bounds (Kreider et al., 2020); or (4) leveraging variation in WIC program parameters by state (Brien & Swann, 1997; Chatterji et al., 2002; Swann, 2010). While selective participation calls into question whether the benefits of the program are received by those at greatest risk, none of the approaches taken in the existing literature address how the effects of the program vary, particularly how WIC affects those at greatest risk of negative infant health outcomes.

We employ a novel approach that accounts for selective participation and allows us to estimate the variation in WIC's infant health effects by predicted risk of poor outcomes. To characterize infant health, we examine (1) preterm birth (PTB); (2) LBW; (3) small size for gestational age (SGA); and (4) infant mortality (IM). Key to our approach is a unique dataset that combines vital statistics natality data (the basis for many studies of WIC) with detailed electronic health records (e.g., diagnostic history, biomarkers), and social service records (e.g., housing assistance, substance abuse treatment) for each mother-infant dyad. With this rich, time-varying data we employ the double/debiased machine learning (DML) method of Chernozhukov et al. (2018), a doubly robust estimator and one of a growing number of machine learning tools in the econometric toolbox for program evaluation (Abadie & Cattaneo, 2018; Athey & Imbens, 2017). We use machine learning algorithms to predict each infant health outcome and WIC participation, then, per the DML method, we combine each prediction to estimate the infant health effects of WIC. The predictions of WIC participation are used to account for selection and compare the infant health outcomes of those predicted at risk and predicted to participate in WIC at similar levels. Additionally, we incorporate repeated sample splitting within the DML procedure so that we can stratify by predicted risk while avoiding the overfitting bias that typically accompanies such efforts. In this way, we separately identify the effects of WIC at each category of predicted risk (Abadie et al., 2018).

Our estimates of average treatment effects show that WIC affects the risk of: (1) PTB by −0.04% points (pp) (not significant); (2) LBW by −0.43 pp (significant at 90% level); (3) SGA by −0.51 pp (significant at 90% level); and (4) IM by 0.39 pp (not significant). These estimates lie within the range of existing estimates in the literature. However, these average treatment effects do not tell the most interesting or important story because the vast majority of infants observed are at low risk of these negative infant health outcomes. The more important results show how WIC affects those at greatest risk of these negative infant health outcomes. By predicted risk category, WIC is shown to significantly (at the 95% level) reduce the risk of: (1) PTB by 2.19 pp and 3.52 pp among the two highest risk categories; (2) LBW by 4.43 pp in the second highest risk category; (3) SGA by 3.72 pp in the highest risk category; and (4) IM by 8.74 pp in the highest risk category. For each outcome, all other estimates for risk categories are small and statistically insignificant. These results are shown to be robust to model specification. In the lead up to these estimated effects of WIC, we also show that our models are able to accurately predict the risk of each of the infant health outcomes and WIC participation. These results have important implications for current maternal and child health policy. WIC participation has been declining nationally leading to reduced funding (USDA, 2022a, 2022b) and various reforms (Partnerships, 2021). The results of our study provide evidence that reforms, while costly, if targeted to individuals likely to participate and at high risk of negative infant health outcomes, would produce larger than average benefits to infant health.

The remainder of the paper proceeds as follows. In Section 2 we provide background both for the WIC program and the prior WIC literature. In Section 3 we describe our approach, including the data and methods. In Section 4 we detail the results of our main analyses. In Section 5 we provide placebo tests, and in Section 6 we explore robustness checks. Finally, we conclude with a discussion of the results and their implications in Section 7.

2 | BACKGROUND

2.1 | Women, infants, and children (WIC)

The Supplemental Nutrition Program for WIC was piloted in 1972 and became permanent in 1975 with the goal of improving the nutritional well-being of low income pregnant and postpartum women and their children (Chorniy et al., 2020). When WIC was introduced, the motivation stemmed from the belief that a variety of health problems could be avoided by providing food to pregnant and postpartum women and their children at "critical times" (Oliveira, 2002). Women, infants, and children participants receive monthly checks or vouchers or, in some states, monthly direct deposits to debit cards, that can be used to purchase specific types and brands of food at participating retailers (Currie & Rajani, 2015). The packages of food that participants can purchase originally included juice, fortified cereal, eggs, cheese, milk, dried beans, tuna, carrots, and iron-fortified infant formula (Hoynes et al., 2009). In October 2009, the US Department of Agriculture revised the WIC food packages to include more whole grains, fruits, and vegetables in order to address concerns surrounding obesity and chronic disease (Whaley et al., 2012). The food packages are tailored by group—packages for breastfeeding women include a large variety and quantity of foods and for other women the main component is infant formula (Chorniy et al., 2020). Women, infants, and children also provides its participants with breastfeeding information and counseling, nutritional education classes, referrals to other social services, and referrals to healthcare providers for services such as immunization screenings (Chorniy et al., 2020).

Eligibility for WIC is determined by income (less than 185% of the federal poverty line) or participation in other entitlement programs such as Temporary Assistance for Needy Families (TANF), the Supplemental Nutrition Assistance Program (SNAP), or Medicaid. Additionally, WIC professional staff must evaluate and determine "nutritional risk." Evidence suggests that structural barriers (e.g., transportation) and rules regarding eligibility maintenance (Liu & Liu, 2016) lower participation among the eligible (estimated 71% of all eligible pregnant women during this study's time period) (Johnson et al., 2015). However, the requirements to become and remain eligible—such as attending regular nutritional and breastfeeding counseling sessions—provide a mechanism beyond nutrition through which WIC may impact maternal and infant health. For instance, it is postulated that counseling sessions "nudge" participants' diets in a healthier direction and facilitate access to health services (Currie & Rajani, 2015).

2.2 | Existing evidence of WIC effectiveness

The most robust set of evidence for WIC's effect on infant health surrounds birthweight. For example, Hoynes et al. (2009) examine the historical introduction of WIC across geographic areas over time and show that WIC has a modest but significant positive effect on birthweight. Using maternal fixed effects, Currie and Rajani (2015) find that WIC significantly reduces LBW by 0.41 pp. Figlio et al. (2009) focus on families in a narrow income band around the WIC eligibility threshold to identify large significant effects of WIC on LBW—a 0.75 pp reduction. By instrumenting for WIC participation with the presence of a WIC clinic in the mother's zipcode during pregnancy, Rossin-Slater (2013) finds that WIC participation increases birthweight (though changes to the incidence of LBW are insignificant on average). Alternatively, by examining birthweight before and after changes in WIC food packages, Hamad et al. (2019) show that WIC significantly reduces the incidence of LBW by 0.2 pp. Furthermore, using monotonicity restrictions that are relatively weak compared with the strict orthogonality assumptions of other approaches, Kreider et al. (2020) use partial identification and find that WIC reduces the incidence of LBW by at least 4.8%.

The birthweight evidence has been critiqued because the measure does not disentangle gestational length, which is mechanically correlated with greater birthweight and more time to enroll in WIC. The focus on LBW is due to its availability in vital statistics natality data, but SGA is a more precise measure. Hamad et al. (2019) show that WIC significantly reduced SGA by 0.4 pp. Again focusing on families in a narrow income band around the WIC eligibility threshold, Figlio et al. (2009) estimate a 1.34 pp reduction in SGA. Additionally, Sonchak (2016) uses maternal fixed effects and addresses the gestational age bias concerns by conditioning on gestational age to find that WIC reduces the incidence of SGA by 0.9 pp. Using a similar approach in a different context, Currie and Rajani (2015) find that WIC significantly reduces SGA by 0.29 pp.

The evidence surrounding WIC's effects on PTB (<37 weeks of gestation) has been mixed with some studies indicating that WIC is protective and other studies showing no effects. On the one hand, Bitler and Currie (2005), show that prenatal WIC participation reduces the likelihood of a PTB by 29%. The partial identification approach of Kreider et al. (2020) shows that WIC reduces the incidence of PTB by at least 3.7%. Using maternal fixed effects to account for unobservable,

time-invariant characteristics of mothers, Currie and Rajani (2015) show that WIC reduces PTB by 0.8 pp. On the other hand, Joyce et al. (2008) show no evidence of WIC affecting the incidence of PTB and contend that, as the Institute of Medicine has summarized using evidence from randomized control trials, there is no clinical relationship between dietary supplementation and PTB (Berkowitz & Papiernik, 1993). Based on this lack of evidence in the clinical literature, Joyce et al. (2008) conclude that suggestions of WIC's effects on PTB are likely due to gestational age bias. However, WIC involves more than nutritional supplementation—it also involves counseling, education, and referrals to prenatal care and other social and health services which could affect gestational age (Corman et al., 2018; El-Bastawissi et al., 2007).

While PTB, LBW, and SGA are each strongly correlated with IM, there is less evidence on WIC's effects on IM. However, Khanani et al. (2010) found significant reductions in IM from WIC, particularly among infants of Black women. This is similar to the work of El-Bastawissi et al. (2007) that found significant associations between prenatal WIC participation and fetal death. However, neither of these studies involve research designs which attempt to identify causal effects.

Among the aforementioned studies and others, various approaches have been employed in an attempt to address the potential biases introduced by selective participation in WIC. These include standard ordinary least squares (OLS) regression with controls and/or restricted samples (Bitler & Currie, 2005; El-Bastawissi et al., 2007; Figlio et al., 2009; Joyce et al., 2005; Khanani et al., 2010), differences-in-differences (Hamad et al., 2019; Hoynes et al., 2009), instrumental variables (Bitler & Currie, 2005; Brien & Swann, 1997; Figlio et al., 2009; Rossin-Slater, 2013), partial identification (Kreider et al., 2020), and maternal fixed effects (Brien & Swann, 1997; Chorniy et al., 2020; Currie & Rajani, 2015; Foster et al., 2010; Kowaleski-Jones & Duncan, 2002; Rossin-Slater, 2013; Sonchak, 2016). Each of these approaches have issues ranging from weak instruments, to small sample sizes, to the inability to account for time varying observable and unobservable maternal characteristics. However, there is also evidence that the unobserved determinants of selection may actually understate the estimated effects of WIC. For example, when considering only Medicaid-eligible mothers, WIC participants have been shown to be younger, less educated, and are more likely to be single, obese, and smokers (Bitler & Currie, 2005). Longitudinally, among mothers whose WIC status changed over time, evidence indicates that more often women participated in WIC when they were younger, single, and unemployed (Currie & Rajani, 2015). If, as this evidence suggests, WIC participants are more disadvantaged and less healthy across metrics than non-participants, the selection bias would actually underestimate WIC's effects.

Our study adds to the existing evidence of WIC's effect to a range of infant health outcomes, including PTB, LBW, SGA, and IM. Our study's main contribution is the unique evidence regarding the variation in WIC's effects by the predicted risk level for negative infant health outcomes. To do so, we employ a novel method, DML with random sample splitting, and leverage predictions of PTB, LBW, SGA, and IM to examine variation in WIC's effects by predicted risk level. This evidence is critically important to policymakers who must make decisions about how to allocate scarce resources since it addresses the question of how the program affects those at greatest risk of negative infant health outcomes.

3 | APPROACH

3.1 | Data sources

To examine the infant health effects of WIC, we used a database called IM Prediction System with the Intervention Management (or IMPreSIv) which contains a unique combination of data linked at the individual-level (Tsui et al., 2019; Tsui & Shi, 2019; Yang et al., 2017). IMPreSIv links multiple administrative data sources to describe the outcomes, characteristics, and risk factors for each maternal-infant dyad for each birth and infant death that occurred within Allegheny County, Pennsylvania (where Pittsburgh is located) between 2003 and 2013 (Peet et al., 2021; Schultz et al., 2020).

