Economic Research Initiative on the Uninsured
CONFERENCE DRAFT

HEALTH CARE UTILIZATION PATTERNS AMONG THE UNINSURED: WHAT CAN BE LEARNED FROM DATA WITH ARBITRARY INSURANCE REPORTING ERROR?

Draft: Please do not cite or quote without permission.

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

Brent Kreider
Department of Economics
Iowa State University
Ames, IA 50011
bkreider@iastate.edu

Economic Research Initiative on the Uninsured
University of Michigan
555 South Forest Street, 3rd Floor
Ann Arbor, MI 49104-2531

July, 2004

Acknowledgements: Brent Kreider was provided generous financial support by a grant from the Robert Wood Johnson Foundation through the Economic Research Initiative on the Uninsured at the University of Michigan. Kathleen McMillan provided expert programming and Dzmitry Asinski provided valuable assistance with the data. The views expressed in this paper are those of the authors, and no official endorsement by the U.S. Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services is intended or should be inferred. All errors remain our own.

Abstract: Numerous studies have investigated relationships between health insurance status and a wide variety of outcome measures of interest such as the use of health services, health status, labor supply decisions, and participation in public assistance programs. Accurately identifying the uninsured is essential for making reliable inferences about these relationships, yet studies find a wide range of insurance classification errors rates in surveys.

Using detailed data from the 1996 Medical Expenditure Panel Survey (MEPS), we develop a formal nonparametric framework for assessing what can be learned in the presence of arbitrary insurance classification errors about (a) the gap in the probability of health care use between the insured and uninsured and (b) the impact of universal health insurance on health care utilization. The MEPS data include information from insurance cards and follow-back interviews with employers and insurance companies that allow us to confirm the insurance status of about two-thirds of the sample. We allow for the possibility of classification errors in the remainder of the sample and provide a range of results under various assumptions.

Under our strongest verification and monotonicity assumptions, we can tightly bound the increase in the fraction of the population using health care in a month following the introduction of universal health insurance to lie between 0 to 2.5 percentage points for adults and between 0 and 0.8 percentage points for children. However, identification of the parameters of interest deteriorates rapidly with the degree of uncertainty associated with insurance classifications.

1 Introduction

Policymakers have long been interested in identifying the number of people lacking health insurance, the consequences of uninsurance for access to health care, and the potential cost of covering the uninsured (for example, Institute of Medicine 2003). Rhoades and Cohen (2003) report that about a quarter of the U.S. nonelderly population was uninsured during at least part of 2001, with 13 percent of the population uncovered for the entire year. Numerous academic studies have investigated relationships between health insurance status and a wide variety of outcome measures of interest such as the use of health services, health status, labor supply decisions, and participation in public assistance programs.[1]

To make reliable inferences about these relationships, it is important to accurately measure insurance status. For example, researchers cannot accurately estimate differences in health care utilization between the insured and uninsured without accurately identifying which people are insured,[2] and it is well known that even random classification errors tend to bias parameters of interest. Uncertainty about these classifications also confounds the especially difficult problem of estimating casual effects, such as the projected impact of national health insurance on budgetary costs and health outcomes.[3]

The primary source of information about the uninsured is household surveys. Highlighting surprising degrees of insurance classification error in many popular national surveys (see Section 2) − along with dramatic inconsistencies in responses when experimental follow-up insurance questions have been posed − Czajka and Lewis (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."

Given the relatively large fraction of insured individuals, even small degrees of classification error in identifying the uninsured can lead to unreliable inferences. Using the Survey of Income and

[1] See, e.g., Gruber and Madrian (2001) and Levy (2001) for critical reviews of these literatures.
[2] Throughout this analysis, we assume that the outcome event, such as seeing a physician, is measured without error.
[3] 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). Estimates of incremental spending range from $34 to $69 billion per year, depending on the functional forms for expenditure regressions and assumptions about comparison groups.

Program Participation (SIPP) to measure uninsurance among children, Czajka and Lewis report that a 6-7 percentage point error in the reporting of children ever covered by health insurance during the year can double the estimated number of children without insurance the entire year. Small classification error rates can lead to large errors in estimates of the effects of insurance, in part because the insured and uninsured populations are denominators in comparisons of service use rates, costs, access to care, and other outcomes. Berger and colleagues (1998) find evidence that misreporting can seriously bias estimates of the effects of insurance on wage growth. Otherwise, nearly all econometric analyses have implicitly assumed fully accurate reporting of insurance status, despite evidence of problems measuring insurance status.

This paper provides the first formal framework for investigating relationships between health insurance status and other outcomes, such as medical service use, when insurance classifications are contaminated with arbitrary (non-classical) measurement error. Building on nonparametric methodologies developed by Molinari (2002) and Kreider and Pepper (2003),[4] we lay the groundwork for investigating what can be learned in the presence of arbitrary insurance classification errors about:

(a) the existing gap in the probability of health care use between the insured and uninsured, and

(b) the impact of universal health insurance on the probability of health care use.

Berger et al. (1998) assess the implications of insurance measurement error in a classical parametric errors-in-variables framework (modified to account for the binary nature of the variable measured with error) in which misclassifications are assumed to be random. Our framework allows for general unobserved patterns of classification errors.

Our primitive assumptions on classification errors are not strong enough to identify point estimates of the utilization gap or the impact of universal health insurance on utilization patterns, but we can provide informative nonparametric bounds without relying on the strong parametric assumptions that have been a source of controversy in econometric studies on other topics, such as disability classification

[4] These lines of research emanate from the work by Horowitz and Manski (1995).

errors.[5] In more than 50 articles pertaining to health insurance surveyed by Gruber and Madrian (2001), nearly all parameters of interest in the studies were identified using a variety of parametric approaches ranging from difference-in-differences to simultaneous equations limited-dependent variables models to highly sophisticated and parameterized structural dynamic programming models.[6] We investigate what can be learned in the absence of parametric assumptions. (In a future version of this paper, we will compare the identifying power of nonparametric and parametric assumptions.)

Our methodology separately generalizes parts of Kreider and Pepper's (2003) and Molinari's (2003) frameworks to allow researchers to impose priors on the degree of potential misclassifications within subpopulations suspected of providing some inaccurate responses.[7] We unify some of the previous nonparametric bounds literature and provide tighter bounds on the parameters of interest under the maintained assumptions than possible using previous studies.[8]

Our analysis exploits detailed data in the 1996 Medical Expenditure Panel Survey (MEPS), including information from insurance cards and follow-back interviews with employers and insurance companies. We use these additional sources of information to confirm families' reported insurance status for about two-thirds of the individuals in the sample. The data analysis presents a unique perspective on what can be inferred about insurance coverage patterns among MEPS sample members. In particular, we use conservative approaches to determine agreement between reported insured status, insurance cards produced by some respondents at the time of the interview, and subsequent confirmation of insurance

---

[5] See, e.g., Benitez-Silva et al. (forthcoming) and Kreider and Pepper (2004) for discussion. There is a large and growing literature on measurement error related to health status and labor force activity. See Bound and Burkhauser (1999) for a review. Some of the more recent articles include Maag and Wittenburg (2003), Kreider and Pepper (2003), and Benitez-Silva et al. (forthcoming).

[6] An exception is the study by Olson (1998) who uses semiparametric techniques to estimate the relationship between women's labor hours and the availability of health insurance through a spouse.

[7] These subpopulations pertain to nonresponses in Molinari's framework.

[8] Much attention has focused on inferring a linear mean regression when one or more of the regressors may be measured with error. For the most part, this literature has considered the classical errors-in-variables model where the regressors are continuous and the measurement error is independent of the structural error. More recently, researchers have relaxed these restrictions. Bollinger (1996), for example, bounds the mean regression in the classical setting when the mismeasured regressor is a binary variable.

coverage by employers, insurance companies, and providers. We also describe apparent inconsistencies between reported insured status and sources of medical payments.[9]

The next section describes potential sources of health insurance classification errors in surveys. Section 3 discusses the MEPS data and our health insurance verification strategies. Section 4 presents the statistical identification problem associated with estimating the health care utilization gap between the insured and uninsured when insurance status is subject to arbitrary classification error. This part of the analysis generalizes some of the statistical methodology in Kreider and Pepper (2003). Section 5 generalizes parts of Molinari's (2002) methodology to investigate what can be learned about changes in the probability of health services use under mandatory national health insurance. Section 6 concludes.

2 Sources of health insurance misclassifications

Misclassification of insurance status in survey data can arise from a variety of sources.[10] Apart from random coding errors and household nonresponse, some respondents are mistaken about their insured status or their family members' insured status, and they may imperfectly recall past insurance status. Some respondents may know Medicaid only by another name (such as a generic state name) which is not asked about in the survey, although some datasets which specialize in gathering health-related data, such as the MEPS, go to substantial lengths to mitigate such problems.

Studies find a wide range of rates at which insurance status is misclassified in surveys. The largest estimated misclassifications involve Medicaid. By benchmarking survey responses to administrative data, Bennefield (1996) estimates that Medicaid coverage is underreported in the Current Population Survey (CPS) and the SIPP by about 15 percent. More detailed analyses compare survey responses to administrative records. For example, Blumberg and Cynamon (2001) estimated that survey respondents did not report Medicaid coverage for 21 percent of children with Medicaid in the

[9] While we describe inconsistencies, in no case do we assume that a respondent's report is inaccurate; we merely fail to confirm its accuracy.

[10] See, e.g., Swartz (1986, 1987) and Czajka and Lewis (1999) for a more detailed discussion of the difficulties in interpreting estimates from national surveys of the number of uninsured.

administrative records. Call and colleagues (2001) found that in Minnesota, many of those not reporting Medicaid report other public programs or private coverage, so uninsurance does not appear to be as overreported as Medicaid is underreported.

There is also evidence of misreporting private insurance. Using the CPS, Berger and colleagues (1998) find that 11 percent of full-time full-year workers with at least two years job tenure who participated in both the widely-used March survey and supplemental April/May survey (in 1988 and 1993) gave inconsistent responses regarding their employer-provided insurance coverage, apparently due in part to differences in the wording of the health insurance questions. Although an individual's true insurance status may change over time, the authors argue convincingly that the fraction of discrepancies over the short horizon is too large to be attributed mostly to changes in actual insurance status. Using a separate unique survey of employers and their employees, Berger and colleagues find that 21 percent of workers and their employers disagree about whether the worker was eligible for insurance. Assuming random misreporting in that survey, they find that classification errors in employment-related insurance significantly biases estimates of the effects of insurance on wage growth toward zero.

Despite higher error rates in reporting the *type* of insurance, Pascale (2002) reviews both published and unpublished survey methods research studies and finds there is less misreporting about having *any* insurance. In particular, she finds in cognitive studies that nearly everyone appears to understand and correctly answer questions about having any insurance. These types of studies involve small samples, however, which may not be representative of the general population. In a larger study, Nelson and colleagues (2000) compared reports by 351 nonelderly Wisconsin adults and their insurers and found 97.8 percent of insurers agreed with the adults' reporting insurance. The level of misreporting in cases that could not be verified, however, is unknown and may be higher than among those who cooperated with the verification study. Other studies reviewed by Monheit (2003) suggest reported insurance status can be sensitive to how questions are asked. For persons who do not indicate insurance coverage during the sequence of questions asking about specific sources of coverage, some surveys specifically ask whether the person is insured. Monheit found that the effect of these questions varied across surveys. In telephone surveys with a shorter set of questions about types of insurance, these questions reduce the percent reportedly uninsured, but there is much less effect on in-person surveys that ask more questions about health insurance.

Evidence also suggests recall bias may cause classification errors in retrospective reports of insurance status. In their Wisconsin study, Nelson and colleagues (2000) asked insurers about duration of insurance coverage within weeks after family members reported insurance status. They found adults reported they had their insurance longer than insurers said they did: when an insurer said an adult was insured for a year or less, the adult agreed only 40 percent of the time.

More generally, the magnitudes of the classification errors and estimates of the uninsured may be related to the basic survey design features (for example, in-person vs. telephone interviews, rules for when proxy respondents are permitted, interviewer training, questionnaire wording and flow), answers varying with context, data editing and imputation methods, and errors in variables used to post-stratify sampling weights (Czajka and Lewis 1999; Hess et al. 2002; Swartz 1987).[11] Moreover, surveys often do not cover the entire population of interest (for example, they may exclude the homeless). The identification analysis which follows assumes that the researcher's obtained (weighted) sample is representative of the population of interest.

3 The Medical Expenditure Panel Survey

The data are 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 has 18,851 individuals.

[11] We thank Tom Selden for pointing out the source of error associated with sampling weights.

We study insurance and service use in July 1996. Studying insurance and service use in one month reduces the likelihood of confounding the dynamics of insurance status with misreported insurance status because employment-related insurance typically covers an entire month. 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 1st, 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.[12]

3.1 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 about all jobs held by co-residing 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[13]

---

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

[13] 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 schools, the Veterans Administration, and the Indian Health Services, are not included in measures of hospital/physician insurance.

- 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.[14]

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. 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 78.4 percent of the nonelderly population was insured in July 1996 and 21.6 percent were uninsured (Table 1). Uninsurance was lower for children than adults.

3.2 Probability of Service Use

In each interview, the MEPS asks about health care services used by all co-residing family members since the last interview. We create an indicator for whether the sample person had a hospital stay or ambulatory medical care (a medical provider visit, hospital outpatient visit, or emergency room visit) in July 1996. Persons that the family respondent said were insured in July were nearly 80 percent more likely to have used medical care in that month (22.5 percent of the insured versus 12.7 percent of the uninsured, Table 1). The gap was smaller for children (17.4 percent of the insured versus 12.1 percent of the uninsured).

[14] State-specific program names are used in the questions. Single-service and dread disease plans are not included in measures of hospital/physician insurance. Insurance status is not imputed to families with missing data, which are rare.

3.3 Verification Data

We use detailed data to identify sample members for whom there is evidence confirming their insurance status. The 1996 MEPS includes three sources that can be used to confirm or contradict health insurance reported by households: (1) the HC interviewers ask respondents to show insurance cards, (2) separate interviews with employers, unions, and insurance companies of household members, and (3) information collected from providers about sources of payment for services. The employer, insurance company, and provider can confirm or contradict the family's response. Respondents for the family, for employers, for insurance companies, and for providers could err in reporting a person's insurance status; none provides a gold standard of information. Nonetheless, we use confirmations of insurance status as verification of insurance status. In particular, if any source contradicted the person's insurance status, then it was not verified. If any source confirmed the person was insured, then the person's insurance status was verified. It is difficult to verify the lack of insurance: if a source confirmed the person was not insured through that source, and the family did not report the person had insurance from any other source, then the person was verified as uninsured. Finally, for those lacking verification data, or lacking sufficient verification data, the person's insurance status was not verified.

Table 2 summarizes key features of each source of verification.[15] In the first interview, when a family reported private insurance, Medicare, or Medicaid (79.5 percent of the nonelderly population), the interviewer asked to see insurance cards (Table 2). In subsequent interviews, the interviewer did not ask to see cards, and for most people, July was asked about in the second interview. Hence, we assume verification in the first interview expended to the second interview. Since insurance cards are provided only if insurance was reported, they can only confirm insurance. The lack of insurance card may not indicate lack of insurance; the respondent may have been unwilling to provide the card, or the respondent may not have had another family member's card.

