ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

The impact of the National School Lunch Program on child health: A nonparametric bounds analysis[☆]

Craig Gundersen[a,*], Brent Kreider[b], John Pepper[c]

[a] Department of Agricultural and Consumer Economics, University of Illinois, United States

[b] Department of Economics, Iowa State University, United States

[c] Department of Economics, University of Virginia, United States

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C14
C21
I38

Keywords:
Partial identification
Selection problem
Classification error
Monotone instrumental variable
Regression discontinuity
National School Lunch Program
Food insecurity
Obesity

ABSTRACT

Children in households reporting the receipt of free or reduced-price school meals through the National School Lunch Program (NSLP) are more likely to have negative health outcomes than observationally similar nonparticipants. Assessing causal effects of the program is made difficult, however, by missing counterfactuals and systematic underreporting of program participation. Combining survey data with auxiliary administrative information on the size of the NSLP caseload, we extend nonparametric partial identification methods that account for endogenous selection and nonrandom classification error in a single framework. Similar to a regression discontinuity design, we introduce a new way to conceptualize the monotone instrumental variable (MIV) assumption using eligibility criteria as monotone instruments. Under relatively weak assumptions, we find evidence that the receipt of free and reduced-price lunches improves the health outcomes of children.

1. Introduction

Every school day, more than 19 million children in the United States receive free or reduced-price lunches through the National School Lunch Program (NSLP).[1] With expenditures approaching $10 billion in fiscal year 2009, the NSLP is a large and important child nutrition program that is a vital component of the social safety net for children in low-income households. As such, policymakers expect the program to have a positive impact on the health of this vulnerable population. Yet, the existing empirical literature reveals little supporting evidence and, in some cases, appears to find deleterious effects. In particular, the literature has found that children receiving free or reduced-price lunches are more likely to have negative health outcomes than observationally similar eligible nonparticipants.

Despite these findings, the causal effects of the National School Lunch Program remain uncertain. Assessing the effects of the program is made difficult by the presence of two fundamental identification problems. First, children receiving free or reduced-price meals through the NSLP (hereafter referred to as the free lunch program) are likely to differ from eligible nonparticipants in ways that are not observed in the data. Second, the association between participation in the NSLP and poor nutritional health may be at least partly an artifact of household misreporting of

☆ The authors received valuable comments from Helen Jensen, Peter Orazem, Nese Yildiz, and seminar participants at Iowa State University, Purdue University, University of Rochester, the Southern Economic Association Meetings, the Survey Response Error and Statistical Modeling Workshop, and the conference on Identification and Decisions in honor of Charles Manski. Gundersen acknowledges financial support from the US Department of Agriculture (USDA), Cooperative State Research, Education and Extension Service (CSREES), Hatch project no. ILLU-470-331; Gundersen and Kreider acknowledge financial support from the Food Assistance and Nutrition Research Program of the USDA, Economic Research Service, contract number 59-5000-7-0109; and Pepper acknowledges financial support from the Bankard Fund for Political Economy and the Quantitative Collaborative at UVa.

* Corresponding author.
E-mail addresses: cggunder@illinois.edu (C. Gundersen), bkreider@iastate.edu (B. Kreider), jvpepper@virginia.edu (J. Pepper).

1 The NSLP is based on three categories of payments: free, reduced price, and full price. The former two categories apply for children with household incomes below 185% of the poverty line. While those between 130% and 185% of the poverty line receive reduced-price meals, the costs (no more than 40 cents per meal) are substantially less than the full-price meals. See http://www.fns.usda.gov/cnd/lunch/ for administrative details about the program.

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2011.06.007

participation in the program. Meyer et al. (2009), for example, find evidence of aggregate underreporting rates of 45% in the Current Population Survey and 27% in the Panel Study of Income Dynamics.

While these identification problems have been known to confound inferences on the impact of the free lunch program, credible solutions remain elusive. Most studies treat selection as exogenous and, to our knowledge, all studies assume that the classification of receipt is accurately reported.[2] Moreover, classical prescriptions for addressing these problems, namely linear instrumental variable models, may be untenable. State variation in program rules, which often serve as instruments to study the impact of other means-tested programs in the United States, are less useful as instrumental variables in this setting where most program rules are set by the federal government and have not changed substantially over time. In addition, as with much of the empirical literature on means-tested assistance programs, conventional parametric restrictions imposed to identify treatment effects – for example, the classical linear response model assumption – are difficult to justify when considering programs that are thought to have heterogeneous effects. Finally, classical measurement error models do not apply when the inaccurately measured covariate is discrete (see, e.g., Bollinger, 1996), when measurement error is thought to be systematic in a particular direction (e.g., underreporting of public transfers as documented in Meyer et al., 2009), or when the errors may be correlated with other characteristics of the respondents.

In this paper, we evaluate the impact of the free lunch program in light of the ambiguity created by the selection and measurement problems. Extending a recent analysis of the Supplemental Nutrition Assistance Program (SNAP, formerly known as the Food Stamp Program) in Kreider et al. (2011; KPGJ hereafter), we apply partial identification bounding methods that allow one to consider weaker assumptions than utilized in conventional parametric approaches. The primary methodological innovation is to introduce a new way to conceptualize the monotone instrumental variable (MIV) assumption using eligibility criteria as monotone instruments. We use these partial identification methods coupled with data from the 2001–2004 waves of the National Health and Nutrition Examination Survey (NHANES) to assess the impact of the free lunch program on the nutritional well-being of children. Specifically, we focus on studying what can be learned about the average treatment effect (ATE) of the free lunch program on food insecurity, poor general health, and obesity under various sets of assumptions about the selection and measurement error processes.

After describing the data in Section 2, we formally define the empirical questions and identification problems in Section 3. The usual program evaluation literature formally acknowledges uncertainty associated with counterfactuals but not uncertainty associated with misreporting.[3] As a departure from this literature, we account for both identification problems in a single framework. To do so, we begin by modifying the worst-case Manski (1995) selection bounds to account for classification error in the treatment.[4] While the data are uninformative in the absence of prior information on classification errors, we show how administrative information on the size of the NSLP caseload inherently places informative constraints on the classification error problem in the NHANES (see also KPGJ).

We then introduce a number of monotonicity assumptions that tighten inferences by addressing the selection problem. In Section 4, we consider the identifying power of a *monotone instrumental variable* (MIV) assumption that certain observed covariates are known to be monotonically related to the latent response variable (Manski and Pepper, 2000). Requiring no *a priori* exclusion restriction, the MIV assumption can be plausible in many applications where the standard independence assumption is a matter of considerable controversy. In our application, we maintain the assumption that the latent probability of a poor health outcome is nonincreasing with reported income.

In addition, we also consider two important variations of the MIV. The first variation, introduced in Manski and Pepper (2000), replaces the exogenous treatment selection assumption implicitly imposed in much of the literature with a weaker *monotone treatment selection* (MTS) restriction. This self-selection model formalizes the commonplace explanation for why recipients may have poorer health outcomes than nonrecipients — namely, that the decision to participate in the free lunch program is presumed to be (weakly) monotonically related to poor latent health outcomes.

Our second variation, which extends the methods applied in KPGJ, introduces a new way to conceptualize the MIV assumption by using eligibility criteria as monotone instruments. There is a long history of using ineligible respondents to identify the impact of a wide array public policies, including several recent studies that assess the impacts of the NSLP and the School Breakfast Program. Schanzenbach (2009), for example, uses a regression discontinuity design that exploits the income eligibility cutoff as an instrument, Bhattacharya et al. (2006) use children attending schools that do not offer meal programs, and Gleason and Suitor (2003) use a comparison of days when children participate and do not participate in the NSLP. While the basic idea of the discontinuity design is appealing, in practice there can be several limitations. Considerable disagreement often arises, for example, over the implicit assumption that ineligible respondents reveal the counterfactual outcome distribution for participants. Moreover, even if the comparison group is credible, these designs generally identify the effect only for persons near the eligibility cutoff and are not robust to classification error.

In contrast, the MIV assumption introduced in this paper allows us to relax this traditional identifying assumption by holding that mean outcomes among subgroups of ineligible respondents bound, instead of identify, the counterfactual outcome distribution. In our application, we focus on three ineligible subgroups: (1) children in households with incomes above the income eligibility threshold for free or reduced-price lunches (185% of the federal poverty line), (2) children enrolled in schools that do not participate in the NSLP (primarily private schools), and (3) children who have dropped out of school. Children in the first two groups are presumed to have no worse latent health outcomes on average than eligible children, while children in the third group are presumed to have no better outcomes on average. In this application, these ineligible-MIV assumptions provide substantial identifying information for the ATE.

Finally, in developing this new ineligibility-MIV assumption, we formally allow for the possibility of mislabeling some eligible households as ineligible. Specifically, in cases where self-reports of participation status are inconsistent with our assessment of ineligibility, we remain agnostic about whether it is ineligibility or participation that is misclassified.

Layering successively stronger assumptions, an objective of our analysis is to make transparent how the strength of the conclusions

2 Recent studies that address the possibility of endogenous selection include Gleason and Suitor (2003), Bhattacharya et al. (2006), Schanzenbach (2009), and Millimet et al. (2010). Evidence on selection is mixed. See Currie (2003) and Millimet et al. (2010) for thorough overviews of the literature.