IMPreSIv begins with vital statistics records (i.e., personally identifiable birth and death certificate data) obtained from the Allegheny County Health Department and describing all children born in Allegheny County between 2003 and 2013 (totaling 155,218). The vital statistics data also describes prenatal participation in WIC. From that foundation, we linked electronic health records and social service administrative data from the following sources:

(1) The University of Pittsburgh Medical Center (UPMC) Medical Archival Retrieval System (MARS) electronic health records describing individual health services and claims data (e.g., medical interventions, diagnoses, prescriptions);
(2) The Magee Obstetrical Maternal and Infant (MOMI) database containing additional detailed data describing labor and delivery characteristics, biomarkers, test results, ultrasound measurements, and procedures;
(3) Allegheny County Data Warehouse of social and behavioral services containing information describing individual involvement with these services offered by the county such as housing assistance and substance abuse programs;

(4) Additional services records describing individual involvement with maternal and child health programs offered by the county (e.g., Nurse Family Partnership), or involvement with third party programs such as Healthy Start.

The links between each of these administrative datasets were established by a certified honest broker with access to identifying information who then produced Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor research datasets (Tsui & Shi, 2019). The process undertaken by the honest broker to produce the de-identified, HIPAA Safe Harbor research datasets is described in greater detail in Tsui et al. (2017). Data quality assessments examining completeness, validity, consistency, and currency of the data are also described in Yang et al. (2017). With the de-identified research data, we also linked socioenvironmental data (e.g., local poverty, unemployment, air pollution) from various sources such as American Community Survey, the Environmental Protection Agency, and others using the mother's residential zip code (3 digit) information. In this data we observe a total of 155,218 births or mother-infant dyads.

We then curated the data, first determining the full set of potential infant health determinants. The full data include multiple thousands of variables, but not all relevant. Irrelevant information includes such maternal health diagnoses as broken bones that took place prior to pregnancy. Additionally, other potential outcomes for both mother and infant like appearance, pulse, grimace, activity, and respiration (APGAR) score, neonatal intensive care unit (NICU) admission, and postpartum depression were excluded.

Then, to assess robustness and model fit, we defined further restricted data sets by type of information. The first data type we refer to as "reportable". The name is not meant to imply that the data is self-reported (*all* the data we use is administrative), rather it describes the set of information that could feasibly be self-reported by the mother. This includes the information contained within the birth/death certificate data as well as information about the mother's participation in other medical services (e.g., first trimester prenatal care) and social services (e.g., assisted housing). The data also describe the timing and duration of participation in each service. Additionally, these data include socioenvironmental data. This type of data includes over 900 variables and describes the full sample of 155,218 births or mother-infant dyads.

The second type of data is called "potentially reportable." Again, all data is administrative, the name describes information that the mother could potentially self-report. These data include all of the "reportable" information as well as major maternal diagnoses in the electronic health records (e.g., cardiovascular disease, endocrine disease). This type of data includes over 1200 variables. However, this data includes information only available for births that occurred within the UPMC hospital system, totaling 75,842 births or 49% of the full sample.

The third type of data is called "healthcare" and it includes all the "reportable" and "potentially reportable" information as well as additional, detailed electronic health record data that is only available to healthcare providers in healthcare settings. This includes prescription records, and extensive biomarkers, blood test results, ultrasound measurements, and toxicology reports. This type of data includes over 1800 variables. As with the "potentially reportable" data type, the "healthcare" data type includes information only available for the 75,842 births that occurred within the UPMC hospital system.

The "healthcare" data type is used to the estimate the study's main results. Results using the "potentially reportable" and "reportable" data types are included in the appendix and describe how robust the estimates are to the number of variable inputs to the machine learning algorithms.

3.2 | Summary statistics

We first describe summary statistics of the outcomes of interest (PTB, LBW, SGA, and IM), WIC participation, and a selection of sociodemographic, maternal health, and maternal reproductive history characteristics. Table 1 describes these summary statistics for the full sample as well as the sample that is comprised of infants born at a UPMC hospital (for which the more extensive "healthcare" type of data is available). The average outcomes for both samples generally align with the national averages—approximately 11% of births are PTB, 8.7% are LBW, 5.5% are SGA, and IM rates are 6.5 per 1000 live births. Women, Infants, and Children participation is roughly 30% in each sample. The UPMC sample is comprised of a greater percentage of Black mothers, teen mothers, single mothers, as well as a greater percentage of WIC and Medicaid participants.

Figure 1 describes trends in WIC participation. While relatively constant throughout the period, WIC participation peaked in 2009 (similar to the nation as a whole) and then continually declined until 2013. The bottom panel of Figure 1 shows the racial/ethnic breakdown of WIC participants with roughly 65% of Black, non-Hispanic women participating in WIC, compared to only 20%–25% of White, non-Hispanic women. Table 2 shows the breakdowns in the infant health outcomes by WIC participation. Generally, the infant health outcomes are worse among those who participate in WIC, indicating that WIC participants are generally at greater risk of negative infant health outcomes than non-participants. Relative to the nation as a whole,

TABLE 1 Full and university of Pittsburgh medical center (UPMC) sample summary statistics

	Full sample		UPMC sample	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev**
Outcomes				
Preterm birth	10.96%	31.24	9.50%	29.32
Low birthweight	8.65%	28.11	7.20%	25.85
Small size for gestational age	5.46%	22.73	5.15%	22.10
Infant mortality (per 1000 live births)	6.49	80.32	6.53	80.52
Intervention				
WIC participation	29.64%	39.00	31.07%	46.28
Characteristics				
White	73.82%	9.71	70.70%	45.51
Black	18.71%	25.15	23.40%	42.34
Hispanic	2.34%	38.26	1.11%	10.47
Asian	4.19%	44.72	4.11%	19.86
Other race/ethnicity	0.95%	45.74	0.68%	8.20
Maternal age: 13–19	6.79%	35.46	7.60%	26.50
Maternal age: 20–24	17.81%	17.49	19.00%	39.23
Maternal age: 25–29	27.64%	48.05	27.06%	44.43
Maternal age: 30–34	29.80%	7.40	28.98%	45.37
Maternal age: 35–39	14.75%	26.49	14.27%	34.97
Maternal age: 40+	3.16%	40.35	3.09%	17.30
Married	63.82%	37.36	60.90%	48.80
Maternal education: 8th or less	0.55%	29.71	0.50%	7.08
Maternal education: no high school or equivalent diploma	7.59%	44.29	8.30%	27.59
Maternal education: HS or GED	20.47%	37.96	22.11%	41.50
Maternal education: some college	16.77%	12.22	14.69%	35.40
Maternal education: associates	9.79%	49.25	8.40%	27.74
Maternal education: bachelors	26.80%	42.32	26.33%	44.04
Maternal education: post-graduate	17.45%	7.79	19.39%	39.53
No SSN	1.52%	49.20	1.41%	11.81
Private insurance	58.62%	34.89	51.15%	49.99
Medicaid	23.37%	32.87	25.49%	43.58
Other payment	18.62%	38.93	23.37%	42.32
First pregnancy	41.06%	39.86	44.72%	49.72
Last pregnancy ended: <2 years ago	14.19%	14.22	12.90%	33.53
Last pregnancy ended: 2–3 years ago	12.32%	33.49	13.69%	34.37
Last pregnancy ended: 3+yrs ago	23.91%	16.60	27.00%	44.40
Maternal smoking: ever	19.82%	12.77	16.87%	37.45
Previous child death(s)	1.12%	21.16	1.05%	14.83
Previous C-section(s)	12.87%	9.14	13.50%	34.17
Previous preterm birth(s)	2.84%	19.55	2.08%	14.26
Pre-pregnancy hypertension	1.66%	12.77	1.84%	13.46
Gestational hypertension	4.70%	21.16	4.91%	21.62
Pre-pregnancy diabetes	0.84%	9.14	1.02%	10.06

TABLE 1 (Continued)

	Full sample		UPMC sample	
	Mean	Std. Dev.	Mean	Std. Dev
Gestational diabetes	3.98%	19.55	3.65%	18.77
Pre-pregnancy obesity	18.82%	39.09	20.14%	40.10

Note: The full sample is comprised of 155,218 births, while the UPMC sample is comprised of 75,842 births.

Abbreviations: GED, equivalent diploma; HS, high school; SSN, social security number; WIC, women, infants, and children.

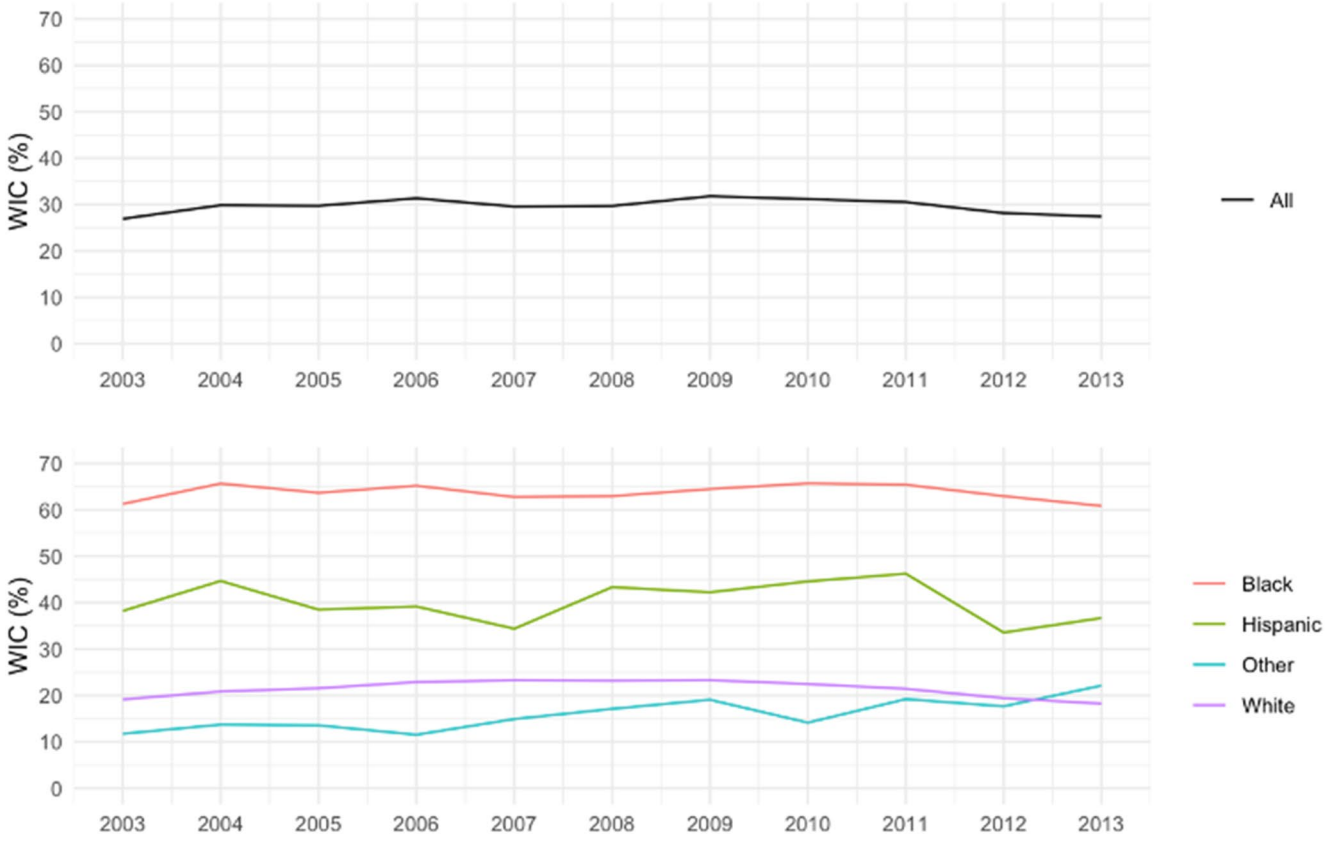

FIGURE 1 Trends in women, infants, and children (WIC) participation. Summary statistics are based on the full sample of 155,218 infants. [Colour figure can be viewed at wileyonlinelibrary.com]

Allegheny County, PA has experienced similar trends in infant health outcomes. Figure 2 shows the percentages of PTB and LBW and the rate of IM in both Allegheny County, PA and the nation between 2003 and 2013. Additionally, like other areas of the country, Allegheny County has stark racial disparities across infant health outcomes. Figure 3 shows the trends in PTB, LBW, and IM by race/ethnicity in Allegheny County which illustrate how Black infants have experienced much worse outcomes than other groups.

