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 Public Policy
University of North Carolina at Chapel Hill
mananroy@email.unc.edu

September 2016

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. To do so, we extend existing partial identification methods to reflect the institutional details of the WIC program by allowing for a much richer measurement error model and by applying a modified regression discontinuity design. Combining survey data from the Early Childhood Longitudinal Study Birth Cohort (ECLS-B) with administrative data from the USDA, we estimate 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). By providing access to nutritious food supplements, education, and preventative health services, the prenatal program aims to improve fetal develop and reduce the incidence of low birth weight and short gestation (USDA, 2003; Johnson et al., 2013).\(^1\) In fact, 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 (Currie, 2003).

However, despite the existing literature, the causal effects of prenatal WIC participation remain uncertain. While participating in WIC might improve fetal development, there may also exist unobserved factors that jointly influence the decision to take up WIC and birth outcomes. For instance, unobserved motivation, health status, emotional well-being, and family resources might all jointly affect both an expectant mother’s decision to participate in WIC and fetal development. Thus, observed relationships between WIC participation and birth outcomes might be spurious. A selection problem arises from the fact that data alone cannot reveal what the birth weight of an infant whose mother participated in WIC would have been had she not participated, or vice versa. While this fundamental identification problem is widely recognized, for the most part it has not been addressed in the literature: nearly every study treats selection as (conditionally) exogenous, and the few studies applying linear instrumental variables models tend to find that the instruments (typically variation in

\(^1\) 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).
state program regulations) are not strongly associated with WIC receipt (e.g., Bitler and Currie, 2005).²

Moreover, even if selection into WIC is exogenous, evaluating the efficacy of WIC is further complicated by a classification error problem. Households are known to systematically underreport the receipt of food assistance in national surveys (e.g., Bollinger and David, 1997; Bitler et al., 2003; and Meyer et al., 2009). As a result, the difference in average outcomes between those receiving and not receiving assistance – the outcome gap – may not reveal the impact of WIC. All studies evaluating the impact of prenatal WIC on birth outcomes treat reporting as accurate.

In this paper, we reevaluate 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 measurement problems. Using data from the Early Childhood Longitudinal Study-Birth cohort of 2001 (ECLS-B) and auxiliary information from USDA administrative data, we address these two identification problems by applying and extending the nonparametric bounding strategy from Kreider, Pepper, Gundersen, and Jolliffe (2012, hereafter KPGJ). This framework does not require the traditional instrumental variable or measurement error assumptions, or the linear response model.³ Instead, we rely on a number of relatively weak monotonicity restrictions that are straightforward to motivate in this application.

---

² Recent studies address the selection problem by focusing on narrowly defined samples. For example, Bitler and Currie (2005) evaluate the impact of WIC on Medicaid-eligible women, and Figlio et al. (2009) evaluate the impact on infants with older school aged siblings. More recently, Hoynes et al. (2011) evaluate the performance of WIC at the time of its establishment as a pilot program in 1972, exploiting the plausibly exogenous variation in participation due to the staggered introduction of the program in the 1970s. In each of these studies, the authors report results from both reduced form and instrumental variable models. The instrumental variables are generally weak, and the strongest evidence comes from the reduced from analyses. These studies tend to conclude that WIC leads to improved infant health, though evidence is mixed.

³ The classical linear response model assumption is difficult to justify when considering programs that are thought to have heterogeneous effects (Moffitt 2005).
After describing the data in Section 2, we formalize the empirical question and the identification problems in Section 3. Then, in Section 4, we begin by following the tradition in much of the literature by assuming that the decision to take up WIC is exogenous but we allow for the possibility that WIC participation is measured with error. 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 and is an important descriptive measure of the association between WIC participation and birth outcomes. In this section, we extend the partial identification methods used to address classification errors by deriving closed-form bounds under the exogenous selection assumption and by allowing for the possibility that some observed subgroups are known to provide accurate reports of WIC participation.4

In Section 5, we turn our attention to drawing inferences on the Average Treatment Effect (ATE) given the endogenous selection problem, abstracting away from measurement issues. Manski’s (1995) classic worst-case ATE bounds impose no assumptions on the selection process. To tighten these bounds, 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. Compared with the standard exclusion restriction or mean independence assumptions, these less restrictive MIVs are attractive in that they require only that healthy birth probabilities vary monotonically with the instrument. In addition to this standard MIV assumption, we also extend a modified discontinuity design applied in Gundersen et al. (2012) to account for WIC eligibility rules. In particular, we introduce the notion of a partially observed monotone instrument based on a federal law that confers adjunctive eligibility into WIC through participation in other assistance programs.

4 The methods in KPGJ allow for the verification of all positive treatment responses, but not some mixture of responses.
Finally, we consider the identifying power of a monotone treatment response (MTR) assumption (Manski, 1997).

In Section 6, we derive sharp bounds on average treatment effects that simultaneously account for both the classification error and selection problems. 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. Section 7 concludes. Under relatively weak nonparametric assumptions, we find evidence that WIC improves birth outcomes.

2. DATA

To study the impacts 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. 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. This sample reflects two notable sample restrictions. 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

---

5 ECLS-B case-level data are available to researchers who are granted a restricted-use data license. Information about receiving a restricted-use data license can be found at [http://nces.ed.gov/pubsearch/licenses.asp](http://nces.ed.gov/pubsearch/licenses.asp). Information about the survey design can be found at: [https://nces.ed.gov/ecls/birth.asp](https://nces.ed.gov/ecls/birth.asp).

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

7 We also drop about 400 observations that had missing information on age.
WIC benefits, either through income eligibility or adjunctive eligibility.\(^8\) 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 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.\(^9\)

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

\(^8\) 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).

\(^9\) According to records assembled by the USDA, about two-thirds of WIC recipients reported participating in at least one public assistance program conferring adjunctively 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.
eligibility for assistance programs using survey data and, in the case of WIC, has been found to underestimate 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 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 applied in Gundersen et al. (2012) and Schanzenbach (2009), we also use information on 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. As an innovation to Gundersen et al.’s (2012) discontinuity approach, we explicitly allow for the possibility that, like WIC status, participation in other programs may also be misclassified.

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). The last two columns report differences in means between those reporting prenatal WIC receipt and those classified as eligible nonparticipants.

---

10 All analyses are performed using survey weights available in the ECLS-B.
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. 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 points. We also evaluate other 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. 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.  

---

11 About 8 percent of birth weights in the sample are clinically macrosomic. 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.  

12 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).
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 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 0.69.\textsuperscript{13} Hausman et al.’s (1998) parametric model of asymmetric misclassification provides an alternative approach for estimating the extent of misreporting. 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 16 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) (Ludwig and Miller, 2005; Joyce, Gibson and Colman, 2005). Women whose pregnancies last longer have a longer period to take up 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 derived bounds on average treatment effects of WIC on birth outcomes.

