Diagnosis Measurement Error and Instrumental Variables

Donna Chen, Brent Kreider, Elizabeth Merwin, and Steven Stern

November 22, 1999

1 Introduction

As is well understood among clinical researchers, medical diagnoses often suffer from substantial measurement error. This is particularly true for mental health diagnoses.[1] Reliance on inaccurate diagnosis indicators can substantially compromise researchers' and clinicians' ability to reliably evaluate treatment approaches, measure clinical progress, or estimate relationships between patient attributes and various outcome measures of interest (such as expected treatment costs, hospital and community tenure, or ability to participate in the labor force). Assessing the generalizability of treatment interventions across patient types, for example, is problematic given limitations in the accuracy of diagnosis information.

Although much effort has been put into developing better diagnosis instruments and into encouraging the use of standardized codes, the nature of the measurement error problem is such that it will always exist as long as researchers must rely on clinical evaluation or research instruments. While "gold standard" references in the literature (e.g., Roy, et. al, 1997) provide useful benchmarks for comparing measures of "reliability" and "validity" across diagnostic instruments, even instruments meeting these gold standards are likely contaminated with substantial errors. Moreover, even if all diagnosis variables were measured without error, any measurement error in auxiliary control variables correlated with the diagnosis indicators (such as comorbidity or severity) would spill over into inferences about the health effects.

The standard approach for handling measurement error in statistical models involves instrumental variables (IV) estimation.[2] Suppose we wish to estimate

[1] For the case of research instruments, there are a number of sources of measurement error including incorrectly answered questions, subjective decision-making or interpretation on the part of the professional administering the instrument, incomplete surveys, and suboptimal weighting of answers to construct diagnosis measures.

[2] IV estimation has recently been used in health research to control for endogenous regressors (see, for example, Newhouse and McClellan, 1998). While there are mathematical similarities between the endogenous regressors problem and the mismeasured regressors problem, they have different causes, different effects, instrument selection rules differ.

the effects of health conditions on medical expenditures. In this context, a researcher might instrument one badly measured diagnosis indicator (a vector of possible diagnoses) with another in attempts to obtain consistent estimates of these health effects. In most widely used secondary data sets, however, multiple indicators of a patient's diagnosis are unlikely to be independent of each other as required for the classical IV method to work. For example, two diagnosers may partially rely on the same written medical history when forming their conclusions. A primary contribution of our research involves a proposed generalized IV approach that first identifies and measures the biases resulting from dependent diagnoses indicators, then offers a method for constructing estimators purged of these biases.

Constructing such an estimator requires construction of the "covariance matrix" of the measurement errors, a rich measure of the magnitudes and directions of measurement errors in the diagnosis data. Our approach allows for very general error structures. Implicit in the construction of such a matrix for clinical evaluation is a model of how physicians and other health professionals make diagnoses. We present a series of such models and discuss their implications for the structure of measurement error. While this proposal focuses on the simplest case of a single primary diagnosis, our approach extends in a straightforward manner to allow for comorbidity.

2 Methodology

2.1 Intuition

We start off with an overly simplified example to provide the reader with some basic intuition. Once the basic intuition is developed, we discuss more interesting and realistic cases.

Consider a simple equation where some outcome variable for person i, y_i, is affected by whether person i has a psychiatric illness. Such an outcome variable could be medical expenses in a given year. We assume, for now, that the outcome variable is affected only by psychiatric illness according to the linear equation,

$$y_i = \beta q_i + \eta_i, \quad i = 1, 2, .., N \tag{1}$$

where q_i is a continuous measure of person i's illness, $0 \leq q_i \leq 1$, β is the effect of psychiatric illness on y_i, and η_i is a random error term with zero mean, uncorrelated with q_i. Assume that, instead of being able to observe q_i in the data, the investigator observes only an imperfect measure of q_i, x_{1i}, with

$$\begin{aligned} \Pr\left[x_{1i} = 1 \mid q_i\right] &= q_i; \\ \Pr\left[x_{1i} = 0 \mid q_i\right] &= 1 - q_i. \end{aligned} \tag{2}$$

Then, we can write

$$x_{1i} = q_i + \upsilon_{1i} \tag{3}$$

with

$$\begin{aligned}\Pr\left[\upsilon_{1i} = 1 - q_i \mid q_i\right] &= q_i; \\ \Pr\left[\upsilon_{1i} = -q_i \mid q_i\right] &= 1 - q_i;\end{aligned}$$

$$\begin{aligned}E\left[\upsilon_{1i} \mid q_i\right] &= q_i\left(1-q_i\right) - \left(1-q_i\right)q_i = 0; \\ Var\left[\upsilon_{1i} \mid q_i\right] &= q_i\left(1-q_i\right)^2 + \left(1-q_i\right)q_i^2 \\ &= q_i\left(1-q_i\right).\end{aligned}$$

Then, if the investigator estimates the equation

$$y_i = bx_{1i} + e_i$$

using ordinary least squares (OLS),

$$\widehat{b}_{OLS} = \frac{N^{-1}\sum_{i=1}^{N} x_{1i}y_i}{N^{-1}\sum_{i=1}^{N} x_{1i}^2},$$

the estimator $\widehat{b}_{OLS}$ converges (as $N \to \infty$) to

$$\begin{aligned}\text{plim}\widehat{\beta}_{OLS} &= \frac{\text{plim}\frac{1}{n}\sum_i x_{ji}y_i}{\text{plim}\frac{1}{n}\sum_i x_{ji}^2} \\ &= \frac{\text{plim}\frac{1}{n}\sum_i \left(q_i + e_{ji}\right)\left(\beta q_i + u_i\right)}{\text{plim}\frac{1}{n}\sum_i \left(q_i + e_{ji}\right)^2} \\ &= \frac{\beta\sigma_q^2}{\sigma_q^2 + \int q\left(1-q\right)f\left(q\right)dq}\end{aligned}$$

where σ_q^2 =plim $N^{-1}\sum_{i=1}^{N} q_i^2$. Since $Eq_i\upsilon_{1i} = Eq_i\eta_i = E\upsilon_{1i}\eta_i = 0$; i.e., errors are uncorrelated with true diagnosis and with each other,

$$\text{plim}\ \widehat{b}_{OLS} = \beta\frac{\sigma_q^2}{\sigma_q^2 + \int q\left(1-q\right)f\left(q\right)dq}.$$

Since

$$0 < \frac{\sigma_q^2}{\sigma_q^2 + \int q\left(1-q\right)f\left(q\right)dq} < 1,$$

the OLS estimator converges to a number biased towards zero; $\left|\text{plim}\ \widehat{b}_{OLS}\right| < |\beta|$. Thus, the investigator will tend to underestimate the effect of the health condition on the outcome of interest when relying on OLS. This is the standard problem when explanatory variables are measured with error. The directions of bias cannot easily be determined a priori when equation (1) is generalized to allow for more than one poorly measured explanatory variable. Nevertheless,

in general, OLS estimates will be biased in the presence of measurement error in explanatory variables.