3 Millimet (2010) provides an extensive Monte Carlo analysis of the consequences of nonclassical measurement error in treatment effect models and finds that even infrequent errors can have dramatic consequences for the performance of various estimators. Kreider (2010) comes to similar conclusions using linear and probit specifications. McCarthy and Tchernis (2010) consider the problem of misclassified and endogenous treatments in a Bayesian setting.

4 We draw on the related literature on corrupt samples in Horowitz and Manski (1995) and for addressing missing treatments from Molinari (2008, 2010). We also borrow from the related partial identification work in Manski and Pepper (2000), Pepper (2000), Kreider and Pepper (2007, 2011), Kreider and Hill (2008), and Kreider et al. (2011).

Table 1
Means and standard deviations by National School Lunch Program participation.

	Income-eligible children[a]	Recipients[b]	Nonrecipients[b]
Age in years	11.1 (3.30)	10.7** (3.11)	12.2 (3.56)
Ratio of income to the poverty line	0.980 (0.472)	0.917** (0.453)	1.162 (0.479)
NSLP recipient	0.743 (0.437)	1.000 (0.000)	0.000 (0.000)
Outcomes:			
Food-insecure household	0.364 (0.481)	0.399** (0.490)	0.263 (0.441)
Poor or fair health	0.072 (0.258)	0.071 (0.257)	0.075 (0.263)
Obese (BMI $\geq$ 95th percentile)	0.193 (0.395)	0.196 (0.397)	0.186 (0.389)
N	2693	2077	616

Sample estimates weighted with the medical exam weight. The estimated means for the NSLP recipient population are superscripted with ** or * to indicate that they are statistically significantly different from the means for the nonrecipient population (with p-values less than 0.01 and 0.05, respectively, based on Wald statistics corrected for the complex design).
[a] Includes all children residing in households with income less than 185% of the poverty line, attending schools that offer a lunch provided by the National School Lunch Program, and without any missing information for the variables in the table.
[b] Receipt is based on a self-reported (by the parent) indicator of participation in the free lunch program.

varies with the strength of the identifying assumptions. In Section 5, we highlight the additional identifying power of a *monotone treatment response* (MTR) assumption that participation in the NSLP would not increase the prevalence of poor health (see, e.g., Currie, 2003). Combined with our MIV assumptions, the MTR assumption narrows the range of uncertainty about the average treatment effects and clearly identifies that the free lunch program reduces poor health outcomes. Section 6 draws conclusions.

2. Data

To study the impact of the NSLP on children's nutritional health, we use data from the 2001–2004 NHANES. The NHANES, conducted by the National Center for Health Statistics, Centers for Disease Control (NCHS/CDC), is a program of surveys designed to assess the health and nutritional status of adults and children in the United States through interviews and direct physical examinations.[5] The survey currently examines a national sample of about 5000 persons each year, about half of whom are children. Vulnerable groups, including Hispanics and African–Americans, are oversampled. The NHANES provides detailed and varied information on dietary and health-related outcomes collected from self-reports of health and nutritional well-being, medical and dental examinations, physiological measurements, and laboratory tests. Given the wealth of health-related information, the NHANES has been widely used in previous research on health- and nutrition-related child outcomes.

We restrict attention to households with children who appear eligible to receive free or reduced-price lunches through the NSLP. Specifically, we restrict the analysis to children between the ages of 6 and 17 who are reported to be attending schools with the NSLP and residing in households with income less than 185% of the federal poverty line. This constitutes the full set of information necessary to determine eligibility for the free lunch program and results in a sample of 2693 eligible children.[6] We also utilize information from three groups of children who appear ineligible to receive free or reduced-price lunches: children residing in households with incomes between 185% and 300% of the poverty line ($N = 899$), income-eligible children in this age range who are enrolled in schools without the NSLP ($N = 84$), and income-eligible children who are no longer in school ($N = 120$).

Table 1 displays the means and standard deviations for the variables used in our analysis for the main sample of children classified as eligible to receive free or reduced-price lunches. For each respondent, we observe a limited set of socioeconomic and demographic information, including age and the ratio of income to the poverty line, the ratio of a family's income to the poverty threshold defined by the US Census Bureau accounting for the family's composition. We also observe a self-reported (by the parent) indicator of participation in the free lunch program. In this survey, 74% of the eligible households claim to have children who received free or reduced-price lunches through the NSLP during the school year.

We examine three outcomes: food insecurity, poor health, and obesity.[7] To calculate official food insecurity rates in the US, defined over a 12-month period, a series of 18 questions about food-related needs and resources in the household are posed in the Core Food Security Module (CFSM) for families with children. Examples include "I worried whether our food would run out before we got money to buy more" (the least severe outcome) and "Did a child in the household ever not eat for a full day because you couldn't afford enough food?" A child is considered to reside in a food-insecure household if the respondent answers affirmatively to three or more of these questions. We measure obesity using measures based on a child's body mass index (BMI, kg/m^2) such that a child is classified as obese if his or her BMI is at or above the 95th percentile for age and gender. General health is based on a self-reported measure provide by the child's parent. A child's health under this measure is placed into one of five categories based on responses from the parent: excellent, very good, good, fair, or poor. In this paper, we combine these general health categories into an indicator of fair or poor health.

5 We pool the 2001–2002 and 2003–2004 two-year cycles of the NHANES and use sample weights that are established within the NHANES for use when multiple cycles are combined.

6 As described below, parts of our analysis account for the possibility that eligibility may be measured with error.

7 For tractability, we treat these health outcomes as accurately measured. While errors in measuring obesity are likely to be minimal (data on height and weight were collected by trained personnel), this assumption may be violated for the poor health and food insecurity outcomes. In general, measurement error in the outcome variables would widen the bounds established in this paper.

Please cite this article in press as: Gundersen, C., et al., The impact of the National School Lunch Program on child health: A nonparametric bounds analysis. Journal of Econometrics (2011), doi:10.1016/j.jeconom.2011.06.007

In our data, 36.4% of the respondents are food insecure, 19.3% are obese, and 7.2% report being in poor or fair health. These three measures are considered to be central indicators of the nutritional health and well-being of children and reflect a wide range of health related outcomes that might be affected by the NSLP. All three outcomes are known to be associated with a range of negative physical, psychological, and social consequences that have current and future implications for health. Children in households suffering from food insecurity, for example, are more likely to have poor health, psychosocial problems, frequent stomachaches and headaches, increased odds of being hospitalized, greater propensities to have seen a psychologist, behavior problems, worse developmental outcomes, more chronic illnesses, less mental proficiency, and higher levels of iron deficiency with anemia (for a review, see Gundersen and Kreider, 2009). Likewise, childhood obesity is known to have negative physical, psychological, and social consequences, including reduced life expectancy (Fontaine et al., 2003).

While policymakers have expressed a great deal of interest in understanding the impact of assistance programs on food insecurity, researchers have not, to our knowledge, examined the impact of the NSLP on food insecurity or measures of general health.[8] There is, however, a growing body of literature that examines the program's impact on obesity, a recent concern among policymakers. This literature finds that the NSLP appears to lead to modest increases in obesity rates (see Millimet et al., 2010; Schanzenbach, 2009).

In Table 1, we see that free lunch recipients appear to be somewhat worse off than eligible nonparticipants. Most striking are the outcomes for food insecurity. The rates of food insecurity among self-reported recipients are nearly 14 percentage points higher than for nonparticipants. The obesity rate is slightly higher for children in households claiming to participate, and the rate of poor health is slightly lower. For both obesity and poor health, these differences are not statistically significant.

3. The average treatment effect with endogenous selection and classification errors

Our interest is in learning the average effect of the free lunch program among eligible households:

$$\begin{aligned} ATE(1,0 \mid X \in \Omega) &= E[Y(1) \mid X \in \Omega] - E[Y(0) \mid X \in \Omega] \\ &= P[Y(1) = 1 \mid X \in \Omega] - P[Y(0) = 1 \mid X \in \Omega], \end{aligned} \tag{1}$$

where $Y(1)$ denotes the health of a child if participating in the NSLP, $Y(0)$ denotes the analogous outcome if not participating, and $X \in \Omega$ denotes conditioning on observed covariates whose values lie in the set Ω.[9] Thus, the average treatment effect reveals how the mean outcome would differ if all eligible children would receive assistance versus the mean outcome if all eligible children would not receive assistance. Two forms of uncertainty arise when assessing the impact of the NSLP on children's outcomes. First, even if participation were observed, the outcome $Y(1)$ is counterfactual for all children who did not receive assistance, and $Y(0)$ is counterfactual for all children who did receive assistance. This is referred to as the selection problem. Second, participation may not be accurately observed for all respondents. This is referred to as the classification error problem.

Conditioning on X allows the researcher to focus on specific subpopulations of interest. Following the existing literature, we consider the subgroup of households that appear to be eligible for the free lunch program based on sufficiently low self-reported income and having a child enrolled in a school that offers the program.[10] We refer to these households as eligible, though errors in measuring the variables used to determine eligibility – most notably income – may lead to some contamination of the eligibility indicator.[11] While this is a well-defined population for analysis regardless of the potential for errors in classifying eligibility, we checked the sensitivity of our results to different income thresholds for defining the main sample. Our empirical results are quite robust to alternative income thresholds ranging from 150% to 200% of the poverty line. To simplify exposition, in what follows we suppress the conditioning on X.