\textsuperscript{13} 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.
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 AVERAGE 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. For binary outcomes, these treatment effects can be expressed as

\[
ATE(1, 0 \mid X \in \Omega) = E[H(1) \mid X \in \Omega] - E[H(0) \mid X \in \Omega]
\]

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 \(\Omega\). 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

\[\text{Section 3 closely follows KPGJ’s foundation for discussing partial identification of average treatment effects. Subsequent sections extend these methods to introduce partial verification of treatment status, a potentially mismeasured monotone instrument, and a formal analysis of the exogenous selection assumption.}\]
who did receive WIC. This is referred to as the selection problem. Using the law of total probability, this identification problem can be highlighted by writing the first term of Equation (1) as

$$P[H(1) = 1] = P[H(1) = 1 | W^* = 1]P(W^* = 1) + P[H(1) = 1 | W^* = 0]P(W^* = 0)$$ (2)

where $W^* = 1$ denotes that an infant resides in a household that truly receives WIC and $W^* = 0$ otherwise. If WIC receipt is observed, the sampling process identifies the selection probability $P(W^* = 1)$, the censoring probability $P(W^* = 0)$, and the expectation of outcomes when the outcome is observed, $P[H(1) = 1 | W^* = 1]$. Still, the sampling process cannot reveal the mean outcome conditional on censoring, $P[H(1) = 1 | W^* = 0]$. Given this censoring, $P[H(1) = 1]$ is not point-identified by the sampling process alone. Analogously, the second term in Equation (1), $P[H(0) = 1]$, is not identified.

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, $W^*$, and thus all of the probabilities on the right hand side of Equation (2) are unknown.

To address these two identification problems, we proceed in three steps. First, we focus on the implications of the measurement error problem alone when we have information on the true participation rate, $P^*$, and the self-reported rate, $P$. Second, assuming reports of WIC participation are accurate, we focus on the selection problem. Finally, we assess what can be inferred when accounting for both identification problems simultaneously.
4. EXOGENOUS SELECTION BOUNDS

Much of the literature examining the impact of WIC on health assumes that selection is 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:

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

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, it is widely applied in the literature (see Currie, 2003). Moreover, even if selection is not exogenous, the difference in expected outcomes among WIC recipients and nonrecipients – the mean outcome gap, $\beta$ – 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 Bounds on the Outcome Gap Given Classification Errors

If one allows for the possibility of classification errors in $W^*$, however, this mean outcome gap is not identified. To make progress in partially identifying this gap, we decompose the first term in Equation (3) into identified and unidentified quantities:

$$P(H = 1| W^* = 1) = \frac{P(H = 1, W^* = 1)}{P(W^* = 1)} = \frac{P(H = 1, W = 1) + \theta_j^- - \theta_j^+}{P(W^* = 1)}$$

where $P(H = 1, W = 1)$ is identified by the data and $\theta_j^+ = P(H = j, W = 1, W^* = 0)$ and $\theta_j^- = P(H = j, W = 0, W^* = 1)$ denote the unobserved fraction of false positive and false negative classifications of WIC recipients, respectively, for infants realizing health outcome $j = 1, 0$. In the numerator, $\theta_j^- - \theta_j^+$ reflects the unobserved excess of false negative versus false positive
classifications among those with a favorable birth outcome. The quantity $P(H = 1 \mid W^* = 0)$ can be decomposed analogously. Clearly, without assumptions restricting the nature or degree of classification errors, the data are uninformative.

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 $P = P(W = 1)$. 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, following KPGJ, the auxiliary information identifies the difference in false negative and positive reporting rates:

$$\Delta = P^* - P = (\theta^*_1 + \theta^*_0) - (\theta^*_1 - \theta^*_0).$$

(5)

In our application, $\Delta$ 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 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 consider two additional restrictions on the classification error problem:

**ME1**: Maximum error rate: $P(Z^* = 0) \leq Q_u$ (6a)

**ME2**: Verification: $V = 1 \Rightarrow W^* = W$ (6b)

where $Q_u$ is a known upper bound on the degree of data corruption in the spirit of Horowitz and Manski (1995) and $V$ is an indicator of whether a self-report of WIC participation is treated as accurate.

For Assumption ME1 considered in KPGJ, the value of $Q_u$ must logically lie within the range $[|\Delta|, 1]$. In the polar case that $Q_u$ equals 1, ME1 is uninformative. When no responses are verified to be accurate, we refer to the case $Q_u = 1$ 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 in our application. Validation data suggest very few instances of households falsely claiming to receive food assistance (e.g., Bollinger and David 1997). Moreover, using a parametric measurement error model formalized by Hausman et al. (1998), we estimate that less than one percent of the income-eligible WIC population incorrectly reports receiving benefits: $\theta^+ < 0.01$. 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.

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. Using Hausman et al.’s (1998) model, the true participation rate is specified as $P^* = F(X^T \gamma)$, 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 $\theta^- = (\theta^- + \theta^+_v)$ is 0.159, and our estimate of $\theta^+ = (\theta^+ + \theta^-_v)$ is 0.008. Full estimation results are available from the authors. With verification, $P'$ is logically required to lie within the range $\Lambda = \max\{P - P(W = 1, V = 0), P(H = 1, W = 1) - \theta^-_{1,19}\}$.
These assumptions imply informative upper bounds on the false reporting probabilities $\theta_i^-, \theta_0^-, \theta_i^+$ and $\theta_0^+$. Specifically, it follows that

$$\theta_i^- \leq \min \left\{ P(H = i, W = 0, V = 0), P^*, \frac{1}{2}(Q_u + \Delta) \right\} \equiv \theta_i^{-UB}, \quad i = 1, 0 \quad (7a)$$

$$\theta_i^+ \leq \min \left\{ P(H = i, W = 1, V = 0), 1 - P^*, \frac{1}{2}(Q_u - \Delta) \right\} \equiv \theta_i^{+UB}, \quad i = 1, 0 \quad (7b)$$

For example, the fraction of false negative responses among households with a particular birth outcome cannot exceed the fraction of all negative responses for that outcome. Nor can it exceed the fraction of households truly participating in WIC, $P^*$.

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

**Proposition 1** Given the classification error model restrictions in Equations (7a) and (7b), the outcome gap, $\beta$, under partial verification is bounded as follows:

$$\frac{P(H = 1, W = 1) - P^* P(H = 1) - \min \{\theta_i^{-UB}, \theta_0^{-UB} - \Delta\}}{P^*(1 - P^*)} \leq \beta \leq \frac{P(H = 1, W = 1) - P^* P(H = 1) + \min \{\theta_i^{+UB}, \theta_0^{+UB} + \Delta\}}{P^*(1 - P^*)}$$

for $P^* \in \Lambda \cap (0,1)$.\(^{18}\)

\(^{18}\) 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 $P^* \in [P, 1]$. Under no false positives (and no further verification), this range reduces to $P^* \in [P, 1]$. If the valid range $\Lambda$ 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.
These bounds simplify in an arbitrary errors model:

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

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

See the appendix for a proof of these results. These bounds apply under the exogenous selection assumption, and they will also be directly incorporated in parts of the subsequent analysis under endogenous selection with classification errors.

