

5-2018

Working Paper Number 18013

Does the Women, Infants, and Children Program (WIC) Improve Infant Health Outcomes?

Brent Kreider

Iowa State University, bkreider@iastate.edu

John V. Pepper

University of Virginia

Manan Roy

Appalachian State University

Original Release Date: May 2018

Follow this and additional works at: https://lib.dr.iastate.edu/econ_workingpapers

 Part of the [Behavioral Economics Commons](#), [Econometrics Commons](#), [Health Economics Commons](#), and the [Social Welfare Commons](#)

Recommended Citation

Kreider, Brent; Pepper, John V.; and Roy, Manan, "Does the Women, Infants, and Children Program (WIC) Improve Infant Health Outcomes?" (2018). *Economics Working Papers*: Department of Economics, Iowa State University. 18013.

https://lib.dr.iastate.edu/econ_workingpapers/56

Iowa State University does not discriminate on the basis of race, color, age, ethnicity, religion, national origin, pregnancy, sexual orientation, gender identity, genetic information, sex, marital status, disability, or status as a U.S. veteran. Inquiries regarding non-discrimination policies may be directed to Office of Equal Opportunity, 3350 Beardshear Hall, 515 Morrill Road, Ames, Iowa 50011, Tel. 515 294-7612, Hotline: 515-294-1222, email eooffice@mail.iastate.edu.

This Working Paper is brought to you for free and open access by the Iowa State University Digital Repository. For more information, please visit lib.dr.iastate.edu.

Does the Women, Infants, and Children Program (WIC) Improve Infant Health Outcomes?

Abstract

We evaluate causal impacts of prenatal WIC participation on healthy birth outcomes, simultaneously accounting for self-selection of expectant mothers into WIC and systematic underreporting of program participation. In doing so, we extend existing partial identification methods to reflect the institutional details of the program. In particular, we allow for a richer measurement error model and apply a modified regression discontinuity design. Combining survey data from the Early Childhood Longitudinal Study Birth Cohort (ECLS-B) with administrative data from the USDA, our reduces the prevalence of unhealthy birth weight by at least 21 percent and unhealthy gestation duration by at least 9.9 percent.

Keywords

Special Supplemental Nutrition Program for Women, Infants, and Children, WIC, Treatment Effects, lth Outcomes, Monotone Instrumental Variables, Partial Identification, Nonparametric Bounds, Classification Error

Disciplines

Behavioral Economics | Econometrics | Health Economics | Social Welfare

Does the Women, Infants, and Children Program (WIC) Improve Infant Health Outcomes?*

Brent Kreider
Department of Economics
Iowa State University
bkreider@iastate.edu

John V. Pepper
Department of Economics
University of Virginia
jvp3m@virginia.edu

Manan Roy
Department of Nutrition and Health Care Management
Appalachian State University
roym1@appstate.edu

May 2018

Abstract: We evaluate causal impacts of prenatal WIC participation on healthy birth outcomes, simultaneously accounting for self-selection of expectant mothers into WIC and systematic underreporting of program participation. In doing so, we extend existing partial identification methods to reflect the institutional details of the program. In particular, we allow for a richer measurement error model and apply a modified regression discontinuity design. Combining survey data from the Early Childhood Longitudinal Study Birth Cohort (ECLS-B) with administrative data from the USDA, our preferred estimates imply that WIC reduces the prevalence of unhealthy birth weight by at least 21 percent and unhealthy gestation duration by at least 9.9 percent.

JEL: C14, C21, I12, I31, I38

Keywords: Special Supplemental Nutrition Program for Women, Infants, and Children, WIC, Treatment Effects, Health Outcomes, Monotone Instrumental Variables, Partial Identification, Nonparametric Bounds, Classification Error

*The authors received valuable comments from Craig Gundersen, Helen Jensen, James Ziliak, and participants at the Research Program on Childhood Hunger Progress Report and Southern Economic Association conferences. This project was supported with a grant from the University of Kentucky Center for Poverty Research through funding by the U.S. Department of Agriculture, Food and Nutrition Service, contract number AG-3198-B-10-0028. Pepper also acknowledges financial support from the Bankard Fund for Political Economy. The opinions and conclusions are solely those of the authors and do not necessarily reflect the opinions or policies of the UKCPR or any agency of the federal government.

1. Introduction

The Supplemental Nutrition Program for Women, Infants, and Children (WIC) is an early intervention program that provides benefits to about nine million recipients per year, nearly a million of whom are pregnant women (Johnson et al., 2013). In 2016, 39.6% of women who gave birth in the U.S. received prenatal benefits from WIC (Driscoll and Osterman, 2018). By providing access to nutritious food supplements, education, and preventative health services, the prenatal program aims to improve fetal development and reduce the incidence of low birth weight and short gestation (USDA, 2003; Johnson et al., 2013).¹ In general, a large body of research finds that WIC recipients are much less likely (10%–40%) to have low birth weight babies than non-WIC recipients (see Currie, 2003; Bitler and Currie, 2005; Hoynes et al., 2011; Rossin-Slater, 2013; Currie and Rajani, 2015).

While participating in WIC might improve fetal development, drawing credible inferences is complicated by two fundamental identification problems: unknown counterfactual birth outcomes for infants and underreported WIC participation of mothers. A selection problem arises because the data alone cannot reveal what the birth outcome of an infant whose mother participated in WIC would have been had she not participated, or vice versa. A measurement error problem arises because households are known to systematically underreport the receipt of food assistance in national surveys.

In this paper, we evaluate the effects of prenatal WIC participation on the probability of normal birth weight (between 2500 grams and 4000 grams), full term pregnancy (gestation age between 38 and 42 weeks), and other related birth outcomes, accounting for both the selection and

¹ WIC has more specific nutritional and health related objectives than the Supplemental Nutrition Assistance Program (SNAP). While the WIC food package for pregnant women has a small dollar value (an average of about \$40/month in 2015), it provides specific nutrients that are known to benefit fetal development. Moreover, unlike SNAP, WIC benefits are not limited to food. WIC also provides nutrition education and counseling, preventive health care, and social services – either on-site or through referrals to other agencies (Johnson et al., 2013). For example, at-risk women can receive special health care including enrollment in smoking cessation programs (Bitler and Currie, 2005).

measurement problems. Our analysis complements the recent literature by simultaneously addressing the measurement and selection problems. Although more recent studies have addressed the self-selection problem by applying instrumental variable models (e.g., Bitler and Currie, 2005; Figlio et al., 2009; Hoynes et al., 2011; Currie and Rajani, 2015) and maternal fixed effects models (e.g., Rossin-Slater, 2015; Currie and Rajani, 2015), all previous studies evaluating the impact of a mother's decision to enroll in prenatal WIC on birth outcomes treat indicators of WIC participation as accurate.² Even if the endogenous selection problem is credibly addressed, evaluating the efficacy of WIC is further complicated by a nonclassical measurement error problem in participation status. Households are known to systematically underreport the receipt of food assistance in national surveys, and reporting errors may be related to both birth outcomes and socioeconomic characteristics (e.g., Bollinger and David, 1997; Bitler et al., 2003; Meyer et al., 2009)³ Even small amounts of nonclassical measurement error in a treatment variable can be sufficient to overturn conclusions obtained under an implicit assumption of accurately measured treatments.⁴

The combination of these dual identification problems poses a particular challenge for estimating treatment effects that cannot be overcome by applying standard IV techniques. Such methods are known to lead to inconsistent estimates when an endogenous binary treatment variable is measured with error (e.g., Black et al., 2000; Frazis and Loewenstein, 2003). Even supposing the treatment is perfectly measured, it is difficult in this literature to find instruments that are both valid

² Kreider et. al. (2016) allow for WIC receipt to be misclassified in their analysis of the impact of WIC on food security, but they do not study birth outcomes. They also use a less flexible measurement error model, and the methods do not exploit information on adjunctive eligibility into WIC.

³ The classical measurement error attenuation result does not necessarily hold in this nonclassical environment. Instead, estimated treatment effects may be either underestimated or overestimated depending on application-specific unobserved patterns of classification errors.

⁴ In a regression framework with nonclassical measurement error, Kreider (2010) finds that just a 1.3 percent error rate in reported health insurance status is sufficient to generate a double-digit percentage point range of uncertainty about the variable's true marginal impact on the use of health services. Millimet (2011) reinforces this finding for a broader class of treatment effect models.

and strong.⁵ Beyond the weak instruments problem, variation in policy instruments across states (for example) that affect WIC participation rates may be endogenously related to birth outcomes. These policies are not randomly assigned, and policies targeted towards more WIC participation (e.g., relatively easy recertification) may be correlated with other state policies designed to improve income or health, which may themselves affect an expectant mother's birth outcome.

A number of recent studies have applied creative and credible identification strategies to address the selection problem. For example, Figlio et al. (2009) focus on families in a narrow income range surrounding the WIC income eligibility threshold, comparing birth outcomes of women in Florida who are marginally eligible and marginally ineligible for WIC. Exploiting a policy change involving the tightening of income documentation requirements, they find that WIC participation substantially reduces the likelihood of adverse birth outcomes, despite having little effect on mean outcomes. Hoynes et al. (2011) evaluate the performance of WIC on infant health at the time of its establishment as a pilot program in 1972, exploiting the plausibly exogenous county-level variation in participation due to the staggered rollout of the program in the 1970s. Their analysis provides evidence that WIC initiation led to improved birthweights during that era, especially for mothers with less education. Rossin-Slater (2013) builds on their approach by exploiting a finer variation in WIC clinic access within zip codes and by employing a maternal fixed effects instrumental variable strategy. Her results show that prenatal access to WIC clinics within a mother's zip code of residence leads to a 0.8% increase in average birth weight. While these latter two studies provide important new insights and evidence about the role of local access to WIC in improving birth outcomes, they do not use direct information about a particular mother's WIC participation status.

⁵ As highlighted by Hoynes et al. (2011) and others, food assistance programs like SNAP and WIC exhibit relatively little geographic variation in benefit rules to be leveraged compared with other programs targeted to disadvantaged households, such as TANF and Medicaid.

Using data from the Early Childhood Longitudinal Study-Birth cohort of 2001 (ECLS-B) and auxiliary information from USDA administrative files, we simultaneously address the measurement and endogenous selection identification problems by extending the nonparametric bounding strategy in Kreider, Pepper, Gundersen, and Jolliffe (2012, hereafter KPGJ) to reflect the institutional details of WIC. This framework does not require traditional instrumental variable or classical measurement error assumptions, or a linear response model.⁶ Instead, we rely on a number of monotonicity restrictions that are relatively weak compared with the strict orthogonality assumptions embedded in standard IV or classical measurement error models.

After describing the data in Section 2, we formalize the empirical question and the identification problems in Section 3. Then, in Section 4, we derive bounds on average treatment effects (ATE) for a reference case that the decision to take up WIC is exogenous but allowing for mismeasured WIC participation. Though the exogenous selection assumption is unlikely to hold in our setting, the difference in expected outcomes among WIC recipients and nonrecipients – the mean outcome gap – is a parameter reported in much of the existing literature (see Currie, 2003) and is an important descriptive measure of the association between WIC participation and birth outcomes.

In Section 5, we turn our attention to drawing inferences on the ATE given the endogenous selection problem. We consider the identifying power of several types of monotone instrumental variables (MIV) (Manski and Pepper, 2000) in which we posit monotonic relationships between the latent probabilities of healthy birth outcomes and certain observed covariates, such as household income. We also extend a modified discontinuity design (see Gundersen et al., 2012) to account for WIC eligibility rules that confer adjunctive eligibility through participation in other assistance programs. Finally, we consider the identifying power of a monotone treatment response (MTR) assumption that WIC does not harm expected birth outcomes on average (Manski, 1997).

⁶ The classical linear response model is not compatible with government programs thought to have heterogeneous impacts across households (Moffitt, 2005).

Combining endogenous selection with measurement error naturally weakens what we can conclude about the average treatment effects. Nevertheless, we can still identify beneficial impacts of WIC on healthy birth outcomes under relatively mild assumptions that explicitly account for selection and underreported participation. Section 6 concludes.

2. Data

To study the impact of WIC on birth outcomes, we use data from the first wave of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), a nationally representative cohort of 10,700 children born in 2001.⁷ Assembled by the U.S. Department of Education, the ECLS-B is particularly useful for this study since it focuses on children's early environmental characteristics, including those in the prenatal period, that are thought to play a crucial role in development. The first wave of the survey collects information from the child's mother between Fall 2001 and Fall 2002 when the child is approximately 9 months old.

Our analysis focuses on 4,750 households included in the first wave of the ECLS-B.⁸ This sample reflects two notable restrictions.⁹ First, the sample is restricted to infants who are singletons (e.g., not twins). Second, we focus on households that appear to be eligible for prenatal WIC benefits, either through income eligibility or adjunctive eligibility.¹⁰ To be income eligible, family gross income cannot exceed 185 percent of the U.S. Poverty Income Guidelines (\$32,653 in 2001 for a family of four). Irrespective of income, families are adjunctively (automatically) eligible for WIC if they participate in Medicaid, the Supplemental Nutrition Assistance Program (SNAP), or

⁷ ECLS-B case-level data are available to researchers who request and obtain a restricted-use data license. Information about receiving a restricted-use data license can be found at <http://nces.ed.gov/pubsearch/licenses.asp>. Information about the survey design can be found at <https://nces.ed.gov/ecls/birth.asp>.

⁸ All sample sizes are rounded to the nearest 50 per National Center for Education Statistics (NCES) restricted data regulations for the ECLS-B.

⁹ We also drop about 400 observations that had missing information on age.

¹⁰ Recipients must also be at nutritional risk, as determined by a physician, nutritionist, or nurse. In practice, nearly all income-eligible households satisfy at least one of the nutritional risk criteria (Ver Ploeg and Betson, 2003; Currie, 2003, p. 215).

Temporary Assistance for Needy Families (TANF). The ECLS-B does not ask about prenatal participation in SNAP or TANF (only about postpartum participation), but it does ask about prenatal Medicaid benefits. Our sample therefore consists of households that either met the 185 percent income threshold or reported the receipt of prenatal Medicaid. Nearly all SNAP and TANF households met the income eligibility criterion (Jacknowitz and Tiehen, 2010), and thus are already included in the income-eligible sample.¹¹

Measurement error may lead to some contamination of our eligibility indicator (see Gundersen et al., 2012, and Jacknowitz and Tiehen, 2010). In addition to standard problems in accurately measuring income in self-reported surveys, there may be a mismatch between the income measures reported in the ECLS-B and those used to determine eligibility during the prenatal period. The ECLS-B survey collects information on annual income at the time of the “9-month” survey, thus covering parts of both the prenatal and postnatal periods. True income-eligibility, however, is determined only in the prenatal period and, depending on particular policies of the state WIC agencies, might be based on a monthly rather than annual income measure. This mismatch between the survey reporting periods and the programmatic rules is a common problem in evaluating eligibility for assistance programs using survey data and, in the case of WIC, has been found to understate the number of eligible households (Ver Ploeg and Betson, 2003).