For nearly all sample members with jobs at the time of the first interview, MEPS attempted to interview their employers. For sample members with private insurance, MEPS attempted to interview employers, former

[15] Details are contained in the appendix.

employers, unions, and insurance companies providing insurance. Establishments were interviewed between August 1997 and February 1998 about the employee's or policyholder's insurance in July 1996 (Cooper et al. 1999). Because of the time lag of 13 to 20 months, some establishment respondents may choose to describe current insurance status rather than insurance in July 1996, despite clearly worded questions. Although these data could include 82.0 percent of population (including dependents of employees and policyholders), useful responses are available for 36.5 percent of the population (Table 2). Each establishment interview can confirm or contradict the family's report of private insurance from that source. Interviews with employers can confirm or contradict reported lack of private insurance; hence they can only partially verify that a person is uninsured. The only individuals who are confirmed as uninsured those in families with no employers reporting insurance, and at least one main employer reported a worker did not have insurance. Taking all these factors into account, these data can be used to confirm or contradict insurance status for 27.8 percent of the sample members.

For sample members who used medical services, medical providers and hospitals were interviewed for information about sources of payment, including payments by insurers. A fifth of the sample used services in July, but not all users were sampled for collection of provider data, and nonresponse also limits the availability of these data to 7.7 percent of the sample (Table 2).[16] These interviews can only indicate the person has insurance (5.9 percent of the sample); a variety of reasons (including deductibles) may explain why an insured person's medical care was not paid for by insurance.

Persons can have insurance from multiple sources, and hence confirmations or contradictions may come from multiple and possibly conflicting sources (Table 2). Any confirmation of insurance status is treated as verified as accurate. Uninsurance cannot be confirmed for persons in families where no one is employed.

3.4 Verification Results

About three-quarters of the insured were in families that showed the interviewer an insurance card for that source of insurance (Table 3). At the time of the round 1 interview, conditional on type of insurance, families were about as likely to show interviewers cards for private and public insurance. By

[16] For those not sampled, the MEPS has imputed payments, and the imputed data were not used in this analysis.

July, however, the population with public insurance had churned, so the percent of those publicly insured in July who showed cards during the first interview was lower than for the privately insured.

Among individuals with complete verification data from employers and insurance companies, there was high agreement between families and establishments. For those with insurance reported by the family respondent, 80.3 percent had an establishment reporting insurance, and 19.7 had an establishment reported uninsurance (Table 3). This is less than Nelson's and colleagues' (2000) finding of a high degree of accuracy in reporting any insurance, when compared to interviews with insurers, but their research design had less time between family and establishment interviews than the 1996 MEPS. For the family-reported uninsured, 90.0 percent had no establishment reporting insurance, and 10.0 percent had at least one establishment reporting insurance. Because establishments provided data for only 36.5 percent of sample members, and only three-quarters of those were confirmed or contradicted, however, the overall confirmation rate in the sample was low (23.0 percent of the insured and 22.0 percent of the uninsured).

For the small portion of the sample with source of payment data from at least one provider, 81.5 percent of the insured had a provider report insurance would pay for at least some care. Among those who the family respondent said were uninsured, 26.6 percent had at least some care in July paid for by insurance. While this result is intriguing, only 31 reportedly uninsured sample members have contradictory data in the month, so not much can be reliably said about them.[17] They may also reflect provider misreports.

Among those with enough data from insurance cards, employers, unions, or insurance companies (68.6 percent of the sample), agreement among sources was high. For those with insurance reported by the family respondent, 98.2 percent were verified as insured, mostly reflecting 100 percent confirmation among those showing insurance cards. Among those with cards or establishment data, 4.9 percent

[17] When we pooled these types of contradictions across months, the sample size is larger, and most contradictions are due to providers reporting private insurance. Providers may, however, report insurance other than physician/hospital insurance as private insurance. A small proportion of this also appears to be caused by misreported starting or ending months for coverage.

reported insurance and had cards, but the establishment said they were not insured. We count these cases as verified insured, because they had some physical evidence of insurance. Because the cards do not verify uninsurance, the verification rate for the uninsured is the same as for the establishment data alone. For the whole sample, taking into account those without cards or establishment interviews, 77.8 percent of individuals with insurance were verified.

Among those with verification data from any of the three sources, agreement with family reports was high. Nearly all those with insurance were verified, again reflecting 100 percent confirmation among those showing insurance cards. For uninsured, the provider data added some cases with contradictory information, so that 87.3 percent were confirmed and 12.7 percent contradicted. For the entire sample, more than three-quarters of those with insurance were confirmed, and one-fifth of the uninsured were confirmed.

Among the sample members for whom we had verification data, reported insurance status is pretty accurate. We could not, however, verify insurance status for the entire sample, and we especially had difficulty verifying uninsurance, which is a residual category. Combining verification for the insured and uninsured (and weighted by the percent of the population in each group), about 96 percent accurately reported their insurance status using either (1) card and establishment data or (2) card, establishment, and provider data. This mostly reflects 100 percent confirmation among those showing insurance cards. This is only slightly less than Nelson's and colleagues' (2000) finding of a high degree of accuracy in reporting any insurance. And apparent inaccuracies may be misreporting by employers, unions, insurance companies, and providers. Finally, the sample members we could not verify may have been less accurate in reporting their insurance status than the sample member for whom we had verification data.

3.5 Future Enhancements

Additional efforts with the data could slightly increase the percentage of the sample we verify, but they could also add additional measurement error. First, Czajka and Lewis (1999) suggest that uninsured children with a parent covered by Medicaid are likely covered by Medicaid. We can focus on parents who did not report activity limitations, to remove those who may have qualified for Medicaid based on their own disabilities, rather than as families. Second, we could extend verification based on provider-reports to other HIEU members whom the family respondent said had the same type of insurance. For the uninsured, we could extend providers' reports of Medicaid coverage to other HIEU members, if the person with Medicaid spending is not disabled.[18]

More importantly, we can look for measured characteristics that are correlated with of confirmations and contradictions of reported insurance status and assess their prevalence in the unverified population. Important factors may include the respondent's education and the respondent's relationship with the sample member.

We now turn to the identification problem associated with estimating differences in the probability of health care use between the insured and uninsured given uncertainty about the insurance status for part of the sample. In Section 6, we focus on inferences about changes in the probability of health services use under mandatory national health insurance.

[18] We considered and rejected other options. For example, we could use reports of paying for care out-of-pocket, but we find apparently insured respondents paying for some services entirely out-of-pocket. When a reportedly uninsured child receives care paid for by private insurance, we could assert that a parent is covered, but in two-parent families, both may not be covered, and in single-parent families a parent residing elsewhere may be the source of coverage. We could use family-reported payment sources, but family respondents typically have not used documents as memory aids when answering questions about charges and payments.

4 The Identification Problem

To evaluate the impact of inaccurate insurance classifications, we introduce notation which distinguishes between classified insured status and actual insured status. Let $I = 1$ be an unobserved indicator for whether the individual is truly insured, with $I = 0$ otherwise. Let X be the insured status reported by the family individual (after any editing by the MEPS), with $X = 1$ if the individual is reportedly insured and $X = 0$ otherwise. A latent variable Z indicates whether a report is accurate, with $Z = 1$ if $I = X$ and $Z = 0$ if $I \neq X$. In cases in which the value of X can be confirmed to be accurate via outside information ($Y = 1$) below, the value of Z is known to equal 1. Otherwise, the value of Z is unknown. In no case is the value of Z known to be 0.

Given uncertainty about the accuracy of reported insurance, the data do not identify the true proportion insured, $P(I = 1)$. The true fraction of insured individuals equals the fraction accurately classified as insured, $P(X = 1, Z = 1)$, plus the fraction inaccurately classified as uninsured, $P(X = 0, Z = 0)$. The true and reported insurance rates coincide only if the fraction of false positive classifications exactly offsets the fraction of false negative classifications.

Uncertainty about health insurance classifications exacerbates uncertainty about relationships between health insurance and health care utilization. For a particular type of health care utilization (e.g., inpatient, ambulatory, etc.), let $U = 1$ indicate that the individual used health care services during the reference period, with $U = 0$ otherwise. In this section, we investigate what can be learned about the utilization gap between the insured and uninsured when true insured status is unobserved:

$$\Delta = P(U = 1|I = 1) - P(U = 1|I = 0). \tag{1}$$

Given accurate utilization information, we observe the utilization rate conditional on reported insurance status, $P(U = 1|X)$.[1] In the presence of insurance classification errors, however, we do not observe true insurance status I, so the utilization gap in (1) is not identified by the data. Without theory or restrictions on insurance classification errors, Δ could lie anywhere between -1

[1] In this analysis, the "utilization rate" refers to the probability of medical service use.

and 1.

Figure 1 provides a graphical illustration of the identification problem. The probabilities reflect the MEPS July 1996 nonelderly sample. The circle represents the 21 percent of the population documented as using medical services in July 1996. Those classified as insured are represented in the top portion of the diagram, $X = 1$, with those classified as uninsured represented in the bottom portion, $X = 0$. Accurate insurance classifications are represented in the $Z = 1$ regions, with verified accurate classifications ($Y = 1$) distinguished from unverified classifications ($Y = 0$). The unknown fractions of inaccurate classifications are represented in the $Z = 0$ regions. Truly insured individuals – corresponding to those accurately classified as being insured $\{X = 1$ and $Z = 1\}$ or inaccurately classified as being uninsured $\{X = 0$ and $Z = 0\}$ – are contained in the dotted regions labeled with upper-case letters, (A+G+B+H)+(F+L). The remaining regions, labeled with lower-case letters, represent the truly uninsured. The data identify only the verified $Y = 1$ regions A, d, j, and G; the magnitudes of the remaining regions are unobserved.

Given that we do not observe the true utilization gap between the insured and uninsured, our objective is to provide a range of worst-case bounds on this gap given assumptions on the data-generating process. Using Bayes' rule, the utilization rate among the insured is given by

$$P(U = 1|I = 1) = \frac{P(U = 1, I = 1)}{P(I = 1)}$$

as represented by the ratio (A+B+F)/(A+G+B+H+F+L) in Figure 1. Neither the numerator nor the denominator are identified since only A and G are observed. Assumptions on classification errors, however, place restrictions on relationships between the unobserved quantities. To make progress, we decompose the numerator and denominator into verified and unverified components:

$$\begin{aligned} P(U = 1|I = 1) =& \frac{P(U = 1, I = 1, Y = 1) + P(U = 1, I = 1, Y = 0)}{P(I = 1, Y = 1) + P(I = 1, Y = 0)} \\ =& \frac{P(U = 1, X = 1, Y = 1) + \mathsf{B} + \mathsf{F}}{P(X = 1, Y = 1) + \mathsf{B} + \mathsf{H} + \mathsf{F} + \mathsf{L}}. \end{aligned} \quad (2)$$

In the second line, the joint distributions involving verified classifications, $Y = 1$, have been identified by replacing true insurance status I with reported insurance status X.

We next isolate false positive and false negative components. While the fraction of health care users who were accurately classified as insured without verification $\mathsf{B} = P(U = 1, X = 1, Z = 1, Y = 0)$ is unobserved, the total fraction of such users classified as insured without verification, $\mathsf{B} + \mathsf{c} = P(U = 1, X = 1, Y = 0)$ is observed. Replacing B in the numerator with $P(U = 1, X = 1, Y = 0) - \mathsf{c}$ and combining terms allows us to rewrite the numerator as $P(U = 1, X = 1) + \mathsf{F} - \mathsf{c}$ where the first term is observed and $\mathsf{F} - \mathsf{c}$ reflects the unobserved excess of false negative vs. false positive insurance classifications among those who utilized health care. Similarly, we can rewrite the denominator as $P(X = 1) + [(\mathsf{F} + \mathsf{L}) - (\mathsf{i} + \mathsf{c})]$ where the term in brackets reflects the unobserved excess of false positive vs. false negative insurance classifications among all individuals. Using these results, (2) becomes

$$P(U = 1|I = 1) = \frac{P(U = 1, X = 1) + \mathsf{F} - \mathsf{c}}{P(X = 1) + (\mathsf{F} + \mathsf{L}) - (\mathsf{i} + \mathsf{c})}. \tag{3}$$

The utilization rate among the uninsured, $P(U = 1|I = 0)$, can be decomposed in a similar fashion.

We now propose two sets of bounds on the unknown utilization gap between the insured and uninsured. The first set of bounds (Propositions 1A and 1B) imposes no structure on the distribution of false positives and false negatives. The second set of narrower bounds imposes a strong assumption that a person's health care utilization outcome does not depend on whether insurance status was correctly coded in the data. We will then show how the bounds can be narrowed under monotonicity assumptions that link the utilization rate to other covariates.

4.1 Partial verification bounds

Suppose that the researcher has no information about the distribution of insurance classification errors among unverified cases but is confident that the overall accurate classification rate among unverified cases is at least v_0:

$$P(Z = 1|Y = 0) \geq v_0.$$

If v_0 equals 1, then the utilization gap is exactly identified because all insurance classifications are known to be accurate. If the researcher is unwilling to assume anything about the unverified cases, then v_0 can be set equal to 0; in any case, the sensitivity of the utilization bounds can be examined by varying the value of v_0 between 0 and 1.

To find a lower bound on the utilization rate among the insured, first notice in (3) that $P(U = 1|I = 1)$ is increasing in $\mathsf{i} = P(U = 0, X = 1, Z = 0)$, the unobserved fraction of individuals who did not utilize health services and were misclassified as being insured. As a worst-case possibility for the lower bound, we must set $\mathsf{i} = 0$. Then differentiating (3) with respect to $\mathsf{F} = P(U = 1, X = 0, Z = 0)$ – the fraction of individuals who utilized care and were falsely classified as being uninsured – shows that $P(U = 1|I = 1)$ is rising in F for any values of L and c.[2] Therefore, we must also set F equal to 0 to obtain:

$$P(U = 1|I = 1) \geq \frac{P(U = 1, X = 1) - \mathsf{c}}{P(X = 1) + \mathsf{L} - \mathsf{c}}. \tag{4}$$

While c and L are unobserved, their ranges are restricted. The unobserved fraction of individuals who utilized care and were falsely classified as insured, $\mathsf{c} = P(U = 1, X = 1, Z = 0)$, cannot exceed the observed fraction of individuals who utilized care and were classified as insured but were not verified. Nor can this fraction exceed the total allowed fraction of misclassified cases. Similarly, the unobserved fraction of individuals who did not utilize care and were falsely classified as being uninsured, $\mathsf{L} = P(U = 0, X = 0, Z = 0)$, cannot exceed the observed fraction of individuals who did not utilize care and were classified as being uninsured but were not verified; nor can it exceed the total fraction of misclassified cases:

$$\begin{aligned} 0 \;\leq\; \mathsf{c} &\leq \min\{\phi, P(U = 1, X = 1, Y = 0)\} \equiv \bar{\mathsf{c}} \\ 0 \;\leq\; \mathsf{L} &\leq \min\{\phi, P(U = 0, X = 0, Y = 0)\} \equiv \bar{\mathsf{L}} \end{aligned}$$

where $\phi \equiv (1 - v_0)P(Y = 0)$ represents the fraction of total allowed misclassifications.

To find the lower bound of $P(U = 1|I = 1)$, we need to find the minimum feasible value for the right-hand side (4). Therefore, for any candidate value of c, we need L to attain its maximum

[2] The derivative has the same sign as $P(U = 0, X = 0) + \mathsf{L} > 0$.

allowed value condition on c:

$$\mathsf{L}^* = \min\left\{\phi - \mathsf{c}^*, \overline{\mathsf{L}}\right\} = \min\left\{\phi - \mathsf{c}^*, P(U=0, X=0, Y=0)\right\}$$

Therefore, the objective becomes one of minimizing

$$\frac{P(U=1, X=1) - \mathsf{c}}{P(X=1) + \min\left\{\phi - \mathsf{c}, P(U=0, X=0, Y=0)\right\} - \mathsf{c}} \tag{5}$$

over feasible values of c. It is useful to define $\mathsf{c}^o \equiv \phi - P(U=0, X=0, Y=0)$, the critical value of c which makes the two arguments in the min function equal.