All of the probabilities in Proposition 1 and the corollary are consistently estimated using data from the ECLS-B except \(P^*\), the true WIC participation rate. 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 \(P^*\). The appendix also provides (wider) sharp 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 alternatives (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, \(P^* = P^* = 0.74\), along with Imbens-Manski (2004) confidence intervals that cover the true value of the ATE with 90% probability.

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 $\beta$ 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 (6b). A practitioner’s ability to formally verify some mixture of positive and negative participation responses can dramatically reduce the identification region, especially when the true participation rate is unknown.

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 $[-0.205, 0.584]$ under arbitrary errors to $[-0.188, 0.0592]$ under no false positives to $[-0.0759, 0.0592]$ under stronger verification. 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.

---

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 Sections 5 and 6 when we additionally allow for endogenous selection.
5. ENDOGENOUS SELECTION BOUNDS

In this section, we focus on what can be learned about the ATE when selection is endogenous for the special case of perfectly accurate reporting ($P^* = P$ with full verification). A natural starting point is to ask what can be learned in the absence of any assumptions invoked to address the selection problem (see Manski, 1995 and Pepper, 2000). Since the counterfactual latent probability $P[H(1) = 1 | W^* = 0]$ must lie within [0,1], it follows from Equation (2) that

$$P[H(1) = 1 | W^* = 1]P(W^* = 1) \leq P[H(1) = 1 | W^* = 1]P(W^* = 1) + P(W^* = 0).$$

Intuitively, the width of this bound on $P[H(1) = 1]$ equals the censoring probability, $P(W^* = 0)$.

To narrow the bounds, prior information must be brought to bear. While the exogenous selection assumption seems untenable, a number of middle ground assumptions restrict the relationship between WIC participation, birth outcomes, and observed covariates. We consider the identifying power of several monotonicity assumptions: two using instruments, one on treatment selection, and one on treatment response.

5.1. Monotone Instrumental Variables

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. For example, 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. 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.

To formalize this idea, let $v$ be the monotone instrumental variable such that

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

(8)
That is, the latent positive birth outcome probabilities weakly increase with income.

Let LB($u$) and UB($u$) be the known lower and upper bounds evaluated at $v = u$, respectively, given the available information. Then the MIV assumption formalized in Manski and Pepper (2000, Proposition 1) implies:

$$\sup_{u \geq u_1} LB(u) \leq P[H(j) = 1 | v = u] \leq \inf_{u \geq u_2} UB(u).$$

(9)

Bounds on the unconditional latent probability, $P[H(j) = 1]$, can then be obtained using the law of total probability.

We also use eligibility criteria to construct a monotone instrument by extending an approach in Gundersen et al. (2012) to account for particular institutional features of the eligibility rules of the WIC program. Similar to a regression discontinuity design, suppose the latent probability of a healthy birth outcome among WIC-eligible households is no better than among higher-income ineligible households. 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.

While latent birth outcomes are unlikely to be mean independent of eligibility status, an MIV-ineligibles assumption holding that mean response varies monotonically across these groups seems credible:

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

(10)

---

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.

21 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.
This ineligibility bound on $P[H(j) = 1]$ has the same structure as the income-MIV restriction in Equation (8), but it has the unique property that $W = 0$ among ineligible households. Assuming eligibility is observed, the data point-identify $P[H(0) = 1 | v = \text{ineligible}]$ as $P(H = 1 | v = \text{ineligible})$. This quantity then serves as an upper bound on the potential outcome $P[H(0) = 1]$ for our primary population of income-eligible households (Gundersen et al., 2012):

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

(11)

We learn nothing new about $P[H(1) = 1]$ using this approach because the outcome $H(1)$ is counterfactual within the eligibles population.

To account for the possibility that adjunctive WIC eligibility through other programs may be measured with error, we extend the MIV model in Equation (11). Let $B^* = 0$ indicate a true lack of prenatal benefits from Medicaid, TANF, or SNAP, and denote $B = 0$ its potentially mismeasured counterpart. There are two reasons $B^*$ might be mismeasured. First, households may misreport their participation in programs like Medicaid just as they may misreport their participation in WIC. Second, as described earlier, respondents in the ECLS-B are asked about prenatal participation in Medicaid but are not specifically asked about prenatal participation in SNAP and TANF.22

In Appendix A.2, we generalize Equation (11) given this potential measurement problem:

$$P[H(0) = 1] \leq \frac{P(H = 1 | v = \text{ineligible})}{1 - \theta P(H = 0 | v = \text{ineligible})}$$

(11')

where $\theta = P(B^* = 1 | H = 0, v = \text{ineligible})$ is the unknown fraction of households in the unfavorable health outcome subset of the apparent eligibles that did, in fact, receive benefits from a WIC-

---

22 Given its widely varying income thresholds across states, Medicaid is by far the most important program for conferring WIC eligibility to otherwise ineligible households. Jacknowitz and Tiehen (2010) point out that few households in the ECLS-B otherwise ineligible for WIC would adjuntively qualify through SNAP or TANF which tend to have stricter eligibility criteria. Most WIC eligibility comes through meeting its 185% income threshold or receiving Medicaid benefits.
related program. This fraction should be close to 0 given that the subsample includes only apparently income-ineligible households that reported no public assistance from any WIC-related program, either prenatal or postpartum. If $\theta = 0$ such that $B^*$ is never mismeasured, the upper bound on $P[H(0) = 1]$ is the mean health status in the group of ineligibles, which equals 0.831 for the normal birth weight outcome and 0.809 for the normal gestation duration outcome. In the polar opposite case that $\theta = 1$ such that $B^*$ is always mismeasured, the upper bound is 1 so we cannot learn anything. Below, we assess the sensitivity of our results to departures from Gundersen et al.’s (2012) implicit assumption that $\theta = 0$.

5.2 Monotone Treatment Selection

Information on the selection process by which expectant mothers decide to enroll in WIC can also be used to narrow the bounds. Although unobserved factors associated with the enrollment decision might be positively or negatively associated with birth outcomes, the consensus in the literature is that negative selection dominates (Ludwig and Miller, 2005, p.696; also see Currie, 2003). That is, the latent probability of a favorable birth outcome is no higher for households that self-selected into WIC participation. We know, in fact, that observed factors associated with poor health are positively associated with the decision to take up the program; WIC recipients are known to have unfavorable demographic, socioeconomic, and health characteristics (e.g., Bitler and Currie, 2005; Gundersen, 2005).

