IOWA STATE UNIVERSITY
Digital Repository

11-2007

Partially Identifying Treatment Effects with an Application to Covering the Uninsured

Brent Kreider
Iowa State University, bkreider@iastate.edu

Steven C. Hill
U.S. Department of Health & Human Services

Partially Identifying Treatment Effects with an Application to Covering the Uninsured

Abstract

We extend the nonparametric literature on partially identified probability distributions and use our analytical results to provide sharp bounds on the impact of universal health insurance on provider visits and medical expenditures. Our approach accounts for uncertainty about the reliability of self- reported insurance status as well as uncertainty created by unknown counterfactuals. We construct health insurance validation data using detailed information from the Medical Expenditure Panel Survey. Imposing relatively weak nonparametric assumptions, we estimate that under universal coverage monthly per capita provider visits and expenditures would rise by less than 8% and 16%, respectively, across the nonelderly population.

IOWA STATE UNIVERSITY

**Partially Identifying Treatment Effects with an
Application to Covering the Uninsured**

Brent Kreider, Steven C. Hill

April 2005

Working Paper # 05013

Department of Economics
Working Papers Series

Ames, Iowa 50011

Partially Identifying Treatment Effects with an Application to Covering the Uninsured[*]

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

Steven C. Hill
Center for Financing, Access and Cost Trends
Agency for Healthcare Research and Quality
SHill@AHRQ.gov

November 2007

Abstract: We extend the nonparametric literature on partially identified probability distributions and use our analytical results to provide sharp bounds on the impact of universal health insurance on provider visits and medical expenditures. Our approach accounts for uncertainty about the reliability of self-reported insurance status as well as uncertainty created by unknown counterfactuals. We construct health insurance validation data using detailed information from the Medical Expenditure Panel Survey. Imposing relatively weak nonparametric assumptions, we estimate that under universal coverage monthly per capita provider visits and expenditures would rise by less than 8% and 16%, respectively, across the nonelderly population.

Keywords: Partial identification, treatment effect, national health insurance, universal coverage, nonparametric bounds, classification error

JEL classification numbers: C14, C21, I18

I. Introduction

Numerous academic studies have investigated relationships between health insurance status and a wide variety of outcomes such as health care utilization, health status, labor supply, and participation in public assistance programs. In more than 70 articles surveyed by Gruber and Madrian (2004), Levy and Meltzer (2004), and Buchmueller et al. (2005), nearly all parameters of interest are identified using parametric approaches.[1]

We develop the first nonparametric framework for studying the potential impact of universal health insurance on the nation's use of medical services. Within this framework, we study relationships between insurance status and use of services (expenditures and number of provider visits) in an environment of uncertainty about both counterfactual utilization outcomes and status quo insurance status. Uncertainty about counterfactuals arises because insurance status is not randomly assigned. For example, families that expect to use more health services have more incentive to acquire health insurance. More generally, insurance status depends on individual and family characteristics that may also influence health care use. Utilization patterns among households who become insured though expanded programs would not necessarily resemble observed utilization patterns among households that had already self-selected into insured status on their own.

Beyond the uncertainty created by unknown counterfactuals, validation studies have recently called into question the reliability of households' responses to questions about their current insurance status. Significant misreporting has been documented in several popular surveys including the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP), the Behavioral Risk Factor Surveillance System (BRFSS) survey, the

---

[1] As an exception, Olson (1998) uses semiparametric techniques to estimate the relationship between women's labor hours and the availability of health insurance through a spouse. More recently, Gerfin and

1

Medical Expenditure Panel Survey (MEPS), and others (Davern et al. 2007; Card et al. 2004; Hill 2007/2008; Nelson et al. 2000). Estimated error rates, which vary across surveys, have been linked in part to difficulties in recalling past coverage and difficulties reporting the status of other family members (Nelson et al. 2000; Pascale 2007). The Census Bureau now issues caveats about the accuracy of insurance estimates from the CPS (DeNaves-Walt et al. 2005).

Using matched surveys of employers and their employees, Berger et al. (1998) find that 21% of the workers and their employers disagree about whether the worker was eligible for insurance. Their study appears to represent the only prior analysis of potentially mismeasured insurance status in an econometric framework. Assuming exogenous measurement error in a classical errors-in-variables setting (after accounting for the binary nature of the mismeasured variable), they find that even nonsystematic reporting error seriously biases their estimated effect of insurance eligibility on wage growth.

The presence of reporting errors compromise a researcher's ability to make reliable inferences about the status quo, and it further confounds identification of counterfactual outcomes associated with policies that would alter the distribution of insurance coverage within the population, such as national health insurance.[2] Consistent with such concerns, some advocates have argued that uncertainty about the numbers and characteristics of the uninsured constitutes an important barrier to identifying optimal policy solutions (e.g., Hunter 2004; Woolhandler and Himmelstein 2007). Highlighting surprising degrees of insurance classification error in many popular national surveys, along with dramatic inconsistencies in responses when experimental follow-up insurance questions have been posed, Czajka and Lewis

---

Schellhorn (2006) use nonparametric techniques and Swiss data to bound the effects of deductibles on the probability of visiting a physician.

[2] The extent to which universal coverage would increase use of services and expenditures has been estimated in a variety of parametric studies (Institute of Medicine 2003; Blumberg et al. 2006). Estimates of incremental spending range from $34 to $69 billion per year depending on the statistical assumptions and choice of comparison groups.

(1999) write: "Until we can make progress in separating the measurement error from the reality of uninsurance, our policy solutions will continue to be inefficient, and our ability to measure our successes will continue to be limited."

Our analysis extends the nonparametric literature on partially identified probability distributions in several dimensions. First, we provide sharp bounds on the conditional mean of a random variable (in our case health care visits or expenditures) for the case that a binary conditioning variable (insurance status) is subject to arbitrary endogenous classification error. In this environment, insurance reporting errors are allowed to be arbitrarily related to true insurance status and health care use. These results extend parts of the analyses of Horowitz and Manski (1998) and Kreider and Pepper (2007). Second, we formally assess how statistical identification of a treatment effect decays with the degree of uncertainty about the status quo. Our approach extends the nonparametric treatment effect literature for the case that some treatments are unobserved (especially Molinari, 2007a).[3] Third, we relax the nondifferential errors independence assumption evaluated, for example, by Bollinger (1996) and Bound et al. (2001) embedded in the classical errors-in-variables model. As an alternative, we evaluate the identifying power of a weaker "nonincreasing errors" monotonicity assumption that presumes that misreporting of insurance status does not rise with the level of utilization. This assumption allows for the possibility, for example, that using health services informs a patient of her true insurance status.[4]

---

[3]To isolate identification problems associated with partially unobserved insurance status as a conditioning variable or treatment, we assume that other variables in the analysis are measured without error.

[4]Using methods in Lewbel (2007), the treatment effects could be point-identified in certain cases if we had instruments that affect insurance status but not classification error or the average treatment effect. For related work on potentially endogenous classification errors in a linear regression framework, see Frazis and Loewenstein (2003). Given the difficulty in identifying plausible instruments in our application, we consider what can be identified in the absence of instruments.

We exploit detailed data in the 1996 MEPS to construct health insurance validation data for a nonrandom portion of the sample based on insurance cards, policy booklets, and follow-back interviews with employers and insurance companies. The next section describes these data and our health insurance verification strategy. Section III formalizes the identification problem associated with estimating a particular descriptive statistic, the gap in health service use between the status quo insured and uninsured under existing policies, when insurance status is subject to arbitrary patterns of classification error. We derive bounds on the unknown utilization gap under alternative assumptions about the nature and degree of reporting errors. Extending these results, Section IV investigates what can be learned about the impact of national health insurance on the use of health services. Combining our "nonincreasing errors" assumption with common monotonicity assumptions in the treatment effects literature, such as monotone treatment response (related to moral hazard) and monotone treatment selection (related to adverse selection), we can provide reasonably tight bounds on the impact of universal coverage without relying on some of the more controversial assumptions involving functional forms and independence.

Our primary set of estimates focuses on policies that would extend coverage to the uninsured using the same mixture of private and public insurance that exists under the status quo. Another set of estimates focuses on policies that would extend existing public programs to cover the uninsured. Both types of expansions have been implemented by states. Hawaii, Illinois, Pennsylvania, and Washington created public programs open to all uninsured children but with premiums for higher income families (Kaiser Commission on Medicaid and the Uninsured 2007). Maine, Massachusetts, and Vermont expanded public programs and developed new programs with subsidized premiums and other features (Kaye and Snyder 2007). Massachusetts also mandates that individuals purchase insurance, facilitated with subsidies for low-income adults and fees for employers that do not provide coverage (Holahan and Blumberg 2006). California and other states are considering proposals with features similar to those of

Massachusetts, and prominent presidential candidates also advocate covering the uninsured with a mixture of private and public insurance. Section V compares our nonparametric estimates with those from parametric studies, and Section VI concludes.

II. The Medical Expenditure Panel Survey

The data come from the 1996 Medical Expenditure Panel Survey (MEPS), a nationally representative household survey conducted by the U.S. Agency for Healthcare Research and Quality. In the MEPS Household Component (MEPS HC), each family (reporting unit) was interviewed five times over two and a half years to obtain annual data reflecting a two year reference period (Cohen 1997). This paper focuses on the nonelderly population because almost all adults become eligible for Medicare at age 65. The sample contains 18,851 individuals.

We study insurance and service use in July 1996. We focus on July because the 1996 MEPS has a follow-back survey of employers, unions, and insurance companies which reported insurance information as of July 1, 1996. We use 1996 data because that is the only year for which respondents to the follow-back survey reported on the employees' and policyholders' insurance status rather than whether the establishment offered insurance.[5] Studying insurance and service use in one month also reduces the likelihood of confounding the dynamics of insurance status with misreported insurance status because employment-related insurance typically covers an entire month.

A. Insurance Status Reported in the Household Component

The MEPS HC asks about insurance from a comprehensive list of all possible sources of insurance. In the first interview, conducted between March and August 1996, MEPS HC asked the family respondent about insurance held at any time since January 1st. Because employment-related insurance is the most prevalent source of insurance, the family respondent was asked

---

[5]These data are available at the AHRQ Data Center.

about all jobs held by coresiding family members since January 1st, jobs family members had retired from, and the last job held. The family respondent was asked whether the employee had insurance from each job. Then the family respondent was asked whether anyone had:

- Medicare
- Medicaid
- Champus/Champva
- For those who did not report Medicaid, any other type of health insurance through any state or local government agency which provided hospital and physician benefits
- Health benefits from other state programs or other public programs providing coverage for health care services[6]
- Other sources of private insurance, such as from a group or association, insurance company, previous employer, or union.

For each source of insurance, MEPS HC asked which family members were covered and when.[7]

In the second interview, conducted between August and December 1996, MEPS HC asked questions based on jobs and insurance reported to be held at the time of the first interview to determine whether previously reported insurance was still held or when it ended. MEPS also asked about new jobs and insurance from those jobs, public insurance acquired since the first interview, and insurance acquired from other sources since the first interview. The recall period is not especially long, typically four to seven months. Responses to the questions from the first and second interview were used to construct indicators of insurance coverage at any time during July 1996 and uninsurance, the residual category. Family respondents reported 80.7% of the nonelderly population were insured in July 1996 and 19.3% were uninsured (Table 1).

---

[6]A very small number of individuals are reportedly covered through Aid to Families with Dependent Children (AFDC) or Supplemental Security Income (SSI), and these are counted as Medicaid. Other sources, such as the Veterans Administration and the Indian Health Services, are not included in measures of hospital/physician insurance.

[7]State-specific program names are used in the questions. Insurance status is not imputed to families with missing data, which are rare.

B. Service Use and Expenditures

In each interview, the MEPS asks about health care services used by all coresiding family members since the last interview. The MEPS also obtains permission to interview a sample of the medical providers identified in the Household Component surveys to supplement household-reported health care expenditure and source of payment information. We create measures of service use and expenditures in July 1996: number of provider visits for ambulatory medical care (a medical provider visit, hospital outpatient visit, or emergency room visit), an indicator for whether the sample person had a hospital stay or ambulatory services, and expenditures for hospital stays and ambulatory services. Twenty-one percent of the (weighted) sample used medical care in that month.[8] Persons who the family respondent said were insured in July were nearly 80% more likely to have used medical care (22.5% of the insured versus 12.7% of the uninsured, Table 1). The mean number of provider visits is also much greater for the reportedly insured, as are mean expenditures.

C. Verification Data

We use detailed data to identify sample members for whom there is evidence corroborating their insurance status. The 1996 MEPS includes three sources that can be used to confirm health insurance reported by families: (1) the HC interviewers ask respondents to show insurance cards, (2) the HC interviewers ask respondents to provide policy booklets, and (3) separate interviews were conducted with family members' employers and insurance companies. Respondents for the family, employers, or insurance companies could err in reporting a person's insurance status; none provides a gold standard of information. Nonetheless, we use confirmations of insurance status to formally verify the insurance status of some sample members. This approach represents a compromise between taking reported insurance status at face value for all sample members and

---

[8] In the MEPS, outpatient prescription medications, medical supplies, and durable medical equipment are not linked to specific months; these expenditures are excluded.

discarding valuable family respondents' reports about insurance status.

We label sample members as verified insured if an insurance card was shown at the time of the interview, a policy booklet was given to the interviewer, or if an employer or insurance company confirmed that the person was covered by insurance. We assume that a report that a sample member is uninsured is accurate as long as there is no contradictory information from any family member's employers and all employers provided data. The person's insurance status was not verified (but not assumed to be incorrect) if there were insufficient verification data or if employers or insurance companies contradicted the family respondent. See Hill (2007/2008) for details.

As shown in Table 1, we find that 80.2% of the reportedly insured were confirmed as insured by a card, policy booklet, or an establishment. For the few cases in which a respondent produced an insurance card but the establishment reported that the person was uninsured, we treat these cases as verified insured based on the physical evidence of insurance. Among the reportedly uninsured, 11.7% are verified (Table 1). This relatively low number reflects the lack of an employed family member in some uninsured families and the lack of response by some employers. Recall that uninsurance is verified under this strategy only if all of the family's employers responded and confirmed that they did not provide insurance to the sample member. Overall, 67.0% of the sample was verified.