First consider values of $\mathsf{c} \leq \mathsf{c}^o$. For such values, the derivative of (5) with respect to c is negative; therefore, we can exclude as potential candidates any values of c less than $\mathsf{c}_{\min} \equiv \max\left\{0, \min\left\{\overline{\mathsf{c}}, \mathsf{c}^o\right\}\right\}$. For $\mathsf{c} > \mathsf{c}^o$ the derivative has the same sign as

$$\delta_1 \equiv P(U=1, X=1) - P(U=0, X=1) - \phi. \tag{6}$$

When this quantity is negative, we must raise c to its maximum feasible value, $\mathsf{c}^* = \overline{\mathsf{c}}$; otherwise, we set $\mathsf{c}^* = \mathsf{c}_{\min}$.

Similar logic provides an upper bound on $P(U=1|I=1)$. After defining $\delta_2 \equiv P(U=0, X=1) - P(U=1, X=1) - \phi$, the preceding results establish Proposition 1:

Proposition 1A. *Let $P(Z=1|Y=0) \geq v_0$. Then the following sharp bounds apply:*

$$\frac{P(U=1, X=1) - \theta_1}{P(X=1) + \min\left\{\phi - \theta_1, P(U=0, X=0, Y=0)\right\} - \theta_1} \tag{7}$$

$$\leq P(U=1|I=1) \leq$$

$$\frac{P(U=1, X=1) + \theta_2}{P(X=1) - \min\left\{\phi - \theta_2, P(U=0, X=1, Y=0)\right\} + \theta_2}$$

where

$$\theta_1 = \begin{cases} \min\{\phi, P(U=1, X=1, Y=0)\} & \text{if } \delta_1 \leq 0 \\ \max\{0, \min\left\{P(U=1, X=1, Y=0), \phi - P(U=0, X=0, Y=0)\right\}\} & \text{otherwise} \end{cases}$$

$$\theta_2 = \begin{cases} \min\{\phi, P(U=1, X=0, Y=0)\} & \text{if } \delta_2 \geq 0 \\ \max\{0, \min\left\{P(U=1, X=0, Y=0), \phi - P(U=0, X=1, Y=0)\right\}\} & \text{otherwise} \end{cases}$$

Analogous bounds for $P(U=1|I=0)$ are obtained by replacing $X=1$ with $X=0$ and vice versa.

These bounds are always informative as long as some insurance classifications are verified to be accurate. In particular, note that the lower bound is strictly positive when $P(Y=1)>0$ since $\theta_1 \leq P(U=1, X=1, Y=0) < P(U=1, X=1)$. It can also be shown that the upper bound is strictly less than 1. Note that increasing v_0 narrows the bounds over some ranges of v_0 but not others. Kreider and Pepper's (2003) degree bounds and verification bounds (their Propositions 1 and 2) apply as special cases when $P(Y=1)=1$ or $v_0=0$, respectively. Their partial verification bounds assume that nothing is known about extent of reporting errors within unverified subsamples.

Utilization gap

A naive upper (lower) bound on the utilization gap Δ can be found by subtracting the Proposition 1 lower (upper) bound on $P(U=1|I=0)$ from the Proposition 1 upper (lower) bound on $P(U=1|I=1)$. Although these bounds on the difference between the two conditional probabilities are intuitive and simple to compute, they are not sharp: the constraint on the fraction of misclassifications places further restrictions on Δ. These bounds are naive in the sense that they do not take into account the fact that some of the same unobserved parameters appear in both $P(U=1|I=1)$ and $P(U=1|I=0)$. Using this fact further constrains the upper and lower bound on the gap. Write

$$\Delta = P(U=1|I=1) - P(U=1|I=0) = \frac{P(U=1, X=1) + \mathsf{F} - \mathsf{c}}{P(X=1) + (\mathsf{F}+\mathsf{L}) - (\mathsf{i}+\mathsf{c})} - \frac{P(U=1, X=0) - \mathsf{F} + \mathsf{c}}{P(X=0) - (\mathsf{F}+\mathsf{L}) + (\mathsf{i}+\mathsf{c})}. \quad (8)$$

Using the reasoning above, finding the lower bound on Δ requires setting $\mathsf{i} = \mathsf{F} = 0$ and $\mathsf{L}^* = \min\{\phi - \mathsf{c}^*, P(U=0, X=0, Y=0)\}$ and then minimizing the resulting quantity over feasible values of c. Finding the upper bound requires setting $\mathsf{c} = \mathsf{L} = 0$ and $\mathsf{i} = \min\{\phi - \mathsf{F}^*, P(U=0, X=1, Y=0)\}$ and then maximizing the resulting quantity over feasible values of F. This result is stated as Proposition 1B:

Proposition 1B. *Let $P(Z=1|Y=0) \geq v_0$. Then the following sharp bounds on $\Delta = P(U=1|I=1) - P(U=1|I=0)$ apply:*

$$\inf_{\theta_1 \in [0, \min\{\phi, P(U=1,X=1,Y=0)\}]} \left\{ \frac{P(U=1,X=1)-\theta_1}{P(X=1)+\min\{\phi-\theta_1, P(U=0,X=0,Y=0)\}-\theta_1} - \frac{P(U=1,X=0)+\theta_1}{P(X=0)-\min\{\phi-\theta_1, P(U=0,X=0,Y=0)\}+\theta_1} \right\}$$

$$\leq \Delta \leq$$

$$\sup_{\theta_2 \in [0, \min\{\phi, P(U=1,X=0,Y=0)\}]} \left\{ \frac{P(U=1,X=1)+\theta_2}{P(X=1)-\min\{\phi-\theta_2, P(U=0,X=1,Y=0)\}+\theta_2} - \frac{P(U=1,X=0)-\theta_2}{P(X=0)+\min\{\phi-\theta_2, P(U=0,X=1,Y=0)\}-\theta_2} \right\}$$

Over part of the range of v_0, these bounds differ from the naive bounds obtained from Proposition 1A. Consider, for example, the lower bound in Proposition 1A. If the value of the unknown parameter θ_1 that minimizes the first expression (i.e., the lower bound on $P(U=1|I=1)$) differs from the value of θ_1 that maximizes the second expression (i.e., the upper bound on $P(U=1|I=0)$), the two bounds on Δ will differ and the Proposition 2 bounds will be tighter. The two bounds will be identical when the lower bound on $P(U=1|I=1)$ and the upper bound on $P(U=1|I=0)$ are realized at same value of the unknown parameter θ_1.

4.2 Results

Figure 2 presents Proposition 1A bounds on the probability of health care use $P(U=1|I)$ by true insured status for July 1996. For this figure, true insured status is verified for 67.1 percent of the MEPS sample based on insurance cards and employer confirmation. True insured status of the remaining fraction of individuals is unknown. The horizontal axis measures the assumed lower bound accurate insurance classification rate among the unverified cases, v_0. At one extreme when $v_0 = 1$, all unverified insurance classifications are assumed to be accurate. In this case, the probability of utilizing health services is point-identified for the insured and uninsured as equal to their reported utilization rates, $P(U=1|X=1) = 0.23$ and $P(U=1|X=0) = 0.13$, respectively.

Abstracting from sampling variability, there is no uncertainty about the utilization rates when $v_0 = 1$.[3]

At the other extreme, setting $v_0 = 0$ allows for the possibility that all unverified cases are miscoded. That is, all unverified individuals reporting to be insured may actually be uninsured (false positives), and all unverified individuals reporting to be uninsured may be insured (false negatives). Given this degree of uncertainty about which individuals are truly insured and which are truly uninsured, the bounds on the probability of service use widen to $[0.16, 0.29]$ for the insured and to $[0.02, 0.62]$ for the uninsured.

The bounds remain unchanged over a wide range of v_0. The lower bound on the utilization rate for the insured gradually begins to rise once v_0 exceeds 0.49 and then rises rapidly once v_0 exceeds 0.89. To understand these discrete jumps, refer back to equation (4). As a worst case lower bound on $P(U = 1|I = 1)$, we must make the unknown false positive quantity $\mathsf{c} = P(U = 1, X = 1, Z = 0)$ among health care users in Figure 1 as large as possible and then make the unknown false negative quantity $\mathsf{L} = P(U = 0, X = 0, Z = 0)$ among non-users as large as possible conditional on c.[4] For sufficiently low values of v_0, nothing prevents c from being as large as $(\mathsf{B}+\mathsf{c})$, the observed fraction of health care users who unverifiably report health insurance coverage, and nothing prevents L from being as large as $(\mathsf{k} + \mathsf{L})$, the observed fraction of non-users who unverifiably report being uninsured. Once v_0 exceeds 0.49, the allowed total degree of misclassification is small enough that L must begin declining with v_0. Then once v_0 exceeds 0.89, c must also start declining with v_0. Similar patterns apply to the other sets of bounds.

Note that unless v_0 exceeds 0.94, the upper bound on the probability of utilization among the uninsured exceeds the lower bound on the probability of utilization among the insured. Thus, without further assumptions on the distribution of insurance classification errors, we cannot be

[3] The next step in our analysis is to provide bootstrapped confidence intervals for the presented bounds. The bounds will naturally widen after accounting for sampling variability. Given the large sample size, however, preliminary findings suggest that the uncertainty created by the identification problem is much more extensive than the uncertainty created by sampling variability.

[4] Based on the data, it turns out that δ_1 in (6) is always negative. Therefore, for all values of v_0, the first priority is to make c is as large as possible before making L as large as possible.

confident that the probability of use is higher among the insured than among the uninsured unless most classifications are known to be accurate.

4.3 Independence between utilization and insurance coding errors

Propositions 1A and 1B allow for the possibility that utilization decisions might depend on perceived insured status instead of actual insured status. Here we develop bounds assuming that actual, rather than measured, insurance status affects service use. By comparing these new bounds with the ones developed in Propositions 1A and 1B, we can assess the extent to which uncertainty about the relationship between insurance and service use is related to perceived insurance status.[5]

Assume that service use depends on actual insured status but not on whether insured status is accurately coded:[6]

$$P(U = 1|I) = P(U = 1|I, Z). \tag{9}$$

That is, conditioning on the accuracy of the classifications does not add any additional information about utilization outcomes. In this case, an extra restriction is imposed on the unobservables in equation (3):[7]

$$\frac{P(U = 1, X = 1) - \mathsf{c}}{P(X = 1) - (\mathsf{i} + \mathsf{c})} = \frac{\mathsf{F}}{\mathsf{F} + \mathsf{L}}.$$

When the "irrelevance of coding errors" assumption (9) holds, the Proposition 1A bounds narrow as follows:

Proposition 2A. *Let $P(U = 1|I) = P(U = 1|I, Z)$ and $P(Z = 1|Y = 0) \geq v_0$. Then the following sharp bounds apply:*

$$\frac{P(U = 1, X = 1) - \min\{\phi, P(U = 1, X = 1, Y = 0)\}}{P(X = 1) - \min\{\phi, P(U = 1, X = 1, Y = 0)\}}$$

$$\leq P(U = 1|I = 1) \leq$$

[5] Some research indicates that perceived insured status has a greater effect on behavior than actual insured status (Reschovsky et al. 2002).

[6] It does not appear that this type of assumption has been previously introduced in the health or nonparametric bounds literatures. Note that this assumption does not impose a restriction that the trued insured rate is the same among accurate and inaccurate classifications.

[7] To see this, note that (9) implies $P(U = 1|I, Z = 1) = P(U = 1|I, Z = 0)$.

$$\frac{P(U=1,X=1)}{P(X=1)-\min\{\phi,P(U=0,X=1,Y=0)\}}$$

Proof. See Appendix.

It is easy to see that the bounds converge to the classified utilization rate, $P(U=1|X=1)$, when $v_0=1$ or $P(Y=0)=0$. Bounds for the utilization rate among the uninsured, $P(U=1|I=0)$, are obtained by replacing $X=1$ with $X=0$ and vice versa in Proposition 2A.

Sharp bounds on the utilization gap, Δ, when utilization is independent of insurance coding errors are provided in Proposition 2B:

Proposition 2B. *Let* $P(U=1|I)=P(U=1|I,Z)$ *and* $P(Z=1|Y=0)\geq v_0$*. Then the following sharp bounds on* $\Delta=P(U=1|I=1)-P(U=1|I=0)$ *apply:*

$$\inf_{\theta_1\in[0,\min\{\phi,P(U=1,X=1,Y=0)\}]}\left\{\frac{P(U=1,X=1)-\theta_1}{P(X=1)-\theta_1}-\frac{P(U=1,X=0)}{P(X=0)-\min\{\phi-\theta_1,P(U=0,X=0,Y=0)\}}\right\}$$

$$\leq\Delta\leq$$

$$\sup_{\theta_2\in[0,\min\{\phi,P(U=1,X=0,Y=0)\}]}\left\{\frac{P(U=1,X=1)}{P(X=1)-\min\{\phi-\theta_2,P(U=0,X=1,Y=0)\}}-\frac{P(U=1,X=0)-\theta_2}{P(X=0)-\theta_2}\right\}.$$

Proof. See Appendix.

4.4 Results

Figure 3 illustrates the identifying power of the irrelevance of coding errors assumption for the July 1996 MEPS sample. The solid lines are taken directly from Figure 2 when this assumption is not imposed. When the assumption is imposed, the bounds narrow as indicated by the dashed lines. The most striking improvement is on the upper bound probability of services use among the uninsured. When $v_0=0$, the width of the bounds for the uninsured declines from 0.60 in Figure 2 to 0.37 under the irrelevance assumption. Moreover, the upper bound much more rapidly converges to the classified utilization rate for values of v_0 exceeding 0.62. When $v_0=0$, the bounds for the insured narrow from 0.13 to 0.08. Recall that in Figure 2, the probability of utilization was not known to be higher among the insured than among the uninsured unless v_0 exceeds 0.94. Under the irrelevance assumption, this critical value falls to 0.80.

Sharp bounds on the utilization gap between the insured and uninsured for Propositions 1B and 2B are presented in Figure 4a. When nothing is known about the accuracy of insurance classifications among unverified cases, $v_0 = 0$, the width of the bounds equals 0.72 in the general case and 0.45 under the coding error irrelevance assumption. In both cases, there is not enough information about insured status in the sample to identify whether the insured or uninsured are more likely to use health services.

When $v_0 = 1$ such that all insurance classifications are known to be accurate, the utilization gap is point-identified as $P(U = 1|X = 1) - P(U = 1|X = 0) = 0.13$. Identical to the critical value in Figure 2 (and Figure 3), the probability of health services use among the insured is known to exceed the probability among the uninsured if v_0 exceeds 0.94. In the context of this verification model, a researcher convinced that the insured are more likely to use health services than the uninsured is implicitly assuming that no more than 6 percent of the unverified sample has misreported insured status. Under the irrelevance assumption, this critical value falls to 0.75. Notice that this critical value is smaller than the corresponding critical value in Figure 2 (0.80) obtained by taking differences between the relevant upper and lower bounds. The reason is that the utilization gaps presented in Figure 3 based on Propositions 1B and 2B are sharp in that they exploit all available information, while calculating the utilization gaps using results from Propositions 1A and 2A are not necessarily sharp.

Figure 4b presents corresponding Proposition 1B results separately for children (younger than 18) and adults. Taking the data at face value with $v_0 = 1$, the probability of health services use is 7 percentage points higher among adults (0.12) than children (0.05). Until v_0 approaches 0.5, however, the lower bounds on the two subpopulations are almost identical. Unless v_0 exceeds 0.93, we cannot reject the possibility that uninsured children are more likely to use health services than insured children. Unless v_0 exceeds 0.97, we cannot reject the possibility that uninsured adults are more likely to use health services than insured adults.