To formalize this idea about the selection process, we treat participation status as a monotone instrument such that the latent probability of a positive health outcome is weakly lower for

---

23 On the one hand, relatively motivated and informed women may be 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 – the so called gestational age bias (see Section 2.2 for further details) (Joyce, Gibson and Colman, 2005; Joyce, Racine, and Yunzal-Butler, 2008). On the other hand, women who are most likely to have poor birth outcomes and who are in the precarious socio-economic situations may be the most likely to enroll in WIC.
households that self-selected into WIC participation. In particular, we assume Monotone Treatment Selection (MTS) (Manski and Pepper, 2000) such that:

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

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, \( W^* = 0 \).

### 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 generalized version of the 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:

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

While this MTR assumption precludes a strictly negative average treatment effect, it provides no information on the magnitude of the ATE and does not rule out a value of 0. 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.

### 5.4. Results for the No Errors Case

For each of the outcomes, Table 2 displays the estimated bounds and confidence intervals for the ATE for the no errors case. In row (i), we report the point estimates under the exogenous selection assumption. In row (ii), we make no assumptions about how eligible households select into the WIC program. Consistent with Manski (1995), the width of these “worst case” selection bounds always equals 1, and the bounds always include zero. For example, the ATE bounds are
[-0.358, 0.642] for normal birth weight and [-0.408, 0.592] for normal gestation age. These bounds highlight a researcher’s inability to make strong inferences about the efficacy of WIC without making assumptions that address the problem of unknown counterfactuals. In the absence of restrictions that address the selection problem, we cannot rule out the possibility that WIC has a large positive or negative impact on the likelihood of positive birth outcomes.

These bounds, however, are narrowed substantially under common monotonicity assumptions on treatment selection (MTS) and relationships between the latent outcome and observed instrumental variables (MIV). In row (iii), the MTS assumption alone identifies the sign of the impact of WIC on the probability of normal birth weight. In particular, we estimate that WIC increases the probability of normal birth weight by at least 1.95 percentage points and by as much as 64.2 percentage points. We find similar results for the other birth weight outcomes. For the gestation outcomes, the MTS assumption notably increases the estimated lower bound but does not identify the sign of the effect of WIC. For example, the estimated lower bound on the probability of being in the normal gestation range improves from -0.408 to -0.010, but the bounds still include zero.

These bounds narrow further when we combine the MTS assumption with the two MIV assumptions that the latent probability of a favorable birth outcome weakly increases with family resources (row iv) and is no worse among ineligible households (row v). In fact, 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). For example, the estimates imply that WIC increases the probability of a normal birth weight by at least 6.54 points, and the confidence interval does not include zero. In this no measurement error setting, layering on the MIV eligibles assumption (row v) has no additional identifying power. As discussed below, however, the eligibles MIV assumption has substantial identifying power after accounting for underreported WIC participation.
The last row in Table 2 adds the MTR assumption. In most cases, the MTR assumption does not notably improve the estimated lower bounds on the ATE under no classification errors relative to the estimates derived under the MTS-MIV assumptions alone. Like the ineligibles MIV assumption, however, the MTR assumption has substantial identifying power when WIC participation is underreported.

Overall, the no-errors results indicate that prenatal WIC may lead to substantial increases in the favorable birth outcomes and, at worst, have slightly deleterious effects. These results are qualitatively similar to the findings from the earlier literature which assumed selection is exogenous. This literature finds that WIC leads to substantial improvements in mean birth outcomes (e.g., between a 10 and 40 percent reduction in the probability of low birth weight – see Currie, 2003).

6. INFERENCE WITH MEASUREMENT AND SELECTION PROBLEMS

In this section, we study what can be learned about the effects of the program under these monotonicity assumptions when WIC participation may be measured with error. The selection bounds estimated in Section 5 presume that everyone reports WIC participation accurately. With reporting errors, however, there is uncertainty not only about counterfactuals but also about the reliability of the data on WIC participation, \( W^* \).

6.1. Models

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

\[
\begin{align*}
& -[P(H = 1, W = 0) + P(H = 0, W = 1)] + \Theta \\
& \leq ATE(1,0) \leq \\
& [P(H = 1, W = 1) + P(H = 0, W = 0)] + \Theta.
\end{align*}
\]  \( (14) \)
where $\Theta = (\theta_1^+ + \theta_0^+) - (\theta_1^- + \theta_0^-)$. In the absence of classification errors, $\Theta = 0$, and Equation (14) simplifies to Manski’s (1995) worst-case selection bounds. With classification errors, the models from Section 4 – see Equations (5), (6) and (7) – allow us to place informative restrictions on $\Theta$.

Extending KPGJ’s Proposition 1 to incorporate partial verification, we have:

**Proposition 2:** Given the bounds on the ATE in (3) (see KPGJ), combined with restrictions ME1, ME2, and (11) on classification errors, endogenous selection bounds on the ATE under partial identification are given by:

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

A proof of this proposition is provided in the appendix. Parallel to Proposition 1, the appendix also provides sharp bounds for the case that the researcher has no auxiliary information about $P^*$.

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.

### 6.2. Results for Measurement Error and Endogenous Selection

We present estimated bounds on average treatment effects for the case that WIC participation is endogenous and underreported in three layers. First, 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. Second, Figure 3 displays estimates under the MIV, MTS and MTR assumptions using the arbitrary and no-false positive errors models. Focusing on the normal birth weight outcome, the assumptions addressing the selection problem substantially narrow the bounds but in many cases do not reveal sign of the ATE.

Finally, Table 3 displays estimates using the ME2 assumption that self-reports of WIC participation from households reporting any type of government benefits are assumed to be accurate. This partial verification model, where 84% of respondents are known to provide accurate reports of WIC receipt, has 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.

Importantly, this finding highlights the methodological contribution of the verification assumption ME2 and the resulting Proposition 2. Our methods allow us to sign the ATE without imposing 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. The methods in KPGJ do not afford this flexibility to verify mixtures of positive and negative participation responses.

6.2.1 Worst-Case Bounds

Figure 2 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 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.

6.2.2 Bounds Under MTS, MIV, and MTR

Figure 3, along with the corresponding table, displays estimated lower bounds on the impact of WIC for the normal birth weight outcome under various restrictions on the selection process. Here, we consider identification under arbitrary errors or no false positives. We focus on estimates of the lower bound 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 (only lower bounds are shown in the figure to save space). 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 \( P = 0.692 \). As seen in the figure, 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%. 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.

Finally, adding 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 \( P^* = 0.692 \) by at least 3.86 percentage points, or 21 percent. 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 \( P^* \).

### 6.2.3 Bounds Under Partial Verification Models with MTS, MIV, and MTR

Finally, we focus on the identifying power of the stronger assumption that allows for the verification of some negative WIC participation responses (also see Figure 2). In particular, we treat

\[ \frac{\{1 - P[H(1) = 1] - P[H(0) = 1]\}}{P[H(0) = 1]} \]

24 The lower bound percentage reduction in unfavorable birth outcomes under the program is
responses about WIC receipt as reliable if the household reported any type of government benefits (with unknown reliability if the household claims no benefits from any source). Recall that 84% of households are verified to provide accurate reports in this 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. We continue to estimate a strictly positive average treatment effect even when allowing for errors in measuring adjunctive eligibility (see Section 5.1). 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.