III. Identifying Utilization Differences Between the Insured and Uninsured

We now study what can be learned about differences in health care utilization between the insured and uninsured when we cannot perfectly distinguish between the truly insured and uninsured. In Section IV, we extend the analysis beyond descriptive statistics to assess what can be learned about the potential impacts of universal coverage. We ignore heterogeneity of insurance plans and treat coverage as a binary event. Let $I^* = 1$ indicate that a person is truly insured, with $I^* = 0$ otherwise. Instead of observing I^*, we observe the self-reported counterpart

8

I. A latent variable Z^* indicates whether a report is accurate. If *I* and I^* coincide, then $Z^* = 1$; otherwise, $Z^* = 0$. Let $Y = 1$ indicate that *I* is verified to be accurate (i.e., Z^* is known to equal 1). If $Y = 0$, then Z^* may be either 1 or 0.[9] In no case is the value of Z^* assumed to be 0.[10]

Let *U* denote the amount of health care services consumed during the reference period. Typically, the amount of care is measured as health expenditures or number of provider visits. Policymakers are also interested in the proportion of the population that uses any medical care, in which case *U* can be treated as a binary variable. In this section, we investigate what can be learned about the status quo utilization gap between the insured and uninsured,

$$\Delta = E(U \mid I^* = 1) - E(U \mid I^* = 0), \tag{1}$$

when true insurance status, I^*, is unobserved for part of the sample.[11]

The utilization gap Δ is not identified since we observe $E(U \mid I)$ but not $E(U \mid I^*)$. Our objective is to provide worst-case bounds on Δ. To partially identify $E(U \mid I^*)$, we will first derive bounds on the fraction of the population that consumes no more than a particular amount of care *t* conditional on unobserved insurance status, $P(U \leq t \mid I^*)$. We can then provide bounds on $E(U \mid I^*)$ by integrating over these worst-case probabilities.

We begin by writing

$$P(U \leq t \mid I^* = 1) = \frac{P(U \leq t, I^* = 1)}{P(I^* = 1)}. \tag{2}$$

Neither the numerator nor the denominator is identified, but assumptions on the pattern of

---

[9] In their analysis of testing for environmental pollutants, Dominitz and Sherman (2004) were the first to formalize the idea of distinguishing between "verified" and "unverified" observations in the data.

[10] That is, we conservatively allow for the possibility that the MEPS insurance classification is accurate even if the classification is not formally verified.

[11] Our notation leaves implicit any other covariates of interest. We focus on bounding the utilization gap for the nonelderly population as a whole, but it is straightforward to condition on subpopulations of interest. Note that we are not estimating a regression, and there are no regression orthogonality

classification errors can place restrictions on relationships between the unobserved quantities.

Let $\theta_t^+ \equiv P(U \leq t, I = 1, Z^* = 0)$ and $\theta_t^- \equiv P(U \leq t, I = 0, Z^* = 0)$ denote the fraction of false

positive and false negative insurance classifications, respectively, for those whose medical

consumption did not exceed t. Let $\theta_t'^+ \equiv P(U > t, I = 1, Z^* = 0)$ and $\theta_t'^- \equiv P(U > t, I = 0, Z^* = 0)$

denote the analogous fractions for those whose use of care exceeded t. We can then decompose

the numerator and denominator in (2) into identified and unidentified quantities:

$$P(U \leq t \mid I^* = 1) = \frac{P(U \leq t, I = 1) + \theta_t^- - \theta_t^+}{P(I = 1) + \left(\theta_t^- + \theta_t'^-\right) - \left(\theta_t^+ + \theta_t'^+\right)} \tag{3}$$

where $P(U \leq t, I = 1)$ and $P(I = 1)$ are identified by the data. In the numerator, $\theta_t^- - \theta_t^+$ reflects

the unobserved excess of false negative versus false positive insurance classifications among

those whose use of services did not exceed t. In the denominator, $\left(\theta_t^- + \theta_t'^-\right) - \left(\theta_t^+ + \theta_t'^+\right)$ reflects

the unobserved excess of false negative versus false positive insurance classifications within the

entire population. Utilization among the uninsured, $P(U \leq t \mid I^* = 0)$, can be decomposed in a

similar fashion.

We now assess what can be learned about Δ. First, we present "arbitrary errors" bounds

that impose no structure on the distribution of false positives and false negatives. This

environment that allows for arbitrary error patterns is termed "corrupt sampling" (Horowitz and

Manski 1995). We then consider the identifying power of assumptions that restrict the patterns

of errors. Results in this section also inform our treatment effect bounds in Section IV.

A. Arbitrary error bounds

Our analytic framework allows us to trace out bounds on our unknown parameters of interest as

function of a researcher's confidence in reported insurance status. Following Horowitz and

Manski (1995), we consider a lower bound, v, on the accuracy rate among unverified cases.

conditions to be satisfied.

Setting $v = 1$ corresponds to an assumption that reported insurance status is always accurate, the implicit assumption in previous analyses. Setting $v = 0$ corresponds to an assumption that nothing is known about the accuracy of unverified reports. We will evaluate patterns of identification decay for our parameters of interest as v departs from 1.

Since we do not observe true insurance status for unverified responses, the accuracy of such responses cannot logically be known. Nevertheless, our empirical analysis considers four candidates for v: 0.50, 0.84, 0.92, and 1. For the most conservative value, $v = 0.50,$ the researcher need only make the common classification error assumption that accurate responses are more prevalent than reporting errors (see, e.g., Bollinger, 1996 and Frazis and Loewenstein, 2003).

The other candidate thresholds, $v = 0.84$ and $v = 0.92,$ are derived from studies that have investigated the reliability of self-reported insurance status in the MEPS and other surveys. Among the reportedly insured ($I = 1$), let $R = 1$ and $M = 1$ denote reported private and public coverage, respectively, and let R^* and M^* denote true coverage. If both private and public coverage are reported in the month, which is rare, private coverage is assumed to take precedence; in such cases, we label reported coverage as $R = 1$ and $M = 0$ (and analogously for true coverage). Suppressing the conditioning on $Y = 0,$ the probability of misclassification is given by

$$P(Z^* = 0) = P(I^* = 0 \mid R = 1)P(R = 1) + P(I^* = 0 \mid M = 1)P(M = 1)$$
$$+ P(R^* = 1 \mid I = 0)P(I = 0) + P(M^* = 1 \mid I = 0)P(I = 0).$$

For the first term, we conservatively allow the proportion that is truly uninsured among those reporting private coverage, $P(I^* = 0 \mid R = 1),$ to be as high as the proportion simply lacking private coverage, $P(R^* = 0 \mid R = 1).$ Similarly, we allow the true rate of being uninsured among those reporting public coverage, $P(I^* = 0 \mid M = 1),$ to be as high as the rate of simply lacking public coverage, $P(M^* = 0 \mid M = 1).$ Imposing these inequalities, we can write

$$P(Z^* = 1) \geq v = 1 - P(R^* = 0 \mid R = 1)P(R = 1) - P(M^* = 0 \mid M = 1)P(M = 1)$$
$$- P(R^* = 1 \mid I = 0)P(I = 0) - P(M^* = 1 \mid I = 0)P(I = 0).$$

The observed probabilities among the unverified cases are $P(R = 1) = 0.35$, $P(M = 1) = 0.13$, and $P(I = 0) = 0.51$. Estimates of the unobserved contradiction rates are inferred from other studies. Hill (2007/2008) estimates that the probability of truly lacking private insurance when it was reported, $P(R^* = 0 \mid R = 1)$, is about 0.02, with the possibility that it could be as large as 0.12. Our two estimated thresholds, $v = 0.84$ and $v = 0.92$, incorporate the larger and smaller values, respectively. The probability of truly having private insurance when no insurance was reported, $P(R^* = 1 \mid I = 0)$, is estimated to be no higher than 0.10.

For both values of v, we assume that the rate of false positives among those reporting public coverage, $P(M^* = 0 \mid M = 1)$, does not exceed 0.03. Card et al. (2004) estimate that the rate of false positives for Medicaid in the SIPP ranges from 0.01 to 0.03. Like the MEPS, the SIPP is a longitudinal, household survey. The MEPS and the SIPP also have similar recall periods and yield fairly similar estimates of enrollment in Medicaid and the State Children's Health Insurance Program (SCHIP) (Peterson and Grady 2005). Estimates of $P(M^* = 1 \mid I = 0)$ range from 0.02 (for $v = 0.92$) to 0.12 (for $v = 0.84$). These values are derived from benchmarking studies of the MEPS (Banthin and Sing, forthcoming; Nelson 2003; Peterson and Grady 2005), with the lower value reflecting an estimate that most (83%) of those who failed to report their public insurance instead reported private insurance (Call et al. 2007). Further motivation is provided in an appendix available from the authors.

While we have attempted to be conservative in our derivations of v, these derivations necessarily involve extrapolating information from respondents for whom validation data are available. Misreporting rates in our sample of unverified respondents may not reflect estimated misreporting rates in other samples. While the descriptive utilization gap results presented later in this section are sensitive to values of v, our main conclusions in Section IV about the impact

of universal health insurance on utilization (which impose additional monotonicity assumptions)

are fairly robust to the choice of v across a wide range of values.

In Proposition 1, we formalize what can be known about the utilization gap between the

truly insured and truly uninsured given only a lower bound, v, on the fraction of accurate

insurance classifications among unverified cases:

Proposition 1: *Let* $P(Z^* = 1 | Y = 0) \geq v$, *and suppose that nothing is known about the pattern of*

reporting errors. Then the mean utilization rate among the truly insured is bounded sharply as

follows:

$$\int U dF_H \leq E(U \mid I^* = 1) \leq \int U dF_L$$

using the distribution functions

$$F_L(t) = \frac{P(U \leq t, I = 1) - \alpha_t^+}{P(I = 1) - \alpha_t^+ + \min\left\{\phi(v) - \alpha_t^+, P(U > t, I = 0, Y = 0)\right\}}$$

$$F_H(t) = \frac{P(U \leq t, I = 1) + \alpha_t^-}{P(I = 1) + \alpha_t^- - \min\left\{\phi(v) - \alpha_t^-, P(U > t, I = 1, Y = 0)\right\}}$$

and values

$$\alpha_t^+ = \begin{cases} \min\left\{\phi(v), P(U \leq t, I = 1, Y = 0)\right\} & \text{if } \delta_t^L < 0 \\ \max\left\{0, \min\left\{P(U \leq t, I = 1, Y = 0), \phi(v) - P(U > t, I = 0, Y = 0)\right\}\right\} & \text{otherwise} \end{cases}$$

$$\alpha_t^- = \begin{cases} \min\left\{\phi(v), P(U \leq t, I = 0, Y = 0)\right\} & \text{if } \delta_t^H \geq 0 \\ \max\left\{0, \min\left\{P(U \leq t, I = 0, Y = 0), \phi(v) - P(U > t, I = 1, Y = 0)\right\}\right\} & \text{otherwise.} \end{cases}$$

Analogous bounds for the utilization rate among the uninsured, $E(U \mid I^* = 0)$, are obtained by

replacing $I = 1$ with $I = 0$ and vice versa.

Proof. See Appendix.

Notice that increasing v narrows the bounds over some ranges of v but not others, and

the rate of identification decay can be highly nonlinear as v declines. When $v = 0$, these bounds

can be derived from Horowitz and Manski's (1998) censored regressor bounds (their Section

4.1). In that context, some observations of a conditioning variable are missing in the data.

Kreider and Pepper's (2007) Proposition 2 bounds apply when $v = 0$ and the outcome U is

binary.[12]

   We next bound the difference in use between the insured and uninsured. We could

compute a valid lower (upper) bound on the utilization gap, Δ, by subtracting the Proposition 1

upper (lower) bound on $E(U \mid I^* = 0)$ from the Proposition 1 lower (upper) bound on

$E(U \mid I^* = 1)$. While these bounds on Δ would be valid, they would not necessarily be as tight as

possible. In particular, they would not impose the constraints that the parameters α_t^+ (α_t^-) in the

Proposition 1 bounds on $E(U \mid I^* = 1)$ are identical to the parameters α_t^- (α_t^+) in the bounds on

$E(U \mid I^* = 0)$. Therefore, we compute sharp bounds on Δ using numerical methods that impose

these constraints.[13]

B. Arbitrary error results

Table 2 presents estimated bounds on the utilization gap, Δ, for any use of services, number of

provider visits, and expenditures. The arbitrary error bounds are provided in column (1). When

$v = 1$, Δ is point-identified as the self-reported gap obtained from taking the data at face value.

For example, the gap in the proportion of insured and uninsured that used services in July 1996 is

point-identified as $P(U = 1 \mid I = 1) - P(U = 1 \mid I = 0) = 0.098$. The gaps in the number of provider

visits and expenditures are point identified as 0.186 and \$77, respectively. Ninety-five percent

---

[12]Kreider and Pepper (2007) study how labor force participation varies with disability status given a lack
of knowledge of any particular respondent's true disability status. Our Proposition 1 extends their
Proposition 2 by considering continuous outcomes and by assessing identification for values of v greater
than 0 within unverified classifications. Their proposition, however, is more general in the dimension that
they do not impose our identifying assumption that all verified cases are accurate; they allow for the
possibility of errors within verified cases.

[13] Gauss programs are available from the authors upon request.

confidence intervals for the gap in any use, number of visits, and expenditures are calculated as [0.083, 0.112], [0.136, 0.237], and [$47, $107], respectively.[14]

Under arbitrary errors, identification of these utilization gaps deteriorates rapidly as v departs from 1. For the reference case $v = 0.92$, for example, the difference in the number of provider visits per month may lie anywhere between -0.135 and 0.472, while the difference in expenditures may lie anywhere between -$40 and $117 (Table 2). In neither case is the sign of Δ identified, even ignoring the additional uncertainty associated with sampling variability. In fact, without imposing assumptions on the patterns of errors, we cannot identify whether the insured were more likely to use services than the uninsured unless v exceeds 0.95 (not shown). Similarly, we cannot identify whether the insured had more provider visits unless v exceeds 0.98, nor can we identify whether the insured had greater expenditures unless v exceeds 0.99. This represents an important negative result: being almost fully confident in the accuracy of the data is not enough, by itself, to be informative about even the sign of the utilization gap between the insured and uninsured.

C. Restrictions on reporting error patterns

The parameter bounds thus far have allowed for arbitrary patterns of insurance classification errors, including the possibility that reporting errors are endogenously related to true insurance status or the health care utilization outcome. In contrast, most economic research presumes that measurement error is exogenous to the extent that it exists at all. In this section, we make transparent the identifying power of two common (nonnested) independence assumptions that tighten the Proposition 1 bounds. Then we introduce a weaker alternative assumption that is more plausible in our context.

---

[14] Throughout this analysis, we compute confidence intervals around the estimated identification regions using methods recently developed by Chernozhukov et al. (2007). We use balanced repeated replication methods to account for the complex survey design (Wolter 1985).

First, a researcher might consider an orthogonality assumption that insurance

classification errors arise independently of true insurance status:

$$P(I^* = 1 | Z^*) = P(I^* = 1). \tag{4}$$

This assumption may be relatively innocuous compared with the set of homogeneity and

exogeneity assumptions imposed in standard parametric frameworks. Still, stories can be told in

which this assumption may be violated. Reporting errors may not be orthogonal to true

insurance status if, for example, better educated respondents are both more likely to be insured

and more likely to accurately answer survey questions. Similarly, Card et al. (2004) provide

evidence that errors in reporting Medicaid coverage vary with family income, which is also a key

aspect of Medicaid eligibility.