Still, regardless of the potential for errors in classifying eligibility, this sample restriction generates a well-defined subpopulation of interest – namely singleton infants residing in households that report income less than 185 percent of the poverty line and/or the receipt of Medicaid. We have checked the sensitivity of our results to different income thresholds and find that our qualitative

¹¹ According to records assembled by the USDA, about two-thirds of WIC recipients reported participating in at least one public assistance program conferring adjunctive eligibility in 2002 (USDA, 2003, Exhibit 4.3) but only 1.7 percent of prenatal participants reported income of more than 185 percent of the poverty line (USDA, 2003, Exhibit 4.7). Another 16.5 percent recipients did not report income.

conclusions are robust to income thresholds ranging from 150% to 200% of the poverty line and to the exclusion of Medicaid recipients.

Finally, applying a modified regression discontinuity design similar to the one in Gundersen et al. (2012) and Schanzenbach (2009), we also use data on a subsample of infants residing in households with incomes above the income threshold but below the fourth quintile of the socioeconomic status (SES) distribution (N=1,250). We constrain this subsample to include only households that appear to be ineligible for WIC because their incomes are too high and they are not adjunctively eligible through other programs.

2.1. Birth outcomes

Table 1 displays means and standard deviations for the variables used in our analysis for the main sample of infants whose mothers were eligible to receive prenatal WIC (N=4,750).¹² The last two columns report differences in means between those reporting prenatal WIC receipt and those classified as eligible nonparticipants. We refer to this difference as the expected outcome gap based on self-reports of participation. In Section 3, we examine the implications of classification error for drawing inferences on the true mean outcome gap.

We observe a number of different outcome measures of infant health related to birth weight and gestation age. We focus mostly on indicators of normal birth weight (between 2500 grams and 4000 grams) and normal gestation length (between 38 and 42 weeks). Descriptive statistics in Table 1 reveal that infants whose parents report having received prenatal WIC have slightly better birth weight outcomes on average but worse gestation length outcomes than eligible nonparticipants.¹³ For example, the gap in the probability of a normal birth weight is 1.95 percentage points while the gap in the probability of a normal gestation length is -1.00 percentage point. We also evaluate other

¹² All analyses are performed using survey weights available in the ECLS-B.

¹³ Similar qualitative associations are reported in Bitler and Currie (2005).

measures of favorable birth outcomes including birth weight of at least 1500 grams (not very low birth weight), at least 2500 grams (not low birth weight), no more than 4000 grams (not macrosomic), and indicators for near-term pregnancy – gestation age of at least 33 weeks (not very premature) or at least 37 weeks (not premature).

A large literature documents the importance of early health outcomes, even as early as in utero, in influencing future adult outcomes (see Almond, 2006). That birth weight and gestation length affect future outcomes has been widely documented. Breslau et al. (1994), Brooks-Gunn et al. (1996), and Currie and Hyson (1999), for example, link birth weight to average scores on several different tests of intellectual and social development. Goldenberg and Cullhane (2007) find that low birth weight is strongly associated with later adult chronic medical conditions like diabetes, hypertension, and heart disease. Boulet et al. (2004) find that macrosomia is related to fetal injury, perinatal asphyxia, and fetal death, as well as complications for the mother like increasing the probability of caesarean delivery. Using birth weight, gestation age, and Apgar score as metrics for infant health, Oreopoulos et al. (2008) conclude that poor infant health is a predictor of mortality within one year as well as mortality until age 17.¹⁴

Boyle and Boyle (2013) review and summarize the current available literature on infants born at moderate preterm (32-33 weeks) and late preterm (34-36 weeks) gestations and conclude that preterm infants face significantly greater risks of morbidity and mortality than previously believed. Goldenberg et al. (2008) also identify preterm birth as the leading cause of perinatal morbidity and mortality in developed countries. For even longer term outcomes, Crump et al. (2011) find that shorter gestation is most significantly associated with increased mortality in early childhood and mortality related to congenital anomalies and respiratory, endocrine, and cardiovascular disorders in young adults. Morse et al. (2009) report a positive association between preterm birth and risks of

¹⁴ Other studies exploiting sibling comparisons include Conley and Bennett (2000), Johnson and Schoeni (2007), Lawlor et al. (2006), Black et al. (2007), Royer (2009), and Currie and Moretti (2007).

developmental delay and school-related problems such as risk of suspension and disability in prekindergarten at age three and four, among others.

2.2. WIC participation indicator and misclassification

For each household, we observe a self-reported indicator of prenatal participation in WIC. This binary treatment variable takes a value of one if the mother reports receiving WIC benefits during pregnancy, and zero otherwise. In the survey, 69.2 percent of the classified eligible households report prenatal WIC receipt. This self-reported participation rate is similar to those found in other surveys (e.g., the CPS and SIPP) but lower than analogous rates found using administrative data (Bitler et al., 2003; USDA, 2009). Bitler et al. (2003) and Meyer et al. (2009) find evidence of substantial underreporting of WIC participation in the CPS, SIPP, and PSID.

Direct evidence on the extent of classification errors in the ECLS-B can be found using administrative data from the USDA. In particular, the USDA (2009) estimates that there were about 1.2 million pregnant women who were eligible for prenatal WIC benefits in 2000, while the administrative records revealed about 0.89 million participants in the prenatal program (USDA, 2003). These values imply an estimated true participation rate of about 0.74, which is higher than the self-reported rate of 0.692.¹⁵ Hausman et al.'s (1998) parametric model of asymmetric misclassification provides an alternative approach for estimating the extent of misreporting.

¹⁵ The USDA (2009) reports estimates of the participation rate (which they label the coverage rate) to be 0.67 in 2000 and 0.64 in 2001. However, these estimated participation rates do not directly apply in our analysis due to mismatches between the survey and administrative data. Most notably, the USDA rates are based on the average number of monthly participants (0.841 in 2000) whereas the ECLS-B measures any participation during pregnancy. Thus, rather than using monthly averages, we use data on the total number of participants. Even with this adjustment, however, there are number of reasons why the participation rates estimated using the USDA reports may be inconsistent with the measures derived using data from the ECLS-B. First, the numerator may be biased due to mismatches in the timing of the survey and administrative data (mothers whose children were born in 2001 versus calendar year). Second, the estimated number of eligible households may be biased because of timing differences (see above), of errors in measuring adjunctive and income eligibility, of differences in the surveys (the USDA uses the Current Population Survey), and the USDA uses a monthly rather than annual income measure. We do not directly address these latter measurement issues when using administrative data to estimate the true participation rate. Instead, we trace out the sensitivity of inferences to the true participation rate.

Estimates using this approach imply a true participation rate of 0.84, with fewer than 1 percent of the eligible population falsely reporting WIC participation (see Section 4.1, footnote 18 for additional details). Thus, consistent with the related SNAP literature, estimates from this model suggest large degrees of underreporting of WIC participation in the ECLS-B but negligible rates of false positive reporting.

Some have raised concerns that using a simple indicator of receipt may lead to a mechanical upward bias in the traditional estimates of the effect of WIC on birth outcomes that do not account for the timing of take-up (i.e., a gestational age bias) (see Ludwig and Miller, 2005; Joyce et al., 2005; and Rossin-Slater, 2013). Women whose pregnancies last longer have a longer period to use prenatal WIC benefits. While the ECLS-B does not provide information on prenatal WIC benefits, it does ask respondents to report on the take-up trimester. With these data, the methods developed in this paper could be extended to assess whether the timing of take-up has a distinct impact (see Pepper, 2000). Nevertheless, we do not explicitly model the duration of prenatal WIC benefits. Our partial identification models allow the timing of take-up to be heterogeneous and, as such, do not result in biased estimates of the bounds on the ATEs of WIC on birth outcomes. Moreover, with less than 10% of prenatal recipients taking up WIC in the last trimester (USDA, 2003, Exhibit 3.2), adding multidimensional treatments would result in more ambiguity arising from the selection problem, and classification error problems would be compounded and more difficult to credibly address.

3. Identifying the causal effect of WIC on infant health

Our interest is in learning about average treatment effects (ATE) of prenatal maternal WIC participation on infant health among eligible households.¹⁶ For binary outcomes, these average

¹⁶ Section 3 closely follows KPGJ's foundation for discussing partial identification of average treatment effects. Subsequent sections extend these methods to introduce partial verification of treatment status, a potentially

treatment effects can be expressed as

$$\begin{aligned} ATE(1,0 | X \in \Omega) &= E[H(1) | X \in \Omega] - E[H(0) | X \in \Omega] \\ &= P[H(1) = 1 | X \in \Omega] - P[H(0) = 1 | X \in \Omega] \end{aligned} \quad (1)$$

where H is the realized health outcome, $H(1)$ denotes the infant's health if he or she were to receive WIC, $H(0)$ denotes the analogous outcome if the infant were not to receive WIC, and $X \in \Omega$ denotes conditioning on observed covariates whose values lie in the set Ω . Thus, the average treatment effect reveals the mean health effect of prenatal WIC participation (compared with nonparticipation) for a WIC eligible household chosen randomly from the underlying population. In what follows, we will simplify notation by suppressing the conditioning on subpopulations of interest captured in X . For this analysis, we focus on infants who appear to be eligible for prenatal WIC based on sufficiently low household income or through the receipt of Medicaid.

Two identification problems arise when assessing the impact of WIC on infant health. First, even if WIC participation were observed for all eligible households, the potential outcome $H(1)$ is counterfactual for all infants who did not receive WIC, while $H(0)$ is counterfactual for all infants who did receive WIC. This is referred to as the selection problem. Second, true participation status may not be observed for all respondents. This is referred to as the measurement or classification error problem. Instead of observing W^* , we observe a self-reported indicator, W , where $W = 1$ if an infant resides in a household that reported receiving prenatal WIC and 0 otherwise. Without assumptions restricting the nature or degree of classification errors, the sampling process does not reveal useful information on WIC receipt, W^* .

To address these two identification problems, we proceed in two steps. First, we focus on the implications of the measurement error problem alone when we have information on the true

mismeasured monotone instrument, a formal analysis of the exogenous selection assumption, and parallel sets of bounds for the case that the researcher does not have auxiliary information about the true participation rate.

participation rate, P^* , and the self-reported rate, P . Second, we assess what can be inferred when accounting for both the selection and measurement problems simultaneously.

4. Bounds under exogenous selection with classification errors

Until recently, the literature examining the impact of WIC on health assumed that selection is (conditionally) exogenous so that $P[H(j) | W^*] = P[H(j)]$, $j = 1, 0$. Under this assumption, the average treatment effect can be written as

$$ATE(1,0 | X \in \Omega) = P[H(1) = 1 | W^* = 1] - P[H(0) = 1 | W^* = 0],$$

which in turn can be expressed as the difference in conditional means:

$$\beta \equiv P(H = 1 | W^* = 1) - P(H = 1 | W^* = 0). \quad (2)$$

The appeal of the exogenous selection assumption is obvious: if selection is exogenous and WIC receipt W^* is observed, then the average treatment effect is identified by the sampling process.

Though the exogenous selection assumption is unlikely to hold in our setting, the difference in expected outcomes among WIC recipients and nonrecipients – the mean outcome gap, β – is an important descriptive measure of the association between WIC participation and birth outcomes. The sample means displayed in Table 1, for example, suggest that WIC is associated with a slightly higher probability of a normal birth weight but lower probability of a normal gestation age.

4.1 Classification error model

If one allows for the possibility of classification errors in W^* , the mean outcome gap in Equation (1) is not identified. To address the classification error problem, we combine auxiliary data on the size of the caseload from the USDA with survey data from the ECLS-B to estimate the true and self-reported participation rates, $P^* = P(W^* = 1)$ and $P = P(W = 1)$. Following KPGJ, the auxiliary information identifies the difference in false negative and positive reporting rates:

$$\Delta = P^* - P.$$

As noted above, the USDA data reveal that $P^* = 0.74$ and the self-reported rate reported in the ECLS-B is $P = 0.692$. Thus, in our application, Δ is estimated to equal $0.048 (= 0.74 - 0.692)$. The fraction of false negative reports must exceed the fraction of false positive reports by this quantity. Given the potential mismatch between the survey and administrative reports on eligibility and participation (see above), we will also explicitly allow for the possibility that the estimated true participation rate of 0.74 may be in error by tracing out the sensitivity of inferences to variation in this rate.

We apply two additional restrictions on the classification error problem:

$$\mathbf{ME1:} \text{ Maximum error rate: } P(W^* \neq W) \leq Q_u \quad (3)$$

$$\mathbf{ME2:} \text{ Verification: } V = 1 \Rightarrow W^* = W \quad (4)$$

where Q_u is a known upper bound on the degree of data corruption (see Horowitz and Manski, 1995) and V is an observed indicator of whether a self-report of WIC participation is known to be accurate.

For Assumption ME1 considered in KPGJ, the upper bound error rate Q_u must logically lie within the range $[|\Delta|, 1]$. In the polar case that Q_u equals 1 , ME1 is uninformative. We refer to this case as the “arbitrary errors model.” In the other polar case that Q_u equals $|\Delta|$, the researcher is imposing a “no excess errors” assumption. In the case of net underreporting ($\Delta > 0$), for example, this assumption is equivalent to imposing a “no false positives” assumption that respondents do not falsely claim to participate in WIC. This no excess errors restriction serves as a useful benchmark for the receipt of WIC. Validation data suggest very few instances of households falsely claiming to receive food assistance (e.g., Bollinger and David 1997).¹⁷ Moreover, using a parametric measurement error model formalized by Hausman et al. (1998), we estimate that less than one

¹⁷ An exception is Kirlin and Wiseman (2014) who, examining WIC cases in Texas, find evidence of net underreporting but substantial over-reporting as well. In this case, the arbitrary errors assumption is appropriate.

percent of the income-eligible WIC population incorrectly reports receiving benefits.¹⁸ Middle-ground positions are obtained by setting Q_u between $|P^* - P|$ and 1.