3.3 | Methods

Beginning with notation, let Y_{imt} denote the dependent variable or infant health outcome (PTB, LBW, SGA, or IM). The variable of interest is D_{imt}, which describes WIC participation during the preconception, prenatal, or postnatal time period t corresponding mother m's pregnancy with infant i. Other explanatory variables include a wide variety of potential confounders: $\boldsymbol{S}_{imt}$, which describes parental sociodemographic characteristics (e.g., parental age, parental education, parental race/ethnicity); $\boldsymbol{M}_{imt}$, which describes maternal health characteristics (e.g., hypertension, diabetes, smoking, substance abuse); $\boldsymbol{R}_{imt}$, which describes the mother's reproductive history (e.g., first pregnancy, previous C-section, interpregnancy interval); $\boldsymbol{E}_{imt}$, which describes socio-environmental risk factors (e.g., air quality, median household income, local unemployment rates)

TABLE 2 Infant health by women, infants, and children (WIC) participation

	Risk category 1 (0%–20%)		Risk category 2 (20%–40%)		Risk category 3 (40%–60%)		Risk category 4 (60%–80%)		Risk category 5 (80%–100%)	
	Part. %	**Obs. (*N*)**	**Part. %**	**Obs. (*N*)**	**Part. %**	**Obs. (*N*)**	**Part. %**	**Obs. (*N*)**	**Part. %**	**Obs. (*N*)**
Outcomes										
Preterm birth	30.06%	52,692	29.86%	6344	30.59%	5930	36.68%	5554	35.62%	5322
Low birthweight	24.84%	46,893	39.33%	12,891	42.40%	6243	45.66%	5441	40.42%	4374
Small size for gestational age	16.53%	27,367	29.92%	24,869	47.57%	13,408	53.59%	6733	46.41%	3465
Infant mortality (per 1000 live births)	29.29%	65,902	42.88%	5576	43.74%	1646	42.30%	1402	42.02%	1316

Note: "Part. %" refers to the WIC participation rate among each group. The number of observations across all risk categories totals 75,842, or the UPMC sample.

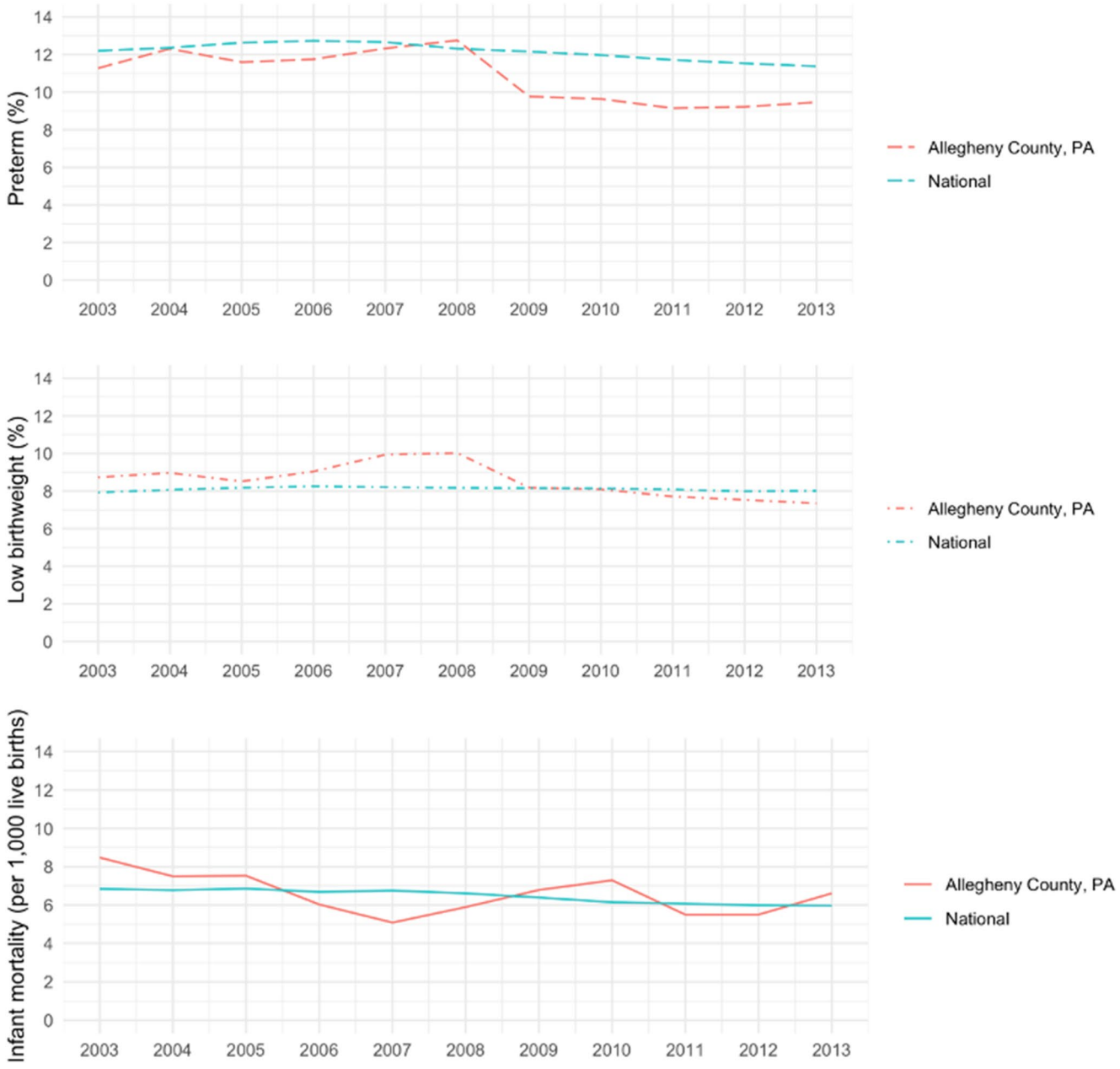

FIGURE 2 Infant health trends 2003–2013, Allegheny County, PA and the nation. Summary statistics are based on the full sample of 155,218 infants. [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 3 Infant health trends by race/ethnicity 2003–2013, Allegheny County, PA. Summary statistics are based on the full sample of 155,218 infants. [Colour figure can be viewed at wileyonlinelibrary.com]

corresponding to the time period t and the three digit zip code of the mother's residence; $\boldsymbol{I}_{imt}$, which describes the participation in other, non-WIC social services (e.g., housing assistance, mental health services, substance abuse programs); $\boldsymbol{H}_{imt}$, which describes the participation in or receipt of health services (e.g., prenatal care frequency, vaccinations, prescriptions, ultrasound measurements); $\boldsymbol{\tau}_t$, which is a vector of year-of-birth t indicators; and $\boldsymbol{\mu}_m$, which is a vector of the mother's location of residence (3 digit zip code) indicators. Let $\boldsymbol{Z}_{imt} = \{\boldsymbol{S}_{imt}, \boldsymbol{M}_{imt}, \boldsymbol{R}_{imt}, \boldsymbol{E}_{imt}, \boldsymbol{I}_{imt}, \boldsymbol{H}_{imt}, \boldsymbol{\tau}_t, \boldsymbol{\mu}_m\}$ denote the set of all covariates. It is important to note (especially as it pertains to gestational age) that the set of variables included in $\boldsymbol{Z}_{imt}$ varies by the outcome Y_{imt}. For $Y_{imt} = \text{LBW}$ and $Y_{imt} = \text{IM}$, $\boldsymbol{Z}_{imt}$ includes gestational age similar to Sonchak (2016) and Currie and Rajani (2015). However, $\boldsymbol{Z}_{imt}$ does not include gestational age when $Y_{imt} = \text{PTB}$ or $Y_{imt} = \text{SGA}$. The full relationship between the dependent and explanatory variables is modeled semi-parametrically as follows (Equation 1):

$$Y_{imt} = \beta D_{imt} + f(\boldsymbol{Z}_{imt}) + \varepsilon_{imt} \tag{1}$$

3.4 | Methods: machine learning

While the econometric method we employ (describe in the next section) will be used for causal inference, the machine learning algorithms that are part of this method are typically used for prediction. In our context, there are several benefits to using machine learning algorithms to predict Y_{imt} and D_{imt} using $\boldsymbol{Z}_{imt}$. First, machine learning algorithms effectively addresses the issues associated with high dimensionality. The blessing of high dimensionality is that the more confounders that are observed, the more likely that conditional unconfoundedness is an appropriate assumption. However, the curse of high dimensionality is that the covariance matrix may become ill-posed and inference will be invalid (Marčenko & Pastur, 1967). In many cases, theory or empirical evidence can be used to select the most relevant confounding variables to be included in the model. However, in some instances, there is no theoretical or empirical basis for model selection. In our case, the unique IMPreSIv database includes variables which have not been examined previously to assess their relationships with the infant health outcomes. Machine learning algorithms address the problem of high dimensionality through regularization (or model selection). There are two main types of regularization: L1, which reduces the complexity of the model by including a penalty term which has the effect of forcing some of the coefficients to be 0 (Park & Hastie, 2007); and L2, which reduces the impact of certain variables by including a term and shrinks the coefficients of less influential variables toward 0 (Demir-Kavuk et al., 2011). Both types of regularization can be used in combination to avoid overfitting.

Second, machine learning algorithms are also adept at identifying unspecified, non-linear and/or interactive relationships in the data. While standard methods, such as OLS, can estimate non-linear and/or interactive relationships between explanatory variables and the outcome, these must be specified. In the case that there are important non-linear and/or interactive relationships between $\boldsymbol{Z}_{imt}$ and Y_{imt} or D_{imt} that are unspecified, OLS will not produce consistent estimates. On the other hand, machine learning algorithms natively estimate non-linear and/or interactive relationships even when unspecified (Vegetabile et al., 2020).

In predicting Y_{imt} and D_{imt} using $\boldsymbol{Z}_{imt}$, we employed multiple machine learning algorithms and then evaluated/compared their performance. Specifically, we used logitboost, neural network, random forest, support vector machine, and gradient boosting machine (GBM) algorithms. In evaluating and comparing these algorithms, we used the area under receiver operating characteristic curve (AUC) scores. AUC combines the true positive rate and false positive rate into one measure which describes the percentage of randomly drawn pairs of one individual who participates in the intervention and one individual who does not participate for which the predictive model is accurate. Among the algorithms, the best performing was the GBM. Intuitively, GBM starts with simple (i.e., linear) models, then the analyzes the error term and learns from it, identifying shortcomings in the simple models by using gradients in a loss function that describes the differences between the predictions and the actual outcomes. Then, GBM updates the models to different functional forms that produce errors that are smaller in magnitude. This iterative process continues until the algorithm cannot find another function that reduces the errors (Friedman, 2002).

3.5 | Methods: double machine learning

Increasingly, machine learning algorithms have been adapted to causal inference and applied to program evaluation in recent years (Bertrand et al., 2017; Davis & Heller, 2017; Farbmacher et al., 2019; Gulyas & Pytka, 2019; Knaus et al., 2020, 2021; Knittel & Stolper, 2019; Strittmatter, 2018). To estimate the effects of WIC on the infant health outcomes of interest (PTB, LBW, SGA, and IM), we employ the double machine learning (DML) introduced by Chernozhukov et al. (2018). At its core, DML is a method for estimating treatment effects when potential confounders are high dimensional (many) but observed. DML integrates machine learning algorithms into causal analysis with observational data by improving the accuracy of treatment participation predictions and thus minimizing the potential for selection bias (Knaus, 2020). In our context, minimizing the potential for selection bias first relies on the IMPreSIv data with a high dimensional set of predictors. Drawing on the high dimensional set of predictors in IMPreSIv, the accuracy of treatment predictions are enhanced by the machine learning algorithms which identify unspecified, non-linear and interactive relationships in the data.