Alternatively, or in combination with (4), a researcher might place restrictions on the

relationship between insurance classification errors and the use of health services. In the popular

classical measurement error framework, reported insurance status does not depend on the level of

health care utilization conditional on true insurance status:

$$P(I = 1 | I^*, U) = P(I = 1 | I^*) \text{ for } I^* = 0,1. \tag{5}$$

Aigner (1973) and Bollinger (1996) study this type of "nondifferential" classification error for

the case of a binary conditioning variable. When the independence assumption (5) holds,

Bollinger's Theorem 1 can be used to show that Δ is bounded below by the reported utilization

gap $E(U | I = 1) - E(U | I = 0)$ (> 0) as long as v exceeds 0.50. Reflecting well-known

attenuation bias associated with random measurement error, the magnitude of the reported

utilization gap represents a downward-biased estimate of the magnitude of the true utilization

gap. Berger et al. (1998) impose the nondifferential errors assumption in the only previous

econometric analysis that allows for misreported insurance status.

Bound et al. (2001, p. 3725) note, however, that in general the nondifferential

measurement error assumption is strong and often implausible. In our context, the

nondifferential assumption is most likely to be violated if using health care informs respondents

about their true insurance status. For example, a health care provider may enroll a patient in Medicaid. More generally, a regular user of health services (or a user with high expenditures) presumably is more likely to know her insurance status than an infrequent user of services.

We propose a weaker alternative assumption on the pattern of reporting errors. Relaxing the nondifferential assumption in (5), we suppose that the probability of misreporting insurance status does not rise with the level of health care utilization:

$$P(I = 1 | I^* = 0, U_1) \le P(I = 1 | I^* = 0, U_0) \tag{6}$$

$$P(I = 0 | I^* = 1, U_1) \le P(I = 0 | I^* = 1, U_0)$$

for $U_1 \ge U_0$. The nondifferential assumption represents a special case. In the next section, we illustrate how the identifying power of this monotone "nonincreasing error rate" assumption compares with the standard nondifferential errors assumption.

D. Restricted error pattern results

Columns (2)-(4) in Table 2 illustrate the identifying power of these stronger assumptions. In each column, restricting the patterns of reporting errors translates into considerably narrower bounds compared with the arbitrary errors case. We focus primarily on the relatively weak "nonincreasing errors" bounds in column (4) that do not require any independence assumptions.[15]

Recall that under arbitrary errors, we cannot identify the sign of Δ for provider visits unless at least 98% of the unverified responses are known to be accurate; for expenditures, the corresponding critical value is 99%. Under the nonincreasing errors assumption, however, the critical values fall dramatically to 78% and 64%, respectively (not shown). For the reference case $v = 0.92$ under nonincreasing errors, Δ is estimated to lie within [0.133, 0.264] for the mean

---

[15]We numerically computed bounds under the various assumptions by searching over logically allowed combinations of false positives and false negatives $\{\theta_t^-, \theta_t^{'-}, \theta_t^+, \theta_t^{'+}\}$.

number of visits per month and within [$68, $85] for expenditures (Table 2). These bounds for visits are 7 percentage points wider than the corresponding nondifferential errors bounds in column (3), but they are 48 percentage points narrower than the arbitrary errors bounds in column (1). Similarly, the nonincreasing errors bounds for expenditures are $14 wider than the nondifferential errors bounds, but they are $140 narrower than the arbitrary errors bounds. The comparisons are similar after accounting for sampling variability.

IV. Utilization under Universal Health Insurance

We now turn to inferences about health care utilization under a hypothetical policy of universal health insurance. Let $U(I^* = 1)$ denote the amount of health services an individual would have used in July 1996 if insured. This outcome is observed in the data only for sample members who are verified to be currently insured; it is unobserved for those verified to be uninsured and for those whose insurance status is not verified. We wish to learn the population's expected utilization if everyone were insured, $E[U(I^* = 1)]$. If current insurance status were randomly assigned, then the utilization among the currently insured, $E(U \mid I^* = 1)$, would represent the best prediction of the utilization rate under universal coverage. Since I^* is not observed for all individuals, we could instead bound $E(U \mid I^* = 1)$ using the methods derived in the previous sections. As discussed earlier, however, the observed distribution of health insurance coverage in the population is not randomly assigned. Instead, insurance status is affected by characteristics potentially related to the use of medical resources.

In the absence of random assignment or other assumptions, the quantity $E[U(I^* = 1)]$ is not identified even if reported insurance status is always accurate. Unlike identification of the conditional utilization rate $E(U \mid I^* = 1)$, identification of the "treatment" outcome $E[U(I^* = 1)]$ requires knowledge about the counterfactual utilization rate of the uninsured had they instead been insured. Uncertainty about the accuracy of reported insurance status, the focus of the

current paper, further complicates identification of counterfactuals.

To bound the impact of universal coverage on utilization, we begin by using the law of total probability to decompose the projected utilization rate under universal coverage into verified and unverified current insurance status:

$$E[U(I^* = 1)] = E[U(I^* = 1) \mid Y = 1]P(Y = 1) + E[U(I^* = 1) \mid Y = 0]P(Y = 0). \qquad (7)$$

The data identify $P(Y = 1)$ and $P(Y = 0)$ but neither utilization term. The first term involving verified insurance status can be written as

$$E[U(I^* = 1) \mid Y = 1] = E(U \mid I^* = 1, Y = 1)P_{11} + E[U(I^* = 1) \mid I^* = 0, Y = 1](1 - P_{11}) \qquad (8)$$

where $P_{11} \equiv P(I^* = 1 \mid Y = 1)$ denotes the status quo insurance rate among verified cases. All the terms in (8) are observed except for the counterfactual expected utilization among the uninsured under the status quo, $E[U(I^* = 1) \mid I^* = 0, Y = 1]$. Without additional assumptions, this quantity may lie anywhere within the support of U, $[0, \sup U]$.

Returning to (7) and decomposing the third term involving the unverified cases obtains

$$E[U(I^* = 1) \mid Y = 0] = E(U \mid I^* = 1, Y = 0)P_{10} + E[U(I^* = 1) \mid I^* = 0, Y = 0](1 - P_{10}) \qquad (9)$$

where $P_{10} \equiv P(I^* = 1 \mid Y = 0)$ is the status quo insurance rate among unverified cases. None of the quantities in (9) are identified. We do not know P_{10}, and we cannot match health care use outcomes to insurance status when insurance status is unknown. Introducing this framework under the implicit assumption that $v = 0$, Molinari (2007a)'s innovative analysis shows that we can learn something about the first term, $E(U \mid I^* = 1, Y = 0)$, if the researcher has outside information restricting the range of P_{10} (denoted p in her framework).[16] She estimates the treatment effect of drug use on employment when drug use is unobserved for part of the sample.

---

[16] In Molinari's framework, $Y = 0$ (our notation) denotes survey nonresponse instead of lack of verification. Molinari (2007b) presents a general treatment of the identification problem for a variety of measurement issues.

We extend her analysis in two dimensions when $v > 0$.

First, an assumption on v translates into internally-generated restrictions on P_{10} as a function of v. We can write the unobserved insurance rate among unverified cases, P_{10}, as a function of the reported rate and unobserved misclassification rates:

$$P_{10} = P(I = 1 | Y = 0) + P(I = 0, Z^* = 0 | Y = 0) - P(I = 1, Z^* = 0 | Y = 0).$$

Allowing the unidentified terms to vary over their feasible ranges obtains $P_{10} \in [\underline{P}_{10}, \overline{P}_{10}]$ where

$$\underline{P}_{10} \equiv P(I = 1 | Y = 0) - \min\{1 - v, P(I = 1 | Y = 0)\} \tag{10}$$

$$\overline{P}_{10} \equiv P(I = 1 | Y = 0) + \min\{1 - v, P(I = 0 | Y = 0)\}.$$

When $v = 0$, P_{10} is trivially bounded to lie within $[0,1]$; at the other extreme when $v = 1$, $P_{10} = P(I = 1 | Y = 0)$.

Second, considering a positive value of v allows us to restrict the expected utilization rate among the unverifiably truly uninsured, $E(U | I^* = 0, Y = 0)$, which in turn allows us to tighten Molinari's bounds on the expected utilization rate among the unverifiably truly insured, $E(U | I^* = 1, Y = 0)$. Her framework can be used to provide sharp bounds at $v = 0$ and $v = 1$. Proposition 2 allows us to fill in identification patterns for values of v between 0 and 1. For a particular value of v, we can bound the population's use of health services under universal coverage as follows:[17]

Proposition 2. *Given $P(Z^* = 1 | Y = 0) \geq v$ and a known value $P_{10} \in [\underline{P}_{10}(v), \overline{P}_{10}(v)]$, the population's health care utilization rate under mandatory universal insurance coverage is*

---

[17] This analysis does not account for potential increases in gross prices for health care resulting from universal coverage. Since such price increases would not increase utilization, these upper bounds on $E(U | I = 1)$ should still apply. For our main analysis, we also assume that insurance coverage to the uninsured would be representative of the current mix of public and private coverage available to the insured.

sharply bounded as follows:

$$E(U \mid I = 1, Y = 1)P(I = 1, Y = 1) + P_{10}P(Y = 0)\int UdG_H$$

$$\leq E\left[U(I^* = 1)\right] \leq \tag{11}$$

$$E(U \mid I = 1, Y = 1)P(I = 1, Y = 1) + P_{10}P(Y = 0)\int UdG_L$$
$$+ \left[P(I = 0, Y = 1) + (1 - P_{10})P(Y = 0)\right]\sup U$$

where

$$G_L(t, v) \equiv \max\left\{0, \frac{P(U \leq t \mid Y = 0) - (1 - P_{10})\Omega_2(t, v)}{P_{10}}\right\}$$

$$G_H(t, v) \equiv \min\left\{\frac{P(U \leq t \mid Y = 0) - (1 - P_{10})\Omega_1(t, v)}{P_{10}}, 1\right\},$$

$$\Omega_1(t, v) \equiv \frac{P(U \leq t, I = 0, Y = 0) - \overline{\Theta_t^-}}{(1 - P_{10})P(Y = 0)}$$

$$\Omega_2(t, v) \equiv \frac{P(U \leq t, I = 0, Y = 0) + \overline{\Theta_t^+}}{(1 - P_{10})P(Y = 0)},$$

$$\overline{\Theta_t^+} \equiv \min\left\{\phi(v), P(U \leq t, I = 1, Y = 0), (1 - P_{10})P(Y = 0) - P(U \leq t, I = 0, Y = 0)\right\}, \text{ and}$$

$$\overline{\Theta_t^-} \equiv \min\left\{\phi(v), P(U \leq t, I = 0, Y = 0)\right\}.$$

If P_{10} is unknown, the lower and upper bounds in (11) are replaced by the infimum and

supremum, respectively, of these bounds over values of $P_{10} \in \left[\underline{P}_{10}, \overline{P}_{10}\right]$.

Proof: See the appendix.

The proof follows the general outline of Molinari's (2007a) Proposition 1 derivation. Her

Proposition 1 is similar except that her counterparts for the probability distributions G_L and G_H

implicitly assume that $v = 0$ such that nothing is known about the reliability of unverified

classifications. For that case, the bounds in (11) collapse to her bounds after setting $\Omega_1(t, v) = 0$

and $\Omega_2(t, v) = 1$ and $\left[\underline{P}_{10}, \overline{P}_{10}\right] = [0,1]$. She also allows for the possibility that the researcher has

21

outside information restricting P_{10} to a range narrower than $[0,1]$, including the possibility that P_{10} is known. In that case, something can be learned about $E[U(I^* = 1)|Y = 0]$ even though $\Omega_1 = 0$ and $\Omega_2 = 1$. Her bounds are as narrow as possible given her imposed assumptions.

In the special case that current insurance status is known to be accurately measured ($v = 1$), the Proposition 2 bounds collapse to the following well-known bounds (Manski, 1995):

$$E(U \mid I = 1)P(I = 1) \leq E\left[U(I^* = 1)\right] \leq E(U \mid I = 1)P(I = 1) + P(I = 0)\sup U.$$

Given the absence of reporting errors in this case, the width of these bounds depends only on the proportion uninsured, $P(I = 0)$, and the uninsured's upper bound use of services in the counterfactual state of being insured, $\sup U$.

For the binary utilization case, $\sup U$ in Proposition 2 is naturally set equal to 1. Yet there is no natural limit to the number of provider visits or dollars spent on medical services. Unless a researcher is nevertheless willing to set an upper bound on U, it must be recognized that an informative upper bound on $E\left[U(I^* = 1)\right]$ cannot be logically identified under the weak conditions specified in Proposition 2. For our Proposition 2 empirical results, we set $\sup U$ to 1.82 for number of visits and to \$862 for expenditures reflecting mean values among individuals who (1) perceived themselves to be in poor health at the time of the first interview and (2) were verified to be privately insured. These values reflect the 92nd percentile for visits and the 98th percentile for expenditures. We do not require any assumptions on $\sup U$ for the Proposition 3 bounds or monotone instrumental variable (MIV) bounds that follow. We conservatively treat the insurance rate among unverified classifications, P_{10}, as unknown. Therefore, we allow this value to lie anywhere within its logically consistent range $[\underline{P}_{10}, \overline{P}_{10}]$, conditional on v (see the last part of Proposition 2).

A. Monotonicity Assumptions

The preceding bounds can be narrowed substantially under common monotonicity assumptions on treatment response and treatment selection. The monotone treatment response assumption (MTR), introduced by Manski (1997), specifies that an individual's utilization is at least as high in the insured state as in the uninsured state:

$$U_i(I^* = 1) \geq U_i(I^* = 0). \tag{12}$$

Given moral hazard, we would expect some individuals to increase their use of health services upon becoming insured; presumably, the use of services would not decline.

Under a monotone treatment selection (MTS) assumption introduced in Manski and Pepper (2000), expected utilization under either "treatment" (insured or uninsured) would be at least as high among the currently insured as among the currently uninsured:

$$E[U(I^* = j) | I^* = 1] \geq E[U(I^* = j) | I^* = 0] \text{ for } j = 0,1. \tag{13}$$

The MTS assumption is related to adverse selection: those who have self-selected themselves into the insured state may tend to be more prone to use health services than their uninsured counterparts.[18] While the MTS assumption would presumably not hold for certain subpopulations, we only require that the tendency holds on average for the nonelderly population as a whole.

