TROMBLEY, MATTHEW J., Ph.D. Examining the Effect of Health Behaviors on Wages and Healthcare Utilization in Models with Endogeneity. (2014)
Directed by Dr. Christopher A. Swann. 186 pp.

This dissertation contains three essays on applied health economics. Although each essay is independent of the others, all three address the issue of estimating models where the relationship of interest is confounded by factors that are unobservable to the researcher. The first essay is an econometric simulation study while essays 2 and 3 address behavioral health topics.

Essay 1 compares the accuracy and efficiency of parametric count data specifications paired with the Extended Olsen Model (EOM; Terza, 1998, 2009). The EOM is a nonlinear instrumental variables approach that allows for consistent estimation of model parameters when the data suffer from binary endogenous switching (e.g., endogenous sample selection or endogenous treatment). Count data models are ubiquitous in the health literature for estimating non-negative, discrete outcomes such as physician visits, hospital admissions, cigarettes smoked, etc. Essay 1 provides insight into the model selection process by informing practitioners which specification is likely to provide the most accurate parameter estimates under a variety of data configurations. Essay 1 also demonstrates the applicability of the Conway-Maxwell Poisson (CMP), a flexible count model developed in the field of industrial engineering that has yet to be utilized in the economic literature.

In Essay 2, I apply a count version of the Extended Olsen Model to estimate the relationship between marijuana use disorder (MUD) and ER visits among US Medicaid recipients. This essay is the first in the literature to estimate the relationship between

marijuana consumption and the demand for ER visits in isolation from other illicit drugs, thus providing an important addition to the ongoing policy regarding the potential relaxation of marijuana regulation. This study is also the first in the illicit substance literature to use an instrumental variables count data model to estimate the full distribution of ER visits, thus accounting for unobserved factors that may be jointly correlated between individual propensity for MUD and demand for ER visits. I fail to find a positive relationship between MUD and ER visits, instead uncovering suggestive, but inconclusive, evidence that MUD and ER visits may rather be negatively correlated.

Essay 3 considers the relationship between wages and obesity. Although prior literature has firmly established a negative relationship between wages and obesity, it is equivocal with regard to the underlying pathway(s