Now, assume that there is another imperfect measure of q_i, x_{2i}, with same properties as x_{1i} described in equation (2) but independent of x_{1i}; i.e., υ_{2i} and υ_{1i} are independent. Then, the investigator can use x_{2i} as an instrument for x_{1i} in an instrumental variables (IV) estimator of β:

$$\widehat{b}_{IV} = \frac{N^{-1}\sum_{i=1}^{N} x_{2i}y_i}{N^{-1}\sum_{i=1}^{N} x_{2i}x_{1i}}.$$

Note that $\widehat{b}_{IV}$ converges to

$$\begin{aligned} \text{plim } \widehat{b}_{IV} &= \frac{\text{plim } N^{-1}\sum_{i=1}^{N} x_{2i}y_i}{\text{plim } N^{-1}\sum_{i=1}^{N} x_{2i}x_{1i}} \\ &= \frac{\text{plim } N^{-1}\sum_{i=1}^{N} (q_i+\upsilon_{2i})(\beta q_i+\eta_i)}{\text{plim } N^{-1}\sum_{i=1}^{N} (q_i+\upsilon_{2i})(q_i+\upsilon_{1i})} \\ &= \frac{\text{plim } N^{-1}\sum_{i=1}^{N} \left[\beta q_i^2 + q_i(\beta\upsilon_{2i}+\eta_i)+\upsilon_{2i}\eta_i\right]}{\text{plim } N^{-1}\sum_{i=1}^{N} \left[q_i^2 + q_i(\upsilon_{2i}+\upsilon_{1i})+\upsilon_{2i}\upsilon_{1i}\right]}. \end{aligned} \tag{4}$$

As long as $Eq_i\upsilon_{1i} = Eq_i\upsilon_{2i} = E\upsilon_{2i}\eta_i = E\upsilon_{2i}\upsilon_{1i} = 0$; i.e., errors are uncorrelated with true diagnosis and with each other, equation (4) simplifies to

$$\text{plim } \widehat{b}_{IV} = \frac{\text{plim } N^{-1}\sum_{i=1}^{N} \beta q_i^2}{\text{plim } N^{-1}\sum_{i=1}^{N} q_i^2} = \beta.$$

Thus, the IV estimator converges to the true effect. This method of dealing with measurement error in explanatory variables is well known in econometrics (e.g., Greene, 1997, p. 440).

2.2 The Basic Problem with Independent Measurement Errors

It is now appropriate to add important characteristics of the problem specific to the application at hand. First of all, it may be the case that, instead of the existence of a psychiatric illness being the explanatory variable of interest, it is the type of psychiatric illness that is of interest. We generalize equation (1) to

$$y_i = \sum_{k=1}^{K} \beta_k g(q_{ik}) + \sum_{k=1}^{M} z_{ik}\gamma_k + \eta_i \tag{5}$$

where $q_i = (q_{i1}, q_{i2}, .., q_{iK})'$ is a vector of size K measuring the severity/existence of different psychiatric conditions, $g(\bullet)$ is some specified function translating diagnoses into outcomes, and $z_i = (z_{i1}, z_{i2}, .., z_{iM})'$ is a vector of other exogenous covariates such as race, sex, and age to control for. We assume throughout that

z_i is measured without error. Assume that x_{ijk}, $j = 1, 2$, is an independent discrete indicator of q_{ik} measured with error such that $\sum_k x_{ijk} = 1$ and

$$p_{ik} = \Pr\left[x_{ijk} = 1 \mid q_i\right] = \frac{\exp\left\{q_{ik}\right\}}{\sum_l \exp\left\{q_{il}\right\}}. \tag{6}$$

Constructing such a vector requires the investigator to commit to a method of aggregating psychiatric diagnoses into a relatively small class of K diagnoses. Though there are important issues to address associated with optimal aggregation, we assume them away and condition on a chosen aggregation method. We can generalize the specification of x_{ijk} easily to allow for multiple (comorbid) diagnoses. Then, as before, we can write

$$x_{ijk} = p_{ik} + \upsilon_{ijk}$$

with

$$\begin{aligned} \Pr\left[\upsilon_{ijk} = 1 - p_{ik} \mid p_{ik}\right] &= p_{ik}; \\ \Pr\left[\upsilon_{ijk} = -p_{ik} \mid p_{ik}\right] &= 1 - p_{ik}; \end{aligned}$$

$$\begin{aligned} E\left[\upsilon_{ijk} \mid p_{ik}\right] &= 0; \\ Cov\left[\upsilon_{ijk}, \upsilon_{ijl} \mid p_i\right] &= \begin{cases} p_{ik}\left(1 - p_{ik}\right) & \text{for } k = l \\ -p_{ik}p_{il} & \text{for } k \neq l \end{cases}; \\ Cov\left[\upsilon_{ijk}, \upsilon_{ij'k'} \mid p_i\right] &= 0 \text{ for all } j \neq j'. \end{aligned} \tag{7}$$

Defining $x_{ij} = \left(x_{ij1}, x_{ij2}, ..., x_{ijK}\right)'$, $X_{ij}^{'} = \left(x_{ij}', z_i'\right)$, $\beta = \left(\beta_1, \beta_2, .., \beta_K\right)'$, $\gamma = \left(\gamma_1, \gamma_2.., \gamma_M\right)$, and $\theta = \left(\beta', \gamma'\right)'$,

$$\widehat{\theta}_{IV} = \left[\frac{1}{n}\sum_i X_{i2}X_{i1}^{'}\right]^{-1}\left[\frac{1}{n}\sum_i X_{i2}y_i\right] \tag{8}$$

with

$$\begin{aligned} \text{plim}\widehat{\theta}_{IV} &= \left[\text{plim}\frac{1}{n}\sum_i X_{i2}X_{i1}^{'}\right]^{-1}\left[\text{plim}\frac{1}{n}\sum_i X_{i2}y_i\right] \\ &= \left[\text{plim}\frac{1}{n}\sum_i \begin{pmatrix} \left(p_i + \upsilon_{i2}\right)\left(p_i + \upsilon_{i1}\right)' & \left(p_i + \upsilon_{i2}\right)z_i' \\ z_i\left(p_i + \upsilon_{i1}\right)' & z_i z_i' \end{pmatrix}\right]^{-1} \bullet \\ &\qquad \left[\text{plim}\frac{1}{n}\sum_i \begin{pmatrix} \left(p_i + \upsilon_{i2}\right)\left[g\left(q_i\right)\beta + z_i\gamma + \eta_i\right] \\ z_i\left[g\left(q_i\right)\beta + z_i\gamma + \eta_i\right] \end{pmatrix}\right] \\ &= \left[\iint \begin{pmatrix} p\left(q\right)p'\left(q\right) & p\left(q\right)z' \\ zp'\left(q\right) & zz' \end{pmatrix} f\left(q, z\right) dqdz\right]^{-1} \bullet \\ &\qquad \left[\iint \begin{pmatrix} p\left(q\right)g'\left(q\right) & p\left(q\right)z' \\ zg'\left(q\right) & zz' \end{pmatrix} f\left(q, z\right) dqdz\right] \begin{pmatrix} \beta \\ \gamma \end{pmatrix}. \end{aligned}$$

If $p(q) = g(q)$, then $\text{plim}\widehat{\theta}_{IV} = \theta$. Otherwise, in general, there is a proportional bias equal to

$$\left[\iint \begin{pmatrix} p(q)\,p^{'}(q) & p(q)\,z^{'} \\ zp^{'}(q) & zz^{'} \end{pmatrix} f(q,z)\,dqdz\right]^{-1}\left[\iint \begin{pmatrix} p(q)\,g'(q) & p(q)\,z' \\ zg'(q) & zz' \end{pmatrix} f(q,z)\,dqdz\right]$$

which is the plim of a regression of $g(q)$ on $p(q)$ and z. In general, since $g(\bullet)$ can not be identified without observing q, we assume that $p(q) = g(q)$ and recognize that results should be interpreted in light of the possibility that it is probably not so.