In support of the aggregate MTS assumption, public insurance programs already cover some of the least healthy populations. For example, people with substantial disabilities (who typically need considerable health care) are often insured by Medicaid or Medicare. Most states also use the Medicaid medically needy option to provide coverage to families with children whose health care expenditures are quite substantial relative to their incomes. More generally, those at relatively low risk of needing health services are less likely to seek coverage, as are

---

[18] The MTS assumption relaxes the commonly imposed "exogenous treatment selection" (ETS) assumption, $E[U(I^* = j) | I^* = 1] = E[U(I^* = j) | I^* = 0]$ for $j = \{0,1\}$, that assumes away the possibility of self-selection (see Manski and Pepper 2000, p. 1001).

people less predisposed to seek health services. For example, low demand for insurance may partly reflect preferences for avoiding medical care (Vistnes and Monheit 2006). Using the Community Tracking Household Survey, Hirth et al. (2006) estimate that 86% of workers who were not offered insurance at their job have low demand for insurance. Compared with workers who are offered insurance at their job, they find that workers who are not offered insurance tend to be younger, have lower family incomes, and have less education – all factors associated with lower health care use. In contrast, however, some of the uninsured may have found it difficult to obtain insurance at affordable prices due to preexisting conditions. Yet while medical underwriting can be a barrier to obtaining private insurance, data from the 1996 MEPS indicate that less than 3% of uninsured adults were "ever denied health insurance because of poor health."

Some evidence of aggregate adverse selection has been reported in parametric studies of the effects of insurance. Using data from the MEPS to estimate a model of private insurance coverage and office-based doctor visits, Deb et al. (2006) find that half of the observed lower health care use among uninsured adults compared with privately insured adults can be attributed to self-selection rather than the lack of insurance. Dor et al. (2006) find similar evidence of selection into private insurance.[19] If aggregate MTS does not hold in the population, however, then the Proposition 3 upper bound (see below) reverts to the Proposition 2 upper bound.

When both MTR and MTS hold, a result in Manski and Pepper (2000, Corollary 2.2) implies

$$E(U) \leq E\left[U(I^* = 1)\right] \leq E(U \mid I^* = 1). \tag{14}$$

The lower bound on the population's use of services under universal coverage is $E(U)$, the status quo national utilization rate in the absence of universal coverage. The upper bound is the status quo utilization rate among those currently insured. This result combined with the upper

---

[19] However, the findings by Deb et al. (2006) and Dor et al. (2006) rely in part on the types of parametric assumptions we are trying to avoid.

bound on $E(U | I^* = 1)$ derived in Proposition 1 leads to the following proposition:

Proposition 3. *Suppose that the MTR and MTS assumptions hold across the population and*

$P(Z^* = 1 | Y = 0) \geq v$. *Then the expected use of services under insurance coverage is bounded*

above by $\int U dF_L$ *where*

$$F_L(t) = \frac{P(U \leq t, I = 1) - \alpha_t^+}{P(I = 1) - \alpha_t^+ + \min\{\phi(v) - \alpha_t^+, P(U > t, I = 0, Y = 0)\}},$$

$$\alpha_t^+ = \begin{cases} \min\{\phi(v), P(U \leq t, I = 1, Y = 0)\} & \text{if } \delta_t^L < 0 \\ \max\{0, \min\{P(U \leq t, I = 1, Y = 0), \phi(v) - P(U > t, I = 0, Y = 0)\}\} & \text{otherwise} \end{cases}$$

and $\delta_t^L \equiv P(U \leq t, I = 1) - P(U > t, I = 1) - \phi(v)$.

In the empirical work that follows, we also consider the additional identifying power of the

independence and nonincreasing errors assumptions considered in Section III.C.

B. Universal Coverage Results

The fraction of the nonelderly population that used health services in July 1996 was 0.206, and

the mean number of provider visits and expenditures were 0.412 and $99, respectively. We are

interested in placing worst-case bounds on average utilization outcomes under a policy of

mandated health insurance coverage. Our main analysis presumes that new coverage extended to

the uninsured would be representative of the current mix of public and private coverage available

to the insured. Later, we consider policies that would cover the uninsured by expanding public

programs.

In the absence of monotonicity or independence assumptions, the Proposition 2 bounds

apply. Point estimates of these bounds, along with 95% confidence intervals, are presented in

Table 3, column (1). Under the standard implicit assumption that insurance status is reported

accurately, $v = 1$, we estimate that the fraction of the nonelderly population using health services

in July if everyone became insured would lie in the range [0.182, 0.374], with 95% confidence

interval [0.173, 0.383]. In percentage terms, the impact would lie within the range [-12%, +82%], with 95% confidence interval [-16%, +86%]. The estimated upper bounds on mean provider visits and expenditures per month are 0.712 (+73%) and $258 (+161%), respectively, which rise to 0.735 (+78%) and $281 (+184%) after accounting for sampling variability. In each case, the identification uncertainty associated with unknown counterfactual outcomes is much greater than the uncertainty associated with sampling variability. Clearly, we cannot learn much about the impact of universal coverage without imposing stronger assumptions, even if there is no uncertainty about the accuracy of status quo classifications. Figures 1-3 trace out the 95% confidence intervals for any use, visits, and expenditures, respectively, across all values of v between 0.5 and 1. As v departs from 1, the Proposition 2 bounds naturally become even wider.

The Proposition 3 bounds apply when MTR and MTS are imposed. As seen in Table 3, column (2) and the figures, the bounds narrow dramatically compared with column (1). The lower bounds rise to the status quo utilization rates of 0.412 visits and $99 per month. Similarly, the upper bounds decline, but they continue to depend on the value of v. When $v = 0.92$, for example, the mean number of visits under universal coverage would rise to no more than 0.503 (a 22% increase), while per capita expenditures would rise to no more than $124 (a 25% increase).

These bounds can be narrowed further under stronger assumptions about the patterns of reporting errors. Columns (3) and (4) in Table 3 present bounds on $E\left[U(I^* = 1)\right]$ under the orthogonality and nondifferential errors assumptions, respectively, discussed in Section III.C. We focus especially on the more plausible "nonincreasing errors" assumption in which the prevalence of insurance status misreporting falls weakly with the level of utilization. These results are presented in Column (5). When $v = 0.92$, the mean number of visits per month among the nonelderly population would rise to no more than 0.463 (a 12% increase) under universal coverage, while per capita monthly expenditures would rise to no more than $117 (an 18% increase).

26

C. Monotone Instrumental Variables

We next use monotone instrumental variables (MIV) techniques developed by Manski and Pepper (2000) and extended by Kreider and Pepper (2007) to assess how the Proposition 3 bounds can be narrowed when combined with monotonicity assumptions linking utilization outcomes and observed covariates such as age or health status. Consider age and use of health services. The incidence of many health conditions rises with age, and many health conditions are persistent once developed. These tendencies suggest that, on average, utilization among adults under universal coverage would be nondecreasing in age. If this assumption holds, then we can improve upon the previously derived bounds by enforcing the restriction that upper bounds identified for younger groups cannot exceed upper bounds identified for older groups.

We treat age and general health status as MIVs. We divide the population into 18 age groups: 0-30, 31-32, 33-34, ..., 63-64. [20] Within each age group, we assume that use of services under universal coverage would be nondecreasing in reported worse general health across the following categories: poor/fair, good, very good, and excellent. Formally, consider the utilization rate within some age group, age' (the extension to multiple MIV dimensions is straightforward). The age MIV assumption implies the following inequality restriction:

$$age_1 \leq age' \leq age_2 \Rightarrow E\left[U\left(I^* = 1\right) | age_1\right] \leq E\left[U\left(I^* = 1\right) | age'\right] \leq E\left[U\left(I^* = 1\right) | age_2\right]. \quad (15)$$

This mean monotonicity condition for an instrument relaxes the more typical (and stronger) mean independence assumption. Under mean independence, the inequalities across the expectations in (15) would be replaced with equalities (Manski and Pepper 2000). In our application, however, it is not obvious where to find instruments that would satisfy mean

---

[20] For the youngest group, we conservatively choose an age range that extends well into adulthood. Newborns tend to use substantial care, but then the use of services tends to decline with a child's age as the frequency of recommended preventive care visits decreases. Use of services tends to rise again in adulthood with the onset of chronic conditions.

independence.

The conditional expectations in (15) are not identified, but they can be bounded using the methods described above. Let $LB(age)$ and $UB(age)$ be the known lower and upper bounds, respectively, given the available information on $E(U \mid I^*, age)$; in computing these bounds, we assume that MTR and MTS continue to hold. (Under the MTS assumption, note that the treatment I^* is itself an MIV.) Then using Manski and Pepper (2000, Proposition 1), we have

$$\sup_{age_1 \leq age'} LB(age_1) \leq E\left[U\left(I^* = 1\right) \mid age'\right] \leq \inf_{age_2 \geq age'} UB(age_2). \qquad (16)$$

The MIV bound on expected utilization under universal coverage is obtained using the law of total probability:

$$\sum_{age'} P(age = age')\{ \sup_{age_1 \leq age'} LB(age_1)\} \qquad (17)$$

$$\leq E\left[U\left(I^* = 1\right)\right] \leq$$

$$\sum_{age'} P(age = age')\{ \inf_{age_2 \geq age'} UB(age_2)\}.$$

Thus, to find the MIV bounds on the utilization rate, one takes the appropriate weighted average of the lower and upper bounds across the different values of the instrument. This MIV estimator is consistent but biased in finite samples. To account for this bias, we employ Kreider and Pepper's (2007) modified MIV estimator that estimates and adjusts for finite-sample bias using Efron and Tibshirani's (1993a) nonparametric bootstrap correction method.[21]

While there is substantial debate in the literature about the appropriateness of self-reported health variables in regression models of employment (see, e.g., Bound 1991), the MIV assumption does not require reported health status to be an unbiased indicator of true health

---

[21] See Kreider and Pepper (2007) for estimation details.

status or to be exogenously reported. What is required is that utilization under universal coverage would be nondecreasing in age and reported health status.[22]

The MIV assumptions cannot be verified by the data because the expectations in (15) involve unobserved counterfactual utilization outcomes; moreover, we do not observe status quo insurance status in the presence of reporting errors. The joint assumption that $v = 1$ and MIV-MTS hold, however, potentially can be rejected by the data using a test similar to one suggested by Manski and Pepper (2000, footnote 9). For each age group, age_j, and health category, H_k, the following inequalities must hold:

$$E\left(U \mid I = 1, age_j, H_k\right) \equiv E\left[U(I = 1) \mid I = 1, age_j, H_k\right]$$
$$\leq E\left[U(I = 1) \mid I = 1, age_{j'}, H_{k'}\right] \equiv E\left(U \mid I = 1, age_{j'}, H_{k'}\right) \text{ for } j \leq j', k \leq k'.$$

Thus, in the absence of measurement error, $E\left(U \mid I = 1, age_j, H_k\right)$ should be monotonic in age and reported health. For each pair of adjacent age and health status groups among the reportedly insured, we conducted a one-sided test of whether utilization is decreasing in age and/or in worse health. We conducted identical tests for the probability of service use, number of visits, and expenditures. One test for visits was rejected at the 5% significance level, but this is not statistically significant after accounting for the multiplicity of tests. As seen in Table 4, the nonelderly population's use of hospital or ambulatory services among the reportedly insured is roughly monotonic in age and perceived health status. After accounting for sampling variability,

---

[22] Health status is routinely included as an explanatory variable in parametric studies of the effects of insurance on service use and expenditures, along with an implicit assumption of homogeneity of effects across individuals. While reported health status is difficult to quantify in surveys and may be mismeasured, reported general health has external validity in that it predicts mortality and changes in functioning (Idler and Benyamini 1997; Idler and Kasl 1995). We do not use reports of chronic conditions, which tend to be misreported; the uninsured are especially likely to underreport conditions because they are less likely to seek care and acquire diagnoses (Baker et al. 2004; Miller et al. 2004).

we cannot reject the hypothesis that utilization is monotonic in these attributes. This is consistent with the many studies that have found that reported health status strongly predicts service use and expenditures (Balkrishnan et al. 2000; Bierman et al. 1999; DeSalvo et al. 2005; Miilupalo et al. 1997).

D. MIV Results

MIV results are presented in the last two columns of Table 3, assuming MTR and MTS continue to hold. The identifying power of the MIV assumption can be assessed by comparing columns (2) and (6) for the case of arbitrary error patterns and by comparing columns (5) and (7) for the case of "nonincreasing errors."   We focus on the latter comparison for $v = 0.92$. In column (5), we estimate that the fraction of the nonelderly population using health services in a month would rise no more than 13% above the status quo to 0.232. Under the additional MIV assumption in column (7), we estimate that this fraction would rise no more than 7% to 0.226. Improvements in the upper bounds for mean number of visits and expenditures are similar. Under the MIV assumption, the upper bound on the number of visits per month improves from 0.463 to 0.444, and the upper bound on expenditures improves from $117 to $115. Thus, in this setting, visits would increase no more than 8% under universal coverage and expenditures per capita would increase no more than 16%. In what follows, we refer to these estimates as our preferred bounds.

E. Expanding Public Coverage

To this point, our analysis presumes that the package of health benefits made available to the uninsured under universal coverage would reflect the mixture of benefits available to the privately and publicly insured under the status quo. Many recent proposals for extending health insurance to the uninsured, however, involve expansions of existing public programs like Medicaid. As part of our sensitivity analysis, we investigate how the estimated bounds in Table 3 would change if we focus attention on policies that would cover the uninsured through

expansions of public insurance. To do so, we repeat the preceding nonparametric bounds analysis over the subsample of individuals classified as uninsured or publicly insured. Within this subsample, we estimate upper bounds on utilization under a policy that extends public insurance to the uninsured. Upper bound utilization rates (for any use, visits, and expenditures) for the entire nonelderly population are then computed as a weighted average of these upper bounds for the non-privately insured and the status quo utilization rates of the privately insured.

This approach requires two simplifying assumptions. First, we assume that reports of private coverage are accurate; any insurance classification errors are confined to errors among the reportedly uninsured and publicly insured. For cases in which validation data are available in the MEPS, Hill (2007/2008) finds that private insurance is very accurately reported – most respondents are aware of whether or not they have private coverage. In contrast, there is evidence of more extensive misreporting of public coverage in surveys (Call et al. 2007; Card et al. 2004; Davern et al. 2007). In our restricted sample that excludes the privately insured, the proportion of inaccurate insurance classifications is likely to be higher. To account for this greater uncertainty, we replace the thresholds $v = 0.92$ and $v = 0.84$ considered in the main analysis with the lower thresholds $v = 0.89$ and $v = 0.77$.