Assumption ME2 extends KPGJ by allowing a researcher to formally verify the accuracy of some mixture of positive and negative WIC responses. Verified responses are denoted $V = 1$. When $V = 0$, a report may be either accurate or inaccurate. In the no false positives model, for example, respondents reporting the receipt of WIC are validated to provide accurate reports. Such respondents have revealed their willingness to report the receipt of WIC benefits despite any potential stigma. We also consider stronger verification that treats a response as accurate if the household is willing to report the receipt of benefits from any related program asked about in the ECLS-B (WIC, SNAP, TANF, or Medicaid). Such households have revealed their general willingness to report benefits, even if they did not specifically report WIC benefits. Under this verification assumption, 84 percent of the WIC responses are treated as accurate. In the traditional literature, all WIC responses are implicitly treated as accurate.

In Appendix A1, we derive closed-form analytical bounds on the outcome gap, β , as Proposition 1. The bounds take a particularly simple form for the arbitrary errors case ($Q_u = 1$), which we present as Corollary 1:

Corollary 1 *The outcome gap, β , under arbitrary errors is bounded sharply as follows:*

$$-\min\left\{\frac{P(H=1)}{1-P^*}, \frac{1-P(H=1)}{P^*}\right\} \leq \beta \leq \min\left\{\frac{P(H=1)}{P^*}, \frac{1-P(H=1)}{1-P^*}\right\}.$$

To estimate these bounds, all we require is an estimate of $P(H = 1)$, the fraction of healthy birth outcomes, and an estimate of P^* , the true WIC participation rate. The fraction of healthy birth

¹⁸ Using Hausman et al.'s (1998) model, the true participation rate is specified as $P^* = F(X'\gamma)$, where $F(\cdot)$ is the standard normal CDF and is a vector of covariates. Given data on the self-reported rate, $P = P(W = 1)$, the model identifies the conditional false negative reporting rate, $P(W = 0 | W^* = 1)$, and the conditional false positive reporting rate, $P(W = 1 | W^* = 0)$. We estimate these rates to be 0.19 and 0.05, respectively. Based on these estimates, we are able to back out an estimate of P^* and the unconditional misreporting rates using $P^* = [P - P(W = 1 | W^* = 0)] / [1 - P(W = 1 | W^* = 0) - P(W = 0 | W^* = 1)]$. Our estimate of the false negative rate is 0.159, and our estimate of false positive rate is 0.008. Full estimation results are available from the authors.

outcomes is consistently estimated using data from the ECLS-B. We estimate $P^* = 0.74$ using administrative data on the size of the WIC caseload as described above, but we also assess the sensitivity of the bounds to variation in P^* . Appendix A1 also provides parallel sets of bounds for the case that the researcher has no auxiliary information about the value of P^* .

4.2. Results under exogenous selection

Figure 1 traces out the estimated Proposition 1 bounds on the outcome gap for normal birth weight as the true participation rate P^* varies between 0 and 1 under (a) arbitrary classification errors (no verification), (b) no false positive reports of WIC receipt, and (c) verified WIC status responses for households that report benefits from any government program. The accompanying table beneath the figure displays the bounds at the self-reported participation rate, $P^* = P = 0.692$, and the true participation rate, $P^* = P^o = 0.74$, along with Imbens-Manski (2004) confidence intervals. Strictly positive estimated average treatment effects are highlighted in bold.

If all WIC responses are known to be accurate, the mean outcome gap is point-identified at $P^* = P$ as $\beta = 0.854 - 0.835 = 0.0195$, consistent with the descriptive statistic in the second row of Table 1. Otherwise, the difference in mean outcomes can only be partially identified. If nothing is known about the accuracy of individual responses, the figure reveals that little can be inferred about the true health outcome gap even if the researcher has knowledge that WIC respondents do not systematically over- or underreport benefits (no net misreporting). Specifically, the true value of β at $P^* = P$ under arbitrary errors could lie anywhere within $[-0.219, 0.494]$, a 71 percentage point range of uncertainty even prior to accounting for sampling variability. These arbitrary errors bounds expand to $[-0.205, 0.584]$ at the estimated true participation rate $P^* = 0.74$.

Partial verification narrows these bounds to $[-0.205, 0.0592]$ under no false positives (a two-thirds reduction in the width of the bounds) and further to $[-0.0759, 0.0592]$ when responses are treated as accurate if the household reported any government benefits (an 82% reduction in the width). Graphically, the small trapezoid-like region in Figure 1 (dotted region) highlights the

methodological contribution of verification assumption ME2 in Equation (4). A practitioner’s ability to formally verify some mixture of positive and negative participation responses can dramatically reduce the identification region.

Still, without further assumptions, identification of the birth outcome gap deteriorates sufficiently rapidly that small degrees of WIC classification error preclude us from identifying the sign of this gap. The conclusion that normal birth outcomes are more (or less) prevalent among WIC recipients than among eligible nonrecipients requires a large degree of confidence in self-reported WIC participation status, an assumption not supported by validation studies. Results are similar for the other birth outcomes. For example, analogous bounds for normal gestation duration (not shown) at $P^* = 0.74$ range from $[-0.335, 0.953]$ under arbitrary errors to $[-0.202, 0.0518]$ under no false positives to $[-0.112, 0.0518]$ under stronger verification.¹⁹ Thus, with even small amounts of classification error, the results from the earlier empirical literature evaluating the impact of WIC on birth weight appear tenuous.

5. Bounds with selection and classification error problems

We now study what can be learned about the ATE when selection is endogenous. With endogenous selection and reporting errors, there is uncertainty about counterfactual birth outcomes and about the reliability of the data on WIC participation, W^* . Using the classification error model presented in Section 4, we derive partial verification identification bounds on the ATE in Equation (1) in Appendix A2. These bounds are formalized as Proposition 2.

In Section 5.1, we present bounds without imposing any restrictions to address the selection problem. Then, to narrow these worst-case selection bounds, we apply a number of middle ground assumptions restricting the relationship between WIC participation, birth outcomes, and observed

¹⁹ Estimation results for all outcomes (normal birth weight, not low birth weight, not very low birth weight, not macrosomic weight, normal gestation, not premature, not very premature) will be provided in Section 5.4 when we additionally allow for endogenous selection.

covariates. In Section 5.2, we consider the identifying power of several monotone instrumental variable assumptions and in Section 5.3 we examine the monotone treatment response assumption that, on average, WIC cannot lead to worse birth outcomes. In these subsections, we present results for the normal birth weight outcome under the arbitrary error and no-false positive error models. In this case, our assumptions addressing the selection problem substantially narrow the bounds but do not necessarily reveal the sign of the ATE unless one is willing to impose the MTR assumption.

Finally, in Section 5.4, we estimate bounds under for a number of different birth outcomes under two stronger verification assumptions: (1) the ME2 partial verification assumption that self-reports of WIC participation from households reporting any type of government benefits are assumed to be accurate, and (2) the assumption of fully accurate reporting. These verification models, where between 84% and 100% of respondents are assumed to provide accurate reports of WIC receipt, have substantial identifying power in this application. In particular, we find that when combined with the other assumptions used to address the selection problem, WIC is estimated to have a beneficial effect on numerous birth outcomes.

5.1. *Worst-Case Selection Bounds*

Figure 2 displays the estimated bounds for the effects of WIC on the normal birth weight probability without imposing assumptions to address the selection problem. This extends Figure 1 to the case that WIC participation is endogenous. As before, the corresponding table presents estimates if the true WIC participation rate P^* is equal to the self-reported rate of 0.692 or the USDA administrative rate of 0.74. Given uncertainty about the true participation rate, the figures trace out bounds on the ATE as P^* varies.

Figure 2 highlights Manski's (1995) worst case average treatment effect bounds on the normal birth weight probability, $[-0.358, 642]$, with width 1 for the case of no measurement error at $P^* = 0.692$ given no false positives (or stronger verification). These bounds expand to $[-0.459, 0.844]$ with width 1.3 at $P^* = 0.692$ under arbitrary errors (no net errors in this case). The

bounds shift to $[-0.412, 0.892]$ at $P^* = 0.74$ under arbitrary errors, then narrow to $[-0.405, 0.690]$ or $[-0.362, 0.690]$ under no false positives or stronger verification, respectively. Naturally, these bounds are quite wide in the absence of restrictions on the selection process into WIC participation.

5.2 *Monotone instrumental variables*

5.2.1 *Model Assumptions*

The Monotone Instrumental Variable (MIV) assumption (Manski and Pepper, 2000) formalizes the notion that the latent probability of a positive health outcome, $P[H(j) = 1]$, varies monotonically with certain observed covariates. To formalize this idea, let v be the monotone instrumental variable such that

$$u_1 \leq u \leq u_2 \Rightarrow P[H(j) = 1 | v = u_1] \leq P[H(j) = 1 | v = u] \leq P[H(j) = 1 | v = u_2]. \quad (5)$$

That is, the latent positive birth outcome probabilities weakly increase with the instrument v .

We apply three MIV assumptions. First, similar to KPGJ, we assume that the latent probability of a favorable birth outcome increases with the ratio of household income to the poverty line accounting for family composition.²⁰ A positive association between health and income has been clearly established in the literature. Chen et al. (2002), for example, reports that child health improves monotonically with socioeconomic status, and Deaton (2002) provides evidence of a negative income gradient in realized health outcomes. Importantly, the MIV assumption in Equation (5) does not require the relationship between income and expected potential birth outcomes to be linear. Indeed, there are known nonlinearities in the gradient between income and health (Deaton, 2002), and the tax and transfer system introduces nonlinearity between income and net resources that are especially salient for lower income households. If the relationship between income and net

²⁰ To estimate these MIV bounds, we first divide the sample into nine income categories provided in the ECLS-B. We assume that the ratio of actual to potential net underreporting does not vary across MIV groups. To find the MIV bounds on the rates of favorable health outcomes, one takes the appropriate weighted average of the plug-in estimators of lower and upper bounds across the groups. As discussed in Manski and Pepper (2000), this MIV estimator is consistent but biased in finite samples. We employ Kreider and Pepper's (2007) modified MIV estimator that accounts for the finite sample bias using a nonparametric bootstrap correction method.

resources is monotonic, this MIV assumption seems credible. However, programs like Medicaid, which introduce a discontinuous notch in the budget constraint, could lead the relationship between income and resources to be negative over part of the range, and thus potentially violate the MIV assumption. Our preferred set of bounds on the ATE imposes this monotonicity assumption, but we also provide alternative estimates that do not.

Second, we use eligibility criteria of WIC to construct a monotone instrument (see Gundersen et al. 2012). Similar to a regression discontinuity design, we compare birth outcomes among WIC-eligible households to those of higher-income ineligible households.²¹ While latent birth outcomes are unlikely to be mean independent of eligibility status (as assumed in a traditional discontinuity design), an MIV-ineligibles assumption holding that mean response varies monotonically across these groups seems credible. In particular, we assume that the latent probability of a healthy birth outcome among WIC-eligible households is no better than among higher-income ineligible households:

$$P[H(j) = 1] \leq P[H(j) = 1 | v = \textit{ineligible}], j = 1, 0. \quad (6)$$

This ineligibility bound on the mean potential health outcome, $P[H(j) = 1]$, has the unique property that ineligible households cannot participate in WIC. Thus, $W^* = 0$ among ineligible households. Assuming eligibility is observed, the data point-identify the mean outcome if all ineligible households were to have not participated in WIC, $P[H(0) = 1 | v = \textit{ineligible}]$, as the mean health status in the group of ineligibles, $P(H = 1 | v = \textit{ineligible})$. This quantity then serves as an upper bound on the potential outcome $P[H(0) = 1]$ for the ATE in Equation (1) within our primary population of income-eligible households (Gundersen et al., 2012):

$$P[H(0) = 1] \leq P(H = 1 | v = \textit{ineligible}). \quad (7)$$

²¹ In particular, we focus on households with socioeconomic status in the third or lower income quintile that were ineligible for WIC because their income exceeded 185 percent of the Poverty Income Guidelines ($inc > 185\%$) and they were not adjunctively eligible via participation in other programs. As noted in Section 2, we do not directly observe adjunctive eligibility status. Instead, to be conservative in our approach, we exclude households in this higher income comparison group that ever reported benefits from WIC, Medicaid, TANF or SNAP.

We learn nothing new about the mean outcome that would be realized if all households participated in WIC, $P[H(1) = 1]$, using this approach because the outcome $H(1)$ is counterfactual within the ineligible population. In Appendix A3, we generalize Equation (7) to account for the possibility that adjunctive WIC eligibility through other programs may be measured with error.

Finally, we consider the *monotone treatment selection (MTS)* assumption that the potential birth outcomes weakly decrease with the true WIC participation indicator such that (Manski and Pepper, 2000):

$$P[H(j) = 1 | W^* = 1] \leq P[H(j) = 1 | W^* = 0] \quad \text{for } j = 1, 0. \quad (8)$$

That is, on average, eligible households that chose to receive WIC, $W^* = 1$, have no better latent birth outcomes than eligible households that did not take up WIC, $W^* = 0$.²²

This assumption formalizes the consensus in the literature that there is negative selection into WIC; unobserved factors associated with poor birth outcomes are positively associated with WIC participation. In fact, it is well established that WIC recipients have unfavorable demographic, socioeconomic, and health characteristics (e.g., Currie, 2003; Bitler and Currie, 2005; Ludwig and Miller, 2005; Gundersen, 2005; Currie and Rajani, 2015). For example, in the ECLS-B sample used in this paper, respondents reporting to receive WIC reside in less educated households with fewer resources than those reporting not to receive WIC. Likewise, Bitler and Currie (2005) find that those who chose to participate in WIC are negatively selected regarding education, age, marital status, father involvement with the birth, smoking behavior, obesity, use of public assistance last year, having wage income last year, having a bathroom in the household, and having had a previous low birth weight or premature infant (if not a first birth). These findings hold across and within racial groups. Thus, for selection on unobserved factors to be positive on average, WIC mothers would

²² Holding only on average, the MTS assumption in (8) does not rule out the possibility that selection is positive for some women. Positive selection could result, for example, if relatively motivated and informed women are more likely to enroll and have better birth outcomes. Likewise, positive spurious associations might arise from the fact that women with longer pregnancies have more opportunity to enroll in WIC – sometimes referred to as gestational age bias (see Section 2.2 for further details) (Joyce, Gibson and Colman, 2005; Joyce, Racine, and Yunzal-Butler, 2008).

need to be negatively selected across this wide range of observables yet positively selected on unobservables. Based on the totality of this evidence, Bitler and Currie (2005) find it implausible that potentially positive selection on unobservables (e.g., nutrition knowledge or desire to be a good mother) could outweigh known strong negative selection across a wide spectrum of observed characteristics: “While it remains theoretically possible that WIC mothers might be positively selected, all the evidence points in the opposite direction” (p. 88).