2.3 The Basic Problem with Correlated Measurement Errors

Now, consider the case where $Cov[\upsilon_{ijk}, \upsilon_{ij'k'} \mid p_i] \neq 0$ for all $j \neq j'$. Mechanically, we could cause correlation by adjusting equation (6) to

$$p_{ik} = \Pr[x_{ijk} = 1 \mid q_i] = \frac{\exp\{q_{ik} + u_{ik}\}}{\sum_l \exp\{q_{il} + u_{il}\}} \tag{9}$$

where $u_i = (u_{i1}, u_{i2}, .., u_{iK})'$ is a vector of errors constant across j. Let

$$m(\theta) = \frac{1}{n}\sum_i X_{i2}(y_i - X_{i1}'\theta)$$

be the set of moment conditions for instrumental variables estimation. Then

$$\begin{aligned} \text{plim}\; m(\theta) &= \text{plim}\frac{1}{n}\sum_i \begin{bmatrix} p(q_i) + \upsilon_{i2} \\ z_i \end{bmatrix}[p'(q_i)\beta + z_i'\gamma + \eta_i - p'(q_i)\beta - \upsilon_{i1}'\beta - z_i'\gamma]' \\ &= \text{plim}\frac{1}{n}\sum_i \begin{bmatrix} p(q_i) + \upsilon_{i2} \\ z_i \end{bmatrix}[-\upsilon_{i1}'\beta + \eta_i]' \\ &= \begin{pmatrix} -\text{plim}\frac{1}{n}\sum_i \upsilon_{i2}\upsilon_{i1}'\beta \\ 0 \end{pmatrix} = \begin{pmatrix} -\Omega\beta \\ 0 \end{pmatrix} \end{aligned}$$

where

$$\Omega = \int E[\upsilon_{i2}\upsilon_{i1}' \mid q]\, f(q)\, dq.$$

If we could estimate Ω consistently, then we could redefine our moment conditions as

$$\widetilde{m}(\theta) = \frac{1}{n}\sum_i X_{i2}(y_i - X_{i1}'\theta) + \begin{pmatrix} \Omega\beta \\ 0 \end{pmatrix}, \tag{10}$$

and the value of θ that set $\widetilde{m}(\theta) = 0$ would be a consistent estimate of θ. In particular,

$$\widehat{\theta}_{IV} = \left[\frac{1}{n}\sum_i \begin{pmatrix} x_{i2}x_{i1}' & x_{i2}z_i' \\ z_i x_{i1}' & z_i z_i' \end{pmatrix} - \begin{pmatrix} \widehat{\Omega} & 0 \\ 0 & 0 \end{pmatrix}\right]^{-1} \left[\frac{1}{n}\sum_i \begin{pmatrix} x_{i2} \\ z_i \end{pmatrix} y_i\right]. \tag{11}$$

Note that, in general, the adjustment term in the inverse, $\begin{pmatrix} \widehat{\Omega} & 0 \\ 0 & 0 \end{pmatrix}$, decreases the inverse and therefore increases the magnitude of the IV estimator (counteracting the effect of correlated measurement error). Also note that, if $\widehat{\Omega} = 0$, the $\widehat{\theta}_{IV}$ is equivalent to the IV estimator in equation (8) which is consistent when measurement errors are uncorrelated.

2.4 Estimating the Measurement Error Covariance Matrix

The method discussed above requires having a consistent estimate of the covariance matrix of the diagnosis measurement errors, Ω. In this section, first we discuss what is meant by this covariance matrix, and then we discuss various ways that one might estimate it. One way to obtain an estimate of Ω is to estimate the covariance matrix of the elements of x_{ij}:

$$V = \begin{pmatrix} Var(x_{ij1}) & Cov(x_{ij1},x_{ij2}) & \cdots & Cov(x_{ij1},x_{ijK}) \\ Cov(x_{ij1},x_{ij2}) & Var(x_{ij2}) & \cdots & Cov(x_{ij2},x_{ijK}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(x_{ij1},x_{ijK}) & Cov(x_{ij2},x_{ijK}) & \cdots & Var(x_{ijK}) \end{pmatrix} \tag{12}$$

with a separate source of data. We distinguish between this special data source ("special data source") and the data used to estimate θ ("estimation data"). The ideal special data source would be one where we have N individuals each of whom is independently diagnosed by H "diagnosers." A diagnoser is a mental health professional (e.g., psychiatrist, psychologist, or social worker) who diagnoses individuals similar in characteristics to the diagnosers in the data used to estimate equation (5).[3] A consistent estimate of V when $H = 2$ is

$$\widehat{V} = \frac{1}{N}\sum_{i,h} \omega_i (x_{i2} - x_{\bullet 2})(x_{i1} - x_{\bullet 1})'$$

where ω_i is a weight given to observation i so that the weighted special data has the same distribution of observed explanatory variables as the estimation

[3] It is important to use the same kind of diagnosers in the special data source as those used in the estimation data. Characteristics of measurement probably vary significantly across different types of diagnosers.

data and

$$x_{\bullet h} = \frac{1}{N} \sum_i \omega_i x_{ih}$$

is the average diagnosis in the special sample for diagnoser h.[4] With a consistent estimate of V, we can compute the covariance matrix for equation (11) by

$$\widehat{\Omega} = \frac{1}{n} \sum_i \omega_i \left(x_{i2} - x_{\bullet 2}\right) \left(x_{i1} - x_{\bullet 1}\right)' - \widehat{V}. \tag{13}$$

Intuitively, there are three sources of variation in x_{ij}: a) variation due to variation in true condition q_i, b) variation due to measurement error independent across diagnosers, and c) measurement error common across diagnosers. The first term in equation (13) includes variation due to variation of types (a) and (c), and the second term includes variation of type (a). Subtraction leaves only randomness due to variation of type (c).

[Discussion of how to collect special sample: do such data sets already exist; problem of independence; cost of collecting data; data specific to diagnoser type]

Alternatively, one could possibly estimate Ω by using a panel of "experts" to construct Ω using the "Delphi method" (e.g., Kahan, et. al 1994). One could measure how much confidence to put in the "Delphi estimate" of Ω by measuring the variance of the estimate across the experts in early rounds of the process.