Second, our approach presumes that service use by the currently privately insured would not rise due to the expansion of public insurance. Like the studies by Hadley and Holahan (2003) and Miller et al. (2004), we do not attempt to estimate the potential consequences of "crowd out" in which some people with private coverage might switch to the newly available public coverage. Based on the literature reviewed by Duchovny and Nelson (2007), switchers from private policies would likely comprise between a quarter to a half of new enrollees under an expanded public insurance program. Our derived upper bounds in this section are valid if the switchers would use no more care, on average, under their new public policies. While public programs that require small or no premium payments are potentially attractive to many private policyholders, *a priori* it is not clear whether on balance differences in benefits and providers

would differentially attract less healthy, privately insured people to public coverage or repel them. Medicaid and SCHIP programs tend to have generous cost sharing and cover more services than private insurance, which would disproportionately attract the less healthy. On the other hand, Medicaid typically pays providers lower fees than private insurance, and lower fees reduce provider willingness to participate in the Medicaid program and reduce enrollees' access to care (Cohen 1993; Mitchell 1991).[23] Econometric studies find similar access to care in public and private programs (Long et al. 2005; Selden and Hudson 2006), which suggests that any incentive or disincentive for less healthy individuals to switch from private to public coverage is likely small.[24] If, on balance, healthier privately insured people (or people otherwise less prone to use care) would be attracted to public coverage, then our estimates remain valid upper bounds. If the opposite is true, then the bounds depend on $\sup U$ and hence would be wide.

Table 5 presents our estimated upper bounds on the impacts of expanding public coverage. The estimates are broadly similar to those presented in Table 3, but the upper bounds that incorporate MTR and MTS are uniformly higher for a given degree of confidence in the data. These higher values do not imply that health care expenses would necessarily be higher under public expansions than under a mixture of public and private expansions. Instead, the higher upper bounds reflect more uncertainty about the impacts of public expansions. Recall from Table 1 that the status quo use of health services is greater among the publicly insured than among the privately insured. These differences are reflected in the upper bounds. Moreover, the confidence intervals are wider for expanding public coverage than for covering the uninsured with a mix of private and public insurance. These wider intervals reflect smaller samples of

---

[23] Hadley and Holahan (2003/2004) provide evidence that public coverage reduces health expenditures relative to private coverage, consistent with lower fees paid to providers by Medicaid than by private insurance.

[24] Several econometric studies of service use find no differential impact of private relative to public coverage for a variety of populations and measures of service use (Glied et al. 1998; Hadley and Holahan 2003; Kaestner 1999; Long et al. 2005; Selden and Hudson 2006).

people reporting public coverage than people reporting any insurance.

When $v = 1$ and the MIV assumption is not imposed, the upper bound on the mean number of monthly provider visits per month by the nonelderly population under universal coverage rises from 0.448 with a mix of private and public coverage (Table 3) to 0.473 with expanded public coverage (Table 5). The upper bound on mean monthly expenditures rises from $114 to $121. When the MIV assumption is additionally imposed, the upper bound on visits rises from 0.440 to 0.457, and the upper bound on expenditures rises from $114 to $120. Our preferred results presented in column (7) impose MIV with nonincreasing errors. When $v = 0.89$, the fraction of the nonelderly population using services during a month under expanded public insurance would rise no more than 14% to 0.235, the mean number of monthly visits would rise no more than 22% to 0.503, and mean monthly expenditures would rise no more than 26% to $125.

V. Comparisons with Parametric Studies

Parametric point estimates of increased health care utilization under universal coverage lie within our preferred estimated bounds. These estimates, which impose MTR, MTS, MIV and "nonincreasing errors" are reported in Table 3, column (7) with $v = 0.92$ when expanded coverage involves a mixture of public and private insurance and in Table 5; they are reported in column (7) with $v = 0.89$ when public coverage is extended to the uninsured.

In parametric studies, Miller et al. (2004) and Hadley and Holahan (2003) estimate that expanding public programs to cover all the uninsured would increase annual total expenditures between 9% and 10%. Our corresponding nonparametric upper bound when $v = 0.89$ is 26%. Their estimates, like our own, do not account for any crowding out of private coverage. Their models assume no measurement error, and self-selection into insured status is allowed only through a set of observed characteristics. One likely reason our worst-case upper bounds are substantially higher than their point estimates is that we do not impose the homogeneity

assumption that the impact of insurance coverage on utilization is identical across individuals with the same observed characteristics. Miller et al. and Hadley, and Holahan also estimate that covering the uninsured with private insurance would increase annual total expenditures between 11% and 17%, depending on their specific assumptions. For Massachusetts, Blumberg et al. (2006) estimate the impact of potential expansions that would resemble the mixture of private and public coverage expansions enacted in that state. They estimate that expenditures would increase by about 12%.[25] Our corresponding upper bound when $v = 0.92$ is 16%.

Buchmueller et al. (2005) review studies of the impacts of insurance on the amount of service use. Relying on a variety of comparison methods, parametric studies find that having insurance increases visits among the uninsured between 16% and 106%. More recent studies, however, find a narrower range of effects. Selden and Hudson (2006) estimate that expanding public insurance to cover uninsured children would increase their annual probability of having an ambulatory visit by 54%, corresponding to an increase of about 7% across the nonelderly population as a whole. They find that private coverage would increase the annual probability by 61%, corresponding to an increase of about 8% across the population as a whole. Estimates in Deb et al. (2006) suggest that expanding private insurance to uninsured adults would increase their number of office-based visits by about 46%, corresponding to an increase for the adult population of about 5%. Hadley and Holahan (2003), who do not account for unobserved factors, estimate that public coverage would increase the annual number of office-based visits among the uninsured by 41%, while private insurance would increase visits by 30%. Their estimates correspond to increases of 5% and 4%, respectively, among all nonelderly. For expansions involving a mixture of private and public insurance, our estimated upper bound on the increase in monthly office-based and hospital outpatient visits is 8%.

---

[25] They model premiums, rather than health care expenditures, so their estimates implicitly include insurance loading as well as expenditures for health care services.

VI. Conclusion

Policymakers have long been interested in identifying the consequences of uninsurance for access to health care and the potential impacts of universal coverage (e.g., Institute of Medicine 2003). Identification of policy outcomes, however, is confounded by both the unobservability of counterfactuals and the potential unreliability of self-reported insurance status. To account for these two distinct types of uncertainty, we developed a nonparametric framework that extends the literature on partially identified probability distributions and treatment effects. Using the new analytical results, we provided tight bounds on the impact of universal health insurance on provider visits and medical expenditures. As part of the paper's contribution, we showed how to partially identify the conditional mean of a random variable for the case that a binary conditioning variable – in our case health insurance – is subject to arbitrary endogenous measurement error.

Our conservative statistical approach provides informative bounds on parameters of interest without imposing parametric assumptions. We began by corroborating self-reported insurance status for a nonrandom portion of the MEPS sample using outside information from insurance cards and follow-back interviews with employers and insurance companies. We allowed for the possibility of insurance reporting errors within the remainder of the sample and illustrated the sensitivity of our empirical results to alternative verification, monotonicity, and independence assumptions. For our preferred estimates, we introduced a "nonincreasing errors" assumption that relaxes the strict nondifferential independence assumption embodied in the classical errors-in-variables framework. In our application, the weaker monotonicity assumption retains much of the identifying power of the independence assumption while allowing for the possibility that using health services may inform a patient of her true insurance status.

Our primary analysis considers the impact of extending insurance to the uninsured using the mix of public and private coverage prevalent under the status quo. Programs intended to cover the uninsured in Massachusetts, Maine, and Vermont, and proposals by some presidential

35

candidates, would cover the uninsured with both private and public insurance. We estimate that the fraction of the nonelderly population using ambulatory or hospital services would rise no more than 9% under universal coverage. We further estimate that per capita monthly provider visits would rise by no more than 8%, and mean expenditures per month would rise no more than 16%. These estimated upper bounds rely on an assumption that no more than 8% of unverified insurance classifications are misreported. While the accuracy of unverified insurance responses cannot logically be known, we have relaxed the standard implicit assumption that all classifications are known to be accurate. Under our preferred monotonicity assumptions, we find that our estimates are not particularly sensitive to the degree of reporting error within plausible ranges. This suggests that, contrary to some assertions, uncertainty about the number of uninsured may not be a major impediment to developing programs to cover the uninsured. Our estimated upper bounds on the effects of expanding public coverage, however, are not as tight.

Our analysis has several limitations. First, the cost of covering the uninsured will depend in part on the generosity of benefits, but our treatment effect approach ignores heterogeneity across plan types in treating insurance as binary. Second, our most informative bounds rely on an unverifiable assumption that, on average, households who tend to use more health services have already self-selected themselves into the insured state. While we impose this assumption only across the nonelderly population as a whole (allowing for the possibility that this tendency is reversed within some subpopulations), the validity of the assumption cannot be directly verified. Third, our methods are less useful for estimating the cost of covering an additional person. Marginal analyses require stronger assumptions about the characteristics of the newly insured. Finally, the health care system continues to evolve, and our 1996 data (which contain the best available validation information) were gathered before SCHIP was implemented.

The methods developed in this paper can be applied to a wide range of topics that involve identification of conditional expectations or treatment effects given uncertainty about the accuracy of the conditioning variable. Our framework, for example, offers an alternative

approach to Blau and Gilleskie's (2001) parametric analysis of the impact of employer-provided retiree health insurance on retirement outcomes. In the Health and Retirement Study data used in their study, about 13 percent of the respondents nearing retirement age said they were unsure about whether they had retiree insurance – thus forming a natural subpopulation of respondents to be characterized as providing unreliable treatment information. The methods in this paper could be used to bound the effects of retiree insurance on employment, informing policymakers about the potential consequences of allowing retirees younger than 65 to purchase Medicare coverage. More generally, we expect this developing line of research to improve researchers' understanding of the consequences of nonclassical measurement error for inferences, which should in turn yield more informed policy analyses.

References

Aigner, Dennis J. 1973. "Regression with a Binary Independent Variable Subject to Errors of
  Observation." *Journal of Econometrics*, 2, 49-59.

Baker, Michael, Mark Stabile, and Catherine Deri. 2004. "What Do Self-Reported, Objective,
  Measures of Health Measure?" *Journal of Human Resources* 39(4): 1067-1093.

Banthin, Jessica S., and Merrile Sing. "How Medicaid Enrollment Estimates from MEPS
  Compare with Administrative Totals." Agency for Health Research and Quality Working
  Paper, forthcoming.

Balkrishnan, Rajesh, Roger T. Anderson, and David Bowton. 2000. "Self-reported Health Status
  Predictors of Healthcare Services Utilization and Charges in Elderly Asthmatic Patients."
  Journal of Asthma 37(5): 415–23.

Berger, Mark C., Dan A. Black and Frank A. Scott. 1998. "How Well Do We Measure
  Employer-Provided Health Insurance Coverage?" *Contemporary Economic Policy*, 16:
  356-367.

Bierman, Arlene S., Thomas A. Bubolz, Elliott S. Fisher, and John H. Wasson. 1999. "How Well
  Does a Single Question about Health Predict the Financial Health of Medicare Managed
  Care Plans?" *Effective Clinical Practice* 2(2): 56–62.

Blau, David M., and Donna B. Gilleskie. 2001. "Retiree Health Insurance and Labor Force
  Behavior of Older Men in the 1990s." *Review of Economics and Statistics*, 83(1): 64-80.

Blumberg, Linda J., John Holahan, Alan Weil, Lisa Clemans-Cope, Matthew Buettgens, Fredric
  Blavin, and Stephen Zuckerman. 2006. "Toward Universal Coverage in Massachusetts."
  Inquiry 43(2): 102–121.

Bollinger, Christopher R. 1996. "Bounding Mean Regressions When a Binary Variable is
  Mismeasured." *Journal of Econometrics*, 73(2): 387-99.

Bound, John. 1991. "Self-Reported Versus Objective Measures of Health in Retirement
  Models." *Journal of Human Resources*, 26(1): 106-138.

Bound, John, and Richard V. Burkhauser. 1999. "Economic Analysis of Transfer Programs
  Targeted on People with Disabilities." In *Handbook of Labor Economics*, vol. 3C, ed.
  Orley Ashenfelter and David Card. Amsterdam: Elsevier Science: 3417-3528.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error in Survey
  Data." In *Handbook of Econometrics*, vol. 5, ed. James J. Heckman and Edward Leamer.
  Amsterdam: Elsevier Science: 3705-3843.

Buchmueller, Thomas C., Kevin Grumbach, Richard Kronick and James G. Kahn. 2005. "The
  Effect of Health Insurance on Medical Care Utilization and Implications for Insurance
  Expansion: A Review of the Literature." *Medical Care Research and Review*, 62(1): 3-

30.

Call, Kathleen Thiede, Michael Davern, Gestur Davidson, et al. 2007b. "Accuracy in Self-Reported Health Insurance Coverage and Bias to Survey Estimates of Uninsurance." Presentation at the Academy Health Annual Research Meeting, Orlando, FL.

Card, David, Andrew K.G. Hildreth, and Lara D. Shore-Sheppard. 2004. "The Measurement of Medicaid Coverage in the SIPP: Evidence From a Comparison of Matched Records." *Journal of Business & Economic Statistics* 22(4): 410-420.

Chernozhukov, Victor, Han Hong, and Elie Tamer. 2007. "Estimation and Confidence Regions for Parameter Sets in Econometric Models." *Econometrica* 75(5): 1243–1284.

Cohen, Joel. 1993. "Medicaid Physician Fees and Use of Physician and Hospital Services." *Inquiry* 30(3): 281-292.

_____. 1997. "Design and Methods of the Medical Expenditure Panel Survey Household Component." MEPS Methodology Report no. 1, AHCPR Pub. no. 97-0026. Rockville, MD: Agency for Health Care Policy Research.

Czajka, John L., and Kimball Lewis. 1999. "Using Universal Survey Data to Analyze Children's Health Insurance Coverage: An Assessment of Issues." Washington, DC: Mathematica Policy Research, Inc., http://aspe.os.dhhs.gov/health/reports/Survey%20Data.htm.

Davern, Michael, Jacob Klerman, David Bough, et al. 2007. "Fitting Square Pegs Into Round Holes: Linking Medicaid and Current Population Survey Data to Understand the 'Medicaid Undercount.'" Presentation at the Centers for Medicare and Medicaid Services, Baltimore, MD.

Deb, Partha, Murat K. Munkin, and Pravin K. Trivedi. 2006. "Private Insurance, Selection, and Health Care Use: A Bayesian Analysis of a Roy-Type Model." *Journal of Business & Economic Statistics* 24(4): 403-415.

DeNaves-Walt, Carmen, Bernadette D. Proctor, and Cheryl Hill Lee. 2005. Income, Poverty, and Health Insurance Coverage in the United States, 2004. U.S. Census Bureau, Current Population Reports P60-229. U.S. Government Printing Office, Washington DC.

DeSalvo, Karen B., Vincent S. Fan, Mary B. McDonell, and Stephan D. Fihn. 2005. "Predicting Mortality and Healthcare Utilization with a Single Question." *HSR: Health Services Research* 40(4):1234-1246.