5.2.2 *Results for normal birth weight with arbitrary errors and no-false positive errors*

Figure 3, along with the corresponding table, displays estimated lower bounds (only lower bounds are shown in the figure to save space) on the impact of WIC for the normal birth weight outcome under various MIV restrictions. Here, we consider identification under arbitrary errors or no false positives. We focus on estimates of the ATE at $P^* = P^o = 0.74$, the administrative WIC participation rate.

Recall that the worst-case endogenous selection bounds at $P^* = 0.74$ in Figure 2 revealed that the ATE could lie anywhere in the range $[-0.412, 0.892]$ under arbitrary errors, which contracted modestly to $[-0.405, 0.690]$ under no false positives. In Figure 3 and corresponding table, these ranges narrow substantially to $[-0.191, 0.886]$ and $[-0.129, 0.678]$, respectively, under the MTS and income-MIV monotonicity assumptions. Still, we cannot rule out the possibility that WIC participation has either a strong negative or positive effect on birth weight.

The MIV-ineligibles assumption imposes the restriction that, on average, infants born into higher-income ineligible households have no worse latent birth outcomes than infants born into WIC eligible households. Adding this assumption to the MTS + income MIV assumptions dramatically improves the lower bound at $P^* = 0.74$ to -0.0221 under arbitrary errors and to -0.0179 under no false positive classifications. At this preferred value of P^* , notice that it makes little difference whether WIC classification errors follow an arbitrary process or are confined to false negatives: the no false positives assumption has little additional identifying power.

This is not the case, however, if P^* is closer to the self-reported rate of $P = 0.692$. Returning to Figure 3, the ATE can be identified as strictly positive under this joint MTS + income MIV + ineligible MIV model if P^* lies between 0.692 and 0.72 – but only under the no false positives assumption (dashed line), not under arbitrary errors (solid line). Under no false positives at $P^* = P = 0.692$ (no misreporting at all), we can identify that WIC leads to at least a 6.64 percentage point improvement in the probability of having a normal birth weight, or by at least 33%.²³ As shown in the table beneath Figure 3, the confidence interval for the ATE lies strictly above 0. Identification then deteriorates as P^* rises. When P^* is about 0.72, we can no longer identify whether the ATE is positive or negative. And with arbitrary errors, we cannot identify the sign of the ATE for any value of P^* without additional assumptions.

5.3. *Monotone Treatment Response*

There is a general consensus among policymakers and researchers that prenatal WIC participation should not lead to worse birth outcomes (Currie, 2003; Ludwig and Miller, 2005). Given this consensus, we consider the identifying power of a Monotone Treatment Response (MTR) assumption (see Manski, 1995 and 1997; Pepper 2000) that formalizes the idea that WIC participation would not lead to a reduction in favorable birth outcomes, on average, conditional on true participation status:

$$P[H(1) = 1 | W^*] \geq P[H(0) = 1 | W^*]. \quad (9)$$

This assumption seems relatively innocuous in our application since WIC restricts purchased food products to be nutritionally sound while providing education and preventative health services. Notice that the assumption, which restricts the average potential outcome, allows for the possibility that WIC could be harmful for some household.

²³ The lower bound percentage reduction in unfavorable birth outcomes under the program is $-\{P[H(1) = 1] - P[H(0) = 1]\} / P[H(0) = 1]$.

Since MTR precludes harmful effects of WIC on average, it would be circular reasoning to use this assumption in support of evidence that the average treatment effect is nonnegative: indeed, $ATE \geq 0$ by construction. Nevertheless, MTR does not rule out a value of 0 (even when combined with the MTS and MIV assumptions), especially after accounting for sampling variability. We impose this assumption because it can be informative in further bounding the magnitude of the ATE when combined with other identifying assumptions; if MTR on average is plausible, then it makes sense to impose the assumption.

The results displayed in Figure 3 show that the MTR assumption (that WIC is not harmful on average) has strong identifying power in this application. Combined with the preceding monotonicity assumptions, we can identify that WIC improves the probability of a healthy birth weight outcome at $P^* = 0.74$ by at least 3.86 percentage points, or 21 percent, under the arbitrary error model. The average treatment effect is strictly positive, and statistically significant, even for large degrees of arbitrary WIC misreporting. We find similar results (not shown) when we replace the normal birth weight outcome with normal gestation duration. Specifically, we can identify that WIC decreases the probability of an unfavorable gestation duration by at least 2.69 percentage points, or 9.9 percent. As above, this estimate is statistically significant and robust to large values of P^* .

5.4. *Verifying those reporting government benefits*

Finally, we examine the identifying power of the stronger classification error assumption that allows for the verification of a mixture of positive and negative WIC participation responses. As a reference point, findings for the special case of full verification (no misreporting) are summarized in Tables 2A and 2B. In that case, the MTS-MIV model (row iv) identifies the ATE as positive for all seven outcomes, though the confidence interval includes zero for several outcomes (not low birth weight, not premature and not very premature).

Our preferred findings allow for WIC classification errors consistent with $P^* = 0.74$. We treat responses about WIC receipt as reliable if the household reported any type of government benefits, remaining agnostic about the reliability of the responses for households that claim no benefits from any source. Recall that 84% of households are verified to provide accurate reports in this partial verification model.

Tables 3A and 3B summarize our findings at $P^* = 0.74$ for all six birth outcomes across the spectrum of monotonicity assumptions under the stronger verification assumption. Strictly positive estimated average treatment effects are highlighted in bold. Given this model, the key finding is that WIC is found to have a strictly positive impact on birth weight outcomes under the MTS and MIV assumptions alone, without imposing MTR. Row (iv) reveals that the ATE of prenatal WIC participation can be identified as strictly positive for the “not very low birth weight” and “not macrosomic birth weight” outcomes when imposing only the MTS and income MIV monotonicity assumptions. Row (v) reveals that we can further identify a strictly positive impact of the program for the normal birth weight outcome after additionally imposing the MIV-ineligibles assumption. Specifically, we estimate that WIC improves the probability of a normal birth weight by at least 1.76 percentage points, or 10.4%. This estimate is statistically significantly different from zero.²⁴

Row (vi) reveals that we can identify a strictly positive impact for all birth outcomes except “not low birth weight” when also imposing the MTR assumption. For example, the probability of a normal birth weight is estimated to increase by at least 3.92 percentage points and the probability of a normal gestation age is estimated to increase by at least 2.69 percentage points.

Importantly, this finding highlights the methodological contribution of the ME2 verification assumption and the Proposition 2 bounds. Our methods allow us to sign the ATE without imposing

²⁴We continue to estimate a strictly positive average treatment effect even when allowing for errors in measuring adjunctive eligibility (see Section 5.2 and Appendix A3). In particular, the estimated bounds are positive as long as less than 16 percent of our labeled “ineligible” households with an unfavorable birth weight outcome were in fact adjunctively eligible.

the MTR assumption if and only if we can treat WIC participation responses as accurate for some households that claim not to have received WIC benefits. In contrast, the methods in KPGJ do not have this flexibility to verify mixtures of positive and negative participation responses.

6. Conclusion

Driven in part by the growing perception that early life conditions have long term impacts on adult life outcomes (e.g., Heckman and Carneiro, 2003; Almond and Currie, 2011) and in part by concerns over discretionary government spending, there has been a renewed interest in understanding the impact of WIC on infant health. Researchers have struggled to draw credible inferences of causal impacts in the presence of unobserved counterfactual outcomes and underreported WIC participation. Even small amounts of nonclassical measurement error in a treatment variable can be sufficient to overturn conclusions that rely on the treatment being accurately measured. Moreover, prior studies tend to find that the instruments (typically variation in state program regulations) are not strongly associated with WIC receipt. It is difficult to find strong instruments for food assistance programs like SNAP and WIC since these programs exhibit relatively little geographic variation in benefit rules compared with programs like Medicaid and TANF.²⁵

In this paper, we complement the existing literature by considering what can be learned about the efficacy of prenatal WIC in improving birth outcomes when program participation may be endogenous and underreported. While our framework does not allow us to point-identify average treatment effects, we derive sharp bounds under arbitrary misclassification of the treatment indicator combined with relatively weak monotonicity assumptions on the selection process. For example, standard IVs are replaced with monotone IVs. Throughout, we assess the sensitivity of inferences to variations in assumptions about the selection and measurement error problems. These monotonicity assumptions are not innocuous, but by considering different assumptions in a layered transparent fashion we illustrate clear tradeoffs between the strength of the assumptions and the findings.

²⁵ Bitler and Currie (2005, p. 75) conclude that variation in WIC program regulations are likely to make poor instruments for WIC participation, rendering estimates based on them potentially unreliable.

Overall, these estimated partial identification bounds highlight how inferences are sensitive to assumptions. Under the weakest models allowing for both the selection and classification error problems, the bounds do not identify whether WIC increases or decreases the probability of healthy birth outcomes. However, under stronger but credible models, we find that WIC improves infant health across a spectrum of outcomes. When combining all of our nonparametric identifying assumptions in a model with partial verification of WIC classification (84% verified), we find that WIC has a substantial impact on birth outcomes, increasing the likelihood of normal birth weight by at least 3.92 percentage points, or 21.3 percent, and normal gestation age by at least 2.69 percentage points, or 9.9 percent. Under full verification where all responses are known to be accurate, these lower bounds increase to 6.54 and 9.51 percentage points, respectively. Using these alternative partial identification methods that allow for underreported WIC participation, our findings lend further support to recent studies that have concluded that WIC does work in achieving its main objectives.

Finally, although the contributions of this paper are largely empirical in nature, we also derive new closed-form bounds on average treatment effects that extend the nonparametric methods developed in KPGJ. Specifically, we extend their approach to provide sharp bounds on average treatment effects under a “partial verification” assumption that allows a researcher to explicitly be confident in the reliability of the treatment variable (e.g., WIC participation) for some respondents but not others. The methods in KPGJ do not afford this flexibility to verify mixtures of positive and negative participation responses. In our application, we combine the verification assumption with auxiliary data from the USDA that places restrictions on the net WIC misreporting rate. We derive sharp bounds on the average treatment effect for both the reference case of exogenous selection and endogenous nonrandom selection. We also derive parallel sets of bounds for the case that the researcher has no auxiliary information about the true participation rate. These new methods will hopefully prove especially useful to researchers who have available some form of validation data, even when the validation data do not comprise a random sample of the primary data.

References

- Almond, D. (2006). "Is the 1918 Influenza Pandemic Over? Long-term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population." *Journal of Political Economy*, 114, 672-712.
- Almond, D. and J. Currie. (2011). "Human Capital Development Before Age Five." *Handbook of Labor Economics*, Volume 4b: Orley Ashenfelter and David Card, editors, Elsevier, Ch. 15, 1315-1486.
- Bitler, M., J. Currie, and J. Scholz. (2003). "WIC Eligibility and Participation." *Journal of Human Resources*, 38(5), 1139-1179.
- Bitler, M. and J. Currie. (2005). "Does WIC Work? The Effects of WIC on Pregnancy and Birth Outcomes." *Journal of Policy Analysis and Management*, 24, 73-91.
- Black, D., Berger, M., and Scott, F. 2000. Bounding Parameter Estimates with Non-Classical Measurement Error, *Journal of the American Statistical Association*, 95 (451), 739-748.
- Black, S.E., P.J. Devereux, and K.G. Salvanes. (2007). "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes." *The Quarterly Journal of Economics*, 122, 409-439.
- Bollinger, C. and M.H. David. (1997). "Modeling Food Stamp Program Participation in the Presence of Reporting Errors." *Journal of the American Statistical Association*, 92(439), 827-835.
- Boulet, S.L., H.M. Salihu, and G.R. Alexander. (2004). "Mode of Delivery and Birth Outcomes of Macrosomic Infants." *Journal of Obstetrics and Gynecology*, 24, 622-629.
- Boyle, J.D. and E.M. Boyle. (2013). "Born Just a Few Weeks Early: Does it Matter?" *Archives of Disease in Childhood. Fetal and Neonatal Edition*, 98, F85-F88.
- Breslau, N., G. Brown, J. Deldotto, S. Kumar, S. Ezhuthachan, P. Andreski, K. Hufnagle, and E. Peterson. (1994). "A Gradient Relationship Between Low Birth Weight and IQ at Age Six Years." *Archives of Pédiatrie and Adolescent Medicine*, 148, 377-83.
- Brooks-Gunn, J., P. Klebanov, and G. Duncan. (1996). "Ethnic Differences in Children's Intelligence Test Scores: Role of Economic Deprivation, Home Environment, and Maternal Characteristics." *Child Development*, 67, 396-408.
- Chen, E., K. A. Matthews, and T. W. Boyce. (2002). "Socioeconomic Differences in Children's Health: How and Why Do These Relationships Change With Age?" *Psychological Bulletin*, 128(2), 295-329.
- Conley, D. and N. Bennett. (2000). "Is Biology Destiny? Birth Weight and Life Chances." *American Sociological Review*, 65, 458-67.
- Crump, C., K. Sundquist, J. Sundquist, and M.A. Winkleby. (2011). "Gestational Age at Birth and Mortality in Young Adulthood." *Journal of the American Medical Association*, 306, 1233-1240.