In general, to the degree on replaces Ω in equation (10) with a consistent estimate of Ω, one adds randomness to the estimator of θ in equation (11). In fact, with an imprecise estimate of Ω, it is easy for the confidence region of $\widehat{\theta}_{IV}$ to explode. The Monte Carlo experiments reported below are suggestive on this point.

3 Models of Diagnosis

The structure of the covariance matrix of diagnosis error may depend upon our model of how diagnosers make diagnoses. Basically, our model may impose some structure on the covariance matrix and help us get more precise estimates of its elements. Below we consider two such models and describe how they affect our estimates.

[4] When $H > 2$, one has to decide how to use the extra diagnosers to increase precision of the estimator of V.

3.1 A Simple Model of Diagnosis

Let $x_{i1} = 1$ if and only if doctor 1 diagnoses a health condition for person i. Assume x_{i1}^* is a latent random variable such that

$$x_{i1} = 1 \text{ iff } x_{i1}^* > 0. \tag{14}$$

Let

$$x_{i1}^* = q_i + \xi_{i1} \tag{15}$$

where q_i is the "truth" and x_{i1}^* is doctor 1's diagnosis. Assume

$$\xi_{i1} = \zeta_i + \varepsilon_{i1}$$

where

$$\begin{aligned} \zeta_i &\sim \text{iid N}\left(0, \sigma_\zeta^2\right), \\ \varepsilon_{i1} &\sim \text{iid Extreme Value.} \end{aligned}$$

Consider the properties of x_{i1}. First,

$$\begin{aligned} p_{i1} &= E\left(x_{i1} \mid q_i\right) = \iint 1\left(x_{i1}^* > 0\right) dF_\varepsilon(\varepsilon_{i1}) dF_\zeta\left(\zeta_i\right) \\ &= \iint 1\left(q_i + \zeta_i + \varepsilon_{i1} > 0\right) dF_\varepsilon(\varepsilon_{i1}) dF_\zeta\left(\zeta_i\right) \\ &= \int \frac{\exp\left(q_i + \zeta_i\right)}{1 + \exp\left(q_i + \zeta_i\right)} dF_\zeta\left(\zeta_i\right). \end{aligned} \tag{16}$$

We have to make an assumption about the distribution of q_i: $q_i \sim$iid N(μ, Σ). We can also easily simulate

$$E\left(x_{i1} - p_{i1}\right)\left(x_{i2} - p_{i2}\right).$$

Consider a representative element,

$$\begin{aligned} & E\left(x_{i1} - p_{i1}\right)\left(x_{i2} - p_{i2}\right) \\ = & \iiint \left[1\left(x_{i1}^* > 0\right) - p_{i1}\right]\left[1\left(x_{i2}^* > 0\right) - p_{i2}\right] dF_\varepsilon(\varepsilon_{i1}) dF_\varepsilon\left(\varepsilon_{i2}\right) dF_\zeta\left(\zeta_i\right) \\ = & \int \left[\frac{\exp\left(q_i + \zeta_i\right)}{1 + \exp\left(q_i + \zeta_i\right)} - p_{i1}\right]\left[\frac{\exp\left(q_i + \zeta_i\right)}{1 + \exp\left(q_i + \zeta_i\right)} - p_{i2}\right] dF_\zeta\left(\zeta_i\right) \\ & - \int \frac{\left[\exp\left(q_i + \zeta_i\right)\right]^2}{\left[1 + \exp\left(q_i + \zeta_i\right)\right]^2} dF_\zeta\left(\zeta_i\right). \end{aligned} \tag{17}$$

In our instrumental variables estimation method, we need to know

$$\begin{aligned} & E\left(x_{i1} - q_i\right)\left(x_{i2} - q_i\right) \\ = & E\left(x_{i1} - p_{i1} + p_{i1} - q_i\right)\left(x_{i2} - p_{i2} + p_{i2} - q_i\right) \\ = & E\left(x_{i1} - p_{i1}\right)\left(x_{i2} - p_{i2}\right) + 2E\left(x_{i1} - p_{i1}\right)\left(p_{i2} - q_i\right) + E\left(p_{i1} - q_i\right)\left(p_{i2} - q_i\right). \end{aligned} \tag{18}$$

Each of these terms is easy to simulate. But we cannot identify σ_ζ^2 because we never see the true diagnosis. However, if we also observe a set of data where the random component ζ is independent over doctors, i.e.,

$$\xi_{ij} = \zeta_{ij} + \varepsilon_{ij},$$

then equation (16) does not change, but equation (17) becomes

$$\begin{aligned} & E\left(x_{i1} - p_{i1}\right)\left(x_{i2} - p_{i2}\right) \\ = & \iiint \left[1\left(x_{i1}^* > 0\right) - p_{i1}\right]\left[1\left(x_{i2}^* > 0\right) - p_{i2}\right] dF_\varepsilon(\varepsilon_{i1}) dF_\zeta(\zeta_{i1}) dF_\varepsilon\left(\varepsilon_{i2}\right) dF_\zeta\left(\zeta_{i2}\right) \end{aligned}$$

which is equal to zero if $j \neq k$ and is equal to equation (17) if $j = k$. The differences in covariance matrices between the two samples, one with correlated errors and one without allows us to identify σ_ζ^2 and, therefore, all other terms given the usual qualifications.

3.2 A Better Model of Diagnosis

We can generalize the model to allow for multiple potential diagnoses and co-morbidity. But missing from this structure is any notion of how diagnosers actually use prior information to make new diagnoses. Let x_{ij}^{**} be a latent measure of the information that diagnoser j directly observes about patient i where x_{ij}^{**} is conditionally distributed $iidF_x\left(\bullet \mid q_i\right)$ with $E\left[x_{ij}^{**} \mid q_i\right] = q_i$. Diagnoser j also observes information collected by previous diagnosers $\{x_{ik}^{**}\}_{k=1}^{j-1}$. Given the information available to him, he constructs a continuous measure of his beliefs about patient i captured in x_{ij}^*. We assume that his updating rule can be represented as

$$x_{ij}^* = \alpha_j x_{ij}^{**} + \left(1 - \alpha_j\right) x_{i,j-1}^* \tag{19}$$

with $\alpha_1 = 1$. Special cases of the updating rule in equation (19) include ignoring previous information ($\alpha_j = 1$) and equal weighting of previous information ($\alpha_j = 1/j$).[5] Some updating rules such as $\alpha_j = 1/j$ imply that

$$\lim_{j \to \infty} x_{ij}^* = q_i;$$

others with $\lim_{j \to \infty} \alpha_j > 0$ do not converge as $j \to \infty$. Given x_{ij}^*, diagnoser reports a binary diagnosis according to equation (14).

Now consider the properties of x_{ij}. First,

$$p_{ij} = E\left(x_{ij} \mid q_i\right) = \iiint 1\left(x_{ij}^* > 0\right) \prod_{k=1}^{j} dF_x\left(x_{ik}^* \mid q_i\right). \tag{20}$$

[5] Implicit in the structure of equation (19) is a strong symetry restriction; i.e. every diagnoser has to value all other diagnosers similarly. However, we can generalize at the cost of more cumbersome notation.