Dominitz, Jeff, and Robert P. Sherman. 2004. "Sharp Bounds Under Contaminated or Corrupted Sampling With Verification, With an Application to Environmental Pollutant Data," *Journal of Agricultural, Biological and Environmental Statistics*, 9(3): 319-338.

Dor, Avi, Joseph Sudano, and David W. Baker. 2006. "The Effect of Private Insurance on the Health of Older, Working Age Adults: Evidence from the Health and Retirement Study." *HSR: Health Services Research* 41(3, Part I): 759-787.

Duchovny, Noelia, and Lyle Nelson. 2007. *The State Children's Health Insurance Program.* Washington, DC: Congressional Budget Office.

Efron, Bradley, and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap.* Chapman and Hall. New York: NY.

Frazis, Harley, and Mark A. Loewenstein. 2003. "Estimating Linear Regressions with Mismeasured, Possibly Endogenous, Binary Explanatory Variables," *Journal of Econometrics*, 117: 151-178.

Gerfin, Michael, and Martin Schellhorn. 2006. "Nonparametric Bounds on the Effect of Deductibles in Health Care Insurance on Doctor Visits -- Swiss Evidence." *Health Economics* 15(9):1011–1020.

Glied, Sherry, A. Bowen Garrett, Christina Hoven, Maritza Rubio-Stipec, Darrel Regier, Robert E. Moore, Sherryl Goodman, Ping Wu, and Hector Bird. 1998. "Child Outpatient Mental Health Service Use: Why Doesn't Insurance Matter?" *Journal of Mental Health Policy and Economics* 1(4):173-187.

Gruber, Jonathan, and Brigitte C. Madrian. 2004. "Health Insurance, Labor Supply, and Job Mobility: A Critical Review of the Literature." In *Health Policy and the Uninsured*, ed. Catherine G. McLaughlin. Washington DC: Urban Institute.

Hadley, Jack, and John Holahan. 2003/2004. "Is Health Care Spending Higher under Medicaid or Private Insurance?" *Inquiry* 40(4): 323-42.

_____ and _____. 2003. "Covering the Uninsured: How Much Would It Cost?" *Health Affairs* Supplemental Web Exclusives W3: 250-265.

Hill, Steven C. 2007/2008. "The Accuracy of Reported Insurance Status in the MEPS." *Inquiry* 44(4), forthcoming.

Hirth, Richard A., Reagan A. Baughman, Michael E. Chernew, and Emilie C. Shelton. 2006. "Worker Preferences, Sorting and Aggregate Patterns of Health Insurance Coverage." *International Journal of Health Care Finance and Economics* 6(4): 259-277.

Holahan, John, and Linda Blumberg. 2006. "Massachusetts Health Care Reform: A Look at the Issues." *Health Affairs* 25: w432–w443.

Horowitz, Joel L., and Charles F. Manski. 1995. "Identification and Robustness With Contaminated and Corrupted Data." *Econometrica* 63(2): 281-02.

_____ and _____. 1998. "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations." *Journal of Econometrics*, 84, 37-58.

Hudson, Julie L., and Thomas M. Selden. 2007. "Children's Eligibility and Coverage: Recent Trends and a Look Ahead." *Health* Affairs 26(5): w618-w629.

Hunter, Derek. 2004. Counting the Uninsured: Why Congress Should Look Beyond the Census Figures. WebMemo #555. Washington, DC: Heritage Foundation.

Idler, Ellen L., and Yael Benyamini. 1997. "Self-rated Health and Mortality: A Review of Twenty-seven Community Studies." *Journal of Health and Social Behavior* 38(1): 21-37.

Idler, Ellen L., and Stanislav V. Kasl. 1995. "Self-ratings of Health: Do They Also Predict Change in Functional Ability?" *Journal of Gerontology* 50B(6): S344-S353.

Institute of Medicine. 2003. *Hidden Costs, Lost Value: Uninsurance in America*. Washington, DC: National Academy Press.

Kaestner, Robert. 1999. "Health Insurance, the Quantity and Quality of Prenatal Care, and Infant Health." *Inquiry* 36(2): 162-175.

Kaiser Commission on Medicaid and the Uninsured. 2007. "State Initiatives to Cover Children." Washington DC: Kaiser Family Foundation.

Kaye, Neva, and Andrew Snyder. 2007. "Health Reform in Maine, Massachusetts, and Vermont: An examination of State Strategies to Improve Access to Affordable, Quality Care." Augusta, ME: Maine Health Access Foundation Issue Brief.

Kreider, Brent, and John V. Pepper. 2007. "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors." *Journal of the American Statistical Association*, 102(478): 432-441.

Levy, Helen, and David Meltzer. 2004. "What Do We Really Know About Whether Health Insurance Affects Health?" In *Health Policy and the Uninsured*, ed. Catherine G. McLaughlin. Washington DC: Urban Institute.

Lewbel, Arthur. 2007. "Estimation of Average Treatment Effects with Misclassification." *Econometrica* 75(2): 537-551.

Long, Sharon K., Teresa Coughlin, and Jennifer Kling. 2005. "How Well Does Medicaid Work in Improving Access to Care?" *HSR: Health Services Research* 40(1): 39-58.

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

Manski, Charles F. 1997. "Monotone Treatment Response." *Econometrica* 65(6): 1311-1334.

Manski, Charles F., and John V. Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica* 68(4): 997-1010.

Miilupalo, Seppo, Ilkku Vuori, Pekku Oja, Matti Pasanen, and Helku Urponen. 1997. "Self-rated Health Status as a Health Measure: The Predictive Value of Self-reported Health Status on the Use of Physician Services and on Mortality in the Working-age Population." *Journal of Clinical Epidemiology* 50(5): 517–28.

Miller, G. Edward, Jessica Banthin, and John Moeller. 2004. "Covering the Uninsured: Estimates of the Impact on Total Health Expenditures for 2002." AHRQ Working Paper 04007. Rockville, MD: Agency for Healthcare Research and Quality.

Mitchell, Janet. 1991. "Physician Participation in Medicaid Revisted." *Medical Care* 29(7): 645-653.

Molinari, Francesca. 2007a. "Missing Treatments," mimeo, Department of Economics, Cornell University.

_____. 2007b. "Partial Identification of Probability Distributions with Misclassified Data." mimeo, Department of Economics, Cornell University.

Monheit, Alan C., and Jessica Primoff Vistnes. 2006. "Health Insurance Enrollment Decisions: Preferences for Coverage, Worker Sorting, and Insurance Take Up." *NBER Working Paper 12429.*

Nelson, David E., Betsy L. Thompson, Nancy J. Davenport, and Linda J. Penaloza. 2000. "What People Really Know about Their Health Insurance: A Comparison of Information Obtained from Individuals and Their Health Insurers." *American Journal of Public Health* 90(6): 94-8.

Nelson, Lyle. 2003. *How Many People Lack Health Insurance and For How Long?* Washington, DC: Congressional Budget Office.

Olson, Craig A. 1998. "A Comparison of Parametric and Semiparametric Estimates of the Effect of Spousal Health Insurance Coverage on Weekly Hours Worked by Wives." *Journal of Applied Econometrics*, 13, 543-565.

Pascale, Joanne. 2007. "Measuring Health Insurance in the U.S." Research Report Series (Survey Methodology #2007-11). Washington DC: US Census Bureau, Statistical Research Division.

Peterson, Chris L., and April Grady. 2005. Medicaid/SCHIP Enrollees: Comparison of Counts from Administrative Data and Survey Estimates. Washington, DC: Congressional Research Service.

Selden, Thomas M., and Julie L. Hudson. 2006. "Access to Care and Utilization Among Children: Estimating the Effects of Public and Private Coverage." *Medical Care* 44 (5 Suppl.): I-19-I-26.

Wolter, Kirk M. 1985. *Introduction to Variance Estimation*. New York: Springer-Verlag.

Woolhandler, Steffie, and David U. Himmelstein. 2007. "Health Reform Failure." *Boston Globe,* September 17, 2007.

Appendix.

Proof of Proposition 1. To place bounds on Δ, we begin by logically determining the lowest

feasible value of $P(U \leq t \mid I^* = 1)$. Differentiating the right hand side of (3), we find that this

quantity is increasing in $\theta_t'^+$, the unobserved fraction of individuals with $U > t$ misclassified as

being insured, and in θ_t^-, the unobserved fraction of individuals with $U \leq t$ misclassified as

being uninsured. As a worst-case possibility for the lower bound, we must therefore set

$\theta_t'^+ = \theta_t^- = 0$ to obtain:

$$P(U \leq t \mid I^* = 1) \geq \frac{P(U \leq t, I = 1) - \theta_t^+}{P(I = 1) - \theta_t^+ + \theta_t'^-}. \tag{18}$$

While θ_t^+ and $\theta_t'^-$ are unobserved, their ranges are restricted. The unobserved fraction that was

falsely classified as insured, $\theta_t^+ = P(U \leq t, I = 1, Z^* = 0)$, cannot exceed the observed fraction that

was classified as insured with unknown insured status. Nor can this fraction exceed the total

allowed fraction of misclassified cases, $\phi(v) \equiv (1 - v)P(Y = 0)$. Similarly, the unobserved

fraction of individuals that was falsely classified as being uninsured,

$\theta_t'^- = P(U > t, I = 0, Z^* = 0)$, cannot exceed the observed fraction that was classified as being

uninsured with unknown insured status; nor can it exceed the total fraction of misclassified

cases:

$$0 \leq \theta_t^+ \leq \min\{\phi(v), P(U \leq t, I = 1, Y = 0)\} \equiv \overline{\theta_t^+}$$
$$0 \leq \theta_t'^- \leq \min\{\phi(v), P(U > t, I = 0, Y = 0)\} \equiv \overline{\theta_t'^-}.$$

To find the lower bound of $P(U \leq t \mid I^* = 1)$, we must find the minimum feasible value for the

right-hand side (18). Therefore, for any candidate value of θ_t^+, we need $\theta_t'^-$ to attain its

maximum allowed value conditional on θ_t^+:

$$\theta_t'^- = \min\left\{\phi(v) - \theta_t^+, \overline{\theta_t'^-}\right\} = \min\left\{\phi(v) - \theta_t^+, P(U > t, I = 0, Y = 0)\right\}.$$

The objective then becomes one of minimizing

$$\frac{P(U \leq t, I = 1) - \theta_t^+}{P(I = 1) - \theta_t^+ + \min\left\{\phi(v) - \theta_t^+, P(U > t, I = 0, Y = 0)\right\}} \tag{19}$$

over feasible values of θ_t^+.

Define $\theta_t^{+o} \equiv \phi(v) - P(U > t, I = 0, Y = 0)$, the critical value of θ_t^+ that makes the two

arguments in the min function equal. First consider values of $\theta_t^+ \leq \theta_t^{+o}$. For such values, the

derivative of (19) with respect to θ_t^+ is negative; therefore, we can exclude as potential

candidates any values of θ_t^+ less than $\theta_{t\min}^+ \equiv \max\left\{0, \min\left\{\overline{\theta_t^+}, \theta_t^{+o}\right\}\right\}$. For $\theta_t^+ > \theta_t^{+o}$, the derivative

has the same sign as

$$\delta_t^L \equiv P(U \leq t, I = 1) - P(U > t, I = 1) - \phi(v). \tag{20}$$

When this quantity is negative, we must raise θ_t^+ to its maximum feasible value, $\overline{\theta_t^+}$; otherwise,

we set θ_t^+ equal to $\theta_{t\min}^+$. Similar logic provides an upper bound on $P(U \leq 1 | I^* = 1)$. After

defining $\delta_t^H \equiv P(U > t, I = 1) - P(U \leq t, I = 1) - \phi(v)$, the preceding results establish Proposition 1.

Proof of Proposition 2. Using Molinari's (2007a, p. 9) decomposition for the case of missing

observations, we begin by writing the distribution of U among unverified cases as a weighted

average of the distributions among unverified insured and uninsured cases:

$$P(U \leq t | Y = 0) = P(U \leq t | I^* = 1, Y = 0)P_{10} + P(U \leq t | I^* = 0, Y = 0)(1 - P_{10}). \tag{21}$$

First consider a particular value of $P_{10} \in (0,1)$. Solving for the fraction consuming $U \leq t$ among

the unverified currently insured implies

$$P(U \leq t | I^* = 1, Y = 0) = \frac{P(U \leq t | Y = 0) - P(U \leq t | I^* = 0, Y = 0)(1 - P_{10})}{P_{10}}. \tag{22}$$

The quantity $P(U \leq t | I^* = 0, Y = 0)$ in the right-hand-side can be rewritten as

$$P(U \leq t \mid I^* = 0, Y = 0) = \frac{P(U \leq t, I = 0, Y = 0) + \theta_t^+ - \theta_t^-}{(1 - P_{10}) P(Y = 0)}. \tag{23}$$

The upper bound on this quantity is obtained by setting $\theta_t^- = 0$ and θ_t^+ equal to its maximum

feasible value. From Section 3.A, we know $\theta_t^+ \leq \overline{\theta_t^+}$. Combined with the requirement that

$P(U \leq t \mid I^* = 0, Y = 0) \leq 1$, the value of θ_t^+ is restricted to lie in the range $0 \leq \theta_t^+ \leq \overline{\Theta_t^+}$ (where

$\overline{\Theta_t^+}$ is defined in the proposition). The lower bound is obtained by setting $\theta_t^+ = 0$ and θ_t^- equal

to its maximum feasible value, where θ_t^- is restricted to lie in the range $0 \leq \theta_t^- \leq \overline{\Theta_t^-}$. Thus,

$P(U \leq t \mid I^* = 0, Y = 0)$ in (23) is bounded to lie within $[\Omega_1(t, v), \Omega_2(t, v)]$. Varying

$P(U \leq t \mid I^* = 0, Y = 0)$ in (22) within this feasible range reveals that $P(U \leq t \mid I^* = 1, Y = 0)$ must

lie within the range $[G_L(t, v), G_H(t, v)]$.[26] Integrating across values of t, expected health care

utilization among the unverifiably truly insured is bounded as follows:

$$\int U dG_H \leq E(U \mid I^* = 1, Y = 0) \leq \int U dG_L. \tag{24}$$

Continuing with the case $P_{10} \in (0, 1)$, applying this result to the first term in (9) and varying

$E[U(I^* = 1) \mid I^* = 0, Y = j]$ within $[0, \sup U]$ for $j = 0, 1$ in (8) and (9) yields the Proposition 2

bounds. These bounds also apply when P_{10} is 1 or 0. In the former case, $E\left[U(I^* = 1) \mid Y = 0\right]$

in (9) is identified as $E(U \mid I^* = 1, Y = 0) = E(U \mid Y = 0)$ since

$E(U \mid Y = 0) = E(U \mid I^* = 1, Y = 0) P_{10} + E(U \mid I^* = 0, Y = 0)(1 - P_{10})$. For $P_{10} = 0$, we only know

that $E\left[U(I^* = 1) \mid Y = 0\right] = E\left[U(I^* = 1) \mid I^* = 0, Y = 0\right] \in [0, \sup U]$.