4.5 Monotone Instrumental Variables

In this section, we assess how the bounds obtained in the previous section can be narrowed when combined with monotonicity assumptions linking utilization outcomes and observed covariates such as age or presence of physician-diagnosed health conditions. For illustration, consider age and utilization. The incidence of many debilitating health conditions rises with age, and many health conditions are persistent once developed. The resulting tendency for individuals to accumulate health problems over time suggests that the population health care utilization rate, conditional on true insured status, is nondecreasing in age beyond some age threshold.[8] Conditional on insured status, the fraction of insured 60 year-olds who utilize health care might be assumed to be no smaller than the fraction of insured 59 year-olds who utilize health care, and so on.

Consider, for example, the utilization rate at some specified age, age_0, above some threshold, age'. Formally, the MIV restriction implies the following inequality restriction:

$$age\prime \leq age_1 \leq age_0 \leq age_2 \Longrightarrow P(U=1|I, age_1) \leq P(U=1|I, age_0) \leq P(U=1|I, age_2). \quad (10)$$

The conditional probabilities in Equation (10) are not identified. However, we can bound these probabilities using the methods described above. Let $LB(age)$ and $UB(age)$ be the known lower and upper bounds, respectively, given the available information on $P(U=1|I, age)$. Then, we have

$$\sup_{age_1 \in [age',\, age_0]} LB(age_1) \leq P(U=1|I, age_0) \leq \inf_{age_2 \geq age_0} UB(age_2) \quad (11)$$

(Manski and Pepper, Proposition 1, 2000). The MIV bound on the conditional utilization rate, $P(U=1|I)$, is obtained using the law of total probability. Then the following bounds apply:

Proposition 4. *If the utilization rate is weakly increasing with the monotone instrumental variable*

[8] Among children, utilization may be decreasing in age over some age range.

age beyond the threshold age′, then the following bounds apply:

$$\sum_{age_0 \epsilon U} P(age = age_0)\{\sup_{age_1 \in [age',\, age_0]} LB(age_1)\} \tag{12}$$
$$\leq P(U = 1|I) \leq$$
$$\sum_{age_0 \epsilon U} P(age = age_0)\{\inf_{age_2 \geq age_0} UB(age_2)\}.$$

Thus, to find the MIV bounds on the probability of use, we take the appropriate weighted average of the lower and upper bounds across the different values of the instrument. This proposition generalizes to more than one monotone instrumental variable. For example, one might assume that conditional on true insured status, health care utilization is weakly increasing across the population in both age and number of physician-diagnosed health conditions.

4.6 Results

(Not yet available.)

5 Probability of Use of Services under Universal Health Insurance

We now turn to inferences about utilization rates under a hypothetical policy of universal health insurance.[9] Let $U(I = 1)$ denote whether an individual would have used health services 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 insured status is not verified. We wish to learn the probability that a sample member will use services if insured, $P[U(I = 1) = 1]$. If current insured status were randomly assigned, then the probability of using services among the currently insured, $P(U = 1|I = 1)$, would represent the best prediction of the utilization rate under universal coverage. Since I is not observed for all individuals, we could instead bound $P[U(I = 1) = 1]$ using the methods described in the previous sections. Of course, 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

[9] This analyis is does not account for potential increases in gross prices for health care resulting from universal coverage. Since such price changes would dampen utilization, however, the obtained upper bounds on the change in the probability of health care use should still apply.

resources. For example, families that expect to use health services may be more likely to acquire health insurance. In that case, an observed positive association between insurance coverage and utilization reflects not only the effect of insurance on utilization but also the effect of anticipated service use on insurance status. More generally, insurance status may depend on individual and family characteristics that also determine health care use.

In the absence of random assignment or other assumptions, the quantity in $P[U(I = 1) = 1]$ is not identified even if all insurance classifications are known to be accurate. Unlike identification of the conditional utilization rate $P(U = 1|I = 1)$, identification of the "treatment" outcome $P[U(I = 1) = 1]$ requires knowledge about the counterfactual utilization rate of the uninsured had they instead been insured. Uncertainty about the accuracy of insurance classifications, the focus of the current paper, further complicates identification of counterfactuals.

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

$$P[U(I = 1) = 1] = P[U(I = 1) = 1|Y = 1]P(Y = 1) + P[U(I = 1) = 1|Y = 0]P(Y = 0). \quad (13)$$

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

$$\begin{aligned} P[U(I = 1) = 1|Y = 1] = {} & P(U = 1|I = 1, Y = 1)P(I = 1|Y = 1) \\ & + P[U(I = 1) = 1|I = 0, Y = 1]P(I = 0|Y = 1). \end{aligned} \quad (14)$$

All the terms in (14) are observed, except for the counterfactual probability of health care utilization among those uninsured under the status quo, $P[U(I = 1) = 1|I = 0, Y = 1]$. Varying this quantity within the range $[0, 1]$ obtains:

$$P(U = 1|I = 1, Y = 1)P(I = 1|Y = 1)$$

$$\leq P[U(I = 1) = 1|Y = 1] \leq \quad (15)$$

$$P(U=1|I=1,Y=1)P(I=1|Y=1)+P(I=0|Y=1).$$

Returning to equation (13) and decomposing the third term involving the unverified cases obtains:

$$P[U(I=1)=1|Y=0]=P(U=1|I=1,Y=0)P(I=1|Y=0) \tag{16}$$
$$+P[U(I=1)=1|I=0,Y=0]P(I=0|Y=0).$$

None of these four quantities is identified. We do not know the existing insurance rate among unverified cases, and we cannot match utilization outcomes to insurance status when insurance status is unknown. Extending a result in Molinari (2002), however, we can learn something about $P[U(I=1)=1|Y=0]$ given information about the range of $P_{10}\equiv P(I=1|Y=0)$, the status quo insured rate among unverified cases.[10] Knowledge about P_{10} places restrictions on the utilization patterns of this subpopulation. In particular, we know that the observed probability of use among unverified cases is a weighted average of the probabilities of use among unverified insured and uninsured cases:

$$P(U=1|Y=0)=P(U=1|I=1,Y=0)P_{10}+P(U=1|I=0,Y=0)(1-P_{10}). \tag{17}$$

First note that the feasible range of P_{10} is restricted by the value of v_0. After all, if $v_0=1$ such that all unverified insurance classifications are deemed accurate, then the only value of P_{10} consistent with the data is the observed quantity $P(X=1|Y=0)$. More generally, we can write the insured rate among unverified classifications as a function of false positive and false negative classification rates:

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

Allowing the unidentified terms to vary over their feasible ranges implies $P_{10}\in\left[P_{10}^{a},P_{10}^{b}\right]$ where

$$P_{10}^{a} \equiv P(X=1|Y=0)-\min\{1-v_0,P(X=1|Y=0\} \tag{18}$$
$$P_{10}^{b} \equiv P(X=1|Y=0)+\min\{1-v_0,P(X=0|Y=0\}.$$

[10] Molinari's framework applies to the case in which $Y=0$ (our notation) denotes survey nonresponse. We treat unverified insurance classifications as a type of survey nonresponse.

When $v_0 = 0$, P_{10} is trivially bounded within $[0, 1]$; at the other extreme when $v_0 = 1$, $P_{10} = P(X = 1|Y = 0)$.[11]

For a particular value of P_{10}, solving for the probability of use among the unverified currently insured obtains

$$P(U = 1|I = 1, Y = 0) = \frac{P(U = 1|Y = 0) - P(U = 1|I = 0, Y = 0)(1 - P_{10})}{P_{10}}. \tag{19}$$

The utilization rate among the unverified uninsured, $P(U = 1|I = 0, Y = 0)$, in the right-hand-side is nontrivially bounded if $v_0 > 0$. Decomposing this quantity into identified terms and unidentified false positive and negative misclassified cases yields:

$$\begin{aligned} P(U = 1|I = 0, Y = 0) &= \frac{P(U = 1, I = 0, Y = 0)}{(1 - P_{10})P(Y = 0)} \\ &= \frac{P(U = 1, X = 0, Y = 0) + P(U = 1, X = 1, Z = 0) - P(U = 1, X = 0, Z = 0)}{(1 - P_{10})P(Y = 0)}. \end{aligned}$$

The false positives and false negatives are bounded, respectively, as follows:

$$\begin{aligned} 0 &\leq P(U = 1, X = 1, Z = 0) \leq \min\{\phi, P(U = 1, X = 1, Y = 0)\} \\ 0 &\leq P(U = 1, X = 0, Z = 0) \leq \min\{\phi, P(U = 1, X = 0, Y = 0)\}. \end{aligned}$$

These restrictions imply

$$\begin{aligned} 0 \leq \Omega_1 \equiv \min\left\{1, \frac{P(U = 1, X = 0, Y = 0) - \min\{\phi, P(U = 1, X = 0, Y = 0)\}}{(1 - P_{10})P(Y = 0)}\right\} \\ \leq P(U = 1|I = 0, Y = 0) \leq \\ \Omega_2 \equiv \min\left\{1, \frac{P(U = 1, X = 0, Y = 0) + \min\{\phi, P(U = 1, X = 1, Y = 0)\}}{(1 - P_{10})P(Y = 0)}\right\} \leq 1. \end{aligned} \tag{20}$$

Then varying $P(U = 1|I = 0, Y = 0)$ within the feasible range $[\Omega_1, \Omega_2]$ in (19) shows that $P(U = 1|I = 1, Y = 0)$ must lie within the range[12]

$$\left[\max\left\{0, \frac{P(U = 1|Y = 0) - \Omega_2(1 - P_{10})}{P_{10}}\right\}, \min\left\{1, \frac{P(U = 1|Y = 0) - \Omega_1(1 - P_{10})}{P_{10}}\right\}\right]. \tag{21}$$

[11] Molinari (2002) implicitly assumes $v_0 = 0$, so she places no restrictions on feasible values of $P_{10} \in [0, 1]$ (denoted p in her framework).

[12] Note that $v_0 > 0$ also directly places restrictions on $P(U = 1|I = 1, Y = 0)$. However, it can be shown that the direct restrictions on this quantity represent a subset of the restrictions imposed on it indirectly via (21) through the restrictions on $P(U = 1|I = 0, Y = 0)$.

Applying this result to (16) and varying $P[U(I=1)=1|I=0,Y=0]$ within $[0,1]$ provides the following sharp bounds on the probability of service use among unverified cases:

$$\max\{0, P(U=1|Y=0) - \Omega_2(1-P_{10})\}$$

$$\leq P[U(I=1)=1|Y=0] \leq \tag{22}$$

$$\min\{1, P(U=1|Y=0) + (1-\Omega_1)(1-P_{10})\}.$$

Combining (13), (15), and (22) then yields the following sharp bounds on the proportion of the population using health services under universal health insurance:

Proposition 5. *Given $P(Z=1|Y=0) \geq v_0$ and a particular value of $P_{10} \equiv P(I=1|Y=0) \in \left[P_{10}^a, P_{10}^b\right]$, the proportion of the population using services under universal insurance coverage is bounded sharply as follows:*

$$P(U=1,X=1,Y=1) + \max\{0, P(U=1|Y=0) - \Omega_2(1-P_{10})\} P(Y=0)$$

$$\leq P[U(I=1)=1] \leq \tag{23}$$

$$P(U=1,X=1,Y=1) + P(X=0,Y=1) + \min\{1, P(U=1|Y=0) + (1-\Omega_1)(1-P_{10})\} P(Y=0)$$

where P_{10}^a and P_{10}^b are defined in (18) and Ω_1 and Ω_2 are defined in (20).

In the special case that $v_0 = 0$ such that nothing is assumed about the accuracy of unverified classifications, this proposition reduces to Molinari's (2002) Proposition 2 with $\Omega_1 = 0$, $\Omega_2 = 1$, and $P_{10} \in [0,1]$. The bounds in (23) intuitively narrow for a sufficiently large $v_0 > 0$.

Proposition 5 applies assuming that the proportion of the unverified individuals holding insurance, P_{10}, is known. If P_{10} is instead assumed to lie within some range $\left[P_{10}^c, P_{10}^d\right] \subseteq \left[P_{10}^a, P_{10}^b\right]$, then the lower and upper bounds in (23) are replaced by the infimum and supremum, respectively, of these bounds over values of $P_{10} \in [P_{10}^c, P_{10}^c]$. In the worst case that $v_0 = 0$ and nothing is known about P_{10}, the lower bound falls to $P(U=1,X=1,Y=1)$ and the upper bound rises to $1 - P(U=0,X=1,Y=1)$.

5.1 Monotonicity assumptions

The preceding bounds can be narrowed under monotonicity assumptions on treatment response or treatment selection. A researcher might be willing to assume, for example, that the probability of using services is at least as high in the insured state as in the uninsured state. Formally, for each person i,

$$U_i(I = 1) \geq U_i(I = 0). \tag{24}$$

Under this monotone treatment response assumption (MTR), the lower bound of the population proportion using services under universal coverage intuitively rises to $P(U = 1)$, the existing national utilization rate in the absence of universal coverage:

$$P[U(I = 1) = 1] \geq P(U = 1).$$

Suppose further that among verified cases, the probability of using services under universal coverage would be at least as high among those currently insured as among those currently uninsured:[13,14]

$$P[U(I = 1) = 1|I = 1, Y = 1] \geq P[U(I = 1) = 1|I = 0, Y = 1]. \tag{25}$$

The validity of this monotonicity treatment selection (MTS) assumption depends on the process by which individuals have selected themselves into insured and uninsured status. This assumption is consistent with evidence from Miller et al. (2003) who find that while the uninsured tend to be less healthy than the privately insured, they tend to be healthier than those publicly insured (e.g., through Medicaid). The uninsured may also have unobserved characteristics, such as attitudes toward care, which make them less likely to seek care. To the extent that the uninsured are less health conscious, they may seek less preventive care and wait longer before deciding to seek treatment for an ailment. In general, those who are currently insured may have a greater propensity to use care than those who are currently uninsured. Using data from the SIPP to estimate a

[13] Below, we also present results for the case that this monotonicity assumption is imposed over the entire population.

[14] In the 1996 MEPS sample, the probability of using services among those classified as being insured was 0.23, compared with 0.13 among the uninsured (Table 1).

parametric structural model of health care utilization and insurance coverage, Li and Trivedi (2004) find that a significant part of the observed reduction in health care use among the uninsured can be attributed to adverse selection instead of the lack of insurance.