While equation (20) is very difficult to evaluate analytically, it is straightforward to simulate as soon as we make an assumption about $F_x(\bullet \mid q_i)$.

Neither of the diagnosis models specified in this section led to simple (e.g., factor analytic) structures for Ω. To the degree that one wants to impose restrictions on the structure of Ω, they should be grounded in a theoretically consistent model of diagnosis. Otherwise, there is no reason to believe the restrictions imposed on Ω. Thus, it worth developing better models of diagnosis.

4 Empirical Examples

4.1 First Example

Our first example uses simulated data from a contrived example that is easy to manipulate and examine. We consider a model of the form in equation (5) In particular, $g(q_{ik}) = p_k(q_i)$ and is defined as

$$p_k(q_i) = \frac{\exp\{q_{ik}\}}{\sum_l \exp\{q_{il}\}}$$

with

$$\begin{aligned} q_{ik} &= u_{ik} + 3\iota_{ik}, \\ u_{ik} &\sim U(0,1), \\ \iota_{ik} &= \begin{cases} 1 & \text{with probability } 1/13 \\ 0 & \text{with probability } 12/13 \end{cases}. \end{aligned}$$

The specification for q_{ik} ensures that q_{ik} is continuous and yet that one diagnosis dominates the others. The other explanatory variables

$$z_{ik} \sim U(0,1).$$

Measurement error occurs according to

$$\Pr[x_{ijk} = 1 \mid q_i] = \frac{p_k(q_i) + \sigma_m \vartheta_{ik}}{\sigma_m + \sum_l p_l(q_i)}.$$

Note that ϑ_{ik} does not vary with diagnosers j, and therefore, as long as $\sigma_m > 0$, there is positive correlation in diagnosers' diagnoses. We set σ_m =0.8. When $p_k(q_i)$ is directly observed, there is no measurement error, OLS and IV should both produce consistent estimates of θ, and OLS should be efficient. When only x_{ijk} is observed but $\sigma_m = 0$, there is uncorrelated measurement error, OLS should provide inconsistent estimates, and IV should provide consistent estimates. When only x_{ijk} is observed amd $\sigma_m > 0$, there is correlated measurement error, both OLS and IV should provide inconsistent estimates. But the asymptotic bias for IV should be smaller than for OLS, and the corrected IV estimates should be consistent. Table 1 provides results for a Monte Carlo experiment with 4 z-variables, 13 x-variables, an error with a standard

deviation of 0.05, and a sample size of $10,000$. There are 100 independent draws of the data for each Monte Carlo experiment.

In the Panel A of Table 1, we see that, for this example, OLS provides unbiased and very precise estimates of the parameters when there is no measurement error. However, once we add measurement error, OLS estimates of the x variable coefficients become significantly biased towards zero as seen in Panel B. The estimates of the z variable coefficients are not biased because the z-variables are uncorrelated with the x-variables. Panel C shows that classical IV estimates (without correction for correlation) are also significantly biased towards zero, but that the bias is much smaller than in the OLS estimates.

Panels D, E, and F report results for IV corrected for correlation. In all three, asymptotic biases are very small and statistically insignificant. The three panels vary, however, with respect to characteristics of the special sample used to estimate Ω. In Panel D, $N = 1000$ and $H = 3$;[6] i.e., the sample is of 1000 patients each independently diagnosed by 3 diagnosers. Even though the estimates in Panel D exhibit no bias, 80% confidence intervals are much larger than they were in previous panels. In Panels E and F, we try different methods to increase the accuracy of $\widehat{\Omega}$ and therefore our IV estimators. In Panel E, we increase H from 3 to 5, while, in Panel F, we increase N from 1000 to 5000. Increasing H has no appreciable effect, while increasing N does. This suggests that an optimal design for a special sample used to estimate V in equation (12) should have a large N but does not require a large H.

4.2 Second Example

Our second example uses simulated data to examine how inferences about the effects of mismeasured mental health conditions depend on the estimation approach. Our outcome measure of interest is hospital length of stay. We simulated a large sample ($N = 10,000$) of patients such that their characteristics match well with the set of 6498 patients actually admitted to a Virginia state psychiatric hospital in 1980. Characteristics of the data are reported in Table 2.[7] In particular, our simulated sample matches the characteristics of a subset of 4893 patients for whom information was available on length of stay, initial and final diagnosis, race, gender, age, and health care facility.

A simulated patient is assumed to be diagnosed with one of 13 primary mental health conditions[8] according to equation (9). We constructed a model

[6] The three diagnosers are used so that the first one is a realization of x_{i1} and the next two are averaged for a realization of x_{i2}. The appropriate adjustment is made due to the fact that x_{i2} is an average. No weighting is necessary because the two data sets have the same distribution.

[7] See Holt, Merwin and Stern (1999) for a full description of the Virginia data.

[8] The possible conditions include substance abuse, alcoholism, organic, schizophrenia, schizo-affective, paranoia, other psychological disorders, bipolar, depression, personality, adjustment, dementia, and other.

of true diagnosis where q_{ik} in equation (5) is modeled as

$$q_{ik} = \frac{\exp\{z_i\alpha_k + e_{ik}\}}{\sum_l \exp\{z_i\alpha_l + e_{il}\}} \tag{21}$$

with $e_{ik.} \sim iidN\left(0, \sigma_e^2\right)$. The existence of the $e_{ik.}$ in equation (21) allows for variation in true medical conditions even after conditioning on observed covariates z_i. Equations (9) and (21) are used to simulate x_{ij}, $j = 1, 2$, where it is assumed that $u_{ik} \sim iidN\left(0, \sigma_u^2\right)$. If u_{ik} varied independently over diagnosers j, then OLS estimates would still be inconsistent (because there is measurement error), but classical IV estimators would be consistent (because the measurement error in x_{i1k} would be independent of the measurement error in x_{i2k}).

"True" values of $(\alpha, \beta, \gamma, \sigma_u, \sigma_e)$ are estimated by matching moments of the simulated data to moments in the Virginia state psychiatric hospital data and matching simulated $\varkappa$ (kappa) statistics[9] to $\varkappa$ statistics found in the literature (e.g., Stravynski, Lamontagne, and Lavallee 1986; Riskind et. al 1987; Clark, et. al 1993; Fennig, et. al 1994; Hiller et. al 1994; McGorry et. al 1995; Kelly and Mann 1996; Parker et. al 1997; Rosenman, Korten, and Levings 1997; Roy et. al 1997; Usten et. al 1997; Clarke, Smith, and Hermann 1998).[10] Estimates of the "true" parameters are reported in Table 3. Once we have "true values" of $(\alpha, \beta, \gamma, \sigma_u, \sigma_e)$, we can simulate data samples from the "true distribution" of observations, estimate the model parameters for each draw of the data, and compute the distribution of our estimators.

The results of this Monte Carlo experiment are recorded in Table 4. OLS produces biased estimates when we introduce measurement error into the explanatory variables. Table 4 shows that all of the OLS estimates are severely biased in that the magnitude of the median bias is typically the same size as the "true value" of the parameter and the 80% confidence interval typically does not include the "truth." Classical IV estimates will also produce biased estimates if diagnosis errors for a patient are correlated across physicians, which, we have argued, will be the case in most administrative data sets. As discussed earlier, however, classic IV will still generally lead to better estimates than OLS. This is seen very clearly in Table 4 in that median biases are uniformly the same sign as the OLS median biases, but they are uniformly a fraction of the size of the OLS median biases. Also, frequently the 80% confidence interval includes the "truth."