---

[26]Note that $v > 0$ also directly places restrictions on $P(U \leq t \mid I^* = 1, Y = 0)$. However, we can show that

the direct restrictions on this quantity represent a subset of the restrictions imposed on it indirectly via the

restrictions on $P(U \leq t \mid I^* = 0, Y = 0)$.

Table 1

Reported Insurance Status, Service Use, Expenditures, and Verification of Insurance Status:

Nonelderly Population July 1996

| | Insurance Status Reported by Family | | | | |
| | Insured | | | | |
	Private	Public	Overall	Uninsured	Overall
Percent of Sample	69.1	11.7	80.7	19.3	100.0
Mean Ambulatory Provider Visits	0.425*	0.580*	0.448*	0.262	0.412
Percent Using Hospital or Ambulatory Services	22.1*	25.1*	22.5*	12.7	20.6
Mean Expenditures for Hospital and Ambulatory Services	$107*	$154*	$114*	$36	$99
Percent Verified by Insurance Cards, Policy Booklets, Employers, or Insurance Companies	82.5*	64.3*	80.2*	11.7	67.0
Number of Observations	11,984	2,788	14,772	4,079	18,851

DATA: Medical Expenditure Panel Survey Household Component and linked Insurance Component, 1996. Sample members age 0 to 64 as of July, 1996. Ambulatory provider visits include medical provider office visits, hospital outpatient visits, and emergency room visits.

* Statistically different from uninsured at the 0.01 level, two-tailed test.

Table 2

Bounds on the Monthly Utilization Gap Between the Insured and Uninsured, Nonelderly Population July 1996

Lower Bound on the Proportion of Unverified Cases Reported Accurately (v)	Assuming the Following Patterns of Insurance Classification Errors:			
	(1) Arbitrary Errors [a]	(2) Orthogonal Errors [b]	(3) Nondifferential Errors [c]	(4) Nonincreasing in Utilization [d]
I. Probability of Using Any Hospital or Ambulatory Services				
$v=1$	[0.098, 0.098] [†] [0.083 0.112] [‡]	[0.098, 0.098] [0.083 0.112]	[0.098, 0.098] [0.083 0.112]	[0.098, 0.098] [0.083 0.112]
$v=0.92$	[-0.034, 0.229] [-0.046 0.241]	[0.052, 0.216] [0.035 0.233]	[0.098, 0.113] [0.081 0.130]	[0.071, 0.120] [0.055 0.136]
$v=0.84$	[-0.080, 0.239] [-0.093 0.253]	[0.012, 0.225] [0.002 0.241]	[0.098, 0.133] [0.081 0.150]	[0.037, 0.140] [0.021 0.157]
$v=0.50$	[-0.512, 0.281] [-0.551 0.342]	[-0.384, 0.228] [-0.511 0.333]	[0.098, 0.193] [0.079 0.210]	[-0.348, 0.229] [-0.422 0.292]
II. Mean Number of Visits				
$v=1$	[0.186, 0.186] [0.136 0.237]	[0.186, 0.186] [0.137 0.235]	[0.186, 0.186] [0.137 0.235]	[0.186, 0.186] [0.137 0.235]
$v=0.92$	[-0.135, 0.472] [-0.192 0.529]	[0.024, 0.276] [-0.038 0.338]	[0.186, 0.251] [0.137 0.300]	[0.133, 0.264] [0.083 0.314]
$v=0.84$	[-0.238, 0.493] [-0.304 0.559]	[-0.079, 0.289] [-0.144 0.355]	[0.186, 0.324] [0.137 0.373]	[0.066 0.347] [0.013 0.400]
$v=0.50$	[-1.761, 0.581] [-2.109 0.929]	[-0.676, 0.289] [-0.770 0.389]	[0.186, 0.324] [0.137 0.373]	[-0.989, 0.461] [-1.251 0.723]
III. Mean Hospital and Ambulatory Expenditures ($)				
$v=1$	[77, 77] [47 107]	[77, 77] [45 110]	[77, 77] [45 110]	[77, 77] [45 110]
$v=0.92$	[-40, 117] [-83 160]	[22, 104] [13 139]	[77, 80] [45 112]	[68, 85] [35 118]
$v=0.84$	[-67, 122] [-118 173]	[-9, 105] [-23 139]	[77, 85] [45 117]	[56, 95] [20 131]
$v=0.50$	[-607, 145] [-863 401]	[-71, 105] [-145 179]	[77, 88] [45 121]	[-167, 121] [-297 251]

NOTES: [a] No restrictions; [b] imposes $P(I^*=1/Z^*=0)=P(I^*=1/Z^*=1)$; [c] imposes $P(I=1/I^*) = P(I=1/I^*,U)$; [d] imposes $P(I=1/I^*=0,U_1) \leq P(I=1/I^*=0,U_0)$ and $P(I=1/I^*=0,U_1) \leq P(I=1/I^*=0,U_0)$ for $U_1 \geq U_0$ where U = use, visits, or expenditures; I^* = true insurance status; I = reported insurance status; $Z^*=1$ if $I^*=I$. Insurance status is verified for 67% of the sample.

[†] Point estimates of the population bounds.
[‡] 95% confidence intervals for the identification region estimated using methods in Chernozhukov, Hong, and Tamer (2007)

Table 3

Bounds on the Monthly Utilization Rate of the Nonelderly Population If the Uninsured Became Insured, 1996

| | | Upper Bounds Assuming Monotone Treatment Response (MTR) and Monotone Treatment Selection (MTS) | | | | | | |
| | | With No Monotone Instrumental Variables (MIV) and the Following Patterns of Insurance Classification Errors: | | | | With Age and Health MIV: | | |
Lower Bound on the Proportion of Unverified Cases Reported Accurately (*v*)	(1) Arbitrary Errors, No Monotonicity Assumptions	(2) Arbitrary Errors	(3) Orthogonal Errors	(4) Nondifferential Errors	(5) Nonincreasing in Utilization	(6) Arbitrary Errors	(7) Nonincreasing in Utilization
		I. Probability of Using Any Hospital or Ambulatory Services (status quo = 0.206)					
v =1	[0.182, 0.374][†]	0.225	0.225	0.225	0.225	0.221	0.221
	[0.173 0.383][‡]	0.232	0.232	0.232	0.232	0.228	0.228
v =0.92	[0.155, 0.422]	0.246	0.244	0.228	0.232	0.234	0.224
	[0.145 0.432]	0.254	0.252	0.235	0.240	0.241	0.231
v =0.84	[0.153, 0.448]	0.255	0.245	0.231	0.241	0.241	0.229
	[0.143 0.459]	0.262	0.253	0.240	0.249	0.248	0.237
v =0.50	[0.153, 0.506]	0.291	0.245	0.248	0.268	0.277	0.256
	[0.142 0.516]	0.299	0.253	0.257	0.276	0.285	0.264
		II. Mean Number of Visits (status quo = 0.412)					
v =1	[0.362, 0.712]	0.448	0.448	0.448	0.448	0.440	0.440
	[0.339 0.735]	0.468	0.468	0.468	0.468	0.459	0.460
v =0.92	[0.305, 0.779]	0.503	0.461	0.456	0.463	0.483	0.444
	[0.283 0.801]	0.524	0.481	0.476	0.485	0.504	0.465
v =0.84	[0.305, 0.827]	0.520	0.462	0.465	0.479	0.506	0.469
	[0.283 0.849]	0.541	0.481	0.486	0.499	0.527	0.490
v =0.50	[0.305, 0.985]	0.600	0.462	0.488	0.554	0.577	0.535
	[0.283 1.006]	0.622	0.481	0.509	0.577	0.590	0.559

		Upper Bounds Assuming Monotone Treatment Response (MTR) and Monotone Treatment Selection (MTS)						
		With No Monotone Instrumental Variables (MIV) and the Following Patterns of Insurance Classification Errors:				With Age and Health MIV:	
Lower Bound on the Proportion of Unverified Cases Reported Accurately (v)	(1) Arbitrary Errors, No Monotonicity Assumptions	(2) Arbitrary Errors	(3) Orthogonal Errors	(4) Nondifferential Errors	(5) Nonincreasing in Utilization	(6) Arbitrary Errors	(7) Nonincreasing in Utilization
		III. Mean Hospital and Ambulatory Expenditures (status quo = $99)					
$v =1$	[$92, $258]	$114	$114	$114	$114	$114	$114
	[69 281]	139	139	139	139	140	140
$v =0.92$	[76, 284]	124	118	114	117	120	115
	[52 308]	149	142	139	143	146	141
$v =0.84$	[76, 306]	129	118	114	120	126	117
	[52 331]	154	142	140	148	152	144
$v =0.50$	[76, 394]	149	118	116	136	140	129
	[52 417]	179	142	141	167	170	158

NOTES: Monotone treatment response: an uninsured individual's use would not decline if she became insured; monotone treatment selection: under universal coverage, the currently insured would use at least as much services as the currently uninsured. Contaminated sampling imposes $P(I^*=1/Z^*=0)=P(I^*=1/Z^*=1)$, nondifferential errors imposes $P(I=1/I^*) = P(I=1/I^*,U)$, and nonincreasing error rates imposes $P(I=1/I^*=0,U_1) \leq P(I=1/I^*=0,U_0)$ and $P(I=1/I^*=0,U_1) \leq P(I=1/I^*=0,U_0)$ for $U_1 \geq U_0$ where U = use, visits, or expenditures; I^* = true insurance status; I = reported insurance status; Z^* = 1 if I^* = I. Monotone instrumental variables estimates assume use and expenditures are nondecreasing in age among those older than 30 and nondecreasing in perceived worse health status.

[†]Point estimates of the population bounds
[‡]95% confidence intervals for the identification region estimated using methods in Chernozhukov, Hong, and Tamer (2007)

DATA: Medical Expenditure Panel Survey Household Component and linked Insurance Component, 1996. Sample members age 0 to 64 as of July, 1996.

Table 4

Nonelderly Population's Use of Hospital or Physician Services by Age and

Perceived Health Status: Reportedly Insured in July, 1996

	Perceived Health Status			
Age	Excellent	Very Good	Good	Poor or Fair
0 to 29	16.1	17.4	23.6	29.7
30 to 34	15.2	21.1	21.6	40.5
35 to 40	16.9	20.4	29.4	41.4
40 to 44	18.3	22.8	27.4	42.1
45 to 50	18.4	27.2	26.8	49.8
50 to 54	20.2	28.0	26.5	51.2
55 to 60	30.4	29.3	32.7	46.0
60 to 64	24.2	35.8	38.4	52.6

DATA: Medical Expenditure Panel Survey Household Component, 1996. Sample
members age 0 to 64 as of July, 1996 reportedly covered by insurance.

Table 5

Upper Bounds on the Monthly Utilization Rate of the Nonelderly Population If the Uninsured Became Publicly Insured, July 1996

Lower Bound on the Proportion of Unverified Cases Reported Accurately (*v*)	(1) Arbitrary Errors, No Monotonicity Assumptions	Assuming Monotone Treatment Response (MTR) and Monotone Treatment Selection (MTS)					
		With No Monotone Instrumental Variables (MIV) and the Following Patterns of Insurance Classification Errors:				With Age and Health MIV:	
		(2) Arbitrary Errors	(3) Orthogonal Errors	(4) Nondifferential Errors	(5) Nonincreasing in Utilization	(6) Arbitrary Errors	(7) Nonincreasing in Utilization
I. Probability of Using Any Hospital or Ambulatory Services (status quo = 0.206)							
v =1	[0.182, 0.374][†]	0.230	0.230	0.230	0.230	0.224	0.224
	[0.173 0.383][‡]	0.238	0.238	0.238	0.238	0.234	0.234
v =0.89	[0.171, 0.405]	0.267	0.258	0.239	0.249	0.259	0.235
	[0.161 0.415]	0.275	0.266	0.248	0.258	0.270	0.250
v =0.77	[0.171, 0.405]	0.294	0.287	0.248	0.258	0.281	0.244
	[0.161 0.415]	0.303	0.297	0.257	0.267	0.293	0.256
v =0.50	[0.171, 0.405]	0.299	0.287	0.250	0.258	0.289	0.247
	[0.161 0.415]	0.309	0.297	0.259	0.268	0.302	0.260
II. Mean Number of Visits (status quo = 0.412)							
v =1	[0.358, 0.708]	0.473	0.473	0.473	0.473	0.457	0.457
	[0.335 0.732]	0.497	0.497	0.497	0.497	0.483	0.483
v =0.89	[0.335, 0.770]	0.587	0.534	0.493	0.518	0.572	0.503
	[0.314 0.791]	0.614	0.561	0.517	0.547	0.604	0.536
v =0.77	[0.335, 0.789]	0.683	0.555	0.524	0.563	0.660	0.545
	[0.313 0.812]	0.722	0.583	0.556	0.595	0.703	0.582
v =0.50	[0.335, 0.789]	0.690	0.572	0.525	0.571	0.672	0.556
	[0.313 0.812]	0.729	0.603	0.556	0.601	0.715	0.590

Lower Bound on the Proportion of Unverified Cases Reported Accurately (v)	(1) Arbitrary Errors, No Monotonicity Assumptions	Assuming Monotone Treatment Response (MTR) and Monotone Treatment Selection (MTS)				With Age and Health MIV:	
		With No Monotone Instrumental Variables (MIV) and the Following Patterns of Insurance Classification Errors:					
		(2) Arbitrary Errors	(3) Orthogonal Errors	(4) Nondifferential Errors	(5) Nonincreasing in Utilization	(6) Arbitrary Errors	(7) Nonincreasing in Utilization
III. Mean Hospital and Ambulatory Expenditures (status quo = $99)							
v =1	[$92, $258]	$122	$121	$121	$121	$120	$120
	[69 281]	155	154	154	154	156	155
v =0.89	[83, 281]	149	135	123	131	138	125
	[59 304]	186	167	155	167	177	164
v =0.77	[83, 294]	167	139	124	141	150	133
	[59 317]	208	172	157	180	194	170
v =0.50	[83, 294]	167	141	124	141	152	135
	[59 317]	208	173	157	180	196	179