To bound the probability of service use under universal coverage given these monotonicity assumptions, we decompose the probability of use under the mandated policy into verified and unverified components:

$$P\left[U(I=1)=1\right]=P\left[U(I=1)=1|Y=1\right]P(Y=1)+P\left[U(I=1)=1|Y=0\right]P(Y=0). \quad (26)$$

Using a result in Manski and Pepper (2000, Corollary 2.2), the probability of service use within the verified group is bounded sharply as follows:

$$\begin{aligned} &P(U=1|I=0,Y=1)P(I=0|Y=1)+P(U=1|I=1,Y=1)P(I=1|Y=1) \quad (27)\\ &\leq P\left[U(I=1)=1|Y=1\right]\leq P(U=1|I=1,Y=1). \end{aligned}$$

Combining (26) and (27) obtains:

$$\begin{aligned} &P(U=1)\leq P\left[U(I=1)=1\right] \quad (28)\\ &\leq P(U=1|I=1,Y=1)P(Y=1)+P\left[U(I=1)=1|Y=0\right]P(Y=0). \end{aligned}$$

Each of the terms in (28) is identified except $P\left[U(I=1)=1|Y=0\right]$, the probability of using services under universal coverage among those with unverified current insured status. Decomposing this term into currently insured and currently uninsured groups obtains:

$$P\left[U(I=1)=1|Y=0\right]=P(U=1|I=1,Y=0)P_{10}+P\left[U(I=1)=1|I=0,Y=0\right](1-P_{10}).$$

Recall from (21) that the first term cannot exceed $\min\left\{1,\left[P(U=1|Y=0)-\Omega_1(1-P_{10})\right]/P_{10}\right\}$ when the aggregate insured rate among unverified cases, P_{10}, is known. Applying this result to (28) and varying $P\left[U(I=1)=1|I=0,Y=0\right]$ within $[0,1]$ obtains the following proposition:

Proposition 6A. *Suppose the monotone treatment response assumption in (24) holds, and the monotone treatment selection assumption (25) holds within the subpopulation for which current* *insured status is verified. Given $P(Z=1|Y=0) \geq v_0$ and a particular value of $P_{10} \equiv P(I=1|Y=0) \in \left[P_{10}^a, P_{10}^b\right]$, the following sharp bounds apply to the probability of using services under mandatory universal insurance coverage:*

$$P(U=1) \leq P\left[U(I=1)=1\right]$$
$$\leq P(U=1|X=1,Y=1)P(Y=1) + \min\left\{1, P(U=1|Y=0) + (1-\Omega_1)(1-P_{10})\right\}P(Y=0)$$

where P_{10}^a and P_{10}^b are defined in (18) and Ω_1 is defined in (20). If P_{10} is known to lie within some range $\left[P_{10}^c, P_{10}^d\right] \subseteq \left[P_{10}^a, P_{10}^b\right]$, then the upper bound is replaced by the supremum of this bound over values of $P_{10} \in \left[P_{10}^c, P_{10}^d\right]$.

Molinari's (2002) Proposition 5 applies as a special case when $v_0 = 0$ and nothing is known about P_{10}. This result is stated as a corollary:

Corollary (Molinari, 2002, Proposition 5). *When the monotonicity conditions in Proposition 6 hold with $v_0 = 0$ and $P_{10} \in [0,1]$, the upper bound in Proposition 6 weakly rises as follows:*

$$P\left[U(I=1)=1\right] \leq P(U=1|X=1,Y=1)P(Y=1) + P(Y=0).$$

The upper bound in Proposition 6 can be improved substantially when the monotone treatment selection assumption applies to the entire population instead of only the subpopulation with verified insured status. In that event, Manski and Pepper's (2000) result in (27) implies

$$P\left[U(I=1)=1\right] \leq P(U=1|I=1).$$

This result combined with the upper bound on $P(U=1|I=1)$ derived in Proposition 1 leads to the following proposition:

Proposition 6B. *When monotonicity assumptions (24) and (25) hold jointly over the entire population and $P(Z=1|Y=0) \geq v_0$, the following bounds apply to the utilization rate under mandatory universal insurance coverage:*

$$P(U=1) \leq P\left[U(I=1)=1\right] \leq \frac{P(U=1,X=1)+\theta_2}{P(X=1) - \min\left\{\phi-\theta_2, P(U=0,X=1,Y=0)\right\} + \theta_2}$$

where $\delta_2 \equiv P(U=0, X=1) - P(U=1, X=1) - (1-v_0)P(Y=0)$ and

$$\theta_2 = \begin{cases} \min\{(1-v_0)P(Y=0), P(U=1, X=0, Y=0)\} & \text{if } \delta_2 \geq 0 \\ \max\{0, \min\{P(U=1, X=0, Y=0), \phi - P(U=0, X=1, Y=0)\}\} & \text{otherwise} \end{cases}.$$

5.1.1 Results

Empirical results for the consequences of universal insurance coverage are presented in Figure 5a.[15] Each set of bounds in the figure is drawn under the assumption that the insured rate among unverified classifications, P_{10}, may lie anywhere within its logically consistent range $[P_{10}^a, P_{10}^b]$. When $v_0 = 0$ and no monotonicity assumptions are imposed, Molinari's (2002, Proposition 2) results apply. In this case, the national monthly utilization rate under universal coverage is bounded to lie within the range $[0.15, 0.52]$. When all unverified classifications are assumed to be accurate, $v_0 = 1$, the bounds narrow to $[0.18, 0.37]$. These bounds reflect the well-known classical effect of a mandatory policy (e.g., Manski, 1995). In that case, the only uncertainty involves the currently uninsured's counterfactual probability of using health services if they were to become insured, with no uncertainty about status quo insured status. Middle-ground assumptions about the accuracy of insurance classification within unverified cases are calculated by varying the value of v_0 between 0 and 1. Notice that the upper bound remains flat until v_0 exceeds 0.61 but then improves rapidly. The lower bound remains flat until v_0 exceeds 0.89. Even when $v_0 = 1$, the width of the bounds is large in the absence of additional assumptions because we do not know the counterfactual outcomes for nearly a third of the sample for which $Y = 0$.

The next sets of bounds impose the monotone treatment response (MTR) and monotone treatment selection (MTS) restrictions. Molinari's (2002, Proposition 5) bounds apply when monotone treatment selection is assumed to hold only within the verified cases and nothing is known about the accuracy of the insurance classifications within the unverified group, $v_0 = 0$. Compared with

[15] As above, our results reflect point estimates of the worst-case bounds and do not yet account for sampling variability. We will produce confidence intervals for these estimates.

the preceding $v_0 = 0$ bounds without these monotonicity assumptions, the bounds narrow from $[0.15, 0.52]$ to $[0.21, 0.49]$, a 9 percentage point decline in the width of the bounds. At the other extreme when $v_0 = 1$, the bounds narrow from $[0.18, 0.37]$ to $[0.21, 0.34]$, a 6 percentage point decline in the width.

An assumption that the monotone treatment selection assumption holds for the general population (Proposition 6B) confers substantial identifying power. When nothing is known about the accuracy of classifications within unverified cases, $v_0 = 0$, the monthly utilization rate under universal coverage is bounded to be less than 0.29, representing a maximum 43 percent increase relative to the baseline utilization rate of 0.21. This upper bound declines to 0.23, representing a maximum 9 percent increase relative to the the status quo, when $v_0 = 1$ such that all uncertainty takes the form of uncertainty about counterfactuals and not about the accuracy of insurance classifications.

Figure 5b presents separate results for adults (solid lines) and children (dashed lines). The July 1996 probability of service use among adults classified as being insured and uninsured are $P(U = 1|X = 1) = 0.25$ and $P(U = 1|X = 0) = 0.13$, respectively. The corresponding probabilities for children are $P(U = 1|X = 1) = 0.17$ and $P(U = 1|X = 0) = 0.12$, respectively. When $v_0 = 0$, upper bounds on the probability of utilization under universal insurance coverage range from 0.32 to 0.55 for adults and from 0.23 to 0.46 for children under the previously discussed monotonicity assumptions. When all uncertainty is confined to uncertainty about counterfactuals and not insurance classifications ($v_0 = 1$), the utilization upper bounds range from 0.22 to 0.40 for adults and from 0.17 to 0.30 for children.

Figures 5a and 5b assume that no information is available about P_{10}, the insured rate among unverified classifications, except that its value is consistent with the maintained assumptions restricting the nature and degree of insurance classification errors. For values of v_0 less than one, all of the bounds presented in this section can be improved given knowledge further restricting the range of P_{10}.

6. Conclusion

We developed a formal nonparametric framework for making inferences on (a) differences between the insured and uninsured in the probability of using health services, and (b) the impact of universal health insurance on the probability of using services when true insured status is unknown for part of the sample. Our approach extends recent research that derives nonparametric bounds on parameters of interest that, in general, cannot be point-identified in the absence of strong parametric assumptions restricting the nature of classification errors.

Our framework applies to the case in which a researcher is confident about the accuracy of reported insurance status within some observed subsamples in the data but not others. In the MEPS data, we were able to confirm insurance status for about two-thirds of the sample through insurance cards and follow-up interviews with employers and providers. These supplemental information sources verify individual cases and thereby confer strong identifying power beyond the limits they place on the aggregate degree of misclassification.

Under the strongest verification and monotonicity assumptions, we can tightly bound the increase in the probability of health care use following the introduction of universal health insurance. In particular, 22.4 percent of adults used inpatient or ambulatory medical care in July 1996, and under universal health insurance we estimate that 22.4 to 24.9 percent would have done so – an increase of 0 to 2.5 percentage points, or 0 to 11 percent.[1] In 1996, 16.6 percent of children used care. Under universal health insurance, we estimate that 16.6 to 17.4 percent would have done so, an increase of 0 to 5 percent.

Identification of these parameters, however, deteriorates rapidly with the degree of uncertainty associated with insurance classifications for unverified data. The specific patterns of identification breakdown can provide useful information about the value of devoting more resources toward verifying additional insurance reports. The marginal value of verifying additional cases depends not only on the

[1] Recall that we have not yet accounted for sampling variability; these reported bounds reflect only uncertainty associated with identification. Preliminary evidence from bootstrapping suggests that the bounds will not widen very much when sampling variability is included.

parameter of interest but also on which classifications are verified. For instance, in a subsequent draft we plan to explore whether verifying an additional case of reported uninsurance narrows the estimated bounds more rapidly than verifying an additional case of reported insurance.

Despite being unable to confirm insured status for about a third of the sample, we discovered no evidence of large-scale classification errors in the MEPS. Among cases with data available to confirm or contradict insured status, we found that 96 percent accurately reported their insurance status, primarily reflecting that all sample members were confirmed as insured when the respondent showed an insurance card to the interviewer. We found relatively few inconsistencies between family respondents and employers, insurance companies, or providers. Some of these inconsistencies may be misreports by establishment respondents or reflect differences in timing.

In the absence of strong parametric assumptions restricting the nature of misclassifications, our analysis highlights severe identification problems associated with even small degrees of uncertainty about insured status. In a future version of this paper, we will compare the identifying power of nonparametric and parametric assumptions, allowing researchers to observe tradeoffs between the strength of the identifying assumptions and the consequences for inferences. We expect this developing line of research to improve researchers' understanding of the extent and consequences of measurement error, which will in turn hopefully lead to improved measures and yield more informed policy analyses.

References

[1] Benitez-Silva, H., Buchinsky, M., Chan, H., Sheidvasser, S., and Rust, J. (forthcoming). "How Large is the Bias in Self-Reported Disability." *Journal of Applied Econometrics.*

[2] Berger M., D. Black and F. Scott, 1998. "How Well Do We Measure Employer-Provided Health Insurance Coverage?" *Contemporary Economic Policy*, 16, 1998, 356-367.

[3] Blumberg, S.J., and M.L. Cynamon. 2001. "Misreporting Medicaid Enrollment: Results of Three Studies Linking Telephone Surveys to State Administrative Records." Seventh Conference on Health Survey Research Methods, M. Cynamon and R.Kulka, eds. Hyattsville, MD: National Center for Health Statistics, p. 189-195.

[4] Bollinger, C. (1996). "Bounding Mean Regressions When A Binary Variable is Mismeasured." *Journal of Econometrics*, 73(2), 387-99.

[5] Bound, J. and Burkhauser, R. (1999). "Economic Analysis of Transfer Programs Targeted on People with Disabilities." In Orley Ashenfelter and David Card (Eds.), Handbook of Labor Economics, Vol. 3C. Amsterdam: Elsevier Science, 3417-3528.

[6] Call, K.T., G. Davidson, A.S. Sommers, R. Feldman, P. Farseth, and T. Rockwood. "Uncovering the Missing Medicaid Cases and Assessing their Bias for Estimates of the Uninsured." *Inquiry* 38(4): 396-408.

[7] Cohen, J. Design and Methods of the Medical Expenditure Panel Survey Household Component. MEPS Methodology Report no. 1, AHCPR Pub. no. 97-0026. Rockville, MD AHCPR 1997.

[8] Cooper, P.F., A.P. Hagy, and J.P. Vistnes. 1999. "The Medical Expenditure Panel Survey: Creation of a Linked Employer-Employee Database." In J.C. Haltiwanger, J.I. Lane, J.R. Speltzer, et al., eds., The Creation and Analysis of Employer-Employee Matched Data. New York: Elsevier.

[9] Czajka, J. and K. Lewis, "Using universal Survey Data to Analyze Children's Health Insurance Coverage: An Assessment of Issues, Mathematica Policy Research, Inc., May 21, 1999, http://aspe.os.dhhs.gov/health/reports/Survey%20Data.htm.

[10] Efron, B. and R. J. Tibshirani. (1993). An Introduction to the Bootstrap. Chapman and Hall. New York: NY.

[11] Gruber J. and B. Madrian (2001). "Health Insurance, Labor Supply, and Job Mobility: A Critical Review of the Literature." ERIU Working Paper 4, http://www.umich.edu/~eriu/pdf/wp4.pdf

[12] Hess, J., J. Moore, J. Pascale, J. Rothgeb, and C. Keeley. 2001. The Effects of Person-Level versus Household-Level Questionnaire Design on Survey Estimates and Data Quality. *Public Opinion Quarterly* 65(4): 574-584.

[13] Horowitz, J. and Manski, C. (1995). "Identification and Robustness With Contaminated and Corrupted Data." *Econometrica* 63(2), 281-02.

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

[15] Kreider, B. and J. Pepper (2003). "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors," mimeo.

[16] Kreider, B. and J. Pepper (2004). "Inferring Disability Status from Corrupt Data," mimeo.

[17] Levy, Helen and David Meltzer (2001). "What Do We Really Know About Whether Health Insurance Affects Health?" Economic Research Initiative on the Uninsured, Working Paper 6, http://www.umich.edu/~eriu/pdf/wp6.pdf.

[18] Machlin, S.R., and A.K. Taylor. Design, Methods, and Field Results for the 1996 Medical Expenditure Panel Survey Medical Provider Component. Methodology Report No. 9. Rockville, MD: Agency for Healthcare Research and Quality, 2000. AHRQ Pub. No. 00-0028.

[19] Miller, E.G., J.S. Banthin, and J.F. Moeller. 2003. "Covering the Uninsured: Estimates of the Impact on Total Health Expenditures for 2000." Working paper. Rockville, MD: Agency for Healthcare Research and Quality.

[20] Monheit, A.C. (2003) "Verifying Lack of Health Insurance in the Medical Expenditure Panel Survey: An Assessment." Report submitted to the Agency for Health Care Research and Quality. New Brunswick, NJ: Rutgers University.

[21] Lambert, D. and Tierney, L. (1997). "Nonparametric Maximum Likelihood Estimation from Samples with Irrelevant Data and Verification Bias." *Journal of the American Statistical Association*, 92: 937-944.

[22] Li, Chenghui and Pravin Trivedi. 2004. "Disparities in Medical Care Utilization, Insurance Coverage, and Adverse Selection." Working Paper.

[23] Maag, E. and Wittenburg, D. (2003). "Real Trends or Measurement Problems? Disability and Employment Trends from the Survey of Income and Program." mimeo, Urban Institute.

[24] Molinari, F. (2002). "Missing Treatments." Cornell University ,Working Paper.

[25] Manski, C. and Pepper, J. (2000). "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica*, 68(4): 997-1010

[26] Manski, C. (1995). Identification Problems in the Social Sciences. Harvard University Press, Cambridge, MA.

[27] Nelson, D.E., B.L. Thompson, N.J. Davenport, and L.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.

[28] Olson, C., "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, 1998, 543-565.

[29] Pascale, J. 2002. "A Quantitative and Qualitative Assessment of the Data Quality of Health Insurance Measurement Methodologies." Suitland, MD: Center for Survey Methods Research, U.S. Census Bureau. Working paper.

[30] Rhoades, J.A., and J.W. Cohen. 2003. "The Uninsured in America 1996-2002: Estimates for the Civilian Noninstitutional Population Under Age 65." Statistical Brief #24. Rockville, MD: Agency for Healthcare Research and Quality.