Finally, 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.
7. CONCLUSION

Driven in part by concerns over discretionary government spending and in part by the growing perception that early life conditions have long term impacts on adult life outcomes (Heckman and Carneiro, 2003; and Almond and Currie, 2011), there has been a renewed interest in understanding the impact of WIC on infant health. Despite the long-standing interest in understanding the impact of WIC on health outcomes, researchers have struggled to draw credible inferences.

This paper considers what can be learned about the efficacy of prenatal WIC in improving birth outcomes when program participation is underreported and selection may be endogenous. While our framework does not allow us to point-identify average treatment effects, we derived sharp bounds under arbitrary misclassification of the treatment indicator.

Overall, these estimated partial identification bounds illustrate how inferences are sensitive to assumptions. Under the weakest models allowing for both 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 appears to improve infant health across a spectrum of outcomes. When combining all of our identifying assumptions, we find that WIC has a substantial impact on birth outcomes, increasing the prevalence 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.
References


APPENDICES

A1. Proof of Exogenous Selection Bounds with Classification Errors

**Proposition 1**

We can write the average treatment effect as

\[
ATE = \frac{P(H = 1, W = 1) - P(W = 1)P(H = 1) + \theta^- - \theta^+}{P(W^* = 1)P(W^* = 0)}.
\]

(A1)

Thus, subject to restrictions on the unknown classification error rates in Equations (5)-(7) and the laws of probability, \(^{25}\) the upper bound is found by maximizing \(\theta^-\) and minimizing \(\theta^+\). Likewise, the lower bound is found by minimizing \(\theta^-\) and maximizing \(\theta^+\). Let \(\Psi = \theta^- - \theta^+\).

If \(\Delta \geq 0\) (net false negative reporting): For the *upper bound*, first consider the case that \(\theta_i^{UB} < \Delta\). Here, \(\theta^-\) cannot exceed \(\theta_i^{UB}\) and \(\theta^+\) cannot fall below 0 so that \(\Psi \leq \theta_i^{UB}\). At this upper bound, Equation (5) implies that \(\theta_0^- = \Delta - \theta_i^{UB}\) and \(\theta_0^+ = 0\). If \(\theta_i^{UB} \geq \Delta, \theta_i^-\) cannot exceed \(\min \{\theta_i^{UB}, \theta_0^{UB} + \Delta\}\) and \(\theta_i^+\) cannot fall below 0 so that \(\Psi \leq \min \{\theta_i^{UB}, \theta_0^{UB} + \Delta\}\). At this upper bound, Equation (5) implies \(\theta_0^- = 0\) and \(\theta_0^+ = \min \{\theta_0^{UB}, \theta_i^{UB} - \Delta\}\). For the *lower bound*, first consider the case that \(\theta_0^{UB} < \Delta\). Here, \(\theta^-\) cannot exceed \(\theta_0^{UB}\) so from Equation (5) we know that \(\theta^-\) must be no less than \(\theta_0^{UB} - \Delta\). From Equation (7), we know that \(\theta^+\) 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 \(\theta^+\) above 0. Thus, we have \(\Psi \geq \Delta - \theta_0^{UB}\). If \(\theta_0^{UB} \geq \Delta, \Psi\) is minimized when \(\theta_0^- = \min \{\theta_0^{UB}, \theta_i^{UB} + \Delta\}, \theta_i^- = 0, \theta_i^+ = \min \{\theta_i^{UB}, \theta_0^{UB} - \Delta\}\), and \(\theta_0^+ = 0\) so that \(\Psi \geq -\min \{\theta_i^{UB}, \theta_0^{UB} - \Delta\}\).

\(^{25}\) In particular, \(P(H = 1 | W^* = 1)\) and \(P(H = 1 | W^* = 0)\) lie within \([0,1]\). A more detailed proof showing these restrictions is available from the authors.
If $\Delta \leq 0$ (net false positive reporting): For the upper bound, first consider the case that $\theta_0^{UB} < -\Delta$. Here, $\theta_0^+$ cannot exceed $\theta_0^{UB}$ so from Equation (5) we know that $\theta_1^+$ must be no less than $-\Delta - \theta_0^{UB}$. From Equation (7), we know that $\theta_1^-$ 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 $\theta_1^-$ above 0. Thus, we have

$$\Psi \leq -\Delta - \theta_0^{UB}.$$  
If $\theta_0^{UB} \geq -\Delta$, $\Psi$ is maximized when $\theta_0^+ = \min\{\theta_0^{UB}, \theta_0^{UB} - \Delta\}$, $\theta_1^+ = 0$, $\theta_1^- = \min\{\theta_1^{UB}, \theta_0^{UB} + \Delta\}$, 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, $\theta_1^+$ cannot exceed $\theta_1^{UB}$ and $\theta_1^-$ cannot fall below 0 so that $\Psi \leq -\theta_1^{UB}$. At this lower bound, Equation (5) implies that $\theta_0^+ = -\Delta - \theta_1^{UB}$ and $\theta_0^- = 0$. If $\theta_1^{UB} \geq -\Delta$, $\theta_1^+$ cannot exceed $\min\{\theta_1^{UB}, \theta_0^{UB} - \Delta\}$ and $\theta_1^-$ cannot fall below 0 so that $\Psi \geq -\min\{\theta_1^{UB}, \theta_0^{UB} - \Delta\}$.

At this lower bound, Equation (5) implies $\theta_0^+ = 0$ and $\theta_0^- = \min\{\theta_0^{UB}, \theta_1^{UB} + \Delta\}$. Combining these results, it follows that $-\min\{\theta_1^{UB}, \theta_0^{UB} - \Delta\} \leq \Psi \leq \min\{\theta_1^{UB}, \theta_0^{UB} + \Delta\}$.  

Corollary 1

Under the arbitrary errors model, $Q_a = 1$. From Equation (7), 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

$$\min\{\theta_1^{UB}, \theta_0^{UB} + \Delta\} = \min\{P(W^* = 1) - P(H = 1,W = 1), P(H = 1,W = 0)\}.  \quad \text{(A2)}$$

Likewise, from Equation (7) we know that

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

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

$$\min\{\theta_1^{UB}, \theta_0^{UB} - \Delta\} = \min\{P(W^* = 0) - P(H = 1,W = 0), P(H = 1,W = 1)\}.  \quad \text{(A3)}$$

36
Substituting (A2) into Proposition 1 obtains an upper bound of \( \frac{P(H = 0)}{1 - P^*} \) if \( P^* \leq P(H = 1) \) and \( \frac{P(H = 1)}{P^*} \) if \( P^* \geq P(H = 1) \), which reduces to \( \min \left\{ \frac{P(H = 0)}{1 - P^*}, \frac{P(H = 1)}{P^*} \right\} \). Similarly, the lower bound is obtained using Equation (A3).