The corrected IV estimates should be consistent, but they can have reduced precision because Ω must be estimated. It can be estimated under various

[9] The $\varkappa$ statistic is a measure of agreement and is equal to

$$\frac{p_o - p_c}{1 - p_c}$$

where $.p_o$ is the proportion of observations where there is agreement and p_c is the proportion of observations where there would have been agreement by chance. See Cohen (1960).

[10] σ_u and σ_e are identified by deviations between correlations across diagnoses in the Virginia data and what they would be if $\sigma_u = 0$ and matching $\varkappa$ statistics. We assume $\varkappa = 0.7$ for all conditions.

assumptions about the structure of the errors. For this particular application, the corrected IV approach did not produce worthwhile results when we allowed all the elements of Ω to be unrestricted. In particular, the biases were not much different from the OLS case and the confidence intervals became very large. However, we can improve on these estimates by imposing some structure on Ω.

In many cases, it is feasible to specify Ω as a function of a small number of parameters. In our application, for example, we can write Ω as a function of only two parameters, σ_e and σ_u. Our correction method in this case involves estimating σ_e and σ_u along with the other parameters in θ. To the extent that this structure is correct (which it is by construction in this simulation), we obtain a more efficient estimate of Ω and improve the coefficient estimates.The last column of Table 4 shows that, even in this case, the correction tends to overcorrect so that the median biases are away from zero. Also, 80% confidence intervals are quite large. This suggests that further analysis of small sample properties of the IV correction could lead to a rule of thumb for adjusting $\widehat{\Omega}$ by multiplying by a constant, $0 < \rho < 1$, to both reduce the size of the correction and its affect on the width of the confidence interval. We have not yet performed such an analysis.

5 Conclusions

The analytical results in this paper suggest a feasible approach to correct the bias caused by measurement error in explanatory variables. The results suggest that classical instrumental variables will almost always reduce the bias. However, in the typical case where the instrument is correlated with the explanatory variable measured with error, it will not completely delete the bias. We suggest a feasible correction that deletes the bias asymptotically.

Our empirical results confirm our analytical results with respect to the relationship between OLS and classical IV. They also suggest that estimation error associated with $\widehat{\Omega}$ can have serious effects on small sample bias and confidence interval sizes. They suggest that one might reduce these unfortunate effects by adjusting the correction by a proportionality factor. It is left to future research to determine how to choose a proportionality factor optimally (if at all).

Our empirical results are quite different across examples We do not yet understand what characteristics of the examples caused the corrected IV estimates in Example 1 to perform so much better than those in Example 2. It is also left to further analytical and empirical research to understand better in what circumstances corrected IV will perform well.

Finally, in these artificial examples, we assumed away all real-world problems associated with estimating V in equation (12). In reality, it is impossible for two or more diagnosers to independently diagnose a patient if, for no other reason, they rely on (possibly inaccurate) information provided by the patient. Important topics for future research include the best ways to minimize correlation in diagnoses and to measure the sensitivity of the corrected IV estimates' properties to small, contaminating amounts of correlation in estimates of V.

References

[1] Clark, L., J. McEwen, L. Collard, and L. Hickok (1993). "Symptoms and Traits of Personality Disorder: Two New Methods for Their Assessment." *Psychological Assessments*. 5: 81-91.

[2] Clarke, D., G. Smith, and H. Hermann (1998). "Monash Interview for Liaison Psychiatry (MILP): Development, Reliabilitiy, and Procedural Validity." *Psychosomatics*. 39: 318-328.

[3] Cohen, Jacob (1960). "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement*. 20(1): 37-46.

[4] Fennig, S., J. Craig, M. Tanenberg-Karant, and E. Bromet (1994). "Comparison of Facility and Research Diagnoses in First-Admission Psychotic Patients." *American Journal of Psychiatry*. 151(10): 1423-1429.

[5] Greene, William (1997). *Econometric Analysis*. 3rd Edition. Saddle River, N.J.: Prentice Hall.

[6] Hiller, W., G. Dichtl, H. Hecht, W. Hundt, and D. Zerssen (1994). "Testing the Comparability of Psychiatric Diagnoses in ICD-10 and DSM-III-R." *Psychopathology*. 27: 19-28.

[7] Holt, Fredrick, Elizabeth Merwin, and Steven Stern (1999). "title." University of Virginia, unpublished manuscript.

[8] Kahan, J., S. Bernstein, L. Leape, L. Hilborne, R. Park, L. Parker, C. Kamberg, and R. Brook (1994). "Measuring the Necessity of Medical Procedures." *Medical Care*. 32(4): 357-365.

[9] Kelly, T. and J. Mann (1996). "Validity of DSM-III-R Diagnosis by Psychological Autopsy: A Comparison with Clinician Ante-Mortem Diagnosis." *Acta Psychiatrica Scandinavica*. 94: 337-343.

[10] McGorry, P., C. Mihalopoulos, L. Henry, J. Dakis, H. Jackson, M. Flaum, S. Harrigan, D. McKenzie, J. Kulkarni, and R. Karoly (1995). "Spurious Precision: Procedural Validity of Diagnostic Assessment in Psychotic Disorders." *American Journal of Psychiatry*. 152:2.

[11] Newhouse, Joseph and Mark McClellan (1998). "Econometrics in Outcomes Research: The Use of Instrumental Variables." *Annual Review of Public Health*. 19: 17-34.

[12] Parker, T., P. May, M. maviglia, S. Petrakis, S. Sunde, and S. Gloyd (1997) "PRIME-MD: Its Utility in detecting Mental Disorders in American Indians." *International Journal of Psychiatry in Medicine*. 27(2): 107-128.

[13] Riskind, J., A. Beck, R. Berchick, G. Brown, and R. Steer (1987). "Reliability of DSM-III Diagnoses for Major Depression and Generalized Anxiety Disorder Using the Structured Clinical Interview for DSM-III." *Archives of General Psychiatry.* 44: 817-820.

[14] Rosenman, S., A. Korten, and C. Levings (1997). "Computerized Diagnosis in Acute Psychiatry: Validity of CIDI-Auto Against Routine Clinical Diagnosis." *Journal of Psychist. Res.* 31(5): 581-592.

[15] Roy, M., G. Lanctot, C. Merette, D. Cliché, J. Fournier, P. Boutin, C. Rodrigue, L. Charron, M. Turgeon, M. Hamel, N. Montgrain, L. Nicole, A. Pires, H. Wallot, A. Ponton, Y. Garneau, C. Dion, J. Lavallee, A. Potvin, P. Szatmari, and M. Maziade (1997). "Clinical and Methodological Factors Related to Reliability of the Best-Estimate Diagnostic Procedure." *American Journal of Psychiatry.* 154(12): 1726-1733.