[31] Swartz, K., "Interpreting the Estimates from Four National Surveys of the Number of People Without Health Insurance," *Journal of Economic and Social Measurement*, 14, 1986, 233-242.

7 Appendix: Proofs of Propositions 2A, 2B, and 3B

Proof for Proposition 2A. Using the independence assumption (9) and Bayes' Theorem, write the utilization rate among the insured as

$$
\begin{aligned}
P(U=1|I=1) =& P(U=1|I=1,Z=1) \\
=& \frac{P(U=1,I=1,Z=1)}{P(I=1,Z=1)} = \frac{P(U=1,X=1)-\mathsf{c}}{P(X=1)-\mathsf{c}-\mathsf{i}}.
\end{aligned}
$$

Combining terms,

$$
P(U=1|I=1) = \frac{P(U=1,X=1)-\mathsf{c}}{P(X=1)-\mathsf{c}-\mathsf{i}} \tag{29}
$$

where $\mathsf{c} = P(U=1,X=1,Z=0)$ and $\mathsf{i} = P(U=0,X=1,Z=0)$. The lower bound requires setting $\mathsf{i}=0$. Then differentiating (29) with respect to c shows that c must be set to its largest feasible value, $\min\{\phi, P(U=1,X=1,Y=0)\}$. To find the upper bound, setting i to its maximum feasible value, $\min\{\phi, P(U=0,X=1,Y=0)\}$. Then (29) is decreasing in c which requires that c be set equal to zero. □

Proof for Proposition 2B. Given the independence assumption (9), the utilization rate among the uninsured can be written as

$$
P(U=1|I=0) = P(U=1|I=0,Z=1) = P(U=1|X=0,Z=1).
$$

The utilization gap can then be written as:

$$
\Delta = \frac{P(U=1,X=1)-\mathsf{c}}{P(X=1)-\mathsf{c}-\mathsf{i}} - \frac{P(U=1,X=0)-\mathsf{F}}{P(X=0)-\mathsf{F}-\mathsf{L}} \tag{30}
$$

where the unknown quantities c, i, F, and L have been defined above. Minimizing this difference requires setting $\mathsf{i}=\mathsf{F}=0$ with L set equal to its largest feasible value, $\min\{\phi-\mathsf{c}, P(U=0,X=0,Y=0\}$. The lower bound is then found by minimizing the resulting quantity over feasible values of c. Maximizing the difference requires setting $\mathsf{c}=\mathsf{L}=0$ with i set equal to its largest feasible value, $\min\{\phi-\mathsf{F}, P(U=0,X=1,Y=0\}$. The upper bound is found by maximizing the resulting quantity over feasible values of F. □

Proof for Proposition 3B. The utilization gap is given by

$$
\Delta = \frac{P(U=1,X=1)+\mathsf{F}-\mathsf{c}}{P(X=1,Y=1)+P_{10}P(Y=0)} - \frac{P(U=1,X=0)+\mathsf{c}-\mathsf{F}}{P(X=0,Y=1)+(1-P_{10})P(Y=0)}
$$

The lower bound is found by setting F to zero and finding the infimum over the resulting quantity over feasible values of the remaining unknown parameter c. The upper bound is found similarly by setting c equal to zero and finding the supremum of the resulting quantify over feasible values of F. □

8 Data Appendix

This appendix contains a detailed description of the availability and construction of the insurance verification variables from insurance cards, employers and insurance companies, and providers.

8.1 Insurance Cards

We construct a person-level summary measure of whether the respondent showed the interviewer a Medicare, Medicaid or private insurance card. In the first interview, for each person reportedly covered by Medicare, the interviewer asked to see the card. If anyone in the family was reportedly covered by Medicaid or other state or local government hospital/physician benefit programs, the interviewer asked to see a card.[1] For each reported source of private insurance (for example, through a person's employer), the interviewer asked to see the cards. The interviewer asked for cards for each source of insurance in the household, not for each covered person, except for Medicare, where eligibility is individual rather than family. Cards where not requested for CHAMPUS/CHAMPVA (now called Tricare). We assume that if an insurance card was shown for Medicaid or private insurance, everyone in the family reportedly covered by Medicaid or that source of private insurance had their insurance status verified. We also assume that the coverage period reported by the family respondent was correct, and that verification in the first interview extended to July. Note that uninsurance cannot be verified with insurance cards.

8.2 Linked Household and Insurance Components

The 1996 linked HC and Insurance Component (IC) attempted to collect information from establishments related to the HC sample of households (Cooper et al. 1999). The sample frame for the linked data was the following types of establishments:

[1] For the very few people who reported coverage through AFDC or SSI, the interviewer did not ask about cards, although the respondents were coded as covered by Medicaid during the data editing process.

1. Main employers at the time of the first interview for each worker age 16+. Exceptions were: federal employees, self-employed persons with no employees, and workers employed by someone else but the firm had only 1 employee.

2. Other employers providing physician/hospital insurance at the time of the first interview, such as former employers providing severance packages, COBRA, or retiree health insurance.

3. Unions providing physician/hospital insurance at the time of the first interview.

4. Insurance companies and/or agents providing physician/hospital insurance at the time of the first interview.

HC sample members were asked to sign permission forms to allow MEPS to interview these establishments.

Employers, unions, and insurance companies (hereinafter called employers and insurance companies) were interviewed from August 1997 through February 1998. Employers were asked whether they offered insurance to any employees, how many were eligible and enrolled, and other details. All employers and insurance companies were asked about the status of the employee, union member, retiree, or policyholder from the HC, hereinafter called employee or policyholder, as of July 1st, 1996. If this person was a current or former employee or policyholder, or relative or survivor of an employee or policyholder, then MEPS asked about the person's insurance through the establishment as of July 1st. If the IC respondent said the person was insured, then MEPS asked whether it was single, adult-adult, adult-child, or family coverage. If the IC respondent did not know who the person was or the person was a contract employee or employee of a temporary agency, no questions about the person's insurance status were asked.

8.3 Verification from the Insurance Component

All the interviewed establishments could report insurance held by an employee or policyholder. When the family respondent said an employee, union member, or retiree was not insured, however, then due to the criteria for choosing establishments, only the main employers could verify the person was not insured, at least through that establishment. An employer was coded as not insuring the person if it did not offer insurance, it did not verify the person was a current or former employee or relative of a current or former employee, it reported the person was a contract employee or employed by a temporary agency, or the employee was not eligible or not enrolled. An insurance company was coded as not insuring the person if the person did not have any insurance or did not have hospital/physician insurance but instead had a single service, dread disease, or cash benefit plan.[2]

All these establishments could also provide information verifying other family members' insurance status by reporting whether the person had single, adult-adult, adult-child, or family coverage. We assume that if the family respondent reported a dependent had insurance through a employee or policyholder, and the employer reported the family coverage, then the dependent's insurance was verified. If, however, the family respondent did not report dependent coverage, then we use the health insurance eligibility unit (HIEUs) concept, which reflects typical private insurance eligibility rules for dependents and resembles nuclear families, to determine whose insurance status was verified by the IC respondent. For example, if the family respondent said that no one had insurance and if the employer reported that the employee did not have insurance, then everyone in the HIEU was verified as not insured through that establishment.[3]

Among nonelderly sample members, 36.5 percent had verification information from at least one employer or insurance company. For many sample members, establishments did not verify insurance status, either because the person did not provide a permission form, the establishment did not respond, or the establishment did not provide information about the person. To confirm an individual had insurance, we only need one establishment that reported the person had insurance, even if no information was collected from other establishments. On the other hand, if a person's lack of insurance through an establishment was verified by the IC, but the family reported insurance through another source (including public coverage), then the person's

---

[2] We investigated temporal differences as a potential source of differences between HC and employer reports. Specifically, the HC sought permission forms for establishments as of the first interview, conducted March through August, while the IC asked about insurance status as of July 1st. For many sample members, insurance status in July was collected in the second HC interview. An employee could have gained insurance at an employer if he or she was in the waiting period for coverage at a new job at the time of the first interview. Or an employee might no longer have the job by July 1st. In the data we found the number of these cases (51 policyholders and dependents) was too small to affect the results.

[3] For two-person adult-adult coverage, then insurance status was verified for the spouse. When the establishment reported two-person adult-child coverage, then verification was more complex. If the family respondent reported only one child dependent in the HIEU, then insurance status was verified for that child. When the family respondent reported one covered dependent and other HIEU members not covered through the establishment, then uninsurance was verified. In a small number of cases (16 families), the family respondent reported multiple children had coverage. Then we assumed the coverage of the oldest child was verified by the establishment and the others were verified as uninsured. We focus on the oldest child because typically Medicaid eligibility is more generous for younger children.

insurance status could not be verified. (Only 9.3 percent of the sample had both public and private insurance.) As a result, only 27.8 percent of sample members had information from enough employers to verify insurance status (Table 1).

8.4 Source of Payment Data from Providers

Payment sources for health care services reported by providers in the Medical Provider Component (MPC) are another source of verification data. These data can confirm insurance reported by the family or contradict reported uninsurance. In particular, we may find people with presumptive Medicaid eligibility, an eligibility option for states that allows providers to determine a patient is eligible to receive Medicaid services for a month while the patient formally applies for Medicaid. Upon examining the formal application, the state may determine the person or family is not eligible beyond the initial month. These families may not report their brief Medicaid enrollment.

The MPC provides independent source of payment information for a subsample of households in the HC. In 1996, all inpatient and outpatient hospitals visited by family members were sampled with certainty. For office-based visits, families were the sampling unit. Families with at least one member covered by Medicaid were sampled with certainty, 75 percent of families with at least one member in an HMO were sampled, and 25 percent of remaining families (including the uninsured) were sampled (Machlin and Taylor 2000). Each person in the family who reported service use was asked to complete a permission form for each provider. Fifteen percent of sampled household members did not complete permission forms and about 12 percent of sampled providers did not provide payment information (Machlin and Taylor 2000). Providers reported events, charges, payments, and other information. Events reported by providers were matched to those reported by the family respondent.[4] As a result of all these factors, only 7.7 percent of sample members had provider-reported source of payment data for at least one hospitals stay or ambulatory medical visit in July.

[4] For the matched events, the final expenditures and source of payment variables in the MEPS are from the MPC. For all other events, expenditures and source of payment data were imputed from MPC data, and the imputation process used families' reported insurance status.

Payments by insurance for hospitals stays and ambulatory medical visits indicate insurance coverage in July.[5] We identify sample members with at least one stay or visit that was at least partially paid for by Medicare, Medicaid, private insurance, or CHAMPUS/CHAMPVA (now called Tricare). We exclude care that was paid for out-of-pocket or by other federal (such as the Indian Health Service and military treatment facilities), the Veterans Administration (other than CHAMPVA), other state/local (public clinics, other programs), workers' compensation, and other (such as auto insurance).[6] Note, however, that providers may misreport payment sources. For example, they may report auto insurance as private insurance. They may also report limited insurance, such as school-based insurance, as private hospital/physician insurance, whereas this is not classified as hospital/physician insurance.

[5] Lack of insurance payment for care does not confirm uninsurance because an insured person may pay for services entirely out-of-pocket for several reasons: the plan did not cover the service (for example, cosmetic services) or provider, the plan's deductible hadn't been reached, or the consumer sought to maintain confidentiality (for example, for mental health treatment or HIV testing).

[6] Payments by insurance for a person's care could be used to confirm insurance that the family respondent said covered others family members, in addition to verifying insurance covering the person who received services. However, when the family respondent said the person had no insurance, but insurance paid for care, there is no direct way of inferring which other family members were also insured. Only one family member, rather than the entire nuclear family, may be eligible for Medicaid when eligibility is based on disability. Medicaid income eligibility criteria also vary with a child's age, so Medicaid might cover only some children, depending on their age. When children are covered by private insurance, at least one parent is likely covered, but it is not clear which one, and the covered parent may reside elsewhere, for example in the case of divorce or separation.

TABLE 1

INSURANCE STATUS AND SERVICE USE OF NONELDERLY IN JULY, 1996

	Insurance Status Reported by Family		
	Insured	Uninsured	Overall
Percent of Sample	78.4	21.6	100.0
Adults	79.3	20.7	100.0
Children	84.0	16.0	100.0
Percent Using Hospital or Ambulatory Care	22.5	12.7***	20.6
Adults	24.9	11.9***	22.4
Children	17.4	12.1***	16.6
Number of Observations	14,772	4,079	18,851

SOURCE: Medical Expenditure Panel Survey Household Component, 1996. Sample members age 0 to 64 as of July, 1996.

*** Statistically different from insured at the .01 level, two-tailed test.

TABLE 2

SUMMARY OF VERIFICATION DATA IN MEPS

	Insurance Cards	Employers, Unions, and Insurance Companies	Providers
Subpopulation	Privately insured or have Medicaid or Medicare	Employed or privately insured [a]	Medical service users
Subpopulation as Percent of Nonelderly Population	79.5	82.0	20.6
Percent of Nonelderly with Verification Data			
Any data	60.1	36.5	7.7[b]
Enough data to verify insurance status	60.1	27.8	5.9
Ability to Verify Insurance Status			
Uninsured			
Family member employed		Confirm or contradict	
Used services			Contradict
No family member employed, no services used			
CHAMPUS/CHAMPVA (only)			
Used services			Confirm
No services used			
Medicaid, Medicare, Other Public (only)	Confirm		
Used services	Confirm		Confirm
Private Insurance (only)	Confirm	Confirm or contradict	
Used services	Confirm	Confirm or contradict	Confirm
Public and Private Insurance	Confirm	Confirm	
Used services	Confirm	Confirm	Confirm

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

[a] Employers at the time of the first interview, and unions and insurance companies providing insurance at the time of the first interview. Main employers for each worker age 16+, excluding federal employees, self-employed persons with no employees, and workers employed by someone else but the firm had only 1 employee. Including other employers providing physician/hospital insurance, such as former employers providing severance packages, COBRA, or retiree health insurance.

[b] Providers were interviewed for a subsample of families using services.