**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 \( \beta \) are given as follows:

\[
\beta^{LB} = \inf_{\theta^*_I \in [0, \theta^*_{UB}], \theta^*_0 \in [0, \min\{Q, \theta^*_I, \theta^*_{LB}\}]} \left[ \frac{P(H = 1, W = 1) - \theta^*_I}{P(W = 1) - \theta^*_I + \theta^*_0} \right] = \frac{P(H = 1, W = 0) + \theta^*_I}{P(W = 0) + \theta^*_I - \theta^*_0} \\
\beta^{UB} = \sup_{\theta^*_I \in [0, \theta^*_{UB}], \theta^*_0 \in [0, \min\{Q, \theta^*_I, \theta^*_{LB}\}]} \left[ \frac{P(H = 1, W = 1) + \theta^*_I}{P(W = 1) + \theta^*_I - \theta^*_0} \right] = \frac{P(H = 1, W = 0) - \theta^*_I}{P(W = 0) - \theta^*_I + \theta^*_0}.
\]

These bounds are obtained by minimizing or maximizing \( P(H = 1 | W^* = 1) - P(H = 1 | W^* = 0) \) using Equation (4) and its analogous counterpart for \( P(H = 1 | W^* = 0) \). They can be estimated by conducting separate grid searches over \( \{\theta^*_I, \theta^*_0\} \) and \( \{\theta^*_I, \theta^*_0\} \) in the feasible regions to minimize \( \beta^{LB} \) and maximize \( \beta^{UB} \) 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:

\[
ATE^{LB} = \inf_{\theta^*_0 \in [0, \theta^*_{UB}]} \left[ \frac{P(H = 1, W = 1)}{P(W = 1) + \theta^*_0} \right] = \frac{P(H = 1, W = 0)}{P(W = 0) - \theta^*_0} \\
ATE^{UB} = \sup_{\theta^*_0 \in [0, \theta^*_{UB}]} \left[ \frac{P(H = 1, W = 1) + \theta^*_0}{P(W = 1) + \theta^*_0} \right] = \frac{P(H = 1, W = 0) - \theta^*_0}{P(W = 0) - \theta^*_I}.
\]
A2. MIV Ineligible Bounds With Mislabeled Ineligibility

Since ineligible households did not receive prenatal WIC benefits \((inc > 185\% \text{ and } B^* = 0 \text{ implies } W^* = 0)\),\(^{26}\) it follows that \(P[H(0) = 1] \text{ cannot exceed } P(H = 1 | B^* = 0, v = \text{ineligible})\), where \(v\) denotes selecting the comparison group based on \(inc > 185\%\) (income-ineligible), \(SES \leq 3\), and \(B = 0\).

This conditional probability can be written as

\[
\frac{P(H = 1, B^* = 0|v)}{P(B^* = 0|v)} = \frac{P(H = 1, B = 0 | v) + \tilde{\theta}_i^- - \tilde{\theta}_i^-}{P(B = 0 | v) + (\tilde{\theta}_i^- + \tilde{\theta}_0^-) - (\tilde{\theta}_i^- + \tilde{\theta}_0^-)}
\]

where \(\tilde{\theta}_j^+ \equiv P(H = j, B = 1, B^* = 0 | v)\) and \(\tilde{\theta}_j^- \equiv P(H = j, B = 0, B^* = 1 | v)\) 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

\[
\frac{P(H = 1 | I) - \tilde{\theta}_i^-}{1 - \tilde{\theta}_i^- - \tilde{\theta}_0^-}.
\]

The worst-case upper bound is obtained by setting \(\tilde{\theta}_i^- = 0\):

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

where, as described in the main text, \(\theta \equiv P(B^* = 1 | H = 0, v)\) 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.

---

\(^{26}\) 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.
A3. Proof of the Bounds with Selection and Classification Errors

Proposition 2

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 Equations (5)-(7). The upper bound is found by maximizing \((\theta_i^- + \theta_0^+)^-\) and minimizing \((\theta_0^- + \theta_i^+)^-\), and vice versa for the lower bound.

If \(\Delta \geq 0\): For the upper bound, first consider the case that \(\theta_i^- - \Delta \geq 0\). Then \((\theta_0^+ + \theta_i^+\)) is minimized at 0 and Equation (5) simplifies to \(\theta_i^- = \Delta + \theta_0^+\). It follows that \(\theta_0^+\) cannot exceed \(\min\{\theta_i^{UB}, \theta_i^- - \Delta\}\) and \(\theta_i^+\) cannot exceed \(\min\{\Delta + \theta_i^{UB}, \theta_0^+\}\). The upper bound follows directly.

Second, consider the case that \(\theta_i^- < \Delta\). We know that \(\theta_i^-\) cannot exceed \(\theta_i^{UB}\) and, to satisfy the restriction in Equation (5), \(\theta_0^+\) must be no less than \(\Delta - \theta_i^{UB}\). As before, \(\theta_i^+\) is minimized at 0. While \(\theta_0^+\) 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 \(\theta_0^+\) above 0. Thus, we have the upper bound on \(\Theta\) of \(2\theta_i^{UB} - \Delta\) 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 \((\theta_0^- + \theta_i^+\)) is minimized at 0 and Equation (5) simplifies to \(\theta_i^- = \Delta + \theta_0^-\). It follows that \(\theta_0^-\) cannot exceed \(\min\{\theta_i^{UB}, \theta_0^{UB} - \Delta\}\) and \(\theta_i^+\) cannot exceed \(\min\{\Delta + \theta_i^{UB}, \theta_0^{UB}\}\) so that \(\max\{-2\theta_i^{UB} - \Delta, -2\theta_0^{UB} + \Delta\}\) provides the lower bound on \(\Theta\). Second, consider the case that \(\theta_0^{UB} < \Delta\). We know that \(\theta_0^-\) cannot exceed \(\theta_0^{UB}\) and, to satisfy the restriction in Equation (5), \(\theta_0^+\) must be no less than \(\Delta - \theta_0^{UB}\). As before, \(\theta_0^+\) is minimized at 0. While \(\theta_0^+\) 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 lower bound would be unchanged by increasing \(\theta_0^+\) above 0. Thus, we have the lower bound on \(\Theta\) of \(-2\theta_0^{UB} + \Delta\) which can be shown to be no smaller than \(-2\theta_i^{UB} - \Delta\).

If \(\Delta < 0\): For the upper bound, first consider the case that \(\theta_0^{UB} \geq -\Delta\). Then \((\theta_0^+ + \theta_i^-)\) is minimized at 0 and Equation (5) simplifies to \(\theta_0^- = -\Delta + \theta_i^-\). We know that \(\theta_i^-\) cannot exceed \(\min\{\theta_i^{UB}, \theta_0^{UB} + \Delta\}\) and \(\theta_0^+\) cannot exceed \(\min\{\theta_0^{UB}, -\Delta + \theta_i^{UB}\}\). The upper bound follows directly.
Second, consider the case that $\theta_0^{\text{UB}} < -\Delta$. We know that $\theta_0^+$ cannot exceed $\theta_0^{\text{UB}}$ and, to satisfy the restriction in Equation (8), $\theta_0^+$ must be no less than $-\Delta - \theta_0^{\text{UB}}$. As before, $\theta_0^-$ is minimized at 0. While $\theta_0^+$ 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 upper bound would be unchanged by increasing $\theta_0^+$ above 0. Thus, we have the upper bound on $\Theta$ of $2\theta_0^{\text{UB}} + \Delta$ which can be shown to be no greater than $2\theta_1^{\text{UB}} - \Delta$.