[16] Stravynski, A., Y. Lamontagne, and Y. Lavallee (1986). "Clinical Phobias and Avoidant Personality Disorder Among Alcoholics Admitted to an Alcoholism Rehabilitation Setting." *Canadian Journal of Psychiatry.* 31: 714-719.

[17] Usten, B., W. Compton, D. Mager, T. Babor, O. Baiyewu, S. Chatterji, L. Cottler, A. Gogus, V. Mavreas, L. Peters, C. Pull, J. Saunders, R. Smeets, M. Stipec, R. Vrasit, D. Hasin, R. Room, W. Wan den Brink, D. Regier, J. Blaine, B. Grant, and N. Sartorius (1997) "WHO Study on the Reliability and Validity of the Alcohol and Drug Use Disorder Instruments: Overview of Methods and Results." *Drug and Alcohol Dependence.* 47: 161-169.

6 Tables

Table 1
Monte Carlo Results for the First Experiment
Panel A: OLS Estimates When There is no Measurement Error

Variable	True Value	Median Bias and 80% Confidence Interval		Variable	True Value	Median Bias and 80% Confidence Interval
z_1	0.050	0.000 (-0.003,0.002)		z_2	0.030	0.000 (-0.002,0.002)
z_3	0.120	0.000 (-0.003,0.002)		z_4	0.090	0.000 (-0.002,0.002)
x_1	0.110	0.000 (-0.005,0.005)		x_2	0.120	0.000 (-0.005,0.004)
x_3	0.130	-0.001 (-0.005,0.005)		x_4	0.140	0.000 (-0.005,0.005)
x_5	0.050	0.000 (-0.004,0.005)		x_6	0.060	0.001 (-0.003,0.004)
x_7	0.070	0.000 (-0.005,0.005)		x_8	0.080	-0.001 (-0.005,0.004)
x_9	-0.090	0.000 (-0.005,0.005)		x_{10}	-0.100	0.000 (-0.005,0.004)
x_{11}	-0.110	0.000 (-0.005,0.004)		x_{12}	-0.120	0.000 (-0.005,0.004)
x_{13}	-0.130	0.000 (-0.004,0.004)				

Panel B: OLS Estimates When There is Measurement Error

Variable	True Value	Median Bias and 80% Confidence Interval		Variable	True Value	Median Bias and 80% Confidence Interval
z_1	0.050	0.000 (-0.003,0.002)		z_2	0.030	0.000 (-0.004,0.003)
z_3	0.120	-0.001 (-0.003,0.003)		z_4	0.090	0.000 (-0.003,0.003)
x_1	0.110	-0.064 (-0.068,-0.061)		x_2	0.120	-0.071 (-0.076,-0.066)
x_3	0.130	-0.076 (-0.080,-0.071)		x_4	0.140	-0.084 (-0.089,-0.080)
x_5	0.050	-0.025 (-0.029,-0.021)		x_6	0.060	-0.030 (-0.034,-0.025)
x_7	0.070	-0.037 (-0.042,-0.033)		x_8	0.080	-0.045 (-0.050,-0.041)
x_9	-0.090	0.066 (0.062,0.070)		x_{10}	-0.100	0.075 (0.070,0.080)
x_{11}	-0.110	0.082 (0.077,0.086)		x_{12}	-0.120	0.086 (0.081,0.090)
x_{13}	-0.130	0.085 (0.082,0.089)				

Panel C: IV Estimates When There is Measurement Error Without Correction

Variable	True Value	Median Bias and 80% Confidence Interval		Variable	True Value	Median Bias and 80% Confidence Interval
z_1	0.050	0.000 (-0.003,0.004)		z_2	0.030	0.000 (-0.004,0.004)
z_3	0.120	-0.001 (-0.003,0.003)		z_4	0.090	0.001 (-0.003,0.004)
x_1	0.110	-0.027 (-0.035,-0.021)		x_2	0.120	-0.029 (-0.039,-0.021)
x_3	0.130	-0.031 (-0.038,-0.023)		x_4	0.140	-0.036 (-0.044,-0.030)
x_5	0.050	-0.012 (-0.018,-0.005)		x_6	0.060	-0.011 (-0.018,-0.005)
x_7	0.070	-0.015 (-0.023,-0.008)		x_8	0.080	-0.019 (-0.028,-0.012)
x_9	-0.090	0.028 (0.021,0.033)		x_{10}	-0.100	0.033 (0.024,0.041)
x_{11}	-0.110	0.035 (0.027,0.042)		x_{12}	-0.120	0.035 (0.027,0.044)
x_{13}	-0.130	0.033 (0.027,0.040)				

Panel D: IV Estimates When There is Measurement Error With Correction with $N = 1000$ and $H = 3$

Variable	True Value	Median Bias and 80% Confidence Interval		Variable	True Value	Median Bias and 80% Confidence Interval
z_1	0.050	0.000 (-0.004,0.005)		z_2	0.030	0.000 (-0.004,0.004)
z_3	0.120	0.000 (-0.004,0.004)		z_4	0.090	0.001 (-0.004,0.005)
x_1	0.110	-0.004 (-0.025,0.032)		x_2	0.120	0.001 (-0.031,0.040)
x_3	0.130	0.002 (-0.033,0.035)		x_4	0.140	-0.004 (-0.033,0.036)
x_5	0.050	0.001 (-0.024,0.026)		x_6	0.060	0.002 (-0.025,0.027)
x_7	0.070	0.002 (-0.026,0.029)		x_8	0.080	0.006 (-0.024,0.043)
x_9	-0.090	0.002 (-0.030,0.025)		x_{10}	-0.100	0.001 (-0.041,0.030)
x_{11}	-0.110	-0.006 (-0.034,0.033)		x_{12}	-0.120	-0.006 (-0.037,0.023)
x_{13}	-0.130	-0.002 (-0.034,0.019)				

Panel E: IV Estimates When There is Measurement Error With Correction with $N = 1000$ and $H = 5$

Variable	True Value	Median Bias and 80% Confidence Interval		Variable	True Value	Median Bias and 80% Confidence Interval
z_1	0.050	-0.001 (-0.005,0.005)		z_2	0.030	0.001 (-0.005,0.004)
z_3	0.120	0.000 (-0.005,0.004)		z_4	0.090	0.000 (-0.005,0.004)
x_1	0.110	0.006 (-0.023,0.048)		x_2	0.120	0.000 (-0.034,0.033)
x_3	0.130	0.004 (-0.022,0.035)		x_4	0.140	0.001 (-0.032,0.035)
x_5	0.050	0.001 (-0.017,0.022)		x_6	0.060	-0.002 (-0.025,0.023)
x_7	0.070	0.001 (-0.028,0.027)		x_8	0.080	-0.001 (-0.029,0.025)
x_9	-0.090	-0.001 (-0.027,0.025)		x_{10}	-0.100	-0.002 (-0.035,0.027)
x_{11}	-0.110	0.003 (-0.054,0.029)		x_{12}	-0.120	-0.002 (-0.041,0.032)
x_{13}	-0.130	-0.005 (-0.038,0.023)				