TABLE 3

PERCENT OF NONELDERLY WITH VERIFIED INSURANCE STATUS IN JULY

Source of Confirmation or Contradiction	Insurance Status Reported by Family: Insured	Insurance Status Reported by Family: Uninsured
Insurance Card Confirmed	75.3	N.A.
Among families reporting private insurance	75.4	N.A.
Among families reporting Medicaid	69.7	N.A.
Among families reporting Medicare	64.5	N.A.
Employer, Union, and Insurance Company		
Among those with enough information from employers, unions, and insurance companies to confirm or contradict insurance status		
Confirmed	80.3	90.0
Contradicted	19.7	10.0
All nonelderly		
Confirmed	23.0	22.0
Contradicted	5.5	2.5
Provider Reported Insurance		
Among those with provider data		
Confirmed	81.5	
Contradicted		26.6
All nonelderly		
Confirmed	7.1	
Contradicted		.8
Insurance Card, Employer, Union, or Insurance Company		
Among those with enough information to confirm or contradict insurance status		
Confirmed	98.2	90.0
Contradicted	1.8	10.0
All nonelderly		
Confirmed	77.8	22.0

Source of Confirmation or Contradiction	Insurance Status Reported by Family: Insured	Uninsured
Contradicted	1.4	2.5
All Three Sources		
Among those with enough information to confirm or contradict insurance status		
Confirmed	98.3	87.3
Contradicted	1.7	12.7
All nonelderly		
Confirmed	78.8	21.9
Contradicted	1.3	3.2
Number of Observations	14,772	4,079

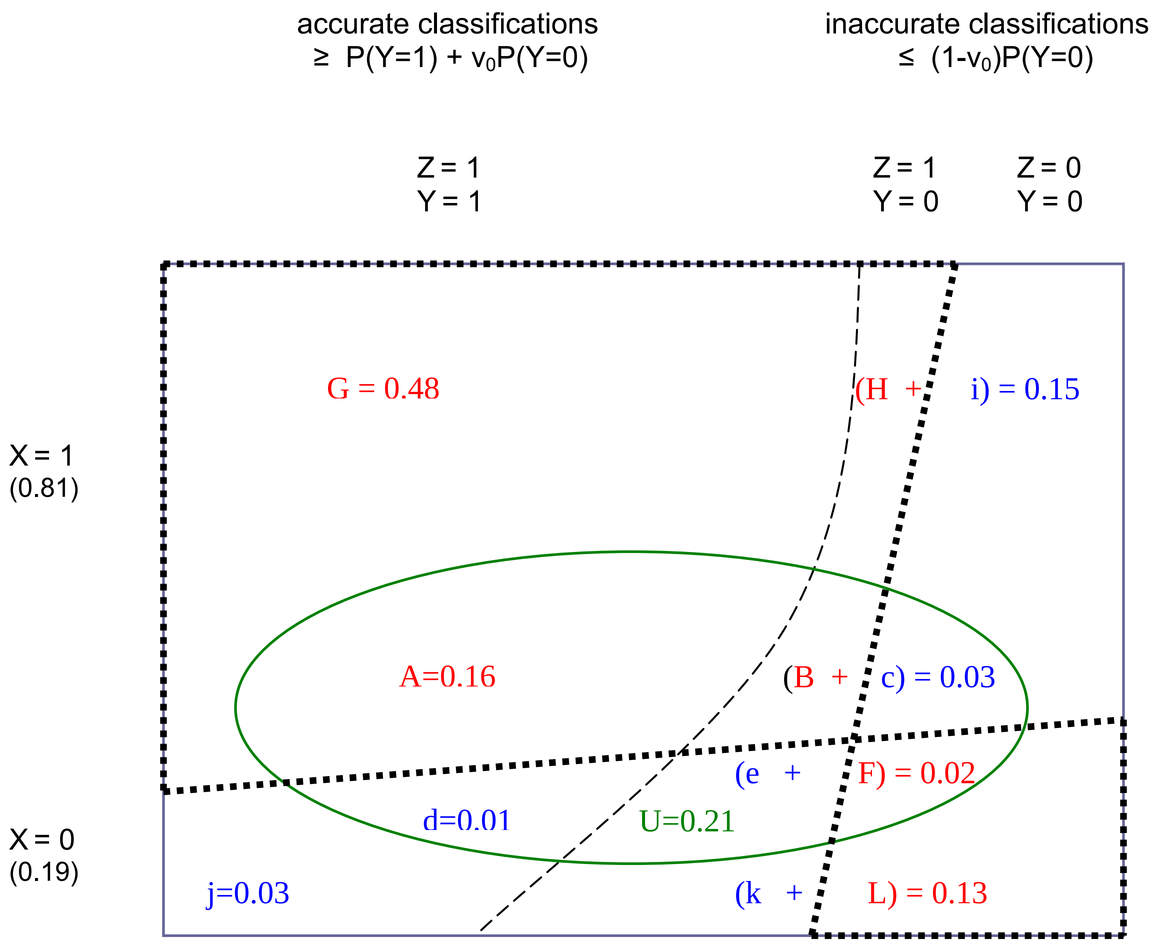
SOURCE: Medical Expenditure Panel Survey Household Component, linked Insurance Component, and Medical Provider Component, 1996. Sample members age 0 to 64 as of July, 1996.

[a] Rows do not sum to 100 percent because families may show cards for multiple types of insurance.

[b] Providers for a subsample of families were interviewed.

FIGURE 1

Venn Diagram of Health Insurance and Use of Services
in the Presence of Insurance Classification Errors

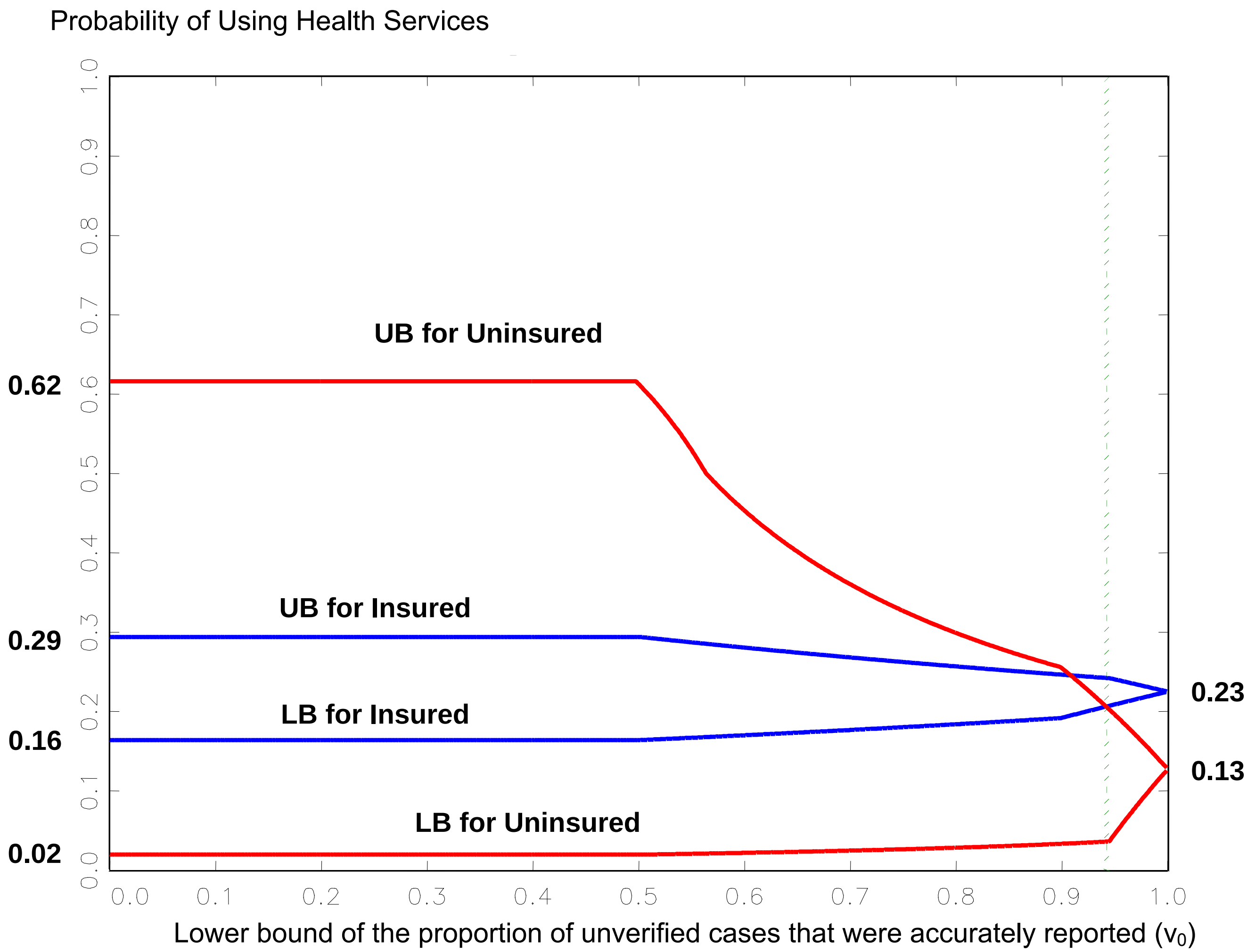


SOURCE: Probabilities are authors' calculations from the Medical Expenditure Panel Survey Household Component and linked Insurance Component, July, 1996. Sample members age 0 to 64 as of July, 1996. Insurance cards and employer confirmations verified 67.1 percent of the sample.

NOTES: U = 1 if the person used any inpatient or ambulatory medical services, and zero otherwise. X = 1 if the family reported the person was insured, and zero otherwise. Y = 1 if person had data verifying insurance status, and 0 otherwise. Z = 1 if the family report is the same as true insurance status, and zero otherwise. v_0 = the lower bound of the proportion of unverified cases that were accurately reported.

FIGURE 2

Proposition 1A Bounds on the Probability of Using Health Services in July, 1996

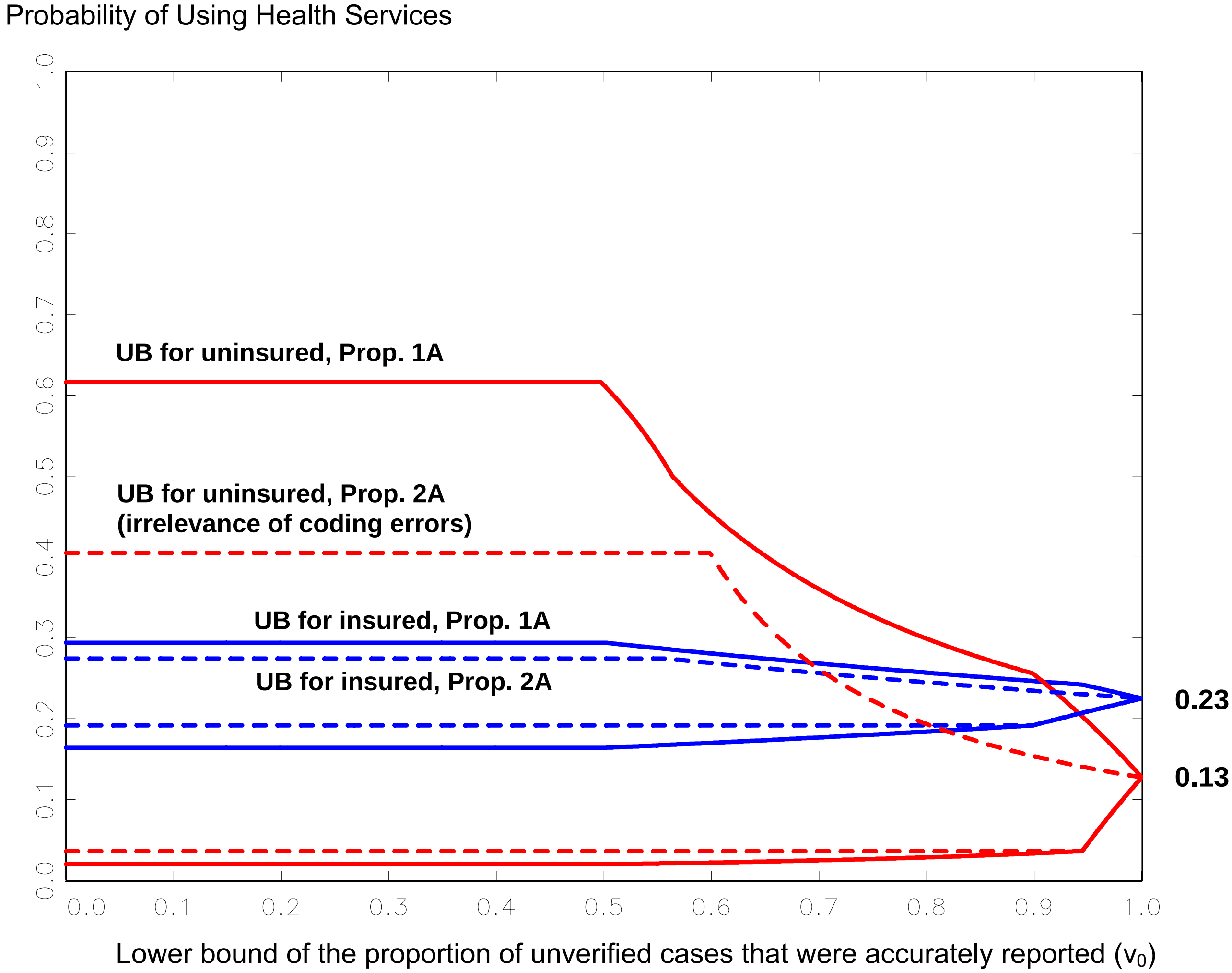


SOURCE: Authors' calculations from the Medical Expenditure Panel Survey Household Component and linked Insurance Component, July, 1996. Sample members age 0 to 64 as of July, 1996. Insurance cards and employer confirmations verified 67.1 percent of the sample.

NOTES: UB = upper bound. LB = lower bound.

FIGURE 3

Comparison Between Proposition 1A and 2A Bounds on Conditional Probabilities of Using Health Services in July, 1996

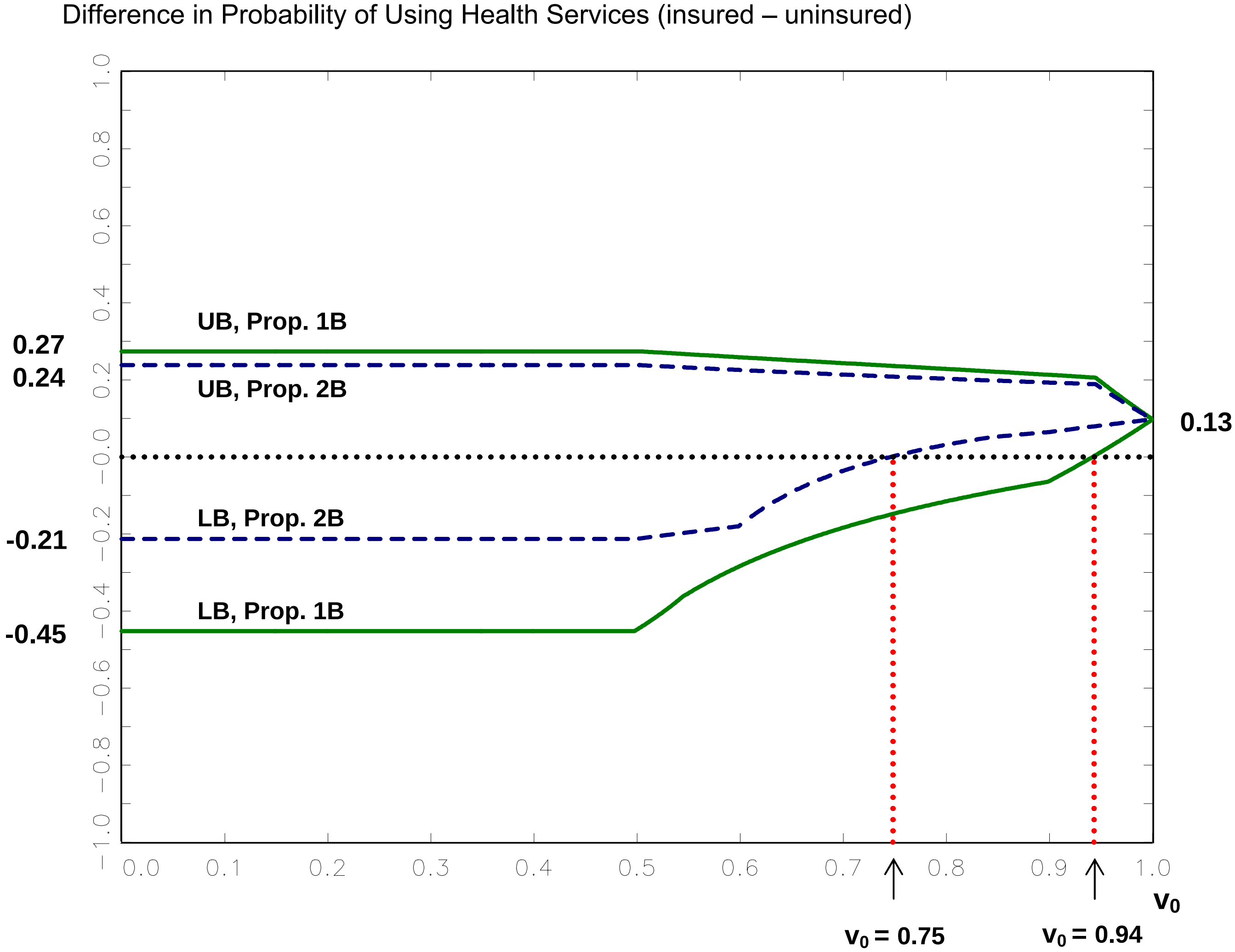


SOURCE: Authors' calculations from the Medical Expenditure Panel Survey Household Component and linked Insurance Component, July, 1996. Sample members age 0 to 64 as of July, 1996. Insurance cards and employer confirmations verified 67.1 percent of the sample.