For the lower bound, first consider the case that $\theta_1^{\text{UB}} \geq -\Delta$. Then $(\theta_1^+ + \theta_0^-)$ is minimized at 0 and Equation (5) simplifies to $\theta_0^+ = -\Delta + \theta_0^-$. We know that $\theta_0^-$ cannot exceed $\theta_0^- + \theta_1^{\text{UB}}$ and $\theta_0^+$ cannot exceed $\min\{\theta_1^{\text{UB}}, -\Delta + \theta_0^{\text{UB}}\}$ so that $-2\theta_1^{\text{UB}} - \Delta, -2\theta_0^{\text{UB}} + \Delta$ provides the lower bound on $\Theta$. Second, consider the case that $\theta_1^{\text{UB}} < -\Delta$. We know that $\theta_1^+$ cannot exceed $\theta_1^{\text{UB}}$ and, to satisfy the restriction in Equation (5), $\theta_0^+$ must be no less than $-\Delta - \theta_1^{\text{UB}}$. As before, $\theta_1^-$ is minimized at 0. While $\theta_0^-$ 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 $\theta_1^-$ above 0. Thus, we have the lower bound of $-2\theta_1^{\text{UB}} - \Delta$ which can be shown to be no smaller than $2\theta_0^{\text{UB}} - \Delta$.

\section*{Endogenous Selection Bounds When $P^*$ Is Not Known:} 27

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: 28

\begin{itemize}
  \item $ATE^{LB} = P(H = 1, W = 1) - P(H = 1, W = 0) - P(W = 1) - \min\{Q_a, \theta_1^{\text{UB}} + \theta_0^{\text{UB}}\}$
  \item $ATE^{UB} = P(H = 1, W = 1) - P(H = 1, W = 0) + P(W = 0) + \min\{Q_a, \theta_1^{\text{UB}} + \theta_0^{\text{UB}}\}$.
\end{itemize}

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{itemize}
  \item $ATE^{LB} = P(H = 1, W = 1) - P(H = 1, W = 0) - P(W = 1) - \theta_0^{\text{UB}}$
  \item $ATE^{UB} = P(H = 1, W = 1) - P(H = 1, W = 0) + P(W = 0) + \theta_1^{\text{UB}}$.
\end{itemize}

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

28 The ATE is given by $P(H = 1) - P(H = 0) = P(H = 1 | W = 1)P(W = 1) + P(H = 1 | W = 0)P(W = 0) - P(H = 1 | W = 1)P(W = 1) - P(H = 1 | W = 0)P(W = 0)$. Letting $P(H = 1 | W = 1)$ vary within [0, 1], the ATE is no smaller than $P(H = 1, W = 1) - P(W = 1) - P(H = 1, W = 0) = P(H = 1, W = 1) + \theta_1^+ + \theta_0^- - \theta_1^- - \theta_0^+$ and $P(H = 1, W = 1) + \theta_1^+ + \theta_0^- - \theta_1^- - \theta_0^+$ for $\theta_1^- = \theta_1^+ = 0$ and recognizing that the sum of errors cannot exceed $Q_a$. The upper bound is obtained by setting $\theta_0^+ = \theta_1^- = 0$ and recognizing that the sum of errors cannot exceed $Q_a$. The upper bound is derived analogously.
Table 1. Reported Prenatal WIC Participation and Birth Outcomes

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

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

Exogenous selection

<table>
<thead>
<tr>
<th>Case Description</th>
<th>Self-reported participation rate: $P^* = P = 0.692$</th>
<th>Administrative participation rate: $P^* = P^o = 0.74$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Arbitrary errors</td>
<td>p.e. $\beta = 0.0195$ CI $\beta = [-0.230, 0.523]$</td>
<td>p.e. $\beta = 0.0195$ CI $\beta = [-0.215, 0.622]$</td>
</tr>
<tr>
<td>(b) No false positives</td>
<td>p.e. $\beta = 0.0195$ CI $\beta = [-0.022, 0.0409]$</td>
<td>p.e. $\beta = 0.0195$ CI $\beta = [-0.0215, 0.0784]$</td>
</tr>
<tr>
<td>(c) Verified if reported any gov’t benefits</td>
<td>p.e. $\beta = 0.0195$ CI $\beta = [-0.022, 0.0409]$</td>
<td>p.e. $\beta = 0.0195$ CI $\beta = [-0.0912, 0.0784]$</td>
</tr>
</tbody>
</table>

$^\dagger$ Point estimates of the population bounds

$^\ddagger$ 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

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

Notes:

a. Bias-corrected point estimates of the bounds
b. 90% Imbens-Manski confidence internals (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 ineligibles monotone instrument
g. MTR denotes Monotone Treatment Response
Table 2B. Sharp Bounds on the ATE of WIC Under No Classification Error, Gestation Age

<table>
<thead>
<tr>
<th>Gestation Age</th>
<th>Normal (38-42 weeks)</th>
<th>Not Premature (≥ 37 weeks)</th>
<th>Not Very Premature (≥ 33 weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) ETS&lt;sup&gt;c&lt;/sup&gt;</td>
<td>p.e.</td>
<td>-0.0100, -0.0100</td>
<td>-0.0188, -0.0188</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>-0.0378, 0.0184</td>
<td>-0.0364, -0.0006</td>
</tr>
<tr>
<td>(ii) Worst Case</td>
<td>p.e.&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.408, 0.592</td>
<td>-0.362, 0.638</td>
</tr>
<tr>
<td></td>
<td>CI&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.419, 0.604</td>
<td>-0.373, 0.650</td>
</tr>
<tr>
<td>(iii) MTS&lt;sup&gt;d&lt;/sup&gt;</td>
<td>p.e.</td>
<td>-0.0100, 0.592</td>
<td>-0.0036, 0.677</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>-0.0321, 0.604</td>
<td>-0.0091, 0.689</td>
</tr>
<tr>
<td>(iv) MTS+MIV&lt;sup&gt;e&lt;/sup&gt;</td>
<td>p.e.</td>
<td>0.0899, 0.576</td>
<td>0.0053, 0.624</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>0.0119, 0.583</td>
<td>-0.0012, 0.634</td>
</tr>
<tr>
<td>(v) MTS+MIV +ineligibles&lt;sup&gt;f&lt;/sup&gt;</td>
<td>p.e.</td>
<td>0.0899, 0.576</td>
<td>0.0053, 0.624</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>0.0119, 0.583</td>
<td>-0.0107, 0.634</td>
</tr>
<tr>
<td>(vi) MTS-MIV+ineligibles +MTR&lt;sup&gt;g&lt;/sup&gt;</td>
<td>p.e.</td>
<td>0.0951, 0.576</td>
<td>0.0083, 0.624</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>0.0285, 0.583</td>
<td>0.0000, 0.634</td>
</tr>
</tbody>
</table>