In Section 3.1, we formalize the identification problems that arise from the selection and participation classification error problems. Then, in Section 3.2 we focus on what can be inferred about the ATE in the worst-case scenario where one has no prior information restricting the selection problem. We show how auxiliary information on the size of the NSLP caseload can be used to constrain the classification error problem. In Sections 4 and 5, we assess the identifying power of monotonicity assumptions used to address the selection problem.

3.1. The identification problem

To disentangle the selection and classification problems, it is useful to introduce notation for accurate and inaccurate reports of participation. In particular, let $S^* = 1$ indicate that the child truly receives free lunches, with $S^* = 0$ otherwise, such that the observed health outcome is $Y = Y(1)S^* + Y(0)(1 - S^*)$. The classification error problem arises because we only observe S, which indicates whether the household reports that a child participates, and not S^*.

We highlight these two identification problems by writing

$$\begin{aligned} P[Y(1) = 1] &= P[Y(1) = 1 \mid S^* = 1]P(S^* = 1) \\ &\quad + P[Y(1) = 1 \mid S^* = 0]P(S^* = 0) \\ &= [P(Y = 1, S = 1) - \theta_1^+ + \theta_1^-] \\ &\quad + P[Y(1) = 1 \mid S^* = 0][P(S = 0) + (\theta_1^+ + \theta_0^+) \\ &\quad - (\theta_1^- + \theta_0^-)], \end{aligned} \tag{2}$$

where $\theta_i^+ = P(Y = i, S = 1, S^* = 0)$ and $\theta_i^- = P(Y = i, S = 0, S^* = 1)$ denote the fraction of false positive and

[8] Food insecurity and self-reported measures of health have been widely used as outcomes in studies of other food assistance programs such as SNAP (e.g., Kreider et al. (2011), DePolt et al. (2009), Gundersen et al. (2008) and Gundersen and Oliveira (2001)).

[9] Previous research on the impact of the NSLP also focuses on estimating the population average treatment effect (see, e.g., Gleason and Suitor (2003), Millimet et al. (2010) and Schanzenbach (2009)). We follow this literature and focus on the ATE parameter since it measures the effect of the program on the entire population of interest to policymakers (Ginther, 2000). Other treatment effects, such as the effect of the treatment on the treated or the status quo treatment effect might also be of interest, and the methods developed in this paper can be modified to evaluate those parameters.

[10] Note that there are no regression orthogonality conditions to be satisfied, and excluding other personal characteristics does not introduce omitted variable bias into the analysis.

[11] In particular, the NSLP defines a child as eligible for free or reduced-price meals based on income information obtained at the start of the school year or, in some cases, during the school year. In contrast, the NHANES asks respondents about their annual income which, due to timing issues, may not coincide with levels reported to the NSLP. Also, these measures may be reported with error. In the context of food stamps, Daponte et al. (1999) highlight the problem that determining whether households are eligible for a program may be conceptually straightforward but operationally difficult, especially when eligibility is determined in part by asset level. We are not aware of research assessing the extent or nature of errors in classifying eligibility status for the NSLP, where eligibility depends on income but not assets. The standard practice in the literature has been to implicitly assume that the information provided by households regarding the relevant eligibility variables is accurate. We generally follow this practice but, as discussed in Section 4, in some cases relax the assumption that eligibility is classified accurately.

Please cite this article in press as: Gundersen, C., et al., The impact of the National School Lunch Program on child health: A nonparametric bounds analysis. Journal of Econometrics (2011), doi:10.1016/j.jeconom.2011.06.007

false negative classifications of NSLP participation, respectively, for children realizing a health outcome of $i = 0$ or 1. Notice that the first term $P[Y(1) = 1 \mid S^* = 1]P(S^* = 1)$ is not identified because of the classification error problem. If we rule out these errors such that $\theta_1^+ = \theta_1^- = 0$, then this term is revealed by the data as $P(Y = 1, S = 1)$. The second term is not identified because of both the selection and classification error problems. The data cannot reveal the counterfactual outcome distribution, $P[Y(1) = 1 \mid S^* = 0]$, regardless of whether participation is measured accurately, and, in the presence of classification errors, the sampling process does not reveal the fraction of participants, $P(S^*)$.

3.2. Worst-case bounds

Suppose that the classification error probabilities, θ, are known, and let $\Theta \equiv (\theta_1^- + \theta_0^+) - (\theta_0^- + \theta_1^+)$. It then follows that

$$[-P(Y=1,S=0) - P(Y=0,S=1)] + \Theta \leq ATE(1,0) \leq [P(Y=1,S=1) + P(Y=0,S=0)] + \Theta. \quad (3)$$

With no classification errors, $\Theta = 0$, and Eq. (3) simplifies to the Manski (1995) worst-case selection bounds. In the absence of assumptions on the selection process, these bounds on the ATE always have a width of 1 and always include 0. The data alone cannot reveal the sign of the treatment effect. In our application, for example, the estimated worst-case bounds on the ATE for the food insecurity outcome are $[-0.514, 0.486]$. With known classification errors, these bounds shift by Θ.

In practice, the classification error probabilities are unknown, and the survey data alone are uninformative.[12] We have found, however, that readily available auxiliary data on the size of the caseload imply meaningful restrictions on these error probabilities. In particular, combining administrative data with survey data from the NHANES allows us to estimate both the true participation rate, $P(S^* = 1)$, and the self-reported rate, $P(S = 1)$. Knowledge of the true and self-reported participation rates bounds the classification error probabilities, θ. For example, if the self-reported rate lies below the true participation rate, we know that the fraction of false negative reports must exceed the fraction of false positive reports. More generally, we exploit the following three restrictions (see KPGJ)[13]:

$$(\theta_1^- + \theta_0^-) - (\theta_1^+ + \theta_0^+) = P(S^* = 1) - P(S = 1) \quad (4a)$$

$$\theta_i^- \leq \min\{P(Y=i, S=0), P(S^*=1)\} \equiv \theta_i^{UB-}, \quad i = 1, 0 \quad (4b)$$

$$\theta_i^+ \leq \min\{P(Y=i, S=1), P(S^*=0)\} \equiv \theta_i^{UB+}, \quad i = 1, 0. \quad (4c)$$

Eq. (4a) restricts the net fraction of false negative reports to equal the difference between the true and self-reported participation rates. Eqs. (4b)–(4c) place meaningful upper bounds on the fractions of false negative and positive reports.

These restrictions on classification error probabilities imply informative bounds on the unknown parameter, Θ. Subject to the restrictions in Eq. (4), the upper bound is found by maximizing $(\theta_1^- + \theta_0^+)$ and minimizing $(\theta_0^- + \theta_1^+)$, and vice versa for the lower bound.

Suppose that $\Delta = P(S^* = 1) - P(S = 1)$ is observed (or can be estimated). Then, for the case where the true participation rate exceeds the self-reported rate (i.e., $\Delta > 0$), KPGJ show that the following bounds apply.

Worst-case bounds on the ATE given restrictions (4a)–(4c) on participation misreporting:

$$-\min\{\theta_0^{UB-}, \Delta + \theta_1^{UB+}\} - \min\{\theta_1^{UB+}, \theta_0^{UB-} - \Delta\} \leq \Theta \leq \min\{\theta_1^{UB-}, \Delta + \theta_0^{UB+}\} + \min\{\theta_0^{UB+}, \theta_1^{UB-} - \Delta\}. \quad (5)$$

See KPGJ for a proof of this result.[14] Combining Eq. (5) with Eq. (3) implies a sharp bound on the average treatment effect. Notice that allowing for ambiguity created by the reporting error problem (weakly) widens the treatment effect bounds in Eq. (3).

Except for the true participation rate, $P(S^* = 1)$, all of the probabilities in Eq. (5) can be consistently estimated using data from the NHANES. To infer $P(S^* = 1)$, we combine auxiliary data on the size of the caseload with data from the NHANES on the size of the eligible population. Administrative data from the US Department of Agriculture (USDA) reveals that from 2001–2005 an average of 16 million children received free or reduced price meals through the NSLP per month (Food and Nutrition Service, 2010). From the NHANES, we estimate that about 18 million children were eligible to receive assistance. Thus, the implied participation rate is about 0.89, 15 percentage points higher than the reported rate of 0.74. We use this estimated participation rate of 0.89 to restrict the classification error probabilities, θ, and to estimate bounds on the ATE. Given that errors in classifying NSLP-eligible children in the NHANES may bias the estimated participation rate, we also assess the sensitivity of the estimated bounds to variation in the true participation rate.