- Currie, J. (2003). "U.S. Food and Nutrition Programs" in: R. Moffitt, (eds.), Means Tested Transfer Programs in the US, Chapter 4. University of Chicago Press.
- Currie, J. and R. Hyson. (1999). "Is the Impact of Health Problems Cushioned by Socioeconomic Status? The Case of Low Birth Weight." *American Economic Review*, 89, 245-50.
- Currie, J. and E. Moretti. (2007). "Biology as Destiny? Short and Long-Run Determinants of Intergenerational Transmission of Birth Weight." *Journal of Labor Economics*, 25, 231-64.
- Currie, J. and I. Rajani. (2015). "Within-Mother Estimates of the Effects of WIC on Birth Outcomes in New York City," *Economic Inquiry*, 53(4), 1691-1701.
- Deaton, A. (2002). "Policy Implications of the Gradient of Health and Wealth." *Health Affairs*, 21(2), 13-30.
- Driscoll, A. K. and M.K.K. Osterman (2018). "Maternal Characteristics of Prenatal WIC Receipt in the United States, 2016." NCHS Data Brief, no 298. Hyattsville, MD: National Center for Health Statistics. 2018.
- Figlio, D., S. Hamersma, and J. Roth. (2009). "Does Prenatal WIC Participation Improve Birth Outcomes? New Evidence from Florida." *Journal of Public Economics*, 93, 235-245.
- Frazis, H., Loewenstein, M.A., 2003. Estimating Linear Regressions With Mismeasured, Possibly Endogenous, Binary Explanatory Variables. *Journal of Econometrics* 117 (1), 151–178.
- Goldenberg, R.L. and J.F. Culhane. (2007). "Low Birth Weight in the United States." *American Journal of Clinical Nutrition*, 85, 584S-590S.
- Goldenberg R.L., J.F. Culhane, J.D. Iams, and R. Romero. (2008). "Epidemiology and Causes of Preterm Birth." *Lancet*, 371, 75-84.
- Gundersen, C. (2005). "A Dynamic Analysis of the Well-Being of WIC Recipients and Eligible Non-Recipients." *Children and Youth Services Review*, 27(1), 99-114.
- Gundersen, C., B. Kreider, and J. Pepper. (2012). "The Impact of the National School Lunch Program on Child Health: A Nonparametric Bounds Analysis." *Journal of Econometrics*, 166, 79-91.
- Hausman, J., J. Abrevaya, and F.M. Scott-Morton. (1998). "Misclassification of the Dependent Variable in a Discrete-response Setting." *Journal of Econometrics*, 97, 239-269.
- Heckman, J. and P. Carneiro, (2003). "*Human Capital Policy.*" in Inequality in America: What Role for Human Capital Policies (edited by James Heckman and Alan Krueger), MIT Press.
- Horowitz, J. and C. Manski. (1995). "Identification and Robustness With Contaminated and Corrupted Data." *Econometrica*, 63(2), 281-302.

- Hoynes, H.W., M. Page, and A.H. Stevens. (2011). "Can Targeted Transfers Improve Birth Outcomes? Evidence from the Introduction of the WIC Program." *Journal of Public Economics*, 95, 813-827.
- Imbens, G. W. and C. F. Manski. (2004). "Confidence Intervals for Partially Identified Parameters." *Econometrica*, 72, 1845-1857.
- Jackowitz, A. and L. Tiehen. (2010). *WIC Participation Patterns: An Investigation of Delayed Entry and Early Exit*, ERR-109, U.S. Department of Agriculture, Economic Research Service.
- Johnson, B., Thorn, B., McGill, B., Suchman, A., Mendelson, M., Patlan, K.L., Freeman, B., Gotlieb, R., & Connor, P. (2013). *WIC Participant and Program Characteristics 2012*. Prepared by Insight Policy Research under Contract No. AG-3198-C-11-0010. Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service.
- Johnson, R. C. and R. F. Schoeni. (2007), "The Influence of Early-Life Events on Human Capital, Health Status, and Labor Market Outcomes Over the Life Course." PSC Research Report No. 07-616.
- Joyce, T., D. Gibson, and S. Colman. (2005). "The Changing Association Between Prenatal Participation in WIC and Birth Outcomes in New York City," *Journal of Policy Analysis and Management*, 24, 661-685.
- Joyce, T., A. Racine, and C. Yunzal-Butler. (2008). "Reassessing the WIC Effect: Evidence from the Pregnancy Nutrition Surveillance System." *Journal of Policy Analysis and Management*, 27, 277-303.
- Kirlin, J.A. and M. Wiseman. (2014). "Getting it Right, or at Least Better: Improving Identification of Food Stamp Participants in the National Health and Nutrition Examination Survey," working paper.
- Kreider, B. (2010). "Regression Coefficient Identification Decay in The Presence of Infrequent Classification Errors," *Review of Economics and Statistics*, 92(4), 1017-1023.
- Kreider, B. and J. Pepper. (2007). "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors." *Journal of the American Statistical Association*, 102(478), 432-441.
- Kreider, B., J. Pepper, C. Gundersen, and D. Jolliffe. (2012). "Identifying the Effects of SNAP (Food Stamps) on Child Health Outcomes When Participation is Endogenous and Misreported." *Journal of the American Statistical Association*, 107, 499, 958-975.
- Kreider, B., J. Pepper, and M. Roy. (2016). "Identifying the Effects of WIC on Food Insecurity Among Infants and Children," *Southern Economic Journal*, 82(4), 1106-1122.
- Lawlor, D., H. Clark, G. D. Smith, and D. Leon. (2006). "Intrauterine Growth and Intelligence Within Sibling Pairs: Findings from the Aberdeen Children of the 1950s Cohort." *Pediatrics*, 117, 894-902.

- Ludwig, J. and M. Miller (2005). "Interpreting the WIC Debate." *Journal of Policy Analysis and Management*, 24(4):691-701.
- Manski, C. (1997). "Monotone Treatment Response." *Econometrica* 65(6), 1311-1334.
- Manski, C. (1995). *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press.
- Manski, C. and J. Pepper. (2000). "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica* 68(4), 997-1010.
- McCarthy, I., D. Millimet, and M. Roy. (2015). "Bounding Treatment Effects: A Command for the Partial Identification of the Average Treatment Effect with Endogenous and Misreported Treatment Assignment," *The Stata Journal*, 15(2), pp. 411–436.
- Meyer, B., D.W. Mok, and J.X. Sullivan. (2009). "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences." National Bureau of Economic Research Working Paper No.15181.
- Millimet D. 2011. "The Elephant in the Corner: A Cautionary Tale about Measurement Error in Treatment Effects Models," in D. Drukker (ed.) *Missing Data Methods: Cross-sectional Methods and Applications* (Advances in Econometrics, Vol. 27), Emerald Group Publishing Limited, 1-39.
- Moffitt, R. (2005). "Remarks on the Analysis of Causal Relationships in Population Research." *Demography*, 42 (1), 91-108.
- Morse, S. B., H. Zheng, Y. Tang, and J. Roth. (2009). "Early School-Age Outcomes of Late Preterm Infants." *Pediatrics*, 123(4), 622 -629.
- Oreopoulos, P., M. Stabile, L. Roos, and R. Walld. (2008), "The Short, Medium, and Long Term Effects of Poor Infant Health." *Journal of Human Resources*, 43, 88-138.
- Pepper, J. (2000). "The Intergenerational Transmission of Welfare Receipt: A Nonparametric Bounds Analysis." *The Review of Economics and Statistics*, 82, 472-88.
- Rossin-Slater, Maya. (2013). "WIC in your neighborhood: New evidence on the impacts of geographical access to clinics," *Journal of Public Economics*, 102, pp. 51-69.
- Royer, H. (2009), "Separated at Girth: Estimating the Long-Run and Intergenerational Effects of Birthweight Using Twins." *American Economic Journal: Applied Economics*, 1(1), 49-85.
- Schanzenbach, D., (2009). "Does the Federal School Lunch Program Contribute to Childhood Obesity?" *Journal of Human Resources* 44 (3), 684-709.
- USDA. (2003). "Nutrition and Evaluation, WIC Participant and Program Characteristics 2002," Food and Nutrition Service, Office of Analysis, Nutrition and Evaluation, WIC-03-PC, by Susan Bartlett, Ramona Olvera, Ellen Bobronnikov, et al. Project Officer, Julie Kresge. Alexandria, VA

USDA Food and Nutrition Service (2009). *WIC Eligibles and Coverage – 1994 to 2007: Estimates of the Population of Women, Infants, and Children Eligible for WIC Benefits*. Authors: E. Harper, J. Hirschman, J. Mabli, S. Nelson, and K. Hourihan.

Ver Ploeg, M. and D. Betson (eds.). (2003). *Report of the Panel to Evaluate the USDA's Methodology for Estimating Eligibility for and Participation in the WIC Program*, National Academies Press.

Appendices

A1. Bounds on the mean outcome gap, β , with classification errors

To make progress in partially identifying this gap, we decompose the first term in Equation (2) into identified and unidentified quantities:

$$P(H = 1 | W^* = 1) = \frac{P(H = 1, W^* = 1)}{P(W^* = 1)} = \frac{P(H = 1, W = 1) + \theta_1^- - \theta_1^+}{P(W^* = 1)} \quad (\text{A1})$$

where the probability of having a healthy birth outcome and self-reporting WIC receipt, $P(H = 1, W = 1)$, is identified by the data with $\theta_j^+ = P(H = j, W = 1, W^* = 0)$ and $\theta_j^- = P(H = j, W = 0, W^* = 1)$ denoting the unobserved fraction of false positive and false negative classifications of WIC recipients, respectively, for infants realizing birth outcome $j = 1, 0$. In the numerator, $\theta_1^- - \theta_1^+$ reflects the unobserved excess of false negative versus false positive classifications among those with a favorable birth outcome. The quantity $P(H = 1 | W^* = 0)$ can be decomposed analogously.

The ME1 and ME2 assumptions imply informative upper bounds on the false reporting probabilities $\theta_1^-, \theta_0^-, \theta_1^+$ and θ_0^+ . Specifically, it follows that

$$\theta_i^- \leq \min \left\{ P(H = i, W = 0, V = 0), P^*, \frac{1}{2}(Q_u + \Delta) \right\} \equiv \theta_i^{-UB}, \quad i = 1, 0 \quad (\text{A2})$$

$$\theta_i^+ \leq \min \left\{ P(H = i, W = 1, V = 0), 1 - P^*, \frac{1}{2}(Q_u - \Delta) \right\} \equiv \theta_i^{+UB}, \quad i = 1, 0. \quad (\text{A3})$$

where Δ is the difference in false negative and positive reporting rates described in Section 3:

$$\Delta = P^* - P = (\theta_1^- + \theta_0^-) - (\theta_1^+ + \theta_0^+).^{26} \quad (\text{A4})$$

Given these restrictions on the degree and nature of classification errors, we have the following proposition:

²⁶ With verification, P^* is logically required to lie within the range $\Lambda = [\max\{P - P(W = 1, V = 0), P(H = 1, W = 1) - \theta_1^{+UB}\}, \min\{P + P(W = 0, V = 0), 1 - P(H = 1, W = 0) + \theta_1^{-UB}\}]$. Intuitively, P^* cannot be too far away from P when some households are known to provide valid responses. Under no false positives (and no further verification), this range reduces to $P^* \in [P, 1]$.

Proposition 1 *Given the classification error model restrictions in Equations (A2) and (A3), the outcome gap, β , under partial verification is bounded as follows:*

$$\begin{aligned} & \frac{P(H = 1, W = 1) - P^* P(H = 1) - \min\{\theta_1^{+UB}, \theta_0^{-UB} - \Delta\}}{P^*(1 - P^*)} \\ & \leq \beta \leq \\ & \frac{P(H = 1, W = 1) - P^* P(H = 1) + \min\{\theta_1^{-UB}, \theta_0^{+UB} + \Delta\}}{P^*(1 - P^*)}. \end{aligned}$$

for $P^* \in \Lambda \cap (0, 1)$.²⁷

Proof: We can write the average treatment effect as

$$ATE = \frac{P(H = 1, W = 1) - P(W = 1)P(H = 1) + \theta_1^- - \theta_1^+}{P(W^* = 1)P(W^* = 0)}. \quad (A5)$$

Thus, subject to restrictions on the unknown classification error rates in ME1, ME2, and Equations (A2)-(A4) and the laws of probability,²⁸ the upper bound is found by maximizing θ_1^- and minimizing θ_1^+ . Likewise, the lower bound is found by minimizing θ_1^- and maximizing θ_1^+ . Let $\Psi = \theta_1^- - \theta_1^+$.

If $\Delta \geq 0$ (net false negative reporting): For the *upper bound*, first consider the case that $\theta_1^{-UB} < \Delta$. Here, θ_1^- cannot exceed θ_1^{-UB} and θ_1^+ cannot fall below 0 so that $\Psi \leq \theta_1^{-UB}$. At this upper bound, Equation (A4) implies that $\theta_0^- = \Delta - \theta_1^{-UB}$ and $\theta_0^+ = 0$. If $\theta_1^{-UB} \geq \Delta$, θ_1^- cannot exceed $\min\{\theta_1^{-UB}, \theta_0^{+UB} + \Delta\}$ and θ_1^+ cannot fall below 0 so that $\Psi \leq \min\{\theta_1^{-UB}, \theta_0^{+UB} + \Delta\}$. At this upper bound, Equation (A4) implies $\theta_0^- = 0$ and $\theta_0^+ = \min\{\theta_0^{+UB}, \theta_1^{-UB} - \Delta\}$. For the *lower bound*, first consider the case that $\theta_0^{-UB} < \Delta$. Here, θ_0^- cannot exceed θ_0^{-UB} so from Equation (A4) we know that θ_1^- must be no less than $\Delta - \theta_0^{-UB}$. From Equation (A3), we know that θ_j^+ can exceed 0 but any conjectured increase in the false positive error rate must be offset by an equivalent increase in the false negative error rate. So, in this case, the lower bound would be unchanged by increasing θ_j^+ above 0. Thus, we have

²⁷ If the valid range Λ includes 1 (in which the denominator in the proposition is 0), the bounds converge to $[-P(H = 0), P(H = 1)]$ as P^* approaches 1. Similarly, if the valid range includes 0, the bounds converge to $[-P(H = 1), P(H = 0)]$ as P^* approaches 0.

²⁸ In particular, $P(H = 1 | W^* = 1)$ and $P(H = 1 | W^* = 0)$ lie within $[0, 1]$.

$\Psi \geq \Delta - \theta_0^{-UB}$. If $\theta_0^{-UB} \geq \Delta$, Ψ is minimized when $\theta_0^- = \min\{\theta_0^{-UB}, \theta_1^{+UB} + \Delta\}$, $\theta_1^- = 0$,

$\theta_1^+ = \min\{\theta_1^{+UB}, \theta_0^{-UB} - \Delta\}$, and $\theta_0^+ = 0$ so that $\Psi \geq -\min\{\theta_1^{+UB}, \theta_0^{-UB} - \Delta\}$.

If $\Delta \leq 0$ (net false positive reporting): For the *upper bound*, first consider the case that $\theta_0^{+UB} < -\Delta$. Here, θ_0^+ cannot exceed θ_0^{+UB} so from Equation (A4) we know that θ_1^+ must be no less than $-\Delta - \theta_0^{+UB}$. From Equation (A2), we know that θ_j^- can exceed 0 but any conjectured increase in the false negative error rate must be offset by an equivalent increase in the false positive error rate.