Notes:

a. Bias-corrected point estimates of the bounds
b. 90% Imbens-Manski confidence internals (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 ineligibles monotone instrument
g. MTR denotes Monotone Treatment Response
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:** $P^o = P^o = 0.74$

### Endogenous selection

<table>
<thead>
<tr>
<th>(a) Arbitrary errors</th>
<th>p.e.</th>
<th>LB [-0.459, 0.844]</th>
<th>UB</th>
<th>width</th>
<th>[1.304]</th>
<th>p.e.</th>
<th>LB [-0.412, 0.892]</th>
<th>UB</th>
<th>width</th>
<th>[1.304]</th>
</tr>
</thead>
<tbody>
<tr>
<td>p.e.</td>
<td>CI [-0.473, 0.858]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p.e.</td>
<td>CI [-0.425, 0.905]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) No false positives</th>
<th>p.e.</th>
<th>LB [-0.358, 0.642]</th>
<th>UB</th>
<th>width</th>
<th>[1.000]</th>
<th>p.e.</th>
<th>LB [-0.405, 0.690]</th>
<th>UB</th>
<th>width</th>
<th>[1.095]</th>
</tr>
</thead>
<tbody>
<tr>
<td>p.e.</td>
<td>CI [-0.369, 0.654]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p.e.</td>
<td>CI [-0.417, 0.701]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(c) Verified if reported any govt’ benefits</th>
<th>p.e.</th>
<th>LB [-0.358, 0.642]</th>
<th>UB</th>
<th>width</th>
<th>[1.000]</th>
<th>p.e.</th>
<th>LB [-0.362, 0.690]</th>
<th>UB</th>
<th>width</th>
<th>[1.052]</th>
</tr>
</thead>
<tbody>
<tr>
<td>p.e.</td>
<td>CI [-0.369, 0.654]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p.e.</td>
<td>CI [-0.375, 0.701]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Point estimates of the population bounds
2. 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, Income MIV, and Ineligibles MIV

\[ \beta = 0.0195 \]

<table>
<thead>
<tr>
<th>MTS+ inc MIV + ineligibles:</th>
<th>width</th>
<th>width</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Arbitrary errors</td>
<td>p.e.</td>
<td>CI</td>
</tr>
<tr>
<td>p.e. ([-0.0328, 0.836])</td>
<td>0.869</td>
<td>([-0.0393, 0.846])</td>
</tr>
<tr>
<td>CI ([-0.0221, 0.886])</td>
<td>0.908</td>
<td>([-0.0285, 0.893])</td>
</tr>
<tr>
<td>(b) No false positives</td>
<td>p.e.</td>
<td>CI</td>
</tr>
<tr>
<td>p.e. ([-0.0221, 0.886])</td>
<td>0.908</td>
<td>([-0.0285, 0.893])</td>
</tr>
<tr>
<td>CI ([-0.0221, 0.886])</td>
<td>0.908</td>
<td>([-0.0285, 0.893])</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MTS+ inc MIV+ineligibles+MTR:</th>
<th>width</th>
<th>width</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Arbitrary errors</td>
<td>p.e.</td>
<td>CI</td>
</tr>
<tr>
<td>p.e. ([-0.0386, 0.836])</td>
<td>0.798</td>
<td>([-0.0386, 0.836])</td>
</tr>
<tr>
<td>CI ([-0.0386, 0.836])</td>
<td>0.798</td>
<td>([-0.0386, 0.836])</td>
</tr>
<tr>
<td>(b) No false positives</td>
<td>p.e.</td>
<td>CI</td>
</tr>
<tr>
<td>p.e. ([-0.0386, 0.836])</td>
<td>0.798</td>
<td>([-0.0386, 0.836])</td>
</tr>
<tr>
<td>CI ([-0.0386, 0.836])</td>
<td>0.798</td>
<td>([-0.0386, 0.836])</td>
</tr>
</tbody>
</table>

† Point estimates of the population bounds corrected for finite sample bias
‡ Imbens-Manski 5th and 95th percentile bounds (1,000 pseudosamples)
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

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

Notes:
- Strictly positive average treatment effects in **bold**
- a. Bias-corrected point estimates of the bounds
- b. 90% Imbens-Manski confidence internals (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

<table>
<thead>
<tr>
<th>Gestation Age</th>
<th>Normal (38-42 weeks)</th>
<th>Not Premature (≥ 37 weeks)</th>
<th>Not Very Premature (≥ 33 weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) ETS&lt;sup&gt;c&lt;/sup&gt;</td>
<td>p.e. [ -0.112, 0.0518 ]</td>
<td>CI [ -0.134, 0.0768 ]</td>
<td>[ -0.0552, 0.0090 ]</td>
</tr>
<tr>
<td>(ii) Worst Case</td>
<td>p.e.&lt;sup&gt;a&lt;/sup&gt; [ -0.422, 0.641 ]</td>
<td>CI&lt;sup&gt;b&lt;/sup&gt; [ -0.435, 0.653 ]</td>
<td>[ -0.338, 0.687 ]</td>
</tr>
<tr>
<td>(iii) MTS&lt;sup&gt;d&lt;/sup&gt;</td>
<td>p.e. [ -0.112, 0.641 ]</td>
<td>CI [ -0.134, 0.653 ]</td>
<td>[ -0.0552, 0.687 ]</td>
</tr>
<tr>
<td>(iv) MTS+MIV&lt;sup&gt;e&lt;/sup&gt;</td>
<td>p.e. [ -0.0563, 0.627 ]</td>
<td>CI [ -0.0876, 0.632 ]</td>
<td>[ -0.0303, 0.675 ]</td>
</tr>
<tr>
<td>(v) MTS+MIV&lt;sup&gt;e&lt;/sup&gt;</td>
<td>p.e. [ -0.0563, 0.627 ]</td>
<td>CI [ -0.0702, 0.632 ]</td>
<td>[ -0.0303, 0.675 ]</td>
</tr>
<tr>
<td>(vi) MTS-MIV&lt;sup&gt;e&lt;/sup&gt; +ineligibles&lt;sup&gt;f&lt;/sup&gt;</td>
<td>p.e. [ <strong>0.0269</strong>, 0.627 ]</td>
<td>CI [ 0.0133, 0.632 ]</td>
<td>[ <strong>0.0041</strong>, 0.675 ]</td>
</tr>
</tbody>
</table>

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 ineligibles monotone instrument
- g. MTR denotes Monotone Treatment Response