Specifically, the solid lines in Figs. 1A–1C trace out the estimated worst-case bounds on the ATE on the household food insecurity rate, fair or poor health, and obesity as the true participation rate, $P^* = P(S^* = 1)$, varies between 0 and 1. The accompanying tables reproduce these results for P^* equal to the self-reported participation rate of 0.74 based on reports in the NHANES data and the estimated true participation rate of 0.89 based on administrative data from the USDA. For the self-reported rate of 0.74, we report the bounds under two scenarios: (1a) no reporting errors (not shown in the figures), and (1b) reporting errors in which false positive and false negative reports exactly cancel, $\Delta = 0$. To account for sampling variability, we also report Imbens and Manski (2004) confidence intervals that cover the ATE with 90% probability.[15]

As noted above, if receipt of reduced-price lunches is known to be accurately reported, the worst-case bounds on the ATE have a width of 1 and always include 0. Allowing for classification errors increases the width of these bounds. For example, suppose that the true participation rate remains at 0.74, but that reporting errors are allowed as long as there is no net misreporting (i.e., the rate of false positives equals the rate of false negatives). Then, as shown in Fig. 1A and the accompanying table, the ATE bounds on the food insecurity rate expand from $[-0.514, 0.486]$, with a width of 1, to $[-0.893, 0.621]$, with a width of 1.5. If the true participation rate is 0.89 (the rate consistent with the USDA administrative data), the bounds are $[-0.747, 0.475]$, with a width of 1.2. Interestingly, the upper bound of 0.475 is improved compared with the no classification error upper bound of 0.486. This occurs because the estimated upper bound on Θ is negative. Similar results for poor general health status and obesity are traced out in Figs. 1B and 1C.

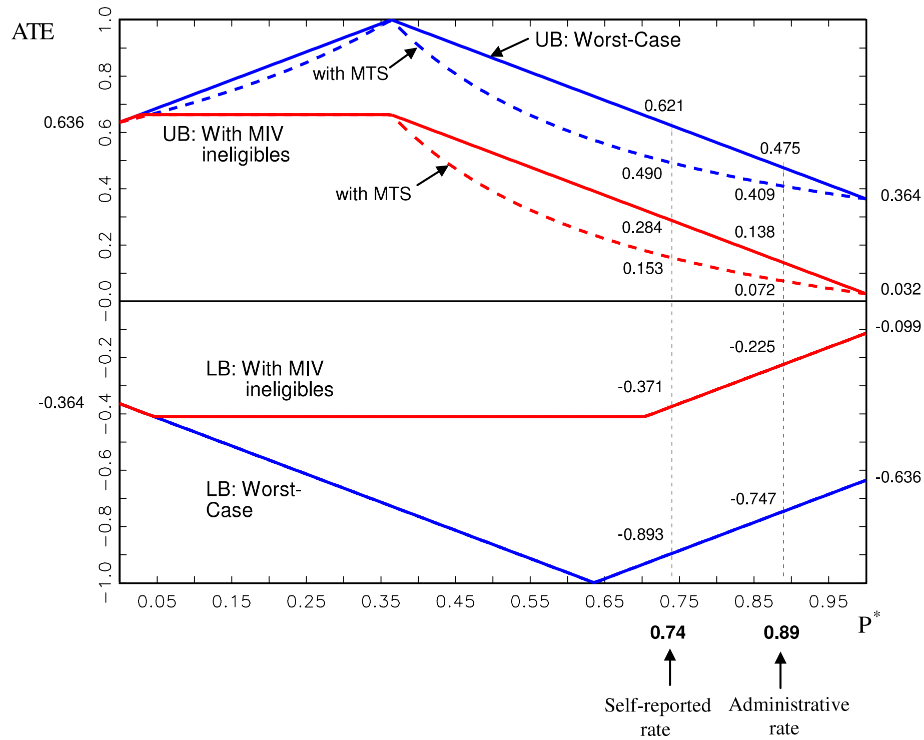
Without additional restrictions to address the selection problem, we cannot rule out the possibility that the free lunch program has large positive or negative effects on health outcomes. To narrow these bounds, we consider a number of additional identifying assumptions.

12 To see this, note that, without any restrictions on classification errors, $\Theta \in [-P(Y=0,S=0) - P(Y=1,S=1), P(Y=1,S=0) + P(Y=0,S=1)]$ and the worst-case bounds are $[-1, 1]$.

13 In addition, the aggregate false positive and false negative rates are bounded as follows: $\max\{0, P(S=1) - P(S^*=1)\} \leq (\theta_1^+ + \theta_0^+) \leq \min\{P(S^*=0), P(S=1)\}$ and $\max\{0, P(S^*=1) - P(S=1)\} \leq (\theta_1^- + \theta_0^-) \leq \min\{P(S=0), P(S^*=1)\}$. The total misreporting rate is bounded to lie within $[|P(S^*=1) - P(S=1)|, \min\{P(S=1) + P(S^*=1), P(S=0) + P(S^*=0)\}]$.

14 KPGJ provide results for all values of Δ.

15 Our confidence sets do not account for the fact that Δ is estimated.



<u>Bounds for selected values of P^*</u>

	$P^* = P = 0.74$: no misreporting (not shown in figure)	$P^* = P = 0.74$: no net misreporting	$P^* = 0.89$: administrative rate (underreporting)
Worst-Case:	[-0.514, 0.486] [-0.534 0.505]	[-0.893, 0.621] [-0.910 0.638]	[-0.747, 0.475] [-0.764 0.492]
+ MTS:	[-0.514, 0.136] [-0.534 0.178]	[-0.893, 0.490] [-0.910 0.512]	[-0.747, 0.409] [-0.764 0.428]
With MIV ineligibles:	[-0.114, 0.216] [-0.187 0.312]	[-0.371, 0.284] [-0.444 0.379]	[-0.225, 0.138] [-0.298 0.233]
+ MTS:	[-0.114, 0.062] [-0.187 0.149]	[-0.371, 0.153] [-0.444 0.250]	[-0.225, 0.072] [-0.298 0.168]
+ Income MIV:	[-0.090, -0.014] [-0.182 0.149]	[-0.266, 0.105] [-0.366 0.215]	[-0.158, 0.037] [-0.251 0.141]

Fig. 1A. Sharp bounds on the ATE for *household food insecurity* as a function of P^*, the unobserved true NSLP participation rate.

4. Monotone instrumental variable assumptions: variations on a theme

Many observed variables are thought to be monotonically related to the latent health outcomes in Eq. (1). In this section, we formalize and assess the identifying power of several different types of monotone instrumental variable (MIV) assumptions. We begin by considering the relatively innocuous assumption that the latent probability of negative health outcomes weakly decreases with income relative to the poverty threshold (which varies by family size), as in KPGJ. A large body of empirical research supports the idea of a negative gradient between reported income and the health outcomes studied in this paper (see, e.g., Nord et al. (2010) for food insecurity, (Case et al., 2002) for general health, (Shrewsbury and Wardle, 2008) for obesity, and (Deaton, 2002) for a broad array of health outcomes).

To formalize the notion that the latent probability of a negative health outcome $P[Y(t) = 1]$ is known to vary monotonically with an observed covariate, let v be a monotone instrumental variable such that

(A1) *Income-MIV*

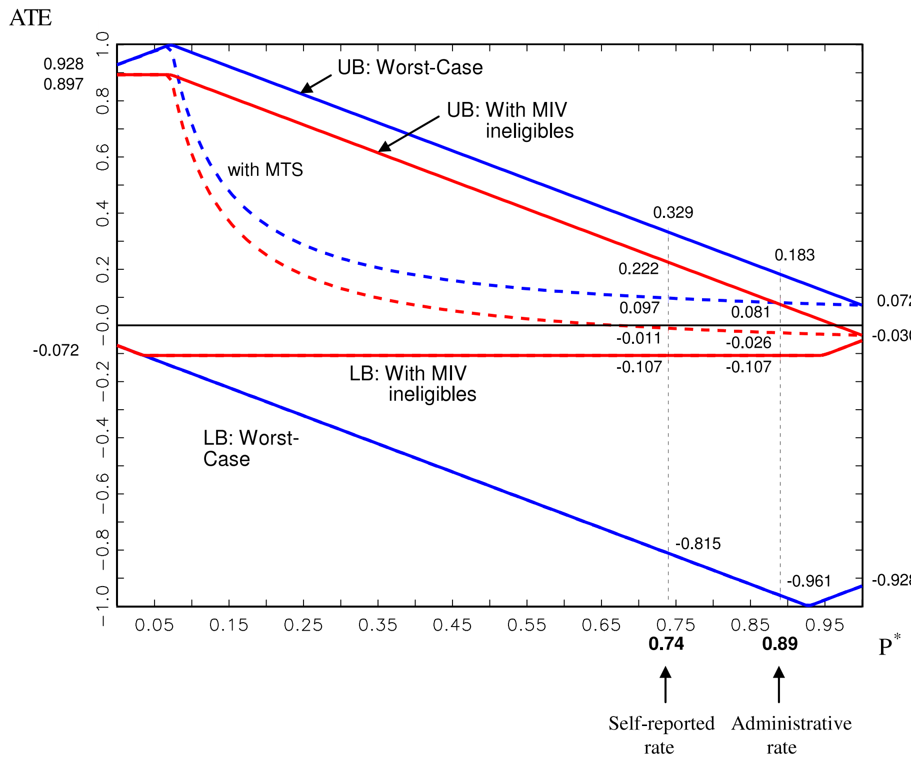
$$u_1 \le u \le u_2 \Rightarrow P[Y(t) = 1 \mid v = u_2] \le P[Y(t) = 1 \mid v = u] \le P[Y(t) = 1 \mid v = u_1].$$

Then the MIV restriction presented in Manski and Pepper (2000, Prop. 1) implies that

$$\sup_{u \le u_2} LB(u_2) \le P[Y(t) = 1 \mid v = u] \le \inf_{u \ge u_1} UB(u_1),$$

where $LB(u)$ and $UB(u)$ are the known lower and upper bounds evaluated at $v = u$, respectively, given the available information. The MIV bound on the unconditional latent probability $P[Y(t) = 1]$ can then be obtained using the law of total probability.