So, in this case, the upper bound would be unchanged by increasing θ_j^- above 0. Thus, we have

$\Psi \leq -\Delta - \theta_0^{+UB}$. If $\theta_0^{+UB} \geq -\Delta$, Ψ is maximized when $\theta_0^+ = \min\{\theta_0^{+UB}, \theta_1^{-UB} - \Delta\}$, $\theta_1^+ = 0$,

$\theta_1^- = \min\{\theta_1^{-UB}, \theta_0^{+UB} + \Delta\}$, and $\theta_0^- = 0$ so that $\Psi \leq \min\{\theta_1^{-UB}, \theta_0^{+UB} + \Delta\}$. For the *lower bound*, first

consider the case that $\theta_1^{+UB} < -\Delta$. Here, θ_1^+ cannot exceed θ_1^{+UB} and θ_1^- cannot fall below 0 so that

$\Psi \geq -\theta_1^{+UB}$. At this lower bound, Equation (A4) implies that $\theta_0^+ = -\Delta - \theta_1^{+UB}$ and $\theta_0^- = 0$. If $\theta_1^{+UB} \geq -\Delta$,

θ_1^+ cannot exceed $\min\{\theta_1^{+UB}, \theta_0^{-UB} - \Delta\}$ and θ_1^- cannot fall below 0 so that $\Psi \geq -\min\{\theta_1^{+UB}, \theta_0^{-UB} - \Delta\}$.

At this lower bound, Equation (A4) implies $\theta_0^+ = 0$ and $\theta_0^- = \min\{\theta_0^{-UB}, \theta_1^{+UB} + \Delta\}$. Combining these

results, it follows that $-\min\{\theta_1^{+UB}, \theta_0^{-UB} - \Delta\} \leq \Psi \leq \min\{\theta_1^{-UB}, \theta_0^{+UB} + \Delta\}$. \square

In an arbitrary errors model ($Q_u = 1$), these bounds simplify:

Corollary 1 *Under arbitrary errors, the Proposition 1 bounds reduce to:*

$$-\min\left\{\frac{P(H=1)}{1-P^*}, \frac{P(H=0)}{P^*}\right\} \leq \beta \leq \min\left\{\frac{P(H=0)}{1-P^*}, \frac{P(H=1)}{P^*}\right\}.^{29}$$

Proof of Corollary 1: Under the arbitrary errors model, $Q_u = 1$. From Equation (A3), we know that

$\theta_0^{+UB} + \Delta = \min\{P(W^* = 1) - P(H = 1, W = 1), P(W = 0), \frac{1}{2}[P(W^* = 1) + P(W = 0)]\}$ and

$\theta_1^{-UB} = \min\{P(H = 1, W = 0), P(W^* = 1), \frac{1}{2}[P(W^* = 1) + P(W = 0)]\}$. Thus, it follows that

$$\min\{\theta_1^{-UB}, \theta_0^{+UB} + \Delta\} = \min\{P(W^* = 1) - P(H = 1, W = 1), P(H = 1, W = 0)\}. \quad (\text{A6})$$

²⁹ As in Proposition 1, these bounds are defined over $P^* \in \Lambda \cap (0, 1)$, where Λ is defined in the previous footnote.

Likewise, from Equations (A2) and (A3) we know that

$$\theta_0^{-UB} - \Delta = \min \left\{ P(W^* = 0) - P(H = 1, W = 0), P(W = 1), \frac{1}{2}[P(W^* = 0) + P(W = 1)] \right\} \text{ and}$$

$$\theta_1^{+UB} = \min \left\{ P(H = 1, W = 1), P(W^* = 0), \frac{1}{2}[P(W^* = 0) + P(W = 1)] \right\}. \text{ Thus, it follows that}$$

$$\min \left\{ \theta_1^{+UB}, \theta_0^{-UB} - \Delta \right\} = \min \left\{ P(W^* = 0) - P(H = 1, W = 0), P(H = 1, W = 1) \right\}. \quad (\text{A7})$$

Substituting (A6) into Proposition 1 obtains an upper bound of $\frac{P(H = 0)}{1 - P^*}$ if $P^* \leq P(H = 1)$ and

$$\frac{P(H = 1)}{P^*} \text{ if } P^* \geq P(H = 1), \text{ which reduces to } \min \left\{ \frac{P(H = 0)}{1 - P^*}, \frac{P(H = 1)}{P^*} \right\}. \text{ Similarly, the lower bound}$$

is obtained using Equation (A7). \square

Exogenous selection bounds when P^* is not known

If Assumptions ME1 and ME2 hold under arbitrary errors but the researcher has no auxiliary information about the value of P^* , sharp bounds on β are given as follows:

$$\beta^{LB} = \inf_{\theta_1^+ \in [0, \theta_1^{+UB}], \theta_0^- \in [0, \min\{Q_u - \theta_1^+, \theta_0^{-UB}\}]} \left[\frac{P(H = 1, W = 1) - \theta_1^+}{P(W = 1) - \theta_1^+ + \theta_0^-} - \frac{P(H = 1, W = 0) + \theta_1^+}{P(W = 0) + \theta_1^+ - \theta_0^-} \right]$$

$$\beta^{UB} = \sup_{\theta_1^- \in [0, \theta_1^{-UB}], \theta_0^+ \in [0, \min\{Q_u - \theta_1^-, \theta_0^{+UB}\}]} \left[\frac{P(H = 1, W = 1) + \theta_1^-}{P(W = 1) + \theta_1^- - \theta_0^+} - \frac{P(H = 1, W = 0) - \theta_1^-}{P(W = 0) - \theta_1^- + \theta_0^+} \right].$$

These bounds are obtained by minimizing or maximizing $P(H = 1 | W^* = 1) - P(H = 1 | W^* = 0)$ using Equation (A1) and its analogous counterpart for $P(H = 1 | W^* = 0)$. They can be estimated by conducting separate grid searches over $\{\theta_1^+, \theta_0^-\}$ and $\{\theta_1^-, \theta_0^+\}$ in the feasible regions to minimize β^{LB} and maximize β^{UB} subject to the constraint that none of the conditional probabilities (given by the ratios) exceed 1. Under the no false positives assumption, these bounds simplify to:

$$ATE^{LB} = \inf_{\theta_0^- \in [0, \theta_0^{-UB}]} \left[\frac{P(H = 1, W = 1)}{P(W = 1) + \theta_0^-} - \frac{P(H = 1, W = 0)}{P(W = 0) - \theta_0^-} \right]$$

$$ATE^{UB} = \sup_{\theta_0^+ \in [0, \theta_0^{+UB}]} \left[\frac{P(H = 1, W = 1) + \theta_1^-}{P(W = 1) + \theta_1^-} - \frac{P(H = 1, W = 0) - \theta_1^-}{P(W = 0) - \theta_1^-} \right].$$

A2. Sharp bounds under endogenous selection and classification errors

KPGJ show that the worst-case selection bounds on the ATE with classification errors can be written as

$$\begin{aligned} & -[P(H = 1, W = 0) + P(H = 0, W = 1)] + \Theta \\ & \leq ATE(1, 0) \leq \\ & [P(H = 1, W = 1) + P(H = 0, W = 0)] + \Theta \end{aligned} \quad (A8)$$

where $\Theta \equiv (\theta_1^- + \theta_0^+) - (\theta_0^- + \theta_1^+)$. In the absence of classification errors, $\Theta = 0$, and Equation (A8) simplifies to Manski's (1995) worst-case selection bounds. With classification errors, the models from Section 4 allow us to place informative restrictions on Θ . Extending KPGJ's Proposition 1 to incorporate partial verification, we have:

Proposition 2: *Given the bounds on the ATE in Equation (A8), combined with restrictions ME1, ME2, and (A2)-(A4) on classification errors, endogenous selection bounds on the ATE under partial verification are given by:*

$$\begin{aligned} & [-P(H = 1, W = 0) - P(H = 0, W = 1)] + \max\{-2\theta_1^{+UB} - \Delta, -2\theta_0^{-UB} + \Delta\} \\ & \leq ATE(1, 0) \leq \\ & [P(H = 1, W = 1) + P(H = 0, W = 0)] + \min\{2\theta_1^{-UB} - \Delta, 2\theta_0^{+UB} + \Delta\}. \end{aligned}$$

Proof: This proof follows the structure of the proof of Proposition 1 in KPGJ but allows for verification and an upper bound error rate. These restrictions are embedded in restrictions ME1, ME2, and (A2)-(A4). The upper bound is found by maximizing $(\theta_1^- + \theta_0^+)$ and minimizing $(\theta_0^- + \theta_1^+)$, and vice versa for the lower bound.

If $\Delta \geq 0$: For the *upper bound*, first consider the case that $\theta_1^{-UB} \geq \Delta$. Then $(\theta_0^- + \theta_1^+)$ is minimized at 0 and Equation (A4) simplifies to $\theta_1^- = \Delta + \theta_0^+$. It follows that θ_0^+ cannot exceed $\min\{\theta_0^{+UB}, \theta_1^{-UB} - \Delta\}$ and θ_1^- cannot exceed $\min\{\Delta + \theta_0^{+UB}, \theta_1^{-UB}\}$. The upper bound follows directly. Second, consider the case that $\theta_1^{-UB} < \Delta$. We know that θ_1^- cannot exceed θ_1^{-UB} and, to satisfy the restriction in Equation (A4), θ_0^- must be no less than $\Delta - \theta_1^{-UB}$. As before, θ_1^+ is minimized at 0.

While θ_0^+ can exceed 0, any conjectured increase in the false positive error rate must be offset by an equivalent increase in the false negative error rate. So, in this case, the upper bound would be unchanged by increasing θ_0^+ above 0. Thus, we have the upper bound on Θ of $2\theta_1^{-UB} - \Delta$ which can be shown to be no greater than $2\theta_0^{+UB} + \Delta$.

For the *lower bound*, first consider the case that $\theta_0^{-UB} \geq \Delta$. Then $(\theta_1^- + \theta_0^+)$ is minimized at 0 and Equation (A4) simplifies to $\theta_0^- = \Delta + \theta_1^+$. It follows that θ_1^+ cannot exceed $\min\{\theta_1^{+UB}, \theta_0^{-UB} - \Delta\}$ and θ_0^- cannot exceed $\min\{\Delta + \theta_1^{+UB}, \theta_0^{-UB}\}$ so that $\max\{-2\theta_1^{+UB} - \Delta, -2\theta_0^{-UB} + \Delta\}$ provides the lower bound on Θ . Second, consider the case that $\theta_0^{-UB} < \Delta$. We know that θ_0^- cannot exceed θ_0^{-UB} and, to satisfy the restriction in Equation (A4), θ_1^- must be no less than $\Delta - \theta_0^{-UB}$. As before, θ_0^+ is minimized at 0. While θ_1^+ can exceed 0, any conjectured increase in the false positive error rate must be offset by an equivalent increase in the false negative error rate. So, in this case, the lower bound would be unchanged by increasing θ_1^+ above 0. Thus, we have the lower bound on Θ of $-2\theta_0^{-UB} + \Delta$ which can be shown to be no smaller than $-2\theta_1^{+UB} - \Delta$.

If $\Delta < 0$: For the *upper bound*, first consider the case that $\theta_0^{+UB} \geq -\Delta$. Then $(\theta_0^- + \theta_1^+)$ is minimized at 0 and Equation (A4) simplifies to $\theta_0^+ = -\Delta + \theta_1^-$. We know that θ_1^- cannot exceed $\min\{\theta_1^{-UB}, \theta_0^{+UB} + \Delta\}$ and θ_0^+ cannot exceed $\min\{\theta_0^{+UB}, -\Delta + \theta_1^{-UB}\}$. The upper bound follows directly. Second, consider the case that $\theta_0^{+UB} < -\Delta$. We know that θ_0^+ cannot exceed θ_0^{+UB} and, to satisfy the restriction in Equation (A4), θ_1^+ must be no less than $-\Delta - \theta_0^{+UB}$. As before, θ_0^- is minimized at 0. While θ_1^- can exceed 0, any conjectured increase in the false negative error rate must be offset by an equivalent increase in the false positive error rate. So, in this case, the upper bound would be unchanged by increasing θ_1^- above 0. Thus, we have the upper bound on Θ of $2\theta_0^{+UB} + \Delta$ which can be shown to be no greater than $2\theta_1^{-UB} - \Delta$.

For the *lower bound*, first consider the case that $\theta_1^{+UB} \geq -\Delta$. Then $(\theta_1^- + \theta_0^+)$ is minimized at 0 and Equation (A4) simplifies to $\theta_1^+ = -\Delta + \theta_0^-$. We know that θ_0^- cannot exceed $\min\{\theta_0^{-UB}, \theta_1^{+UB} + \Delta\}$ and θ_1^+ cannot exceed $\min\{\theta_1^{+UB}, -\Delta + \theta_0^{-UB}\}$ so that $\max\{-2\theta_1^{+UB} - \Delta, -2\theta_0^{-UB} + \Delta\}$ provides the lower

bound on Θ . Second, consider the case that $\theta_1^{+UB} < -\Delta$. We know that θ_1^+ cannot exceed θ_1^{+UB} and, to satisfy the restriction in Equation (A4), θ_0^+ must be no less than $-\Delta - \theta_1^{+UB}$. As before, θ_1^- is minimized at 0. While θ_0^- can exceed 0, any conjectured increase in the false negative error rate must be offset by an equivalent increase in the false positive error rate. So, in this case, the lower bound would be unchanged by increasing θ_0^- above 0. Thus, we have the lower bound of $-2\theta_1^{+UB} - \Delta$ which can be shown to be no smaller than $-2\theta_0^{-UB} + \Delta$. \square

Adding the MTR and income MIV assumptions to this model is straightforward. Under MTS, the ATE is bounded from above by the outcome gap in birth outcomes among recipients and nonrecipients:

$$ATE(1,0) \leq P(H = 1 | W^* = 1) - P(H = 1 | W^* = 0).$$

This upper bound is not operational when W^* is unobserved since these conditional means are not identified. Proposition 1, however, provides bounds on this outcome gap.

Endogenous selection bounds when P^* is not known³⁰

If Assumptions ME1 and ME2 hold under arbitrary errors but the researcher does not know the value of P^* , sharp bounds on the ATE are given as follows:³¹

$$\begin{aligned} ATE^{LB} &= P(H = 1, W = 1) - P(H = 1, W = 0) - P(W = 1) - \min\{Q_u, \theta_1^{+UB} + \theta_0^{-UB}\} \\ ATE^{UB} &= P(H = 1, W = 1) - P(H = 1, W = 0) + P(W = 0) + \min\{Q_u, \theta_0^{+UB} + \theta_1^{-UB}\}. \end{aligned}$$

Naturally, these bounds when P^* is unknown are wider than the Proposition 2 bounds when P^* is known. Under no false positives, these bounds narrow to:

$$\begin{aligned} ATE^{LB} &= P(H = 1, W = 1) - P(H = 1, W = 0) - P(W = 1) - \theta_0^{-UB} \\ ATE^{UB} &= P(H = 1, W = 1) - P(H = 1, W = 0) + P(W = 0) + \theta_1^{-UB}. \end{aligned}$$

³⁰ McCarthy, Millimet, and Roy (2015) develop a Stata command for these bounds.