DML relies on a foundational econometric theory—the Frisch-Waugh-Lovell theorem (Frisch & Waugh, 1933; Lovell, 1963). Frisch-Waugh-Lovell is a decomposition of a regression of a dependent variable on two types of explanatory variables—the explanatory variable of interest and the covariates. Frisch-Waugh-Lovell shows that estimating the relationship between Y_{imt} and D_{imt} conditional on $\boldsymbol{Z}_{imt}$, as described in Equation (1), is the same as the following process. First, model Y_{imt} and D_{imt} separately as functions of $\boldsymbol{Z}_{imt}$, as in the following:

$$D_{imt} = g\left(\boldsymbol{Z}_{imt}\right) + \eta_{imt} \tag{2}$$

$$Y_{imt} = h(\boldsymbol{Z}_{imt}) + v_{imt} \tag{3}$$

Second, estimate Equation (2) and find the residuals (W_{imt}). Third, estimate Equation (3) and find the residuals (V_{imt}). And fourth, regress V_{imt} on W_{imt}. Chernozhukov et al. (2018) showed that using machine learning algorithms to estimate the non-parametric functions $g(\cdot)$ and $h(\cdot)$ in Equations (2) and (3) combines the benefits of machine learning within a causal inference approach. Key among these is the systematic approach to high dimensionality, avoiding overfitting while maximizing model fit and increasing the likelihood that conditional unconfoundedness assumption, required for causal interpretation of the estimates, is appropriate.

Another important feature of our approach is that we also employ crossfitting with random sample splitting. Random sample splitting is important for DML because it eliminates asymptotic bias (Chernozhukov et al., 2018), however, it also allows us to examine heterogeneity by the level of predicted risk for each of the infant health outcomes (Abadie et al., 2018). Random sample splitting means that we estimate Equations (2) and (3) (or train the models) on one split ($s = 10$) of the sample, and to crossfit we use a different sample split ($-s$) to test the models and estimate parameters. Then, by regressing each sample split's residuals, we produce multiple, crossfit estimates of WIC's effects, or $\hat{\beta}^s$. Our final estimate is the average of all the crossfit estimates, or $\frac{1}{S}\sum_s \hat{\beta}^s$. This achieves $\sqrt{n}$ convergence of treatment effect estimates (Chernozhukov et al., 2018). Typically, stratifying the estimates by the predicted level of the outcome using in-sample data runs the risk of overfitting bias. However, Abadie et al. (2018) show that random sample splitting avoids overfitting bias. Using random sample splitting to stratify the effects of WIC by the predicted risk of PTB, LBW, SGA, and IM allows us to examine whether WIC is benefiting those at greatest risk of negative infant health outcomes.

Also note that as a doubly robust estimator, DML naturally extends to the estimation of heterogeneous treatment effects. The benefit of doubly robust estimators which model the outcome and the treatment, as DML does, is that they produce unbiased estimates if at least one of the models are correctly specified (rather than both) (Funk et al., 2011). Foster and Syrgkanis (2019) show that under standard assumptions doubly robust estimators, like DML, efficiently estimate heterogeneity in the treatment effects. This is important for our study as we will estimate the heterogeneous effects of WIC by the predicted risk of the negative infant health outcomes (PTB, LBW, SGA, and IM).

4 | RESULTS

4.1 | Predictive models of WIC participation and infant health outcomes

For the healthcare type of data, the inputs for our predictive models of WIC participation and infant health outcomes include over 1800 variables that were either reported by the mother or are recorded within the MARS and MOMI databases. The sample includes 75,842 mother-infant dyads for which delivery occurred at a UPMC hospital between 2003 and 2013.

However, the final predictive models include only a subsample of all the potential variables. First, as authors, we selected potential variable inputs, specifically those which denote: (1) parental sociodemographic characteristics ($\boldsymbol{S}_{imt}$); (2) maternal health characteristics ($\boldsymbol{M}_{imt}$); (3) maternal reproductive history ($\boldsymbol{R}_{imt}$); (4) socio-environmental risk factors ($\boldsymbol{E}_{imt}$); (5) participation in other, non-WIC social services ($\boldsymbol{I}_{imt}$); and (6) participation or receipt of in health services ($\boldsymbol{H}_{imt}$). Effectively, this meant excluding time-inconsistent variables (for instance, if the outcome is observed at birth, like PTB, then variables observed during the postnatal period are irrelevant), and other potential outcomes for both mother and infant (e.g., APGAR score, NICU admission, postpartum depression). Second, the machine learning algorithms performed regularization. As a result of the regularization, a portion of the variables that we, as authors, selected as inputs to the models were excluded (coefficients of 0).

As a way to inspect the predictive models of WIC participation and infant health outcomes, consider the accumulated local effects (ALEs) described in Table 3. Accumulated local effects are a machine learning interpretability method, a way to unpack what is going on in the "black box" of the machine learning algorithms. Accumulated local effects are one of many such methods (such as partial dependence plots), but ALEs are preferred because if the variables are correlated they remain unbiased. Accumulated local effects are comparable to the coefficients of a multivariate regression in that they describe how each variable conditionally influences the prediction on average. Though comparable, they are different from linear regression coefficients in that ALEs are not necessarily linear, and ALEs incorporate parts of any non-zero interactions with other variables. Note that standard errors are not available for ALEs (Molnar, 2020). While ALEs are typically described graphically in order to capture non-linearities, we limit our description to Table 2 because most of the covariates are binary. The only non-binary covariates are the socio-environmental risk factors ($\boldsymbol{E}_{imt}$). For these socio-environmental risk covariates, the values described in Table 3

TABLE 3 Accumulated local effects (ALEs) for a sample of covariates in the models of women, infants, and children (WIC) participation, preterm birth (PTB), low birthweight (LBW), size for gestational age (SGA), and infant mortality (IM)

	WIC	Preterm birth	Low birthweight	Small Size for gestational Age	Infant mortality
Sociodemographics (*S*)					
Mom age 13–19	5.97	0.10	0	0	0
Mom age 35–39	0.39	0	0.15	0	0
Mom age 40+	0	1.44	0.22	0.62	0
Dad age 13–19	−0.45	−0.03	0	0	0
Dad age 35–39	0	0	−0.05	0	0.09
Dad age 40+	0.25	0.09	0	0	0
Mom edu HS or GED	0.64	0	0	0	0.37
Dad edu HS or GED	0.25	0	0.64	0	0
Mom race: white	−0.69	0	−3.91	0	0
Mom race: black	1.46	0	3.82	4.54	0
Dad race: white	−1.82	0	−0.69	−1.98	0
Dad race: black	1.38	0.36	0	0	0
Married	−2.68	−0.58	−2.8	0	−0.69
No SSN	1.08	4.29	2.16	11.27	13.21
Private	−0.12	0	0	−0.42	0
Medicaid	7.97	0	0.31	0	0
Maternal health (*M*)					
Hypertension	0	−2.71	0	1.14	0
Diabetes	0	6.28	0	0	0.71
BMI: obese	0.16	2.43	0	1.69	0
Smoke: ever	0	0	2.21	0	1.94
Smoke: preconception	0	0.27	3.62	0.74	0
Smoke: trimester 1	0.31	1.01	0	2.07	0
Smoke: trimester 2	0.06	0	1.54	1.57	0
Smoke: trimester3	−0.80	0	7.93	7.96	4.43
Infection: hepatitis C	0	−0.01	0	0	0.17
Infection: herpes	0	0	0.51	−0.04	0
Infection: Chlamydia	3.21	0	0	0	−0.03
Infection: chorioamnionitis	0	0	0.95	0.09	0
Opioid disorder	0.81	0	0	0	0.13
Cocaine disorder	0	1.07	0	−0.17	0
Cannabis disorder	0	0	0	−0.08	−0.02
Alcohol disorder	0	3.24	−0.06	0	0
Sedative disorder	−0.04	0	0	0.03	0
Depression	0.25	0	1.24	0	0
Reproductive history (*R*)					
Previous pregnancies	0.25	0	0.76	0	0
Previous C-section	0	0	4.31	1.96	0
Previous preterm	−2.10	6.38	0	4.64	0
Previous miscarriage	0	0	−0.03	0.11	0
First pregnancy	−0.10	0	0	0	0.22
Last pregnancy ended <1 year	0.51	0	1.97	1.29	3.35
Last pregnancy ended 1–1.5 years	0	0	0.67	1.17	0

(Continues)

TABLE 3 (Continued)

	WIC	Preterm birth	Low birthweight	Small Size for gestational Age	Infant mortality
Last pregnancy ended 1.5–2 years	0	0.51	0	0	0.05
Last pregnancy ended 2–3 years	0	0	0	0.19	−0.12
Last pregnancy ended 3 + yrs	−0.09	0	−0.65	0	0
Contraception: any	−1.39	0	0	0.02	0
Socio-environmental risk factors (*E*)					
Poverty (percentage)	2.26	0	0.73	0	0.37
Unemployment rate	3.41	−0.04	0	2.58	0
High school degree (or more) (percentage)	0	0	−0.18	0	0.02
Median monthly rent	−0.77	−0.15	0	−0.04	0
Foreign-born (percentage)	0	0.02	0	0	0.66
Air quality index	0	−2.60	−3.19	−0.44	0
Number of violent crimes	1.54	0	0.05	0	0.26
Other social services (*I*)					
Assisted housing	0.15	1.42	0	0	0
Children, youth, and families	0	0.09	2.16	0	0
Supplemental nutrition assistance program	2.12	0	0	−0.44	0
Supplemental Security income	0	0	−0.76	0	0.11
Temporary assistance for needy families	1.59	0	0	−0.54	0
Drug & alcohol program	0	1.03	0	0.02	−0.45
Family support center	0	0.13	0.02	0	0
Early headstart	−0.06	0	0	0	−0.22
Mental health services	0	−1.23	0	0.12	0
Healthy start	−0.97	0	0.03	0	0
Health services (*H*)					
First exam: preconception	0	−4.11	0	−3.81	0
First exam: trimester 1	−0.11	−0.25	−0.09	0	0
No exams	17.81	9.58	7.3	12.68	11.88
Less than 1 exam per month	2.71	2.51	1.11	0	3.46
1–2 exams per month	−1.43	0.01	−0.32	−0.04	0
1.5–2 exams per month	0	−3.7	0	−4.02	0
2+exams per month	0	0	6.19	5.45	0
Teratogenic prescription	0	0	0	0.08	0
Vaccine during pregnancy	0	−0.05	0	0	0.01
Ultrasound: abdominal circumference	−0.04	0	0	0.03	0
Ultrasound: biparietal diameter	0	0	−0.06	0	0.04
Ultrasound: head circumference	0	0.03	0.07	0	0

Note: The above represents a sample of the total covariates included in the DML estimation method. Specifically, we include covariates which are non-zero in two or more of the models. Therefore, the coefficients of covariates not included here are either 0 in each model or 0 in 4 of 5 models. Models are estimated using the healthcare data type, or the UPMC sample of 75,842 births. Standard errors are not available.

Abbreviations: BMI, body mass index; GED, equivalent diploma; HS, high school.

represent the ALE at the covariate's median value. Note that we limit the variables described in Table 3 to those with non-zero ALEs in two or more of the models.