To estimate the MIV bounds, we take the appropriate weighted average of the plug-in estimators of lower and upper bounds



Bounds for selected values of P^*

	$P^* = P = 0.74$: no misreporting (not shown in figure)	$P^* = P = 0.74$: no net misreporting	$P^* = 0.89$: administrative rate (underreporting)
Worst-Case:	[-0.710, 0.290]	[-0.815, 0.329]	[-0.961, 0.183]
	[-0.730 0.310]	[-0.824 0.338]	[-0.970 0.192]
+ MTS:	[-0.710, -0.004]	[-0.815, 0.097]	[-0.961, 0.081]
	[-0.730 0.017]	[-0.824 0.109]	[-0.970 0.091]
With MIV ineligibles:	[-0.055, 0.202]	[-0.107, 0.222]	[-0.107, 0.076]
	[-0.109 0.269]	[-0.160 0.290]	[-0.160 0.144]
+ MTS:	[-0.055, -0.036]	[-0.107, -0.011]	[-0.107, -0.026]
	[-0.117 0.029]	[-0.161 0.058]	[-0.161 0.043]
+ Income MIV:	[-0.053, -0.036]	[-0.107, -0.011]	[-0.072, -0.026]
	[-0.117 0.029]	[-0.161 0.058]	[-0.133 0.043]

Fig. 1B. Sharp bounds on the ATE for *child poor health* as a function of P^*, the unobserved true NSLP participation rate.

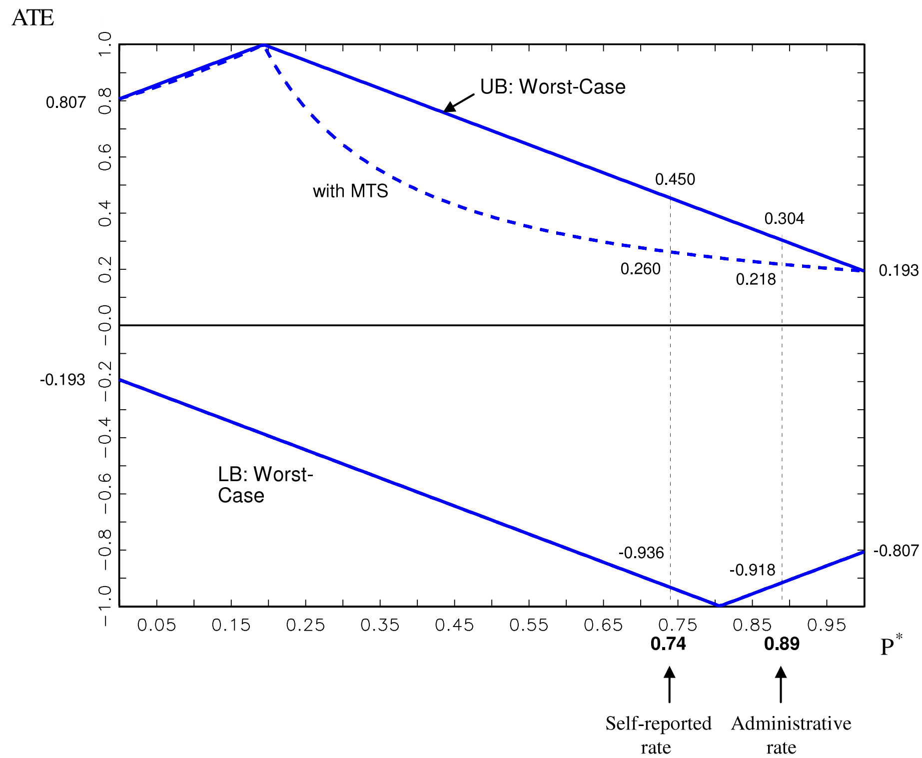
across the different values of the instrument.[16] As discussed in Manski and Pepper (2000, 2009) and Kreider and Pepper (2007), this MIV estimator is consistent but biased in finite samples. We employ Kreider and Pepper (2007)'s modified MIV estimator that accounts for the finite sample bias using a nonparametric bootstrap correction method.

In our application, the income-MIV assumption has some identifying power but does not substantially narrow the worst-case selection bounds. Thus, rather than present bounds under the income-MIV assumption alone, we combine this assumption with two other distinct but related instrumental variable restrictions. In Section 4.1, we apply the monotone treatment selection assumption that participation in the program is (weakly) negatively related to expected health outcomes. In Section 4.2, we introduce and assess the assumption that ineligibility criteria for the NSLP are monotonically related to the latent outcomes. For example, income-ineligible children – i.e., children residing in households with income greater than 185% of the poverty line – are likely to have better average health outcomes than the income-eligible children. In Section 4.3, we present results under these three MIV restrictions.

4.1. Monotone treatment selection

Self-selection into the NSLP is the most common explanation for the positive correlation between participation and poor health (Currie, 2003). Unobserved factors associated with poor health are thought to be positively associated with the decision to take up the program, $S^* = 1$. The *monotone treatment selection* (MTS) assumption, which replaces the exogenous selection assumption

[16] To estimate these income-MIV bounds, we divide the sample into 10 groups defined by the ratio of income to the poverty line.



Bounds for selected values of P*

	$P^* = P = 0.74$: no misreporting (not shown in figure)	$P^* = P = 0.74$: no net misreporting	$P^* = 0.89$: administrative rate (underreporting)
Worst-Case:	[-0.645, 0.355] [-0.665 0.375]	[-0.936, 0.450] [-0.951 0.465]	[-0.918, 0.304] [-0.932 0.319]
+ MTS:	[-0.645, 0.011] [-0.665 0.044]	[-0.936, 0.260] [-0.951 0.279]	[-0.918, 0.218] [-0.932 0.233]
+ Income MIV:	[-0.645, -0.032] [-0.659 0.044]	[-0.934, 0.176] [-0.951 0.232]	[-0.889, 0.152] [-0.939 0.198]

Fig. 1C. Sharp bounds on the ATE for *child obesity* as a function of P^*, the unobserved true NSLP participation rate.

implicit in much of the literature, specifies that children receiving free lunches are likely to have no better latent health outcomes on average than nonparticipants.

(A2) *Monotone treatment selection (MTS)*:

$$P[Y(t) = 1 \mid S^* = 0] \leq P[Y(t) = 1 \mid S^* = 1] \quad \text{for } t = 0, 1.$$

Under this MTS assumption, it follows that

$$ATE(1, 0) \leq P(Y = 1 \mid S^* = 1) - P(Y = 1 \mid S^* = 0).$$

In the absence of classification errors, these probabilities are revealed by the sampling process. In that case, the ATE is identified to be no larger than $P(Y = 1 \mid S = 1) - P(Y = 1 \mid S = 0)$. Otherwise, we can write

$$ATE(1, 0) \leq \frac{P(Y = 1, S = 1) + \theta_1^- - \theta_1^+}{P(S^* = 1)} - \frac{P(Y = 1, S = 0) + \theta_1^+ - \theta_1^-}{P(S^* = 0)}. \tag{6}$$

With information on the true participation rate, $P(S^* = 1)$, we can bound these conditional probabilities under the restrictions in Eq. (4). In particular, the following upper bound applies.

MTS upper bound on the ATE given restrictions (4a)–(4c) on participation misreporting:

$$ATE(1, 0) \leq \begin{cases} \dfrac{P(Y = 0)}{P(S^* = 0)} & \text{if } 0 < P(S^* = 1) < P(Y = 1, S = 1) \\ \dfrac{P(Y = 1, S = 1) + \theta_1^{UB-*}}{P(S^* = 1)} - \dfrac{P(Y = 1, S = 0) - \theta_1^{UB-*}}{P(S^* = 0)} & \text{if } P(Y = 1, S = 1) \leq P(S^* = 1) < 1, \end{cases} \tag{7}$$

where $\theta_1^{UB-*} \equiv \min\{P(Y = 1, S = 0)P(S^* = 1) - P(Y = 1, S = 1)\}$.

See KPGJ for a proof of this result.

This MTS assumption has a notable impact on the upper bound of the ATE. Without classification errors, the upper bound equals the observed difference in the poor health outcomes between recipients and nonrecipients, as revealed in Table 1. For example, the upper bound on the impact of the free lunch program on rates of food insecurity is estimated to be 0.136, notably

Table 2
Means and standard deviations by National School Lunch Program eligibility groups.