³¹ The ATE is given by $P[H(1) = 1] - P[H(0) = 1] = P[H(1) = 1 | W^* = 1]P(W^* = 1) + P[H(1) = 1 | W^* = 0]P(W^* = 0) - P[H(0) = 1 | W^* = 1]P(W^* = 1) - P[H(0) = 1 | W^* = 0]P(W^* = 0)$. Letting $P[H(1) = 1 | W^* = 0]$ and $P[H(0) = 1 | W^* = 1]$ vary within $[0, 1]$, the ATE is no smaller than $P(H = 1, W^* = 1) - P(W^* = 1) - P(H = 1, W^* = 0) = [P(H = 1, W = 1) + \theta_1^- - \theta_1^+] - [P(W = 1) + \theta_1^- + \theta_0^- - \theta_1^+ - \theta_0^+] - [P(H = 1, W = 0) + \theta_1^+ - \theta_1^-] = P(H = 1, W = 1) - P(W = 1) - P(H = 1, W = 0) - \theta_0^- + \theta_0^+ - \theta_1^+ + \theta_1^-$. The lower bound is obtained by setting $\theta_0^+ = \theta_1^- = 0$ and recognizing that the sum of errors cannot exceed Q_u . The upper bound is derived analogously.

A3. MIV ineligible bounds with mislabeled ineligibility

Since ineligible households did not receive prenatal WIC benefits ($inc > 185\%$ and $B^* = 0$ implies $W^* = 0$),³² it follows that $P[H(0) = 1]$ cannot exceed $P(H = 1 | B^* = 0, \nu = \text{ineligible})$, where ν denotes selecting the comparison group based on $inc > 185\%$ (income-ineligible), $SES \leq 3$, and $B = 0$. This conditional probability can be written as

$$\frac{P(H = 1, B^* = 0 | \nu)}{P(B^* = 0 | \nu)} = \frac{P(H = 1, B = 0 | \nu) + \tilde{\theta}_1^+ - \tilde{\theta}_1^-}{P(B = 0 | \nu) + (\tilde{\theta}_1^+ + \tilde{\theta}_0^+) - (\tilde{\theta}_1^- + \tilde{\theta}_0^-)}$$

where $\tilde{\theta}_j^+ \equiv P(H = j, B = 1, B^* = 0 | \nu)$ and $\tilde{\theta}_j^- \equiv P(H = j, B = 0, B^* = 1 | \nu)$ denote false positive and false negative adjunctive eligibility classifications for $j = 1, 0$.

Since we selected our sample of apparent ineligibles based partly on $B = 0$ (no self-reported participation in any WIC-related program), the conditional probability simplifies to

$$\frac{P(H = 1 | I) - \tilde{\theta}_1^-}{1 - \tilde{\theta}_1^- - \tilde{\theta}_0^-}.$$

The worst-case upper bound is obtained by setting $\tilde{\theta}_1^- = 0$:

$$P[H(0) = 1] \leq \frac{P(H = 1 | \nu)}{1 - P(H = 0, B^* = 1 | \nu)} = \frac{P(H = 1 | \nu)}{1 - \theta P(H = 0 | \nu)}$$

where, as described in the main text, $\theta \equiv P(B^* = 1 | H = 0, \nu)$ is the fraction of households in the unfavorable health outcome subset of the apparent ineligibles that did, in fact, receive benefits from a WIC-related program.

³² In some instances, the government might erroneously award benefits to a household that did not meet statutory eligibility criteria (e.g., administrative error). Conceptually, however, we treat any household that was actually awarded benefits as de facto eligible.

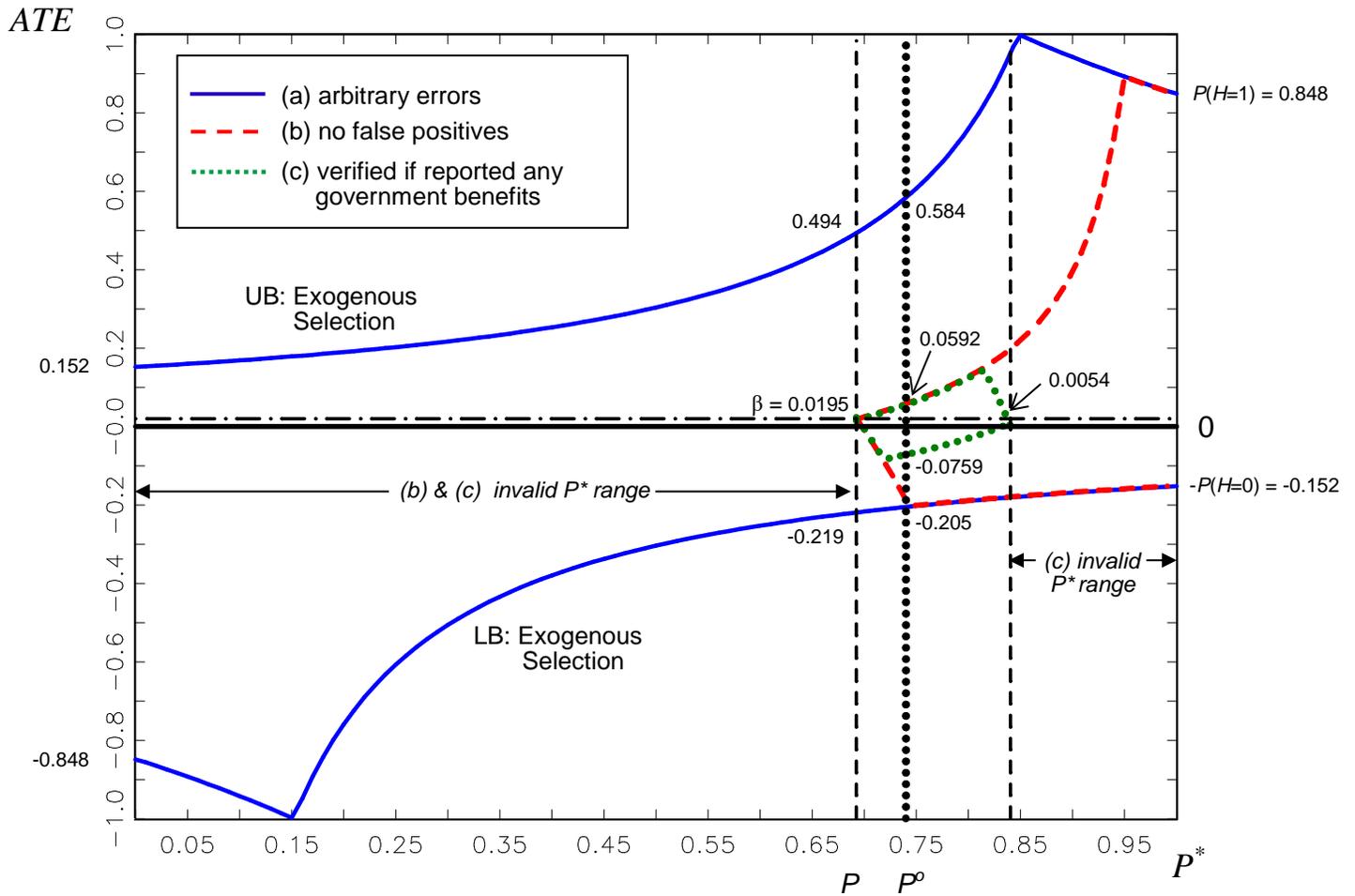
Table 1. Reported Prenatal WIC Participation and Birth Outcomes

	Mean	St. Dev.	Differences:	
			<u>WIC vs. No WIC</u> Mean	p-value
Reported prenatal WIC receipt (1 = Yes)	0.692	0.462		
Normal birthweight: 2500-4000 grams (1=yes)	0.848	0.359	0.0195	0.09
Birthweight \geq 2500 grams, Not low birthweight (1 = yes)	0.928	0.259	0.00246	0.76
Birthweight \geq 1500 grams, Not very low (1 = yes)	0.988	0.110	0.00337	0.33
Birthweight \leq 4000 grams, Not macrosomic (1 = yes)	0.920	0.271	0.0171	0.05
Gestation age: 38-42 weeks, Normal gestation age (1 = yes)	0.752	0.432	-0.0100	0.47
Gestation age \geq 37 weeks, Not premature (1 = yes)	0.882	0.322	-0.0188	0.07
Gestation age \geq 33 weeks, Not very premature (1 = yes)	0.968	0.176	-0.00361	0.52

N = 4750

Note: The sample includes 9-month old children from households with reported income at or below 185% of the Federal Poverty Guidelines along with households reporting prenatal Medicaid participation. All analyses are weighted using Wave 1 specific sample weights.

Figure 1. Sharp Bounds on the ATE for “Normal Birth Weight” (2500-4000 grams) as a Function of P^* , the Unobserved True WIC Participation Rate: **Exogenous Selection**



Self-reported
participation rate:
 $P^* = P = 0.692$

Administrative
participation rate:
 $P^* = P^o = 0.74$

Exogenous selection

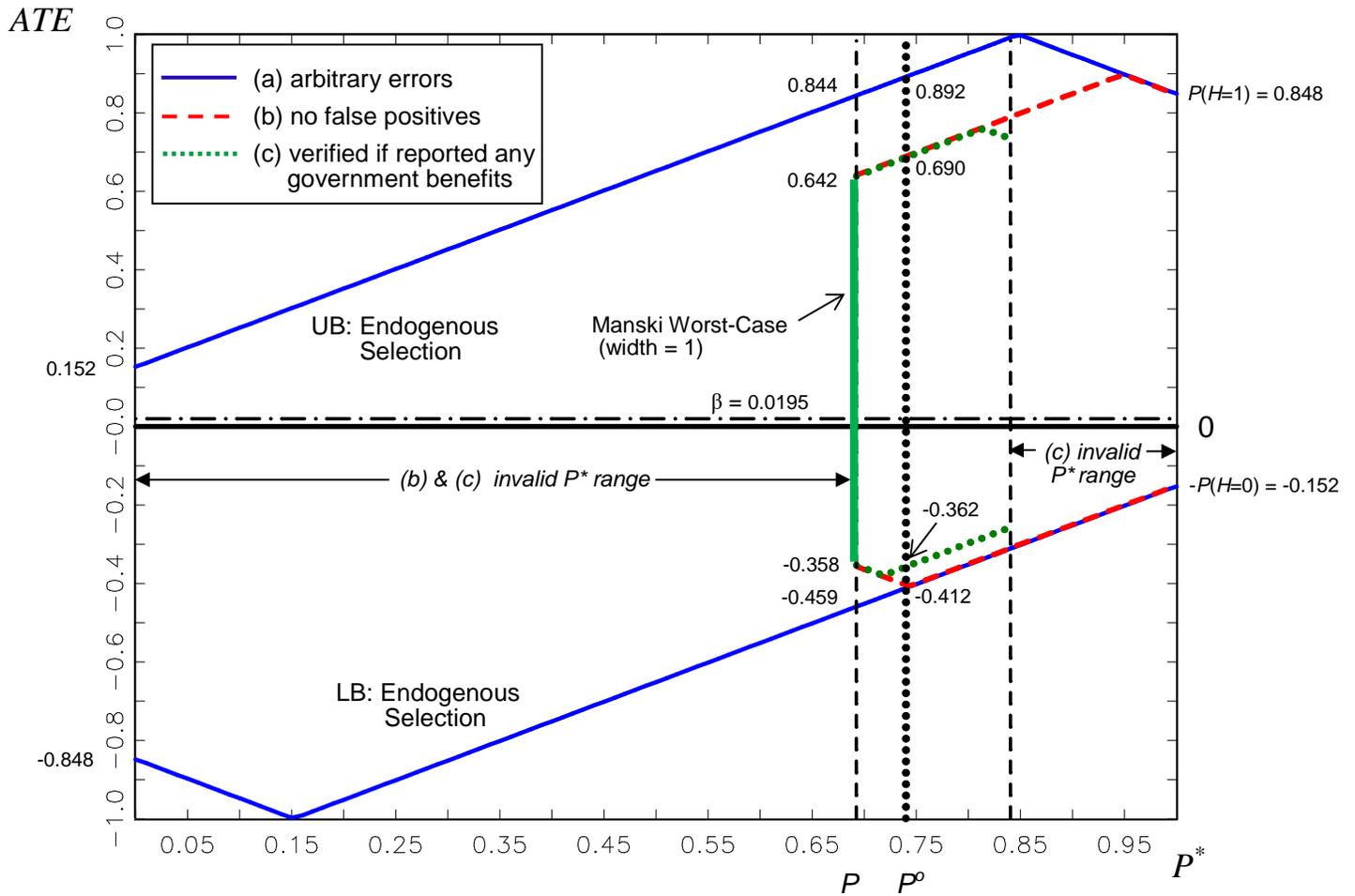
		LB	UB	width	LB	UB	width
(a) Arbitrary errors —	p.e. [†]	[-0.219, 0.494]	0.713		[-0.205, 0.584]	0.789	
	CI [‡]	[-0.230 0.523]			[-0.215 0.622]		
(b) No false positives - - -	p.e.	[0.0195, 0.0195]	0.000		[-0.188, 0.0592]	0.264	
	CI	[-0.0022 0.0409]			[-0.204 0.0784]		
(c) Verified if reported any gov't benefits · · · · ·	p.e.	[0.0195, 0.0195]	0.000		[-0.0759, 0.0592]	0.135	
	CI	[-0.0022 0.0409]			[-0.0912 0.0784]		

Strictly positive average treatment effects in **bold**

[†] Point estimates of the population bounds

[‡] Imbens-Manski 5th and 95th percentile bounds (1,000 pseudosamples)

Figure 2. Sharp Bounds on the ATE for “Normal Birth Weight” (2500-4000 grams) as a Function of P^* , the Unobserved True WIC Participation Rate: **Worst Case Endogenous Selection Bounds**