Figure 4 describes the AUC scores, a measure of model fit and accuracy, for the predictive model of WIC participation, and the predictive models of PTB, LBW, SGA, and IM. As previously noted, AUC scores combine the true positive and false positive rates into one measure which describes the percentage of randomly drawn pairs of one individual who participates in the intervention and one individual who does not participate for which the predictive model is accurate. An AUC score above

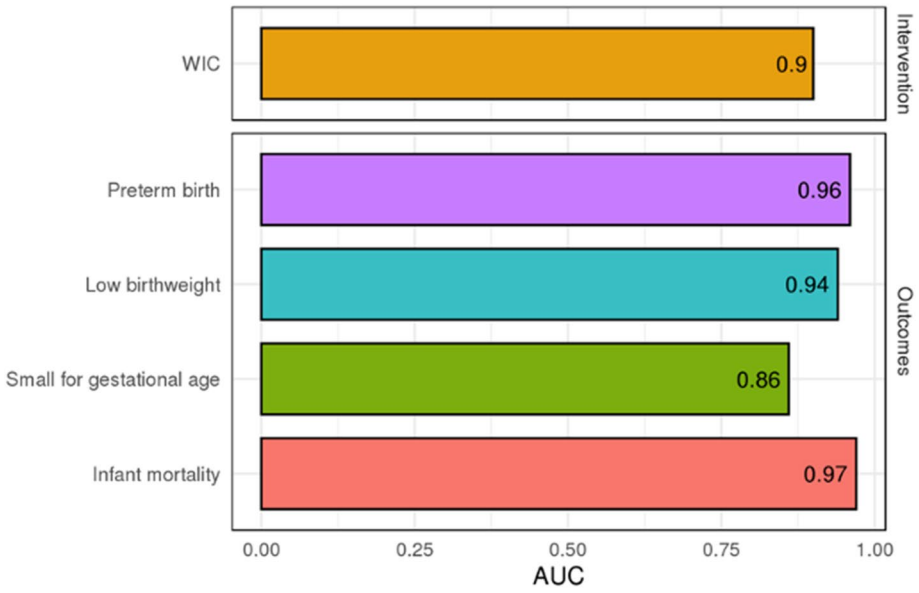

FIGURE 4 Area under the curve (AUC) scores for the predictive models of women, infants, and children (WIC) and infant health outcomes. Predictions are based on the healthcare data type, or the UPMC sample of 75,842 births. [Colour figure can be viewed at wileyonlinelibrary.com]

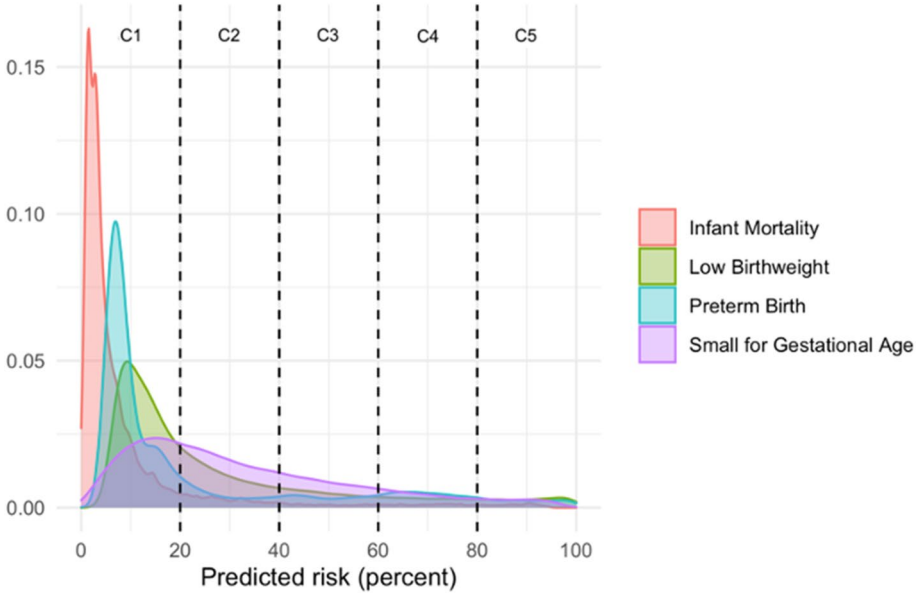

FIGURE 5 Distributions of predicted risk of each infant health outcome. Predictions are based on the healthcare data type, or the UPMC sample of 75,842 births. Dotted lines denote the boundaries of the risk categories. [Colour figure can be viewed at wileyonlinelibrary.com]

0.85 is considered high accuracy, 0.75–0.85 is moderate accuracy, and less than 0.75 is low accuracy (Bowers & Zhou, 2019). Also note that, as is standard practice, each of the models are fit on a training set of data and AUC scores are calculated on a test set, or a random subsample of the data withheld from the training set. For all the models we estimate, the accuracy is high with AUC scores between 0.86 and 0.97. Importantly, these results indicate that participation in WIC is well predicted by the observed data available to healthcare providers. Or, in other words, selection into WIC participation is well accounted for by observed characteristics. Also note that because the healthcare data is the most extensive, the predictive performance of these models is better than similar models with less extensive data (i.e., the reportable and potentially reportable data types).

While the AUC scores are calculated based on discretized predictions, the predictions can also be captured as probabilities on a 0–100 scale. Figure 5 shows the resulting distributions of predicted risk probabilities for each outcome. Due to the relative infrequency of each outcome, particularly IM, each of these distributions is highly skewed. This is important to remember when

considering the estimated average treatment effects. Because the vast majority of infants are low risk (in the first category, 0–20, of predicted risk), the effect of WIC among these observations will heavily influence the estimated *average* treatment effects.

4.2 | Average treatment effects of WIC

The estimates in Figure 6 describe the average treatment effects of WIC on the infant health outcomes PTB, LBW, SGA, and IM. However, these estimates are generally in the range of, or more conservative than the average treatment effects reported in the existing literature. For PTB, we estimate that the average treatment effect of WIC is −0.04 pp, a much smaller estimate of WIC's effects than the −0.8 pp estimate Currie and Rajani (2015) produce using a maternal fixed effects design (Currie & Rajani, 2015). In contrast, we estimate that the average treatment effect of WIC on LBW is −0.43 pp which aligns with the Currie and Rajani (2015) estimate of −0.41 pp and close to the Hamad et al. (2019) estimate of −0.2 pp. For SGA, we estimate that the average treatment effect of WIC is −0.51 pp which is an approximate mid-point between the Currie and Rajani (2015), estimate of −0.29 pp, the Hamad et al. (2019) estimate of −0.4 pp, and the Sonchak (2016) estimate of −0.9 pp. For IM, we estimate that average treatment effect of WIC is 0.39 pp (which has no comparisons in the literature of which we are aware). While none of these estimates are significant at the 95% level (as demonstrated by the 95% confidence intervals in Figure 6), both the LBW and SGA estimates are statistically significant at the 90% level.

4.3 | Effects of WIC by predicted risk category

Moving beyond average treatment effects, we examined heterogeneity in the effects of WIC participation by the category of predicted risk of each infant health outcome (Figures 7–10). These results address the question of whether or not WIC is benefiting those at greatest risk.

First, we considered the effects of WIC by the category of predicted risk for PTB (Figure 7). In the lowest category of predicted risk—which comprises 69% of the observations—the estimated effect of WIC is essentially 0. However, the estimated benefits of WIC increases along with the predicted risk level. Still, among the lowest three categories of predicted risk—which comprises 84% of the observations—the estimated effects of WIC are not statistically significant. Nevertheless, in the top two categories of predicted risk the estimated effects of WIC are statistically significant and large. Among those with a predicted risk of 60%–80% (only 8% of the sample) WIC reduces the risk of PTB by 3.52 pp, and among those with a predicted risk of 80%–100% (also only 8% of the sample) WIC reduces the risk of PTB by 2.19 pp.

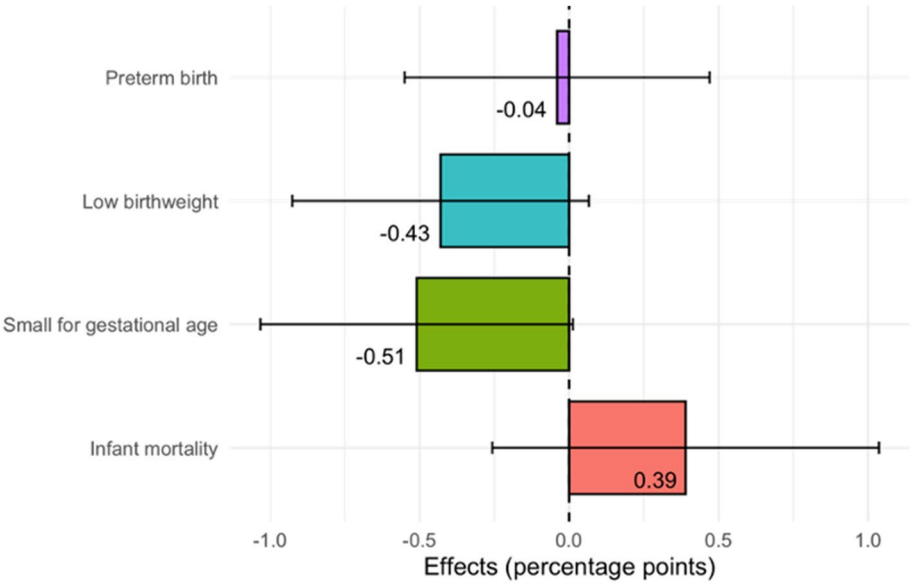

FIGURE 6 Average treatment effects of women, infants, and children (WIC). Estimates are based on the healthcare data type, or the UPMC sample of 75,842 births. For LBW and IM, $\boldsymbol{Z}_{imt}$ includes gestational age; for PTB and SGA, $\boldsymbol{Z}_{imt}$ does not include gestational age. Confidence intervals are calculated using heteroscedasticity robust standard errors (Chernozhukov et al., 2018). [Colour figure can be viewed at wileyonlinelibrary.com]

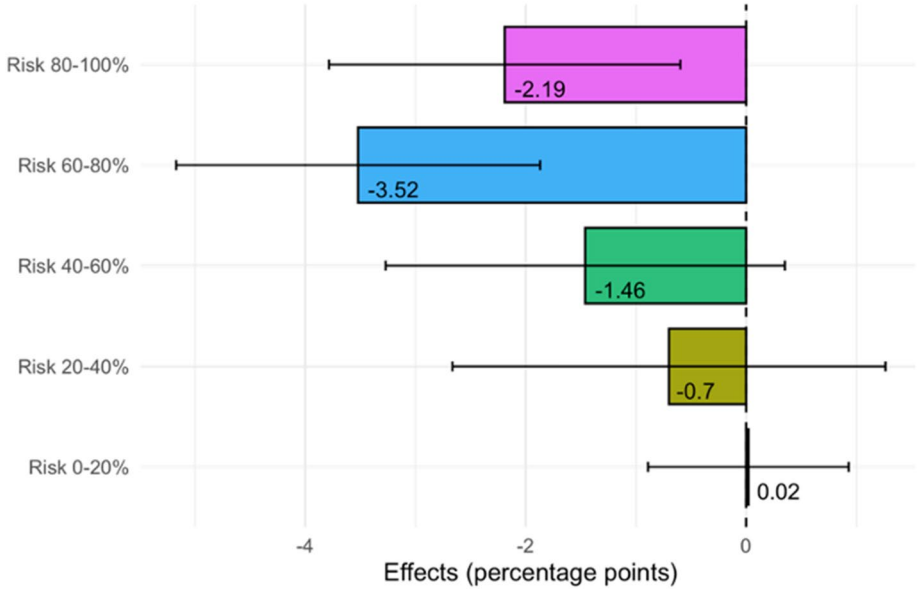

FIGURE 7 Women, infants, and children (WIC) effects on preterm birth (PTB) by the category of predicted risk. Estimates are based on the healthcare data type, or the UPMC sample of 75,842 births. $\boldsymbol{Z}_{imt}$ does not include gestational age. 95% Confidence intervals are calculated using heteroscedasticity robust standard errors (Chernozhukov et al., 2018). [Colour figure can be viewed at wileyonlinelibrary.com]