	Income-eligible Children attending NSLP schools[a] (1)	Income-eligible Children attending non-NSLP schools (2)	Income-eligible Dropouts (3)	Income-ineligible (4)
Age in years	11.1 (3.30)	11.5 (4.01)	14.3** (3.39)	11.4* (3.24)
Ratio of income to the poverty line	0.980 (0.472)	1.059 (0.492)	1.01 (0.467)	2.41** (0.329)
NSLP recipient	0.743 (0.437)	0.000** (0.000)	0.000** (0.000)	0.235** (0.424)
Outcomes:				
Food-insecure Household	0.364 (0.481)	0.337 (0.476)	0.410 (0.494)	0.130** (0.337)
Poor or fair health	0.072 (0.258)	0.107 (0.311)	0.086 (0.281)	0.018** (0.135)
Obese (BMI $\geq$ 95th percentile)	0.193 (0.395)	0.149 (0.358)	0.125 (0.333)	0.191 (0.393)
N	2693	84	120	899

Sample estimates weighted with the medical exam weight. The estimated means for columns (2) through (4) are superscripted with ** or * to indicate that they are statistically significantly different from the means for the income-eligible population in column (1) with p-values less than 0.01 and 0.05, respectively (based on Wald statistics corrected for the complex design).
[a] Children residing in households with income less than 185% of the federal poverty line (FPL) are classified as income eligible, whereas those with income between 185%–300% of the FPL are classified as income ineligible. Among the income-eligible households, some attend schools without the NSLP and some have dropped out of school. All calculations are based on observations without any missing information for the variables in the table.

improved compared with the worst-case upper bound of 0.486. With classification errors, the MTS assumption reduces the upper bound from 0.621 to 0.490 if there is no net misreporting and from 0.475 to 0.409 when there is underreporting with $P^* = 0.89$ (see Fig. 1A and the corresponding table).

4.2. Ineligible comparison groups as MIVs

In this section, we introduce a new way to conceptualize the MIV assumption using eligibility criteria as monotone instruments. Many program evaluations rely on ineligible respondents to reveal the outcome distribution under nonparticipation. In our application, we observe three groups of ineligible respondents based on information in the NHANES: income-eligible children who have dropped out of school, income-eligible children attending schools that do not offer the NSLP, and children whose household income is between 185% and 300% of the poverty line. Table 2 displays descriptive statistics for these ineligible groups of children side-by-side with the group of eligible children. This table reveals that health outcomes of children in schools that do not offer the NSLP are similar to those of eligible children, dropouts are less likely to be obese but are more likely to be food insecure and in poor health, and income-ineligible children are better off with respect to food security and general health.

These comparison groups are unlikely to satisfy the standard instrumental variable restriction that the latent health outcomes are mean independent of eligibility status. However, the MIV assumption holding that mean response varies monotonically across these subgroups seems credible, especially for the food insecurity and poor health outcomes. As a group, children in households with incomes above the eligibility cutoff for the NSLP (i.e., above 185% of the poverty line), for example, are likely to have no worse average latent health outcomes than children below this line. Likewise, children attending schools without the NSLP – which are primarily private schools – are thought to have better outcomes, and dropouts might be assumed to have relatively poor latent health outcomes.[17] We apply these monotonicity restrictions to the food insecurity and poor health outcomes, but not to obesity, since relationships between the three subgroups and latent measures of obesity are less certain.

To formalize the notion that the latent probability of poor health, $P[Y(t) = 1]$, is known to be monotonically related to these observed ineligible subgroups, let υ_2 be the monotone instrumental variable such that:

(A3) *Ineligible comparison group MIV*

i. $P[Y(t) = 1] \geq P[Y(t) = 1 \mid \upsilon_2 = \text{income ineligible}]$,
ii. $P[Y(t) = 1] \geq P[Y(t) = 1 \mid \upsilon_2 = \text{no school lunch program}]$, and
iii. $P[Y(t) = 1] \leq P[Y(t) = 1 \mid \upsilon_2 = \text{dropped out}]$.

Notice that these ineligible-MIV bounds on $P[Y(t) = 1]$ can be rewritten to reflect the same structure as the income-MIV restriction in (A1). The income-ineligible assumption in (A3i), for example, can be rewritten as $P[Y(t) = 1 \mid \upsilon_2 = \text{ income eligible}] \geq P[Y(t) = 1 \mid \upsilon_2 = \text{income ineligible}]$.

If the subgroups defined by υ_2 are known to be accurately classified as ineligible, then by definition $S^* = 0$ within these groups and there is no selection or participation classification error problem. In this case, the data point-identify $P[Y(0) = 1 \mid \upsilon_2]$ but provide no information on $P[Y(1) = 1 \mid \upsilon_2]$. Thus, the MIV restriction in Assumption (A3) implies that

$$\begin{aligned} &\max\{P(Y = 1 \mid \upsilon_2 = \text{income-ineligible}), \\ &\quad P(Y = 1 \mid \upsilon_2 = \text{no school lunch program})\} \\ &\leq P[Y(0) = 1] \leq P(Y = 1 \mid \upsilon_2 = \text{dropped out}). \end{aligned} \tag{8}$$

Errors in classifying income-ineligible respondents, however, may contaminate some portion of this subgroup. As discussed above, income is known to be measured with error and, even if correctly measured, there may be differences in the timing of the survey and eligibility determination. In our data, nearly one-quarter of the households in this subgroup report receiving assistance from the free lunch program. Rather than assuming that all of these reports of receipt are inaccurate, we allow for the possibility that self-reports of participation among this supposedly ineligible subgroup may be accurate; i.e., the ineligibility label may be inaccurate. We assume that reports of nonparticipation are accurate.

[17] The NSLP offers lunches in 99% of US public schools and in 83% of private and public schools combined (USDA/ERS, 2004).

Table 3
Bounds on the ATE with accurately reported participation status.

	Food insecurity	Poor health	Obesity
Worst case:	[−0.514, 0.486][a] [−0.534 0.505][b]	[−0.710, 0.290] [−0.730 0.310]	[−0.645, 0.355] [−0.665 0.375]
+ MTS	[−0.514, 0.136] [−0.534 0.178]	[−0.710, −0.004] [−0.730 0.017]	[−0.645, 0.011] [−0.665 0.044]
+ Income-MIV	[−0.490, −0.014] [−0.533 0.160]	[−0.708, −0.031] [−0.719 0.017]	[−0.645, −0.032] [−0.659 0.044]
With Ineligible-MIV:	[−0.114, 0.216] [−0.187 0.312]	[−0.055, 0.202] [−0.109 0.269]	
+ MTS	[−0.114, 0.062] [−0.187 0.149]	[−0.055, −0.036] [−0.117 0.029]	
+ Income-MIV	[−0.090, −0.014] [−0.182 0.149]	[−0.053, −0.037] [−0.117 0.000]	

[a] Point estimates of the population bounds.
[b] 90% confidence intervals around ATE calculated using methods from Imbens and Manski (2004) with 1000 pseudosamples.

To formally account for measurement error in classifying ineligibles, we begin as in Eq. (2) by decomposing the latent health probability among households classified as ineligible as follows:

$$P[Y(0)=1 \mid v_2] = P(Y=1, S=0 \mid v_2) + \theta^+_{v1} + P[Y(0)=1 \mid S^*=1, v_2][P(S=1 \mid v_2) - \theta^+_{v1} - \theta^+_{v0}],$$

where $\theta^+_{vi} = P(Y=i, S=1, S^*=0 \mid v_2)$. Setting $P[Y(0)=1 \mid S^*=1, v_2]$ to 0 and 1, respectively, for the lower and upper bound implies the following.

Ineligible-MIV bounds on $P[Y(0)=1 \mid v2]$ accounting for potentially mislabeled ineligibility:

$$P(Y=1, S=0 \mid v_2) \le P[Y(0)=1 \mid v_2] \le P(Y=1, S=0 \mid v_2) + P(S=1 \mid v_2). \quad (9)$$

Intuitively, $P[Y(0)=1 \mid v_2]$ is no longer point-identified unless all households in the "ineligible" group confirm that they are not participating in the school lunch program; i.e., $P(S=1 \mid v_2)=0$. In that case, the lower and upper bounds in Eq. (9) revert back to $P(Y=1 \mid v_2)$. This is the case for dropouts and for children attending schools without the NSLP. In contrast, 23.5% of respondents classified as income ineligible report receiving free or reduced-price lunches. Combining Assumption (A3) with the bounds on $P[Y(0)=1 \mid v_2]$ in Eq. (9) weakly narrows the bounds on the ATE.

In our application, this ineligible-MIV assumption has substantial identifying power. For the food insecurity outcome, for example, $P[Y(0)=1]$ is constrained to lie within [0.337, 0.410] with this MIV assumption and, in the case of no classification errors, [0.068, 0.811] otherwise. Thus, this MIV assumption reduces the width of the bounds on $P[Y(0)=1]$ under fully accurate reporting from 0.743 to 0.073 and the width of the bounds on the ATE from 1 to 0.330. For the poor health rate, the width of the bound on the ATE declines by about three-quarters, from 1 to 0.257.

4.3. Main results

The partial identification approach developed above allows us to evaluate bounds on the ATE of the free lunch program under different assumptions about the selection and measurement error problems. This effectively allows one to assess the sensitivity of inferences to different identifying restrictions. We begin by examining the estimated ATE under the assumption of fully accurate reporting of participation in the NSLP. This no-classification error assumption is the norm in the literature, and thus serves as a useful benchmark.