Panel F: IV Estimates When There is Measurement Error With Correction with $N = 5000$ and $H = 3$

Variable	True Value	Median Bias and 80% Confidence Interval		Variable	True Value	Median Bias and 80% Confidence Interval
z_1	0.050	0.000 (-0.005,0.004)		z_2	0.030	0.0000 (-0.005,0.004)
z_3	0.120	0.000 (-0.005,0.004)		z_4	0.090	0.000 (-0.003,0.005)
x_1	0.110	0.000 (-0.015,0.016)		x_2	0.120	0.002 (-0.015,0.020)
x_3	0.130	-0.002 (-0.015,0.017)		x_4	0.140	0.000 (-0.019,0.019)
x_5	0.050	0.000 (-0.013,0.013)		x_6	0.060	0.000 (-0.013,0.013)
x_7	0.070	0.001 (-0.011,0.016)		x_8	0.080	-0.001 (-0.016,0.017)
x_9	-0.090	-0.001 (-0.018,0.013)		x_{10}	-0.100	0.000 (-0.016,0.015)
x_{11}	-0.110	0.000 (-0.022,0.015)		x_{12}	-0.120	0.001 (-0.018,0.016)
x_{13}	-0.130	-0.002 (-0.017,0.011)				

Table 2
Characteristics of 1980 Virginia Data

Variable	Mean	St. Dev.	Definition
LNSPELL	3.27	1.56	Ln Spell Length
BLACK	0.31	0.46	Dummy for Black Race
FEMALE	0.40	0.49	Dummy for Female
FACID1	0.16	0.37	Dummy for Facility 1
FACID2	0.10	0.30	Dummy for Facility 2
FACID3	0.12	0.32	Dummy for Facility 3
FACID4	0.24	0.43	Dummy for Facility 4
FACID5	0.21	0.41	Dummy for Facility 5
FACID6	0.10	0.30	Dummy for Facility 6
FACID7	0.05	0.23	Dummy for Facility 7
FACID8	0.02	0.13	Dummy for Facility 8
MARRY	0.15	0.36	Dummy for married
AGE	39.11	15.19	Age
DEMENT	0.03	0.17	Dummy for diagnosis = Dementia
SUBABU	0.05	0.21	Dummy for diagnosis = Substance Abuse
ALCOHOL	0.15	0.36	Dummy for diagnosis = Alcohol Abuse
ORGANIC	0.06	0.25	Dummy for diagnosis = Organic
SCHIZO	0.22	0.42	Dummy for diagnosis = Schizophrenia
SCHIZAFF	0.08	0.27	Dummy for diagnosis = Schizoaffective

Table 2 (continued)

Variable	Mean	St. Dev.	Definition
PARAN	0.01	0.10	Dummy for diagnosis = Paranoid
OTHPSY	0.06	0.23	Dummy for diagnosis = Other Psychotic
BIPOLAR	0.11	0.31	Dummy for diagnosis = Bipolar
DEPRESS	0.11	0.31	Dummy for diagnosis = Depression
PERSON	0.02	0.14	Dummy for diagnosis = Personality problem
ADJUST	0.07	0.25	Dummy for diagnosis = Adjustment problem
OTHERD	0.04	0.19	Dummy for diagnosis = Other
NPRHST	2.91	3.43	Number of Previous Hospital Stays since 1978

Table 3
IV Estimates of Model Parameters from 1980 Virginia Data

Variable	Estimate		Variable	Estimate
BLACK	-0.057 (0.050)		FEMALE	0.021 (0.047)
FACID1	4.215* (0.274)		FACID2	4.069* (0.279)
FACID3	4.161* (0.279)		FACID4	4.397* (0.271)
FACID5	4.292* (0.278)		FACID6	4.846* (0.280)
FACID7	4.683* (0.270)		FACID8	5.052* (0.302)
MARRY	-0.279* (0.059)		AGE	0.0078* (0.0020)
DEMENT	0.000		SUBABU	-2.538* (0.285)
ALCOHOL	-3.361* (0.234)		ORGANIC	-0.057 (0.341)
SCHIZO	-0.659* (0.238)		SCHIZAFF	-0.656* (0.271)
PARAN	-1.384* (0.633)		OTHPSY	-0.855* (0.351)
BIPOLAR	-1.135* (0.242)		DEPRESS	-0.957* (0.249)
PERSON	-1.208* (0.415)		ADJUST	-2.903* (0.290)
OTHERD	-1.593* (0.361)		NPRHST	0.023* (0.007)
σ_u	$\exp\{2.239\}$		σ_e	$\exp\{1.326\}$

Notes:

1. Starred items are significant at the 5% level.
2. Numbers in parentheses are standard errors.
3. DEMENT is restricted to be zero.
4. No standard errors have yet been estimated for $\widehat{\sigma}_u$ and $\widehat{\sigma}_e$

Table 4
Results of Monte Carlo Experiment

		Median Bias and 80% Confidence Interval for Bias		
Variable	Truth	OLS	Classical IV	Corrected IV
BLACK	-0.057	0.197 (0.159,0.234)	0.033 (-0.015,0.079)	-0.017 (-0.068,0.032)
FEMALE	0.021	0.215 (0.173,0.250)	0.031 (-0.020,0.068)	0.000 (-0.060,0.047)
FACID1	4.215	-1.555 (-1.718,-1.412)	-0.398 (-0.775,-0.066)	0.638 (-0.218,1.367)
FACID2	4.069	-1.562 (-1.720,-1.412)	-0.408 (-0.826,-0.056)	0.637 (-0.225,1.439)
MARRY	-0.279	-0.005 (-0.058,0.057)	-0.002 (-0.064,0.068)	-0.002 (-0.065,0.065)
AGE	0.008	0.008 (0.007,0.010)	0.002 (-0.001,0.004)	-0.002 (-0.005,0.001)
SUBABU	-2.538	1.784 (1.643,1.913)	0.445 (0.063,0.813)	-0.644 (-1.409,0.168)
ALCOHOL	-3.361	2.027 (1.917,2.146)	0.401 (0.134,0.738)	-0.605 (-1.302,0.079)
ORGANIC	-0.057	0.407 (0.285,0.550)	0.239 (-0.093,0.644)	-0.553 (-1.249,0.189)
SCHIZO	-0.659	0.768 (0.657,0.874)	0.299 (-0.001,0.621)	-0.565 (-1.276,0.201)
SCHIZAFF	-0.656	0.692 (0.586,0.815)	0.269 (-0.015,0.607)	-0.595 (-1.143,0.186)
PARAN	-1.384	1.168 (0.998,1.322)	0.354 (-0.358,1.104)	-0.619 (-1.753,0.428)
BIPOLAR	-1.135	0.988 (0.870,1.088)	0.315 (0.014,0.639)	-0.516 (-1.186,0.154)
ADJUST	-2.903	1.966 (1.857,2.084)	0.315 (0.014,0.639)	-0.625 (-1.369,0.072)

Notes:

1. Reported statistics are $\begin{array}{|c|}\hline \text{median bias} \\ \hline (80\%\text{ confidence interval}) \\ \hline \end{array}$.

2. Statistics for some variables are not reported to save space.