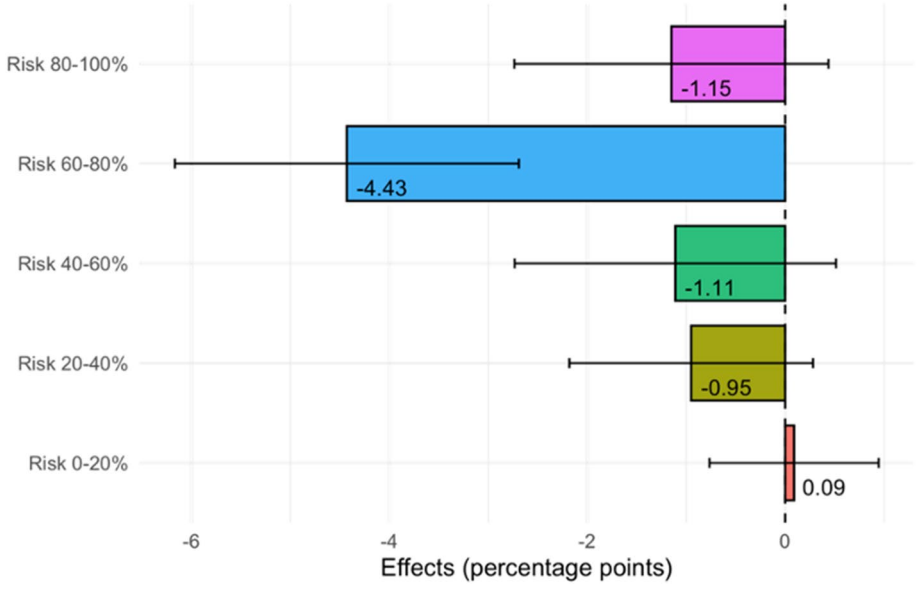

FIGURE 8 Women, infants, and children (WIC) effects on low birthweight (LBW) by the category of predicted risk. Estimates are based on the healthcare data type, or the UPMC sample of 75,842 births. $\boldsymbol{Z}_{imt}$ includes gestational age. 95% Confidence intervals are calculated using heteroscedasticity robust standard errors (Chernozhukov et al., 2018). [Colour figure can be viewed at wileyonlinelibrary.com]

Second, we examined the effects of WIC by predicted risk for LBW (Figure 8). In the lowest category of predicted risk −62% of the observations—the estimated effect of WIC is essentially 0. However, like with PTB, the estimated effects of WIC on LBW increase along with the predicted risk level. If the second and third quartiles were grouped together (predicted risk between 20% and 60%, 24% of the sample) the estimated effect would be 1 pp and statistically significant. This is similar to the estimated effect of WIC effect in the top category of predicted risk. But the largest effect is seen in the 60%–80% category. In that group, which comprises 6% of the sample, WIC is estimated to significantly reduce the risk of LBW by 4.43 pp.

Third, we considered the effects of WIC by predicted risk for SGA (Figure 9). In the lowest two categories—0% to 40% predicted risk and 67% of the sample—the estimated effect of WIC is essentially 0. The effect become marginally significant

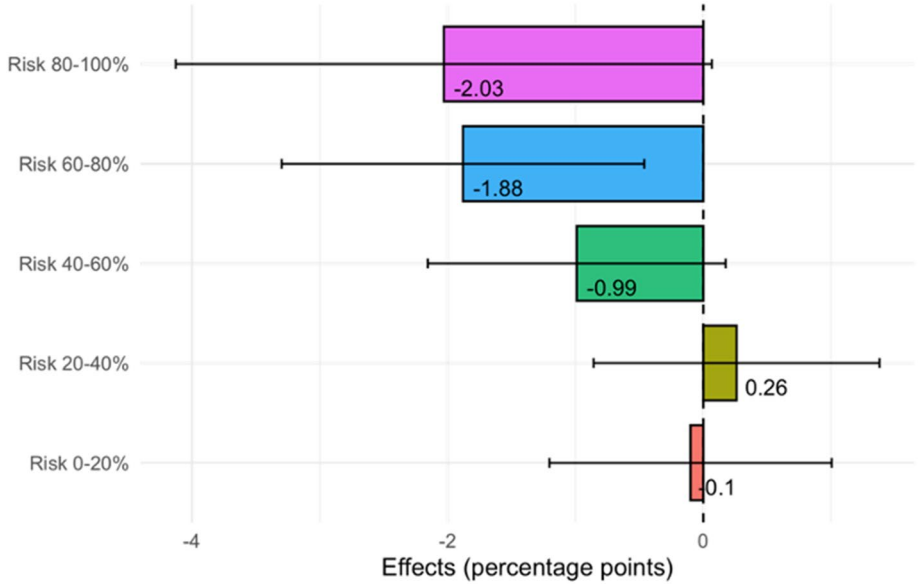

FIGURE 9 Women, infants, and children (WIC) effects on small size for gestational age (SGA) by the category of predicted risk. Estimates are based on the healthcare data type, or the UPMC sample of 75,842 births. $\boldsymbol{Z}_{imt}$ does not include gestational age. 95% confidence intervals are calculated using heteroscedasticity robust standard errors (Chernozhukov et al., 2018). [Colour figure can be viewed at wileyonlinelibrary.com]

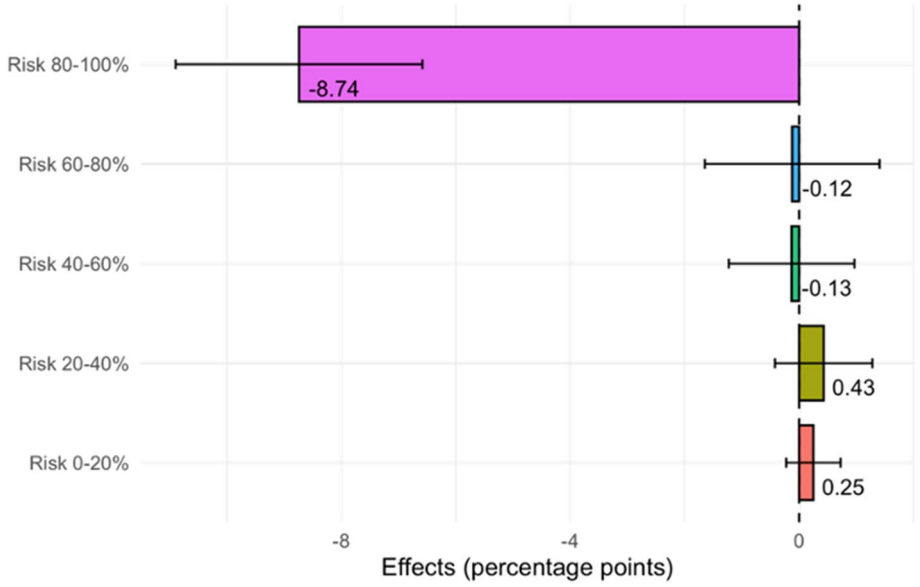

FIGURE 10 Effects of women, infants, and children (WIC) by the category of predicted risk on infant mortality (IM). Estimates are based on the healthcare data type, or the UPMC sample of 75,842 births. $\boldsymbol{Z}_{imt}$ includes gestational age. 95% confidence intervals are calculated using heteroscedasticity robust standard errors (Chernozhukov et al., 2018). [Colour figure can be viewed at wileyonlinelibrary.com]

for the third categories—a 0.99 pp reduction, significant at the 10% level, for 18% of the sample. As with LBW, the effects of WIC on SGA become significant at the 60%–80% risk level, a 1.88 pp reduction for 9% of the sample. The effect is similar in size (a 2.03 pp reduction) for the 80%–100% risk level though less significant due to the smaller size of the subsample (4% of the sample).

Fourth, we examined the effects of WIC by predicted risk for IM (Figure 10). With IM we see the largest variance in the estimated effects of WIC by the predicted risk category. In the bottom four categories of predicted risk—comprising 98% of the sample—the estimated effects of WIC are essentially 0. However, among those at greatest risk of IM in the top category, WIC is estimated to significantly reduce the risk of IM by 8.74 pp. This is a large and important result.

Compare these results with those of previous studies that examine heterogeneity among certain at-risk subgroups, such as (1) Black women, and (2) Medicaid beneficiaries. We describe WIC's estimated average effects among these groups in Figure A.1. While there is heterogeneity by these characteristics, the effects of WIC among these groups are not significantly different from the main sample estimates (refer to Figure 6). In essence, heterogeneity analysis such as this inaccurately defines risk using one or a small number of characteristics. In contrast, our approach uses a high dimensional set of characteristics to accurately define risk and demonstrates that the effects of WIC vary significantly along this gradient.

4.4 | Multiple hypothesis testing

Because we have examined the effects on multiple infant health outcomes and stratified the estimated effects by the predicted risk of each outcome, we have conducted multiple hypothesis tests which require adjustments. Table 4 describes the unadjusted and adjusted *p*-values for each of the tests. Overall, these adjustments do not alter the conclusions we draw.

4.5 | WIC-eligible control group

Here we discuss the estimated effects of WIC when the control group is limited to only those that are WIC eligible. To determine WIC eligibility, we use automatic income eligibility standards based on participation in Medicaid, SNAP, or TANF (USDA, 2022a, 2022b). Table 5 compares the full population estimates to WIC-eligible control group estimates for each of the infant health outcomes. Overall, the results show that while limiting the control group to those that were WIC eligible but did not participate does alter some point estimates, the overall takeaways do not change—WIC is shown to be most effective at the highest levels of predicted risk.

TABLE 4 Multiple hypothesis test adjustments

Outcome	Risk category	Obs. (*N*)	Estimate	Unadjusted *p*-value	Adjusted *p*-value
Preterm birth	80%–100%	5322	−2.19	0.007***	0.035**
	60%–80%	5554	−3.52	0.000***	0.000***
	40%–60%	5930	−1.46	0.114	0.285
	20%–40%	6344	−0.7	0.485	0.693
	0%–20%	52,692	0.02	0.966	0.966
Low birthweight	80%–100%	4374	−1.15	0.156	0.311
	60%–80%	5441	−4.43	0.000***	0.000***
	40%–60%	6243	−1.11	0.180	0.328
	20%–40%	12,891	−0.95	0.131	0.290
	0%–20%	46,893	0.09	0.837	0.924
Small size for gestational age	80%–100%	3465	−2.03	0.058*	0.193
	60%–80%	6733	−1.88	0.009***	0.038**
	40%–60%	13,408	−0.99	0.096*	0.274
	20%–40%	24,869	0.26	0.649	0.865
	0%–20%	27,367	−0.1	0.859	0.924
Infant mortality	80%–100%	1316	−8.74	0.000***	0.000***
	60%–80%	1402	−0.12	0.878	0.924
	40%–60%	1646	−0.13	0.817	0.924
	20%–40%	5576	0.43	0.322	0.496
	0%–20%	65,902	0.25	0.302	0.496

Note: *p*-value adjustments are made using the Benjamini and Hochberg (1995) method.

TABLE 5 Full population and WIC-eligible population effects of women, infants, and children (WIC) by the category of predicted risk

Outcome		Full population			WIC-eligible population		
	Risk category	Estimate	Std. Error	Obs. (*N*)	Estimate	Std. Error	Obs. (*N*)
Preterm birth	80%–100%	−2.19***	0.81	5322	−2.05*	1.15	2371
	60%–80%	−3.52***	0.84	5554	−5.79***	1.30	2514
	40%–60%	−1.46	0.92	5930	−0.61	1.56	3218
	20%–40%	−0.7	1.00	6344	−0.67	1.41	3912
	0%–20%	0.02	0.46	52,692	0.01	0.58	22,760
Low birthweight	80%–100%	−1.15	0.81	4374	−2.57**	1.11	2985
	60%–80%	−4.43***	0.89	5441	−5.94***	1.28	3817
	40%–60%	−1.11	0.83	6243	−1.88	1.24	3799
	20%–40%	−0.95	0.63	12,891	−0.87	0.89	7225
	0%–20%	0.09	0.44	46,893	−0.48	0.62	16,949
Small size for gestational age	80%–100%	−2.03*	1.07	3465	−2.27	1.66	2549
	60%–80%	−1.88***	0.72	6733	−2.36**	1.00	5299
	40%–60%	−0.99*	0.59	13,408	−1.94**	0.82	6986
	20%–40%	0.26	0.57	24,869	0.1	0.79	9084
	0%–20%	−0.1	0.56	27,367	0.11	0.99	10,857
Infant mortality	80%–100%	−8.74***	1.10	1316	−6.39***	1.45	997
	60%–80%	−0.12	0.78	1402	1.61	1.23	1032
	40%–60%	−0.13	0.56	1646	−0.08	0.78	1130
	20%–40%	0.43	0.43	5576	0.64	0.58	3819
	0%–20%	0.25	0.24	65,902	−0.32	0.43	27,797

Note: Estimates are based on the healthcare data type, or the UPMC sample. The full population is comprised of 75,842 births. The WIC-eligible population denotes those participating in Medicaid, SNAP, and/or TANF, and is comprised of 34,775 births. For LBW and IM, Z includes gestational age; for PTB and SGA, Z does not include gestational age. Standard errors are heteroscedasticity robust (Chernozhukov et al., 2018).