Without classification errors, ambiguity still arises from the selection problem. Table 3 displays the estimated MIV bounds for all three outcomes under the various MIV assumptions. For the food insecurity and poor health outcomes, we apply the (A1) income-MIV, (A2) MTS, and (A3) ineligible-MIV assumptions. For obesity, we apply only the income-MIV and MTS assumptions.

By layering on different sets of assumptions, the results clearly illustrate the sensitivity of inferences on the ATE to the different MIV assumptions. In some cases, we learn very little about the ATE, whereas in others the ATE is nearly point-identified. Consider, for example, inferring the impact of the free and reduced-price lunch program on the rate of food insecurity. The estimated bounds are [−0.514, 0.136] under the MTS assumption and [−0.114, 0.216] under the ineligible-MIV assumption. While both of these MIV assumptions substantially reduce the ambiguity created by the selection problem, there still remains much uncertainty about the ATE under both models. Under the combined MTS and ineligible-MIV assumptions, the bounds narrow to [−0.114, 0.062] and then narrow further to [−0.090, −0.014] if we also impose the income-MIV assumption. Thus, given all three MIV assumptions, the bounds on the ATE narrow to an 8 percentage point range with the sign of the ATE identified as negative: that is, the impact of the free lunch program on food insecurity appears to be at least somewhat beneficial. While this estimated bound is strictly negative, the confidence interval covers zero; we cannot reject the hypothesis that the program is ineffective in reducing food insecurity.

The most striking finding revealed in Table 3 is that the joint income-MIV–MTS model identifies the ATE as strictly negative for all three health outcomes. Under the assumption that participation is accurately reported, these estimates suggest that the free lunch program reduces food insecurity by at least 1.4 percentage points, poor health by 3.1 percentage points, and obesity by 3.2 percentage points. These percentage point declines are especially large for poor health and obesity. In the absence of the NSLP, our estimates of $P[Y(0)=1]$ indicate that at least 10.7% of eligible children would be in poor health, and at least 18.5% would be obese. Thus, these estimates indicate that the program has reduced the rate of poor health by at least 29% (=3.1/10.7) and the rate of obesity by at least 17% (=3.2/18.5). The impact on food insecurity is smaller, but these estimates still suggest at least a 3.8% (=1.4/37.2) decline.

While these findings indicate that the NSLP plays an important role in improving children's health, there are two reasons to temper conclusions based on this evidence. First, even with fully accurate reporting, the 90% confidence intervals include zero. Thus, we cannot reject the hypothesis that the program is ineffective in promoting healthy outcomes. Second, allowing for classification errors will increase the width of the bounds. Even though we saw earlier that the presence of classification errors can actually reduce the upper bound given knowledge about the true participation rate, P^*, identification of the sign of the ATE may nevertheless be precluded in models with classification errors.

Returning to Figs. 1A–1C, the estimated bounds on the ATE are traced out for the three outcomes under the different MIV assumptions and for classification error models where the true participation rate, P^*, varies from 0 to 1. These figures reveal the sensitivity of inferences on the ATE to this value. When $P^* = 0.89$, consistent with the administrative data, we can still identify under the joint ineligible-MIV–MTS model that the NSLP reduces poor health by at least 2.6 percentage points, or 24% (compared with a 29% decline above under no misreporting). This result does not depend on whether we impose the income-MIV restriction.

The estimated MIV bounds do not identify the sign of the ATE for the food insecurity and obesity outcomes, however, for any conjectured true participation rate, P^*. The results are especially sensitive for obesity, where the upper bound on the ATE under the MTS model increases from 0.011 when there are no errors to 0.218 when the true participation rate is known to equal 0.89. In contrast, under various ineligible-MIV models, the estimated upper bounds on the food insecurity outcome are not as sensitive to the presence of classification errors when the true participation rate is relatively large. For example, under the joint ineligible-MIV–MTS model, the estimated bounds change from [−0.114, 0.062] when there are no errors to [−0.225, 0.072] when the true participation rate is known to equal 0.89. Thus, while the lower bound is quite sensitive to classification errors, the upper bound rises only slightly. While these findings do not identify the sign of the ATE for poor health and obesity outcomes, the ambiguity created by the selection and measurement problems is notably reduced under these MIV models.

5. Monotone treatment response

Despite the observed positive correlations in the data, there is a general consensus among policymakers and researchers that the food assistance programs such as the NSLP would not increase the rate of food insecurity (Currie, 2003). Long (1991) finds that each additional dollar of benefits leads to about a 50 cent increase in total food expenditures. If food is a normal good, providing in-kind food benefits should weakly increase the consumption of food and, in turn, decrease the prevalence of food insecurity and poor health.

Given this general consensus about the effect of the NSLP, we apply the *monotone treatment response* (MTR) assumption (Manski, 1995, 1997), which formalizes the common idea that the free lunch program would not lead to a reduction in health status.

(A4) *Monotone treatment response (MTR)*: $Y(0) \geq Y(1)$.

The MTR assumption implies that the ATE must be nonnegative: the free lunch program, by assumption, cannot increase the probability of a poor health outcome.

Both the MTS and MTR assumptions reduce the upper bound on $P[Y(1) = 1]$ and the lower bound on $P[Y(0) = 1]$.[18] Combining the MTS and MTR assumptions implies the following.

Bounds on $P[Y(1) = 1]$ *and* $P[Y(0) = 1]$ *under the joint MTR–MTS assumptions:*

$$P[Y(1) = 1] \leq \min\{P[Y(1) = 1 \mid S^* = 1], P(Y = 1)\}$$

and

$$P[Y(0) = 1] \geq \max\{P[Y(0) = 1 \mid S^* = 0], P(Y = 1)\}.$$

Under this joint assumption, which can be combined with the other MIV assumptions, the upper bound on the ATE is nonpositive. The lower bound on the average treatment effect is unaffected.

While the MTR assumption may be relatively noncontroversial in the context of food insecurity and general health, there is much debate about whether the NSLP might decrease obesity. Because NSLP administrators must adhere to nutritional guidelines, one might expect the free lunch program to reduce obesity. Yet, the evidence provides a mixed picture suggesting that school lunches lead to some improved nutrient intake but also a higher portion of fat-related calories associated with obesity (see, e.g., Millimet et al., 2010). Moreover, the receipt of free or reduced-price meals through the NSLP allows families to purchase more food, which in turn could contribute to obesity, but also better quality food, which might lead to reductions in obesity. Overall, we find the MTR assumption to be less credible for the obesity outcome than for the other two outcomes.

Table 4 displays the estimated bounds on the ATE when combining the MTR assumption with the MIV assumptions at selected values of P^* and for the case of no misreporting. Under this joint MIV–MTS–MTR model, the estimated bounds are strictly negative even when we allow for classification errors. For example, the estimated bounds on ATE for the food insecurity rates vary from the 23-point range [−0.266, −0.032] under the assumption of no net misreporting to the 13-point range [−0.158, −0.032] when the true participation rate is known to equal 0.89. Without errors, the ATE is estimated to lie within the 7-point range [−0.090, −0.023]. Under these models, we find that the free lunch program is estimated to reduce food insecurity by at least 2.3 percentage points or 6% (2.3/37.2) and perhaps much more.

The estimated average treatment effects are also negative for both the poor health and obesity outcomes. For poor health outcomes, the ATE is estimated to lie within the narrow range of [−0.053, −0.037] if there are no classification errors and [−0.072, −0.035] when the true participation rate is known to equal 0.89. Thus, we estimate that the free lunch program reduces the incidence of poor health by at least 3.5 percentage points or 33% (3.5/10.7) and as much as 7.2 percentage points or 67% (7.2/10.7). There is much more uncertainty for obesity where the eligibility criteria MIV assumptions are not applied, but we still estimate a strictly beneficial effect of the program under the MIV–MTS–MTR assumption. In particular, we find that the free lunch program reduces obesity by at least 4 percentage points, or 21 percentage (4.0/18.5).

6. Conclusion

Children receiving free or reduced-price school lunches through the National School Lunch Program tend to have worse health outcomes on average than observationally similar children who do not participate, especially in the case of food insecurity. Whether these puzzling correlations reflect causal impacts of the program has become a matter of considerable debate among researchers and policymakers. Much of the empirical literature maintains the untenable exogenous selection assumption, and the systematic classification error problem appears to have been completely ignored. Reviewing the general literature on the causal impacts of food assistance programs, Currie (2003) goes so far as to conclude that "many studies have … simply 'punted' on the issue of identification". Bhattacharya et al. (2006), in studying the National School Breakfast Program, suggest that "no study has dealt convincingly with endogenous participation".

18 Manski and Pepper (2000) show that the joint MTR–MTS assumption is testable when these two restrictions impact different sides of the/ bounds. In their application, for example, the joint MTR–MTS assumption implies that

$$E[Y(0) \mid S = 0] \leq E[Y(1) \mid S = 0] \leq E[Y(1) \mid S = 1],$$

where the observed lower bound reflects the MTR assumption and the observed upper bound reflects the MTS assumption. In that case, the MTR–MTS assumption can be tested by confirming that $E[Y \mid S = 0] \leq E[Y \mid S = 1]$. In our application, however, the joint MTR–MTS model implies that $E[Y(1) \mid S = 0] \leq \min\{E[Y(0) \mid S = 0], E[Y(1) \mid S = 1]\}$, which does not yield a testable restriction.