Self-reported
participation rate:
 $P^* = P = 0.692$

Administrative
participation rate:
 $P^* = P^o = 0.74$

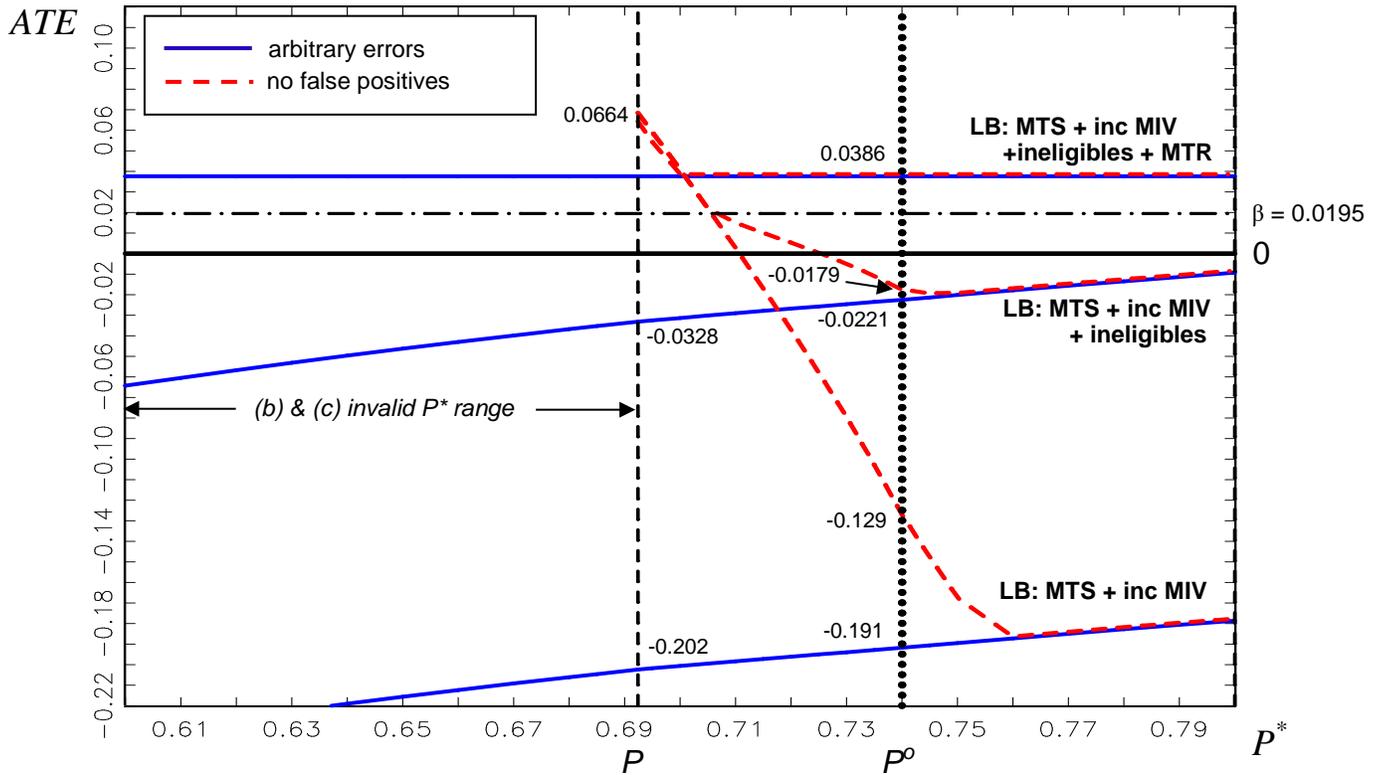
Endogenous selection

		LB	UB	width	LB	UB	width
(a) Arbitrary errors	p.e.	[-0.459, 0.844]	1.304		[-0.412, 0.892]	1.304	
	CI	[-0.473 0.858]			[-0.425 0.905]		
(b) No false positives	p.e.	[-0.358, 0.642]	1.000		[-0.405, 0.690]	1.095	
	CI	[-0.369 0.654]			[-0.417 0.701]		
(c) Verified if reported any gov't benefits	p.e.	[-0.358, 0.642]	1.000		[-0.362, 0.690]	1.052	
	CI	[-0.369 0.654]			[-0.375 0.701]		

† Point estimates of the population bounds

‡ Imbens-Manski 5th and 95th percentile bounds (1,000 pseudosamples)

Figure 3. Sharp Lower Bounds on the ATE for “Normal Birth Weight” as a Function of P^* :
Endogenous Selection Bounds with MTS, MTR, Income MIV, and Ineligibles MIV



		Self-reported rate $P^* = P = 0.692$	Administrative rate $P^* = P^o = 0.74$
MTS+ inc MIV + ineligible:			
(a) Arbitrary errors	p.e.*	[-0.0328, 0.836]	[-0.0221, 0.886]
	CI	[-0.0393, 0.846]	[-0.0285, 0.893]
	width	0.869	0.908
(b) No false positives	p.e.	[0.0664 , 0.630]	[-0.0179, 0.678]
	CI	[0.0438, 0.636]	[-0.0188, 0.685]
	width	0.563	0.696
MTS+ inc MIV+ineligibles+MTR:			
(a) Arbitrary errors	p.e.	[0.0386 , 0.836]	[0.0386 , 0.886]
	CI	[0.0260, 0.846]	[0.0260, 0.893]
	width	0.798	0.847
(b) No false positives	p.e.	[0.0664 , 0.630]	[0.0386 , 0.678]
	CI	[0.0438, 0.636]	[0.0260, 0.685]
	width	0.563	0.640

Strictly positive average treatment effects in **bold**.

† Point estimates of the population bounds corrected for finite sample bias

‡ Imbens-Manski 5th and 95th percentile bounds (1,000 pseudosamples)

Table 2A. Sharp Bounds on the ATE of WIC Under No Classification Error, Birth Weight

		Birth Weight			
		Normal (2500-4000 g.)	Not Low (≥ 2500 g.)	Not Very Low (≥ 1500 g.)	Not Macrosomic (≤ 4000 g.)
(i) ETS ^c	p.e. ^a	[0.0195, 0.195]	[0.0025, 0.0025]	[0.0034,0.0034]	[0.0171, 0.0171]
	CI ^b	[-0.0022,0.392]	[-0.0080, 0.0125]	[0.0011,0.0056]	[-0.0027 0.0359]
(ii) Worst Case	p.e.	[-0.358, 0.642]	[-0.334, 0.667]	[-0.311,0.689]	[-0.331, 0.669]
	CI	[-0.369, 0.654]	[-0.346, 0.677]	[-0.322,0.700]	[-0.342 0.681]
(iii) MTS ^d	p.e.	[0.0195,0.642]	[0.0025, 0.667]	[0.0034,0.689]	[0.0171, 0.669]
	CI	[0.0026,0.654]	[-0.0057, 0.677]	[0.0016,0.700]	[0.0017 0.681]
(iv) MTS+MIV ^e	p.e.	[0.0654, 0.630]	[0.0026, 0.661]	[0.0059,0.683]	[0.0773, 0.640]
	CI	[0.0412, 0.636]	[-0.0057, 0.667]	[0.0046,0.691]	[0.0390 0.660]
(v) MTS+MIV +ineligibles ^f	p.e.	[0.0654, 0.630]	[0.0026, 0.661]	[0.0059,0.683]	[0.0773, 0.640]
	CI	[0.0412 0.636]	[-0.0057, 0.667]	[0.0046,0.691]	[0.0479 0.660]
(vi) MTS-MIV +ineligibles +MTR ^g	p.e.	[0.0654, 0.630]	[0.0026, 0.661]	[0.0065,0.683]	[0.0754, 0.640]
	CI	[0.0412, 0.636]	[0.0000, 0.667]	[0.0050,0.691]	[0.0544 0.660]

Notes: Strictly positive average treatment effects in **bold**.

- a. Bias-corrected point estimates of the bounds
- b. 90% Imbens-Manski confidence intervals (CI) using 1000 pseudosamples
- c. ETS denotes Exogenous Treatment Selection
- d. MTS denotes Monotone Treatment Selection
- e. MIV denotes the income monotone instrument
- f. "ineligibles" denotes the ineligible monotone instrument
- g. MTR denotes Monotone Treatment Response

Table 2B. Sharp Bounds on the ATE of WIC Under No Classification Error, Gestation Age

		Gestation Age		
		Normal (38-42 weeks)	Not Premature (≥ 37 weeks)	Not Very Premature (≥ 33 weeks)
(i) ETS ^c	p.e.	[-0.0100, -0.0100]	[-0.0188, -0.0188]	[-0.0036, -0.0036]
	CI	[-0.0378, 0.0184]	[-0.0364 -0.0006]	[-0.0106, 0.0035]
(ii) Worst Case	p.e. ^a	[-0.408,0.592]	[-0.362, 0.638]	[-0.323,0.667]
	CI ^b	[-0.419,0.604]	[-0.373 0.650]	[-0.334,0.689]
(iii) MTS ^d	p.e.	[-0.0100, 0.592]	[-0.0036, 0.677]	[-0.0036, 0.677]
	CI	[-0.0321, 0.604]	[-0.0091, 0.689]	[-0.0091, 0.689]
(iv) MTS+MIV ^e	p.e.	[0.0899, 0.576]	[0.0053, 0.624]	[0.0107, 0.664]
	CI	[0.0119, 0.583]	[-0.0112 0.634]	[-0.0011, 0.675]
(v) MTS+MIV +ineligibles ^f	p.e.	[0.0899, 0.576]	[0.0053, 0.624]	[0.0107, 0.664]
	CI	[0.0119, 0.583]	[-0.0107 0.634]	[-0.0009, 0.675]
(vi) MTS-MIV +ineligibles +MTR ^g	p.e.	[0.0951, 0.576]	[0.0083, 0.624]	[0.0102, 0.664]
	CI	[0.0285, 0.583]	[0.000 0.634]	[0.0051, 0.675]

Notes: Strictly positive average treatment effects in **bold**.

- a. Bias-corrected point estimates of the bounds
- b. 90% Imbens-Manski confidence intervals (CI) using 1000 pseudosamples
- c. ETS denotes Exogenous Treatment Selection
- d. MTS denotes Monotone Treatment Selection
- e. MIV denotes the income monotone instrument
- f. “ineligibles” denotes the ineligible monotone instrument
- g. MTR denotes Monotone Treatment Response

Table 3A. Sharp Bounds on the ATE of WIC, With Classification Errors and Verification of a Mixture of Responses at $P^* = 0.74$, Birth Weight

		Birth Weight			
		Normal (2500-4000 g.)	Not Low (≥ 2500 g.)	Not Very Low (≥ 1500 g.)	Not Macrosomic (≤ 4000 g.)
(i) ETS ^c	p.e.	[-0.0759, 0.0592]	[-0.0307, 0.0206]	[-0.0006, 0.0067]	[-0.0452, 0.0386]
	CI	[-0.0912 0.0784]	[-0.0382 0.0299]	[-0.0024 0.0088]	[-0.0583 0.0559]
(ii) Worst Case	p.e. ^a	[-0.362, 0.690]	[-0.307, 0.713]	[-0.266, 0.737]	[-0.316, 0.717]
	CI ^b	[-0.375 0.701]	[-0.318 0.725]	[-0.277 0.748]	[-0.328 0.728]
(iii) MTS ^d	p.e.	[-0.0759, 0.690]	[-0.0307, 0.713]	[-0.0006, 0.737]	[-0.0452, 0.717]
	CI	[-0.0912 0.701]	[-0.0382 0.725]	[-0.0024 0.748]	[-0.0583 0.728]
(iv) MTS+MIV ^e	p.e.	[-0.0119, 0.678]	[-0.0213, 0.709]	[0.0011, 0.731]	[0.0295, 0.689]
	CI	[-0.0486 0.684]	[-0.0257 0.714]	[0.0007 0.740]	[-0.0168 0.708]
(v) MTS+MIV +ineligibles ^f	p.e.	[0.0176, 0.678]	[-0.0124, 0.709]	[0.0011, 0.731]	[0.0480, 0.689]
	CI	[0.0095 0.684]	[-0.0166 0.714]	[0.0007 0.740]	[0.0291 0.708]
(vi) MTS-MIV +ineligibles +MTR ^g	p.e.	[0.0392, 0.678]	[0.000, 0.709]	[0.0027, 0.731]	[0.0553, 0.689]
	CI	[0.0316 0.684]	[0.000 0.714]	[0.0007 0.740]	[0.0458 0.708]

Notes:

Strictly positive average treatment effects in **bold**

a. Bias-corrected point estimates of the bounds

b. 90% Imbens-Manski confidence intervals (CI) using 1000 pseudosamples

c. ETS denotes Exogenous Treatment Selection

d. MTS denotes Monotone Treatment Selection

e. MIV denotes the income monotone instrument

f. “ineligibles” denotes the ineligible monotone instrument

g. MTR denotes Monotone Treatment Response

Table 3B. Sharp Bounds on the ATE of WIC, With Classification Errors and Verification of a Mixture of Responses at $P^* = 0.74$, Gestation Age

		Gestation Age		
		Normal (38-42 weeks)	Not Premature (≥ 37 weeks)	Not Very Premature (≥ 33 weeks)
(i) ETS ^c	p.e.	[-0.112, 0.0518]	[-0.0552, 0.0090]	[-0.0113, 0.0041]
	CI	[-0.134 0.0768]	[-0.0690 0.0247]	[-0.0166 0.0102]
(ii) Worst Case	p.e. ^a	[-0.422, 0.641]	[-0.338, 0.687]	[-0.280, 0.726]
	CI ^b	[-0.435 0.653]	[-0.349 0.698]	[-0.291 0.737]
(iii) MTS ^d	p.e.	[-0.112, 0.641]	[-0.0552, 0.687]	[-0.0113, 0.726]
	CI	[-0.134 0.653]	[-0.0690 0.698]	[-0.0166 0.737]
(iv) MTS+MIV ^e	p.e.	[-0.0563, 0.627]	[-0.0303, 0.675]	[-0.0022, 0.715]
	CI	[-0.0876 0.632]	[-0.0487 0.683]	[-0.0084 0.725]
(v) MTS+MIV +ineligibles ^f	p.e.	[-0.0563, 0.627]	[-0.0303, 0.675]	[-0.0022, 0.715]
	CI	[-0.0702 0.632]	[-0.0335 0.683]	[-0.0061 0.725]
(vi) MTS-MIV +ineligibles +MTR ^g	p.e.	[0.0269, 0.627]	[0.0041, 0.675]	[0.0050, 0.715]
	CI	[0.0133 0.632]	[0.0000 0.683]	[0.0008 0.725]

Notes:

Strictly positive average treatment effects in **bold**

- a. Bias-corrected point estimates of the bounds
- b. 90% Imbens-Manski confidence intervals (CI) using 1000 pseudosamples
- c. ETS denotes Exogenous Treatment Selection
- d. MTS denotes Monotone Treatment Selection
- e. MIV denotes the income monotone instrument
- f. “ineligibles” denotes the ineligible monotone instrument
- g. MTR denotes Monotone Treatment Response