5 | PLACEBO OUTCOME TESTS

In this section we discuss placebo outcome tests that we conducted. We replicate the main analysis using different outcome variables which should be unaffected by WIC participation, specifically: (1) maternal education, (2) paternal education, (3) preconception obesity, and (4) contraceptive use. The education outcomes are actually predictors of infant health outcomes, however conditional on other observed confounders the results show no relationship between WIC and these placebo outcomes using the same or slightly smaller sample (missing paternal education observations reduces the sample from 75,842 to 73,967). This suggests that selection related to education or other sociodemographic characteristics is not driving the results. To assess whether selection related to maternal health (both overall and reproductive) is driving the results, we examine preconception obesity and contraceptive use. In these cases, we restrict the sample to women experiencing their first pregnancy because women remain eligible up to the infant's first birthday, thus potentially allowing WIC to impact obesity and contraceptive use in subsequent pregnancies. We observe 33,919 first pregnancies in the healthcare data type (or UPMC sample). Note that for these outcomes, $\boldsymbol{Z}_{imt}$ is restricted to either permanent characteristics (e.g., race/ethnicity) or characteristics observed during the preconception period (e.g., keep non-gestational diabetes but exclude gestational diabetes). The results shown in Figure 11 demonstrate that WIC participation during pregnancy and infancy is unrelated to preconception obesity and contraceptive use. This provides further evidence for the causal interpretation of the results described in Figures 6–10.

6 | ROBUSTNESS

The results presented up to this point use the "healthcare" data type. In this section we discuss robustness checks by which we use the (1). "reportable", and (2). "potentially reportable" data types. In the Appendix, Figure A.2 describes the AUC scores for estimating Equations (2) and (3) using the "reportable" and "potentially reportable" data types. Compared to the models

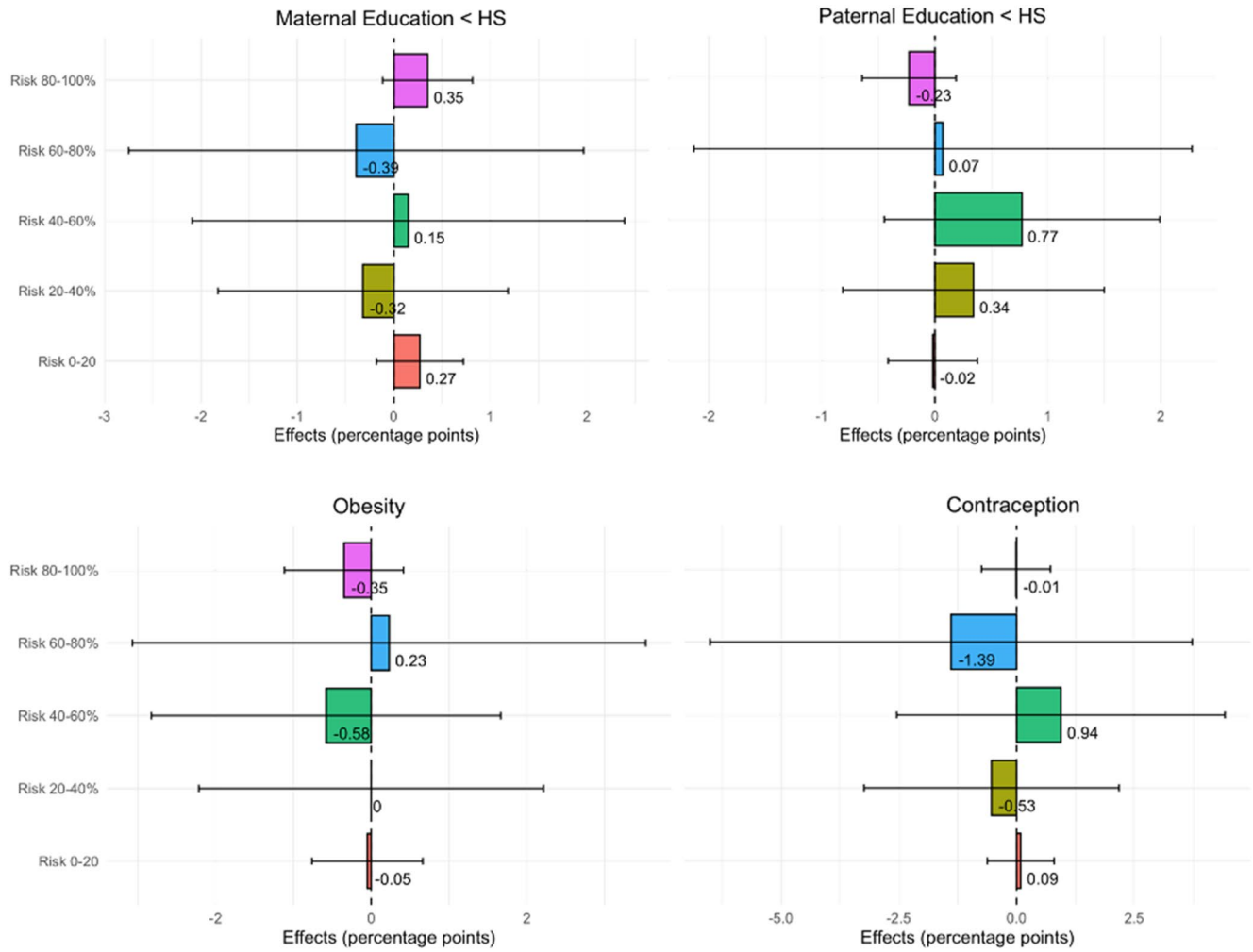

FIGURE 11 Women, infants, and children (WIC's) placebo effects on: (1, top left) maternal education greater than high school or GED; (2, top right) paternal education greater than high school or GED; (3, bottom left) preconception obesity, and (4, bottom right) contraceptive use by the category of predicted probabilities. Estimates are based on the healthcare data type, or the UPMC sample. The sample size for the maternal education outcome is $N = 75{,}842$. Because paternal education has some missing observations, the sample size is $N = 73{,}967$. For pre-pregnancy obesity and contraceptive use, the sample is restricted to women experiencing their first pregnancy, or $N = 33{,}919$. $\boldsymbol{Z}_{imi}$ is restricted to permanent characteristics or those observed during the preconception period. Confidence intervals are calculated using heteroscedasticity robust standard errors (Chernozhukov et al., 2018). GED, equivalent diploma. [Colour figure can be viewed at wileyonlinelibrary.com]

using the healthcare data type (Figure 4), the AUC scores for the "potentially reportable" and "reportable" data types are lower, as expected. However, it is encouraging that the AUC scores are of high or moderate accuracy despite access to less data. Figures A.3–A.6 show the stability of the estimated effects of WIC on each outcome by category of predicted risk. The estimates show some variation but the overall takeaway for the estimates produced by the "potentially reportable" and "reportable" data types is the same as the main estimates (Figures 6–10).

7 | CONCLUSION

In this study, we provide novel evidence of: (1) accurate infant health outcome and WIC participation predictions using machine learning and a unique combination of vital statistics, electronic health records, and other administrative data from various sources; and (2) variation in the infant health effects of WIC depending on the level of predicted risk of each health outcome. Overall, our results show that WIC is most beneficial among those at greatest risk of the negative infant health outcomes. We use a unique and rich database and the novel DML method. Beyond estimates of average treatment effects that align with the existing literature, our approach produces unique insights into variation in the effects of WIC by predicted risk, showing how

WIC affects those with the greatest risk of poor outcomes. Each of the examined infant health outcomes is relatively infrequent (PTB = 11%; LBW = 8.7%, SGA = 5.5%, and IM = 6.5 per 1000 live births), thus average treatment effect can mask how WIC affects those at higher risk. Our main results show that, for each infant health outcome, the greatest benefits of WIC are concentrated among those at greatest predicted risk. Consider the estimated effect of WIC on those at greatest predicted risk of IM (an 8.74 pp reduction). Four hundred ninety five infants died between 2003 and 2013 in the UPMC sample (1009 in the full sample), a rate of 6.5 per 1000 live births. Our results imply that if WIC were not available, there would have been an additional 43 deaths among the UPMC sample (or an additional 88 deaths among the full sample). This variation by predicted risk level calls for a reassessment of the program's cost/benefit calculations, which are typically based solely on average treatment effects. While the groups that are most affected may be small, the benefits received may be large and change the cost/benefit calculation.

A nuanced understanding of the costs/benefits of social programs like WIC is important for policy decisions. On the one hand, a reader may interpret our results as evidence that targeting WIC to those predicted, based on observable characteristics, at risk of negative infant health outcomes could increase the impact of the program. While true that WIC and other social programs have limited resources and targeted distribution may maximize the social benefits, there are a number of caveats to consider. First, WIC is not only intended to improve infant outcomes, but also maternal outcomes. And while maternal and child health outcomes are correlated, there may be discrepancies between the predicted risk of infant outcomes and maternal outcomes. Second, our results echo those of previous studies suggesting mild benefits on average, even for those not at predicted risk of negative infant health outcomes. Since, WIC participation rates are declining among those eligible, any increase in participation will likely be beneficial. Third, as demonstrated in Figure A.2 the accuracy of our predictions decline, though marginally, as the data available to our models declines. In areas with limited data infrastructure it would be impossible to construct a comparable database to generate accurate risk predictions.

However, combining our results with the understanding that WIC participation among the eligible is declining, a nuanced policy recommendation emerges: target efforts to boost participation among those eligible and at greatest predicted risk of negative outcomes. This recommendation involves leveraging data and technology—where available—which certain states have done increasingly. For instance, in Utah, WIC has established a data sharing agreement with other social programs (Partnerships, 2021). With agreements in place to allow local public health departments and social services to share data, predictive models could be developed and integrated to identify those at greatest risk of negative outcomes, and those most likely to participate. With this information, further efforts to boost participation—such as mobile WIC shopping apps, text reminders, and online nutrition education currently offered in Maryland (WHAM Global, 2019) or telehealth visits offered in Georgia (FRAC, 2019)—could be targeted while maintaining broad program eligibility. This avoids adding eligibility restrictions, which research suggests is a cause of declining participation (Partnerships, 2021), while producing large, cost-offsetting benefits among the targeted who enroll.

The main limitation of this study is that the causal interpretation of our findings is dependent on the assumption of conditional unconfoundedness (Chernozhukov et al., 2018). While the data is extensive, there is information that is unobserved. However, where there is missing information, in many cases there are other variables that may be highly correlated or that in combination may capture much of the missing information. For example, while the data does not include a direct measure of income, a number of other variables (e.g., age, education, insurance status/payer, SNAP/TANF participation, poverty/unemployment in the area of the mother's residence) are highly correlated or may combine to adequately proxy for income. Given the extensiveness of the data and because machine learning algorithms can capture unspecified non-linear and/or interactive relationships in the data, there is likely a limited role for unobserved confounders of the infant health outcomes.

Furthermore, while the conditional unconfoundedness assumption is unverifiable, we have offered multiple points of evidence in its support. First, consistent with the literature (Bitler & Currie, 2005; Currie & Rajani, 2015), our sample is comprised of WIC participants who are more disadvantaged than non-participants across a number of metrics. Second, the average effects that we estimate are in line with the many of the previously published average effects of WIC. Third, the predictive model of WIC participation performs very well with observable information, suggesting a limited role for unobserved confounders of WIC participation. Fourth, when we limit the control group to those that were WIC-eligible (based on their use of either Medicaid, SNAP, or TANF) we find that the estimates are similar to the full population estimates. Fifth, we conducted placebo tests for outcomes that should be unaffected by WIC participation and found no effects. Finally, the estimates are shown to be robust to the extensivity of confounding variables included in the models.

ACKNOWLEDGMENTS

While we have received much assistance from colleagues, all errors in this paper are our own. Furthermore, any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of any affiliate. All results have

been reviewed to ensure that no confidential information is disclosed. Financial support via a grant for this research from the Richard King Mellon Foundation is gratefully acknowledged.