Please cite this article in press as: Gundersen, C., et al., The impact of the National School Lunch Program on child health: A nonparametric bounds analysis. Journal of Econometrics (2011), doi:10.1016/j.jeconom.2011.06.007

Table 4
Bounds on the ATE with misreported participation status: joint MTS, MIV, and MTR assumptions.

Selected values of P^*			
	$P^* = P = 0.74$: No misreporting	$P^* = P = 0.74$: No net misreporting	$P^* = 0.89$: Underreporting
Food insecurity[a]:	[−0.090, −0.023] [−0.181 0.000]	[−0.266, −0.032] [−0.366 0.000]	[−0.158,−0.032] [−0.252 0.000]
Poor health[a]:	[−0.053, −0.037] [−0.117 0.000]	[−0.107, −0.035] [−0.161 0.000]	[−0.072, −0.035] [−0.133 0.000]
Obesity[b]:	[−0.645, −0.040] [−0.659 0.000]	[−0.934, −0.040] [−0.951 0.000]	[−0.889, −0.040] [−0.939 0.000]

[a] Includes income-MIV and ineligible-MIV.
[b] Includes income-MIV but not ineligible-MIV.

Extending methods developed in KPGJ (who study the impacts of SNAP), our analysis considers the impact of the free lunch program on health-related outcomes using nonparametric methods that allow us to simultaneously account for the selection and misclassification problems in a single unifying framework. To address the classification error problem, we combine survey data from the NHANES with administrative data from the USDA to place constraints on the magnitudes and patterns of participation reporting errors. To address the selection problem, we apply a range of MIV assumptions as well as the MTR assumption. Most notably, we introduce a new way to conceptualize the MIV assumption using eligibility criteria as monotone instruments. The ineligible-MIV assumption provides substantial identifying power, even when allowing for potential mislabeling of ineligibility status among households claiming to receive benefits. Beyond the school lunch setting, the idea of using ineligible respondents as MIVs may have wide applicability in the program evaluation literature.

By successively layering stronger identifying assumptions into the model, our approach makes transparent how assumptions on the selection and reporting error processes shape inferences about the causal impacts of the program. For our preferred MIV models, the results imply that without classification errors the free lunch program leads to substantial reductions in food insecurity, poor health, and obesity. In particular, estimates from the joint MIV–MTS model reveal that the program reduces the prevalence of food insecurity by at least 3.8%, the rate of poor health by at least 29%, and the rate of obesity by at least 17%. This finding for obesity stands in contrast to much of the food assistance literature that finds that the program is associated with increases in childhood obesity (e.g., Millimet et al. (2010)).

Although these results suggest that the NSLP leads to notable improvements in health outcomes, we cannot reject the hypothesis that the program is ineffective in promoting healthy outcomes. Moreover, administrative data suggests an NSLP participation rate of about 89%, implying systematic underreporting of benefits. Constraining patterns of false reports to be consistent with this participation rate, we can no longer sign the average treatment effect under the joint MIV–MTS model for the food insecurity and obesity outcomes. For the case of poor health, we can still identify that the program reduces the prevalence by at least 24%.

Finally, when we add the MTR assumption, the free lunch program is estimated to reduce the incidence of all three poor health outcomes even in models with classification errors. Under the joint MIV–MTS–MTR model, we find that the program reduces food insecurity by at least 6%, poor health by at least 33%, and obesity by at least 21%.

References

Bhattacharya, J., Currie, J., Haider, S., 2006. Breakfast of champions? the school breakfast program and the nutrition of children and families. Journal of Human Resources 41, 445–466.

Bollinger, C.R., 1996. Bounding mean regressions when a binary variable is mismeasured. Journal of Econometrics 73 (2), 387–399.

Case, A., Lubotsky, D., Paxson, C., 2002. Economic status and health in childhood: The origins of the gradient. American Economic Review 92 (5), 1308–1334.

Currie, J., 2003. US food and nutrition programs. In: Moffitt, R. (Ed.), Means Tested Transfer Programs in the US. University of Chicago Press, (Chapter 4).

Daponte, B., Sanders, S., Taylor, L., 1999. Why do low-income households now use food stamps? evidence from an experiment. Journal of Human Resources 34 (3), 612–628.

Deaton, A., 2002. Policy implications of the gradient of health and wealth. Health Affairs 21 (2), 13–30.

DePolt, R.A., Moffitt, R.A., Ribar, D.C., 2009. Food stamps, temporary assistance for needy families and food hardships in three American cities. Pacific Economic Review 14 (4), 445–4473.

Fontaine, K., Redden, D., Wang, C., Westfall, A., Allison, D., 2003. Years of life lost due to obesity. Journal of the American Medical Association 289 (2), 187–193.

Food and Nutrition Service. 2010. US Department of Agriculture, Food and Nutrition Service. http://www.fns.usda.gov/pd/slsummar.htm (accessed 10/26/10).

Ginther, D., 2000. Alternative estimates of the effect of schooling on earnings. Review of Economics and Statistics 82 (1), 103–116.

Gleason, P., Suitor, C., 2003. Eating at school: How the National School Lunch Program affects children's diets. American Journal of Agricultural Economics 85 (4), 1047–1061.

Gundersen, C., Kreider, B., 2008. Food stamps and food insecurity: what can be learned in the presence of nonclassical measurement error? Journal of Human Resources 43 (2), 352–382.

Gundersen, C., Kreider, B., 2009. Bounding the effects of food insecurity on children's health outcomes. Journal of Health Economics 28, 971–983.

Gundersen, C., Oliveira, V., 2001. The food stamp program and food insufficiency. American Journal of Agricultural Economics 84 (3), 875–887.

Horowitz, J.L., Manski, C.F., 1995. Identification and robustness with contaminated and corrupted data. Econometrica 63 (2), 281–302.

Imbens, G., Manski, C., 2004. Confidence intervals for partially identified parameters. Econometrica 72 (6), 1845–1857.

Kreider, B., 2010. Regression coefficient identification decay in the presence of infrequent classification errors. Review of Economics and Statistics 92 (4), 1017–1023.

Kreider, B., Hill, S.C., 2008. Partially identifying treatment effects with an application to covering the uninsured. Journal of Human Resources 44 (2), 409–449.

Kreider, B., Pepper, J.V., 2007. Disability and employment: reevaluating the evidence in light of reporting errors. Journal of the American Statistical Association 102 (478), 432–441.

Kreider, B., Pepper, J.V., 2011. Identification of expected outcomes in a data error mixing model with multiplicative mean independence. Journal of Business and Economic Statistics 29 (1), 49–60.

Kreider, B., Pepper, J., Gundersen, C., Jolliffe, D., 2011. Identifying the effects of food stamps on child health outcomes. Working Paper.

Long, S., 1991. Do the school nutrition programs supplement household food expenditures? Journal of Human Resources 26 (4), 654–678.

Manski, C.F., 1995. Identification Problems in the Social Sciences. Harvard University Press, Cambridge, MA.

Manski, C.F, 1997. Monotone Treatment Response. Econometrica 65 (6), 1311–1334.

Manski, C.F., Pepper, J.V., 2000. Monotone instrumental variables: With an application to the returns to schooling. Econometrica 68 (4), 997–1010.

Manski, C.F., Pepper, J.V., 2009. More on Monotone Instrumental Variables. The Econometrics Journal 200–216.

McCarthy, I.M., Tchernis, R., 2010. On the estimation and selection models when participation is endogenous and misclassified. Working Paper.

Meyer, B., D W, Mok, Sullivan, J.X., 2009. The under-reporting of transfers in household surveys: its nature and consequences. Working Paper.

Millimet, D., 2010. The elephant in the corner: a cautionary tale about measurement error in treatment effects models, Southern Methodist University and IZA Discussion Paper No. 5140.

Millimet, D.L., Tchernis, R., Husain, M., 2010. School nutrition programs and the incidence of childhood obesity. Journal of Human Resources 45 (3), 640–654.

Molinari, F., 2008. Partial identification of probability distributions with misclassified data. Journal of Econometrics 144 (1), 81–117.
Molinari, F., 2010. Missing treatments. Journal of Economic and Business Statistics 28 (1), 82–95.
Nord, M., Coleman-Jensen, A., Andrews, M., Carlson, S., 2010. Household food security in the united states, 2009. US Department of Agriculture, Economic Research Report 108. Washington, DC: US Government Printing Office.
Pepper, J.V., 2000. The intergenerational transmission of welfare receipt: a nonparametric bounds analysis. The Review of Economics and Statistics 82 (3), 472–488.
Schanzenbach, D., 2009. Does the federal school lunch program contribute to childhood obesity? Journal of Human Resources 44 (3), 684–709.
Shrewsbury, V., Wardle, J., 2008. Socioeconomic status and adiposity in childhood: a systematic review of cross-sectional studies 1990–2005. Obesity 16, 275–284.