NOTES: Monotone treatment response: an uninsured individual's use would not decline if she became insured; monotone treatment selection: under universal coverage, the currently insured would use at least as much services as the currently uninsured. Contaminated sampling imposes $P(I^*=1/Z^*=0)=P(I^*=1/Z^*=1)$, nondifferential errors imposes $P(I=1/I^*) = P(I=1/I^*,U)$, and nonincreasing error rates imposes $P(I=1/I^*=0,U_1) \leq P(I=1/I^*=0,U_0)$ and $P(I=1/I^*=0,U_1) \leq P(I=1/I^*=0,U_0)$ for $U_1 \geq U_0$ where U = use, visits, or expenditures; I^* = true insurance status; I = reported insurance status; $Z^* = 1$ if $I^* = I$. Monotone instrumental variables estimates assume use and expenditures are nondecreasing in age among those older than 30 and nondecreasing in perceived worse health status.
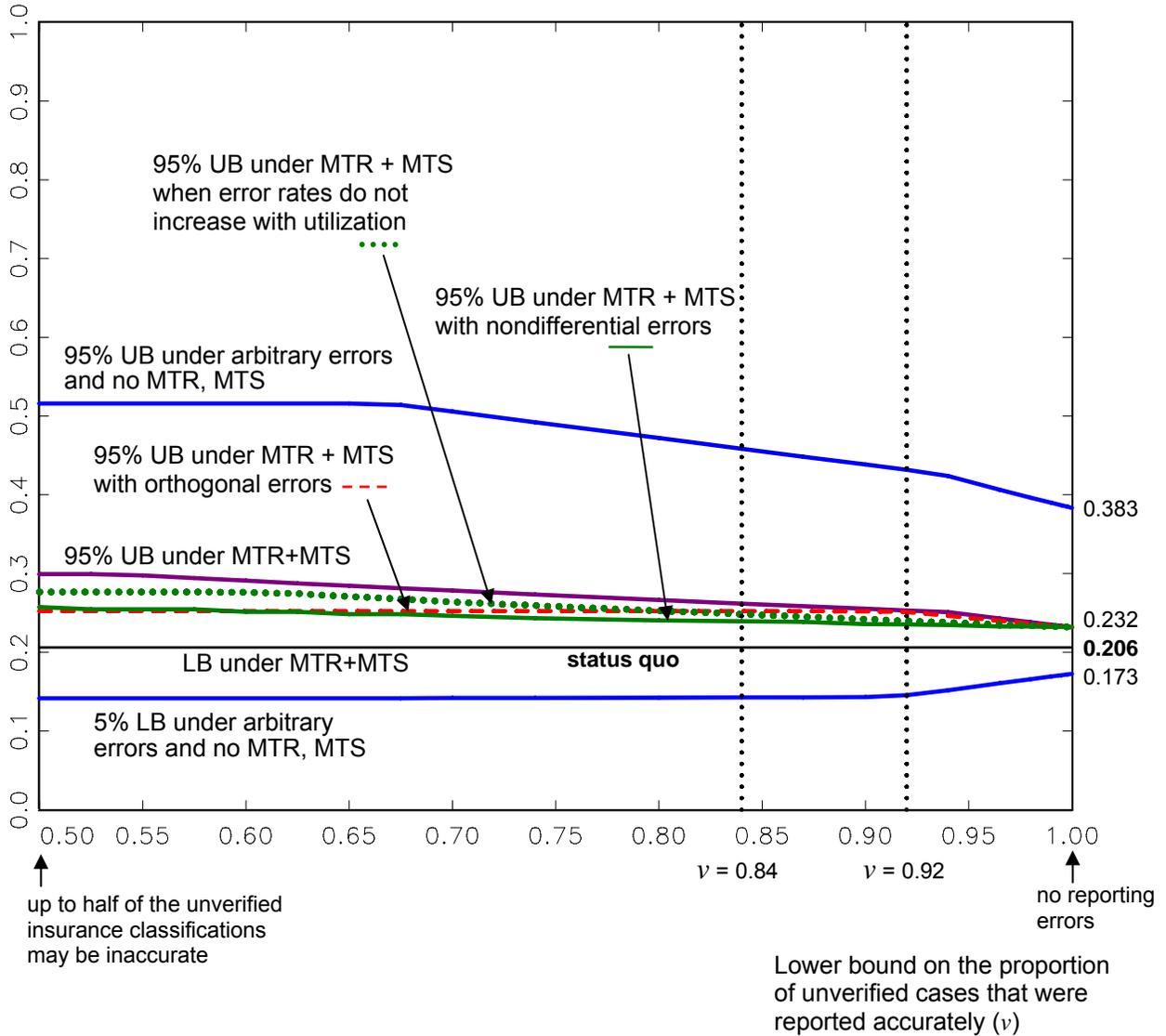
[†]Point estimates of the population bounds
[‡] 95% confidence intervals for the identification region estimated using methods in Chernozhukov, Hong and Tamer (2007)

DATA: Medical Expenditure Panel Survey Household Component and linked Insurance Component, 1996. Sample members age 0 to 64 as of July, 1996.

Figure 1

Bounds on the Fraction of the Nonelderly Population that Would Have Used Any Hospital or Ambulatory Services Nonelderly Population If the Uninsured Had a Mix of Private and Public Insurance, July 1996

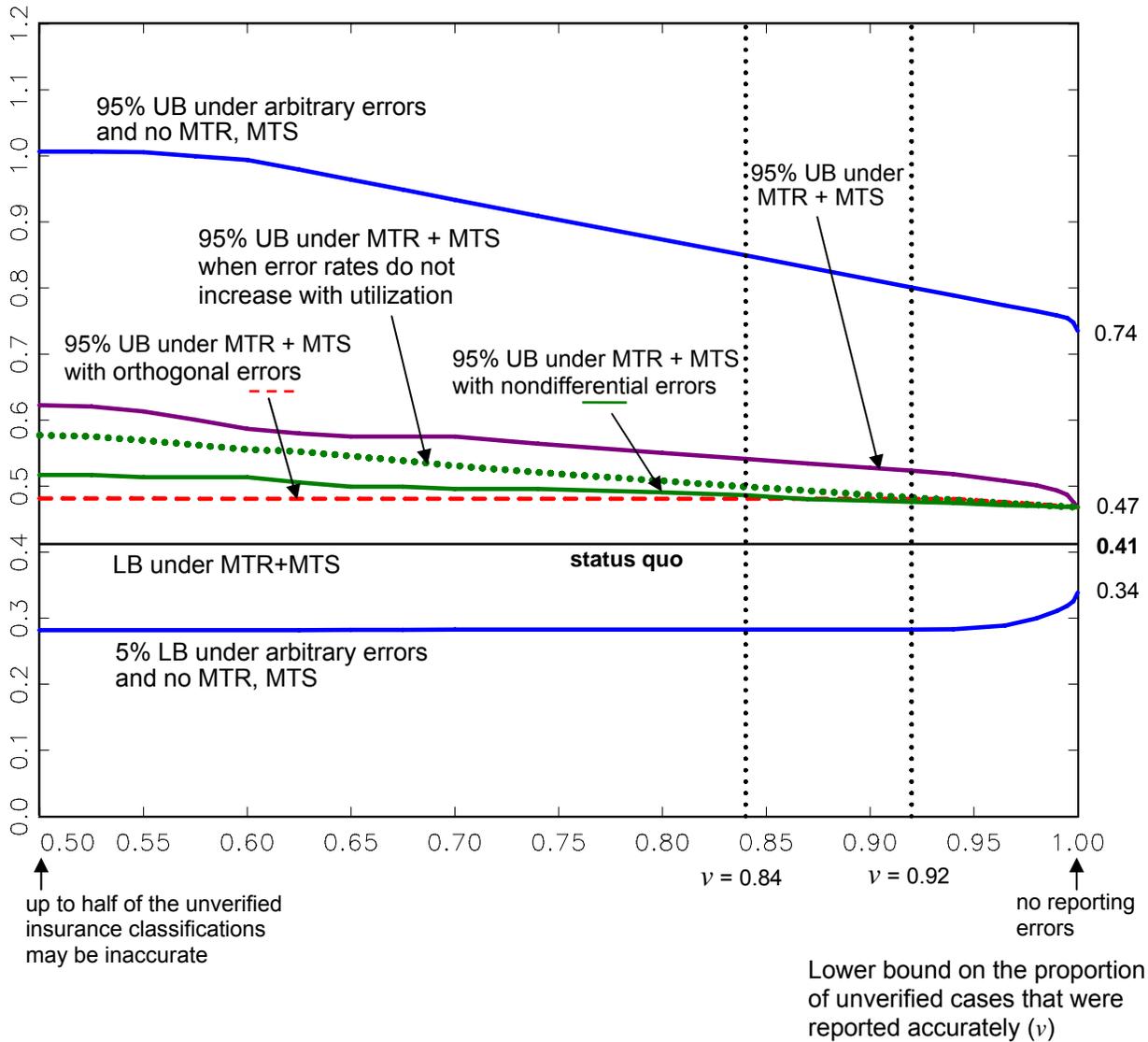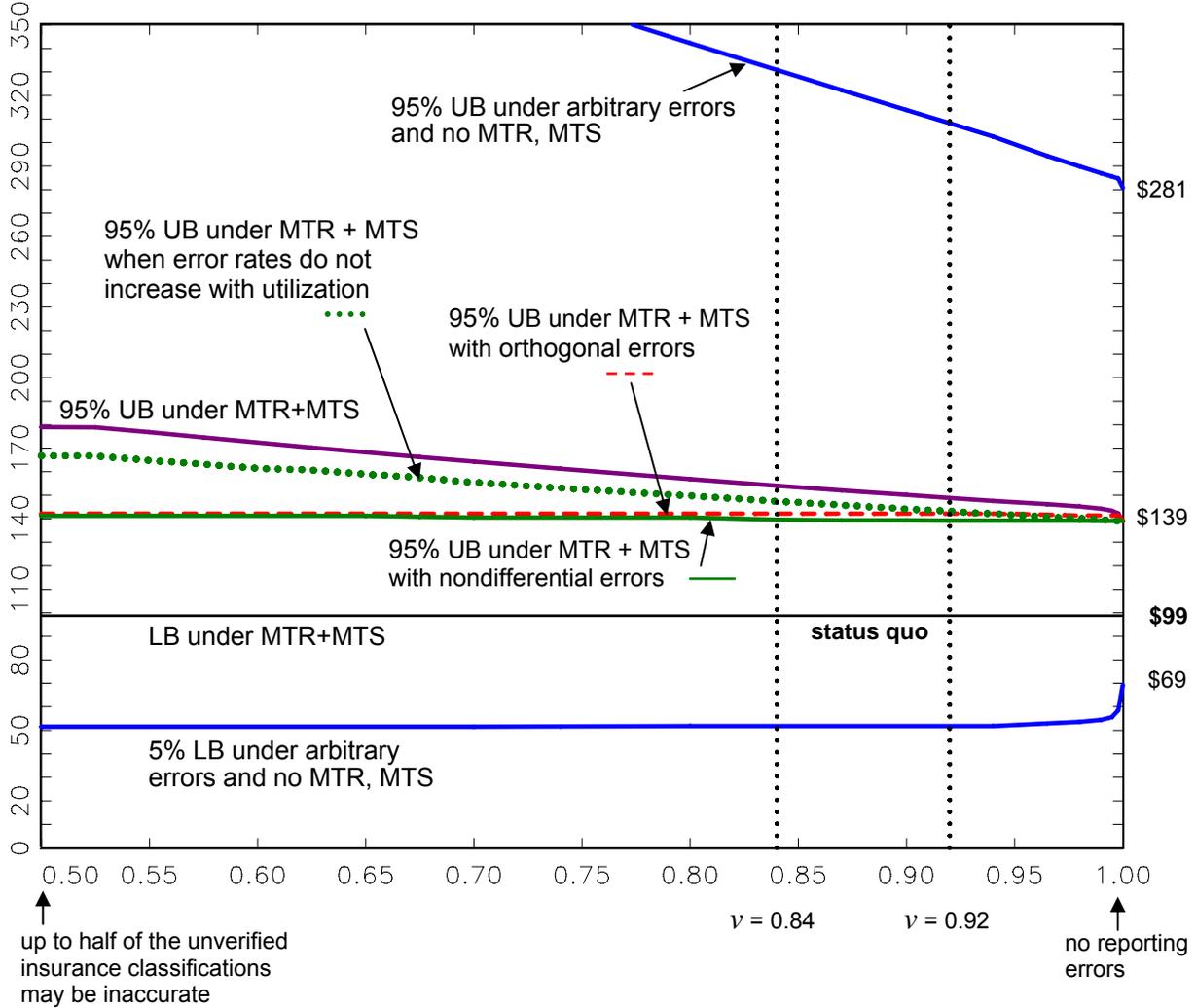**Fraction Using Any Services**

NOTES: MTR = monotone treatment response: an uninsured individual's use would not decline if she became insured. MTS = monotone treatment selection: under universal coverage, the currently insured would use at least as much services as the currently uninsured. Orthogonal errors imposes $P(I^*=1|Z^*=0) = P(I^*=1|Z^*=1)$, nondifferential errors imposes $P(I=1|I^*) = P(I=1|I^*,U)$, and nonincreasing error rates imposes $P(I=1|I^*=0,U_1) \leq P(I=1|I^*=0,U_0)$ and $P(I=1|I^*=0,U_1) \leq P(I=1|I^*=0,U_0)$ for $U_1 \geq U_0$ where U = use, visits, or expenditures; I^* = true insurance status; I = reported insurance status; Z^* = 1 if $I^* = I$. Vertical dotted lines reflect proposed values of v motivated in the text. Insurance status is verified for 67% of the sample. Confidence intervals for the bounds were computed using methods provided by Chernozhukov, Hong, and Tamer (2007).

DATA: Medical Expenditure Panel Survey Household Component and linked Insurance Component, 1996. Sample members age 0 to 64 as of July, 1996.

Figure 2

Bounds on the Nonelderly Population's Mean Number of Provider Visits
Nonelderly Population If the Uninsured Had a Mix of Private and Public Insurance, July 1996

Mean Visits

Figure 3

Bounds on the Nonelderly Population's Mean Hospital and Ambulatory Expenditures
Nonelderly Population If the Uninsured Had a Mix of Private and Public Insurance, July 1996

NOTES: MTR = monotone treatment response: an uninsured individual's use would not decline if she became insured. MTS = monotone treatment selection: under universal coverage, the currently insured would use at least as much services as the currently uninsured. Orthogonal errors imposes $P(I^*=1|Z^*=0) = P(I^*=1|Z^*=1)$, nondifferential errors imposes $P(I=1|I^*) = P(I=1|I^*,U)$, and nonincreasing error rates imposes $P(I=1|I^*=0,U_1) \leq P(I=1|I^*=0,U_0)$ and $P(I=1|I^*=0,U_1) \leq P(I=1|I^*=0,U_0)$ for $U_1 \geq U_0$ where U = use, visits, or expenditures; I^* = true insurance status; I = reported insurance status; Z^* = 1 if $I^* = I$. Vertical dotted lines reflect proposed values of v motivated in the text. Insurance status is verified for 67% of the sample. Confidence intervals for the bounds were computed using methods provided by Chernozhukov, Hong, and Tamer (2007).

DATA: Medical Expenditure Panel Survey Household Component and linked Insurance Component, 1996. Sample members age 0 to 64 as of July, 1996.