CONFLICT OF INTEREST

Each of the authors acknowledge grant funding from the Richard King Mellon Foundation during the conduct of the study.

DATA AVAILABILITY STATEMENT

Our team's use of the data has been governed by strict data use agreements between our team's organizations and the organizations that supplied the data. Moreover, our data is composed of sensitive personal health information. The data use agreements that govern our team's use of the data prohibit us from sharing the data outside of the study team.

ETHICS STATEMENT

This study was approved with a waiver of consent by RAND's Institutional Review Board.

ORCID

Evan D. Peet https://orcid.org/0000-0001-6147-3475

REFERENCES

Abadie, A., & Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, *10*(1), 465–503. https://doi.org/10.1146/annurev-economics-080217-053402

Abadie, A., Chingos, M. M., & West, M. R. (2018). Endogenous stratification in randomized experiments. *The Review of Economics and Statistics*, *100*(4), 567–580. https://doi.org/10.1162/rest_a_00732

Aizer, A., & Currie, J. (2014). The intergenerational transmission of inequality: Maternal disadvantage and health at birth. *Science*, *344*(6186), 856–861. https://doi.org/10.1126/science.1251872

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *The Journal of Economic Perspectives*, *31*(2), 3–32. https://doi.org/10.1257/jep.31.2.3

Berkowitz, G. S., & Papiernik, E. (1993). Epidemiology of preterm birth. *Epidemiologic Reviews*, *15*(2), 414–443. https://doi.org/10.1093/oxfordjournals.epirev.a036128

Bertrand, M., Crépon, B., Marguerie, A., & Premand, P. (2017). Contemporaneous and post-program impacts of a public works program.

Bitler, M. P., & Currie, J. (2005). Does WIC work? The effects of WIC on pregnancy and birth outcomes. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, *24*(1), 73–91. https://doi.org/10.1002/pam.20070

Bowers, A. J., & Zhou, X. (2019). Receiver operating characteristic (ROC) area under the curve (AUC): A diagnostic measure for evaluating the accuracy of predictors of education outcomes. *Journal of Education for Students Placed at Risk*, *24*(1), 20–46. https://doi.org/10.1080/10824669.2018.1523734

Brien, M. J., & Swann, C. A. (1997). *Prenatal WIC participation and infant health: Selection and maternal fixed effects*. Thomas Jefferson Center for Political Economy, University of Virginia.

Chatterji, P., Bonuck, K., Dhawan, S., & Deb, N. (2002). *WIC participation and the initiation and duration of breastfeeding*. Institute for Research on Poverty, University of Wisconsin-Madison.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). *Double/debiased machine learning for treatment and structural parameters*. Oxford University Press.

Chorniy, A., Currie, J., & Sonchak, L. (2020). Does prenatal WIC participation improve child outcomes? *American Journal of Health Economics*, *6*(2), 169–198. https://doi.org/10.1086/707832

Corman, H., Dave, D. M., & Reichman, N. (2018). *Effects of prenatal care on birth outcomes: Reconciling a messy literature*. National Bureau of Economic Research.

Currie, J., & Rajani, I. (2015). Within-mother estimates of the effects of WIC on birth outcomes in New York city. *Economic Inquiry*, *53*(4), 1691–1701. https://doi.org/10.1111/ecin.12219

Davis, J., & Heller, S. B. (2017). Using causal forests to predict treatment heterogeneity: An application to summer jobs. *The American Economic Review*, *107*(5), 546–550. https://doi.org/10.1257/aer.p20171000

Demir-Kavuk, O., Kamada, M., Akutsu, T., & Knapp, E.-W. (2011). Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features. *BMC Bioinformatics*, *12*(1), 1–10. https://doi.org/10.1186/1471-2105-12-412

El-Bastawissi, A. Y., Peters, R., Sasseen, K., Bell, T., & Manolopoulos, R. (2007). Effect of the Washington special supplemental nutrition program for women, infants and children (WIC) on pregnancy outcomes. *Maternal and Child Health Journal*, *11*(6), 611–621. https://doi.org/10.1007/s10995-007-0212-5

Farbmacher, H., Kögel, H. & Spindler, M. (2019). Heterogeneous effects of poverty on cognition.

Figlio, D., Hamersma, S., & Roth, J. (2009). Does prenatal WIC participation improve birth outcomes? New evidence from Florida. *Journal of Public Economics*, *93*(1–2), 235–245. https://doi.org/10.1016/j.jpubeco.2008.08.003

Foster, D. J., & Syrgkanis, V. (2019). Orthogonal statistical learning. arXiv preprint arXiv:1901.09036.

Foster, E. M., Jiang, M., & Gibson-Davis, C. M. (2010). The effect of the WIC program on the health of newborns. *Health Services Research*, *45*(4), 1083–1104. https://doi.org/10.1111/j.1475-6773.2010.01115.x

FRAC. (2019). "Making WIC work better: Strategies to reach more women and children and strengthen benefits use.".

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/s0167-9473(01)00065-2

Frisch, R., & Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, *1*(4), 387–401. https://doi.org/10.2307/1907330

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology*, *173*(7), 761–767. https://doi.org/10.1093/aje/kwq439

Global, W. (2019). Reinvigorating and reimagining the Pennsylvania WIC program.

Gulyas, A., & Pytka, K. (2019). *Understanding the sources of earnings losses after job displacement: A machine-learning approach*. University of Bonn and University of Mannheim.

Hamad, R., Collin, D. F., Baer, R. J., & Jelliffe-Pawlowski, L. L. (2019). Association of revised WIC food package with perinatal and birth outcomes: A quasi-experimental study. *JAMA Pediatrics*, *173*(9), 845–852. https://doi.org/10.1001/jamapediatrics.2019.1706

Hoynes, H. W., Page, M. E., & Stevens, A. H. (2009). *Is a WIC start a better start? Evaluating WIC's impact on infant health using program introduction*. National Bureau of Economic Research.

Hoynes, H. W., & Schanzenbach, D. W. (2009). Consumption responses to in-kind transfers: Evidence from the introduction of the food stamp program. *American Economic Journal: Applied Economics*, *1*(4), 109–139. https://doi.org/10.1257/app.1.4.109

Johnson, P., Giannarelli, L., Huber, E., Betson, D., & Lovellette, G. (2015). *National and state-level estimates of special supplemental nutrition program for women, infants, and children (WIC) eligibles and program reach, 2011*. United States Department of Agriculture, Food and Nutrition Service, Office of Policy Support.

Joyce, T., Gibson, D., & Colman, S. (2005). The changing association between prenatal participation in WIC and birth outcomes in New York City. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, *24*(4), 661–685. https://doi.org/10.1002/pam.20131

Joyce, T., Racine, A., & Yunzal-Butler, C. (2008). Reassessing the WIC effect: Evidence from the pregnancy nutrition surveillance system. *Journal of Policy Analysis and Management*, *27*(2), 277–303. https://doi.org/10.1002/pam.20325

Khanani, I., Elam, J., Hearn, R., Jones, C., & Maseru, N. (2010). The impact of prenatal WIC participation on infant mortality and racial disparities. *American Journal of Public Health*, *100*(S1), S204–S209. https://doi.org/10.2105/ajph.2009.168922

Knaus, M. C. (2020). Double machine learning based program evaluation under unconfoundedness. arXiv preprint arXiv:2003.03191.

Knaus, M. C., Lechner, M., & Strittmatter, A. (2020). Heterogeneous employment effects of job search programmes: A machine learning approach. *Journal of Human Resources*, 0718–9615R0711.

Knaus, M. C., Lechner, M., & Strittmatter, A. (2021). Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence. *The Econometrics Journal*, *24*(1), 134–161. https://doi.org/10.1093/ectj/utaa014

Knittel, C. R., & Stolper, S. (2019). *Using machine learning to target treatment: The case of household energy use*. National Bureau of Economic Research.

Kowaleski-Jones, L., & Duncan, G. J. (2002). Effects of participation in the WIC program on birthweight: Evidence from the national longitudinal survey of youth. *American Journal of Public Health*, *92*(5), 799–804. https://doi.org/10.2105/ajph.92.5.799

Kreider, B., Pepper, J. V., & Roy, M. (2020). Does the women, infants, and children program improve infant health outcomes? *Economic Inquiry*, *58*(4), 1731–1756. https://doi.org/10.1111/ecin.12900

Liu, C. H., & Liu, H. (2016). Concerns and structural barriers associated with WIC participation among WIC-eligible women. *Public Health Nursing*, *33*(5), 395–402. https://doi.org/10.1111/phn.12259

Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, *58*(304), 993–1010. https://doi.org/10.1080/01621459.1963.10480682

Marčenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR: Sbornik*, *1*(4), 457–483. https://doi.org/10.1070/sm1967v001n04abeh001994

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Oliveira, V. J. (2002). *The WIC program: Background, trends, and issues*. US Department of Agriculture, Economic Research Service.

Park, M. Y., & Hastie, T. (2007). L1-regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society: Series B*, *69*(4), 659–677. https://doi.org/10.1111/j.1467-9868.2007.00607.x

Partnerships, P. (2021). A time to thrive: Growing Pennsylvania WIC's impact on children and families.

Peet, E. D. (2021). Early-life environment and human capital: Evidence from the Philippines. *Environment and Development Economics*, *26*(1), 1–25. https://doi.org/10.1017/s1355770x20000224

Peet, E. D., Schultz, D., & Lovejoy, S. L. (2021). "Using an innovative database and machine learning to predict and reduce infant mortality." RAND Report **RB-A858-1**.

Rossin-Slater, M. (2013). WIC in your neighborhood: New evidence on the impacts of geographic access to clinics. *Journal of Public Economics*, *102*, 51–69. https://doi.org/10.1016/j.jpubeco.2013.03.009

Schultz, D., Lovejoy, S. L., & Peet, E. D. (2020). "Examining interventions to address infant mortality in Allegheny county." RAND Report **RR-A858-1**.

Sonchak, L. (2016). The impact of WIC on infant health: Evidence from South Carolina. *Maternal and Child Health Journal*, *20*(7), 1518–1525. https://doi.org/10.1007/s10995-016-1951-y

Strittmatter, A. (2018). What is the value added by using causal machine learning methods in a welfare experiment evaluation? arXiv preprint arXiv:1812.06533.

Swann, C. A. (2010). WIC eligibility and participation: The roles of changing policies, economic conditions, and demographics. *The B.E. Journal of Economic Analysis & Policy*, *10*(1). https://doi.org/10.2202/1935-1682.2352

Tsui, F. R., & Shi, L. (2019). Predicting infant mortality from delivery electronic health records. In *AMIA clinical informatics conference*.

Tsui, F. R., Yang, L., Shi, L., & Wu, L. (2017). *Comprehensive data repository towards infant mortality prediction*. University of Pennsylvania.

Tsui, F. R., Yang, L., Shi, L., & Wu, L. (2019). *A data-driven approach for integrating infant mortality prediction and survival time estimate from electronic health records*. American Medical Informatics Association.

USDA. (2022a). "WIC data tables.".

USDA. (2022b). "WIC eligibility requirements.".

Vegetabile, B. G., Cefalu, M., Peet, E. D., Pane, J. & Damberg, C. L. (2020). Interpretable machine learning models for public policy analysis. RAND Report **IN-A828-1**.

Whaley, S. E., Ritchie, L. D., Spector, P., & Gomez, J. (2012). Revised WIC food package improves diets of WIC families. *Journal of Nutrition Education and Behavior*, *44*(3), 204–209. https://doi.org/10.1016/j.jneb.2011.09.011

Yang, L., Posada, J. D., Su, h.-d., Shi, L., Fan, M., & Tsui, R. (2017). *Using logistic regression to verify completeness of electronic health records for infant mortality analysis*. AMIA.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Peet, E. D., Schultz, D., Lovejoy, S., & Tsui, F. (R.) (2023). Variation in the infant health effects of the women, infants, and children program by predicted risk using novel machine learning methods. *Health Economics*, *32*(1), 194–217. https://doi.org/10.1002/hec.4617