NOTES: UB = upper bound. LB = lower bound.

FIGURE 4a

Proposition 1B and 2B Bounds on the Gap in the Probability of Using Health Services, Δ, in July, 1996

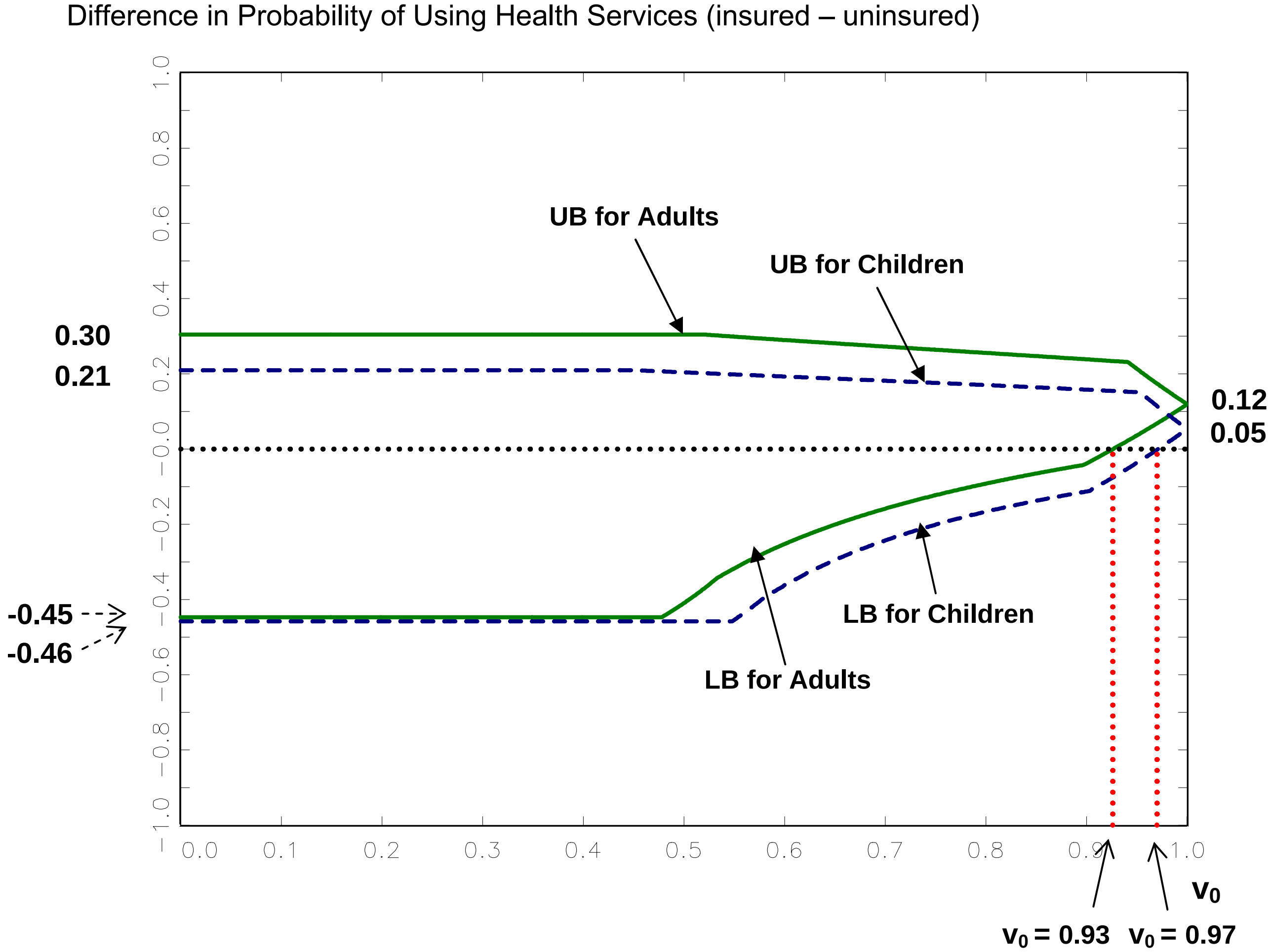


SOURCE: Authors' calculations from the Medical Expenditure Panel Survey Household Component and linked Insurance Component, July, 1996. Sample members age 0 to 64 as of July, 1996. Insurance cards and employer confirmations verified 67.1 percent of the sample.

NOTES: Δ = difference in probability of using health services (insured – uninsured). UB = upper bound. LB = lower bound. v_0 = lower bound of the proportion of unverified cases that were accurately reported.

FIGURE 4b

Proposition 1B Bounds on the Gap in the Probability of Using Health Services, Δ, in July, 1996

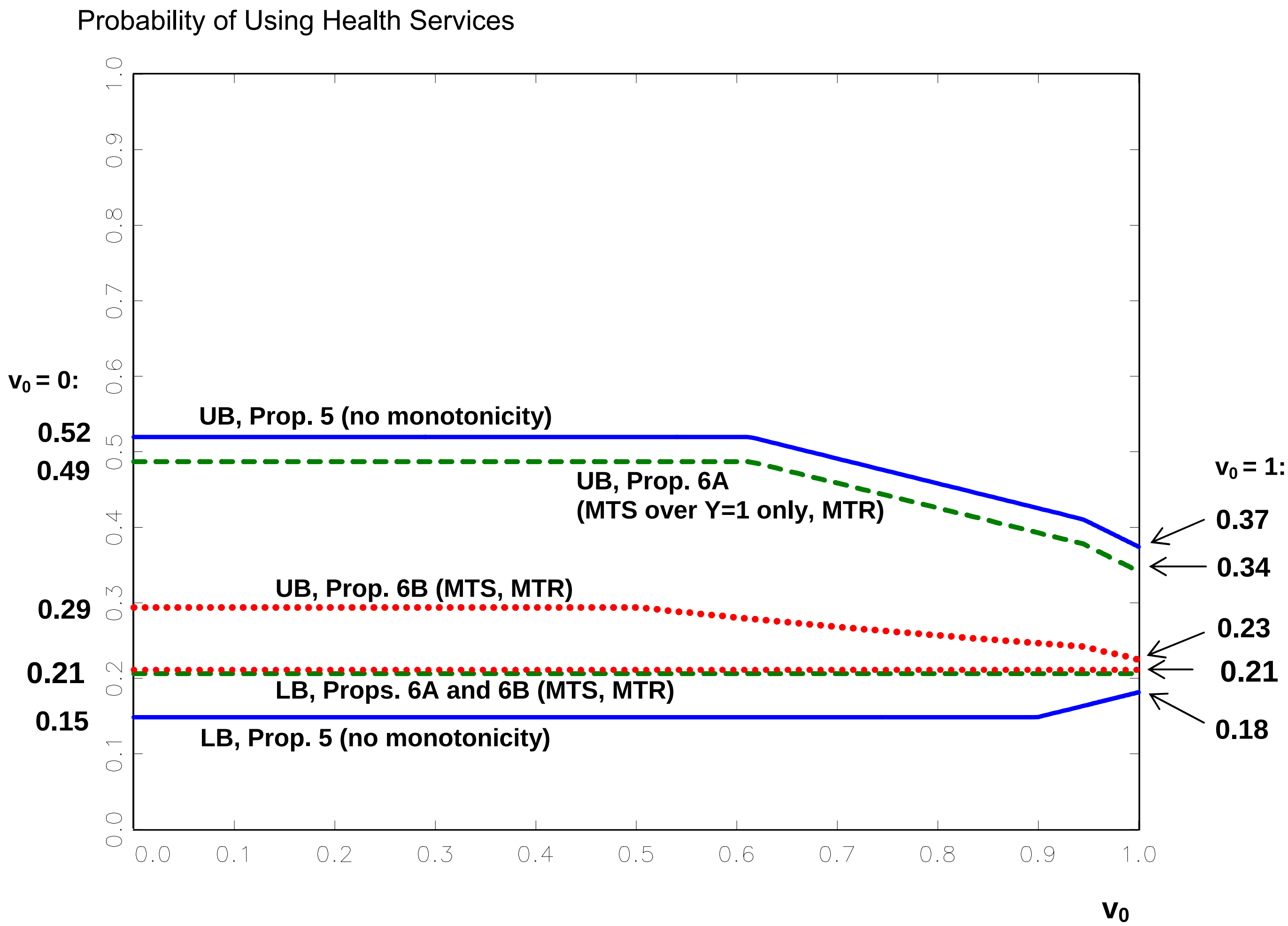


SOURCE: Authors' calculations from the Medical Expenditure Panel Survey Household Component and linked Insurance Component, July, 1996. Adults age 18 to 64 as of July, 1996. Children age 0 to 17 as of July 1996. Insurance cards and employer confirmations verified 66 percent of adults and 70 percent of children.

NOTES: Δ = difference in probability of using health services (insured – uninsured). UB = upper bound. LB = lower bound. v_0 = lower bound of the proportion of unverified cases that were accurately reported. The July 1996 probability of service use among adults classified as being insured and uninsured are $P(U=1|X=1) = 0.25$ and $P(U=1|X=0) = 0.13$, respectively. The corresponding service use probabilities for children are $P(U=1|X=1) = 0.17$ and $P(U=1|X=0) = 0.12$, respectively.

FIGURE 5a

Proposition 5 and 6 Bounds on the Probability of Using Services
Under Universal Coverage with Unknown P(I=1|Y=0)

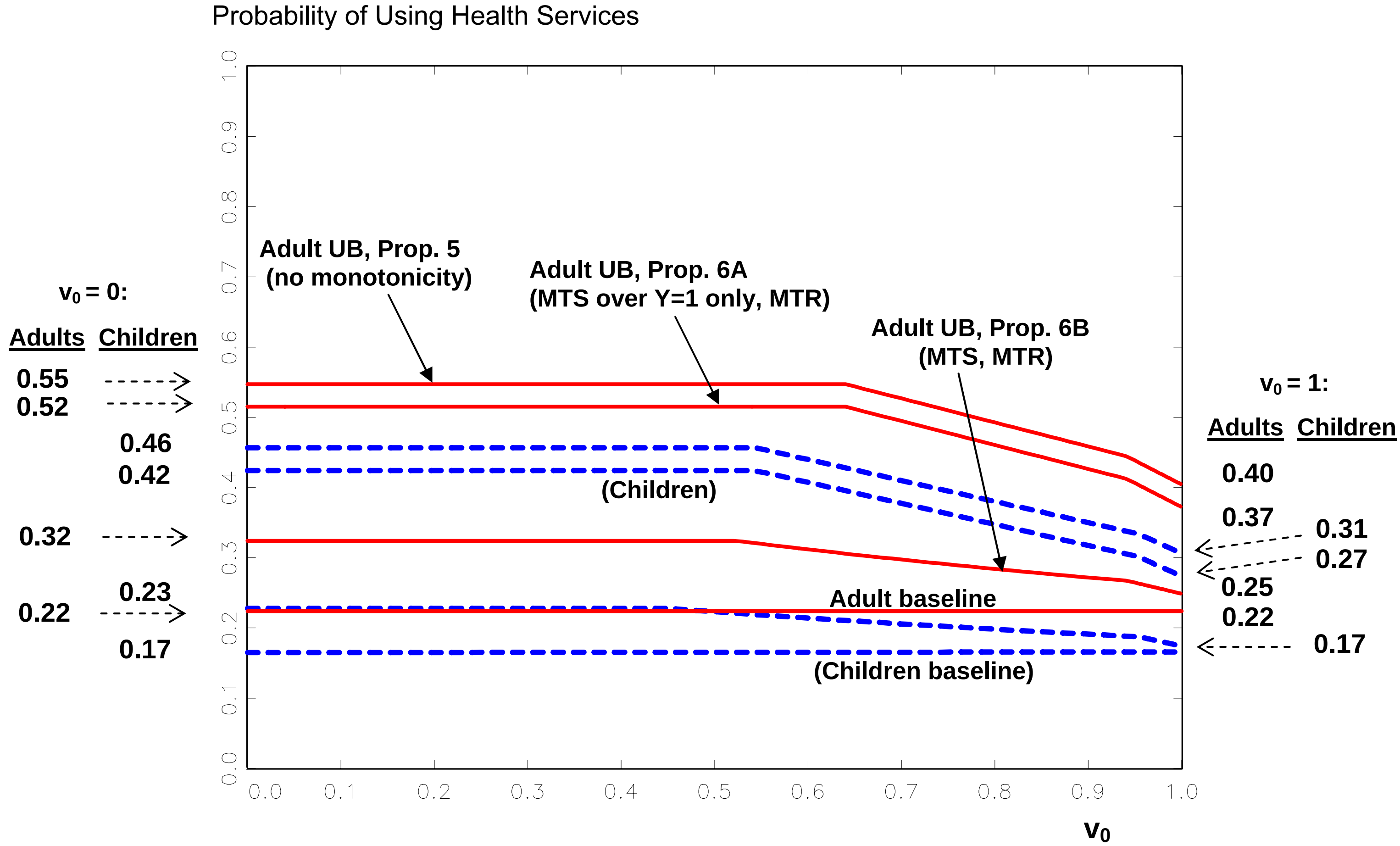


SOURCE: Authors' calculations from the Medical Expenditure Panel Survey Household Component and linked Insurance Component, July, 1996. Sample members age 0 to 64 as of July, 1996. Insurance cards and employer confirmations verified 67.1 percent of the sample.

NOTES: UB = upper bound. LB = lower bound. MTR = monotone treatment response (insurance increases service use). MTS = monotone treatment selection. v_0 = lower bound of the proportion of unverified cases that were accurately reported. For values of v_0<1, these bounds can be narrowed by incorporating possible outside information restricting the range of P(I=1|Y=0), the insured rate among unverified cases. In this figure, nothing is assumed about P(I=1|Y=0) except that its value is logically consistent with the value of v_0.

FIGURE 5b

Proposition 5 and 6 Upper Bounds on the Probability of Using Services
Under Universal Coverage with Unknown P(*I*=1|*Y*=0)



SOURCE: Authors' calculations from the Medical Expenditure Panel Survey Household Component and linked Insurance Component, July, 1996. Adults age 18 to 64 as of July, 1996. Children age 0 to 17 as of July 1996. Insurance cards and employer confirmations verified 66 percent of adults and 70 percent of children.

NOTES: UB = upper bound. LB = lower bound. MTR = monotone treatment response (insurance increases service use). MTS = monotone treatment selection. v_0 = lower bound of the proportion of unverified cases that were accurately reported. The July 1996 probability of service use among adults classified as being insured and uninsured are P(U=1|X=1) = 0.25 and P(U=1|X=0) = 0.13, respectively. The corresponding service use probabilities for children are P(U=1|X=1) = 0.17 and P(U=1|X=0) = 0.12, respectively. For values of v_0<1, these bounds can be narrowed by incorporating possible outside information restricting the range of P(I=1|Y=0), the insured rate among unverified cases. In this figure, nothing is assumed about P(I=1|Y=0) except that its value is logically consistent with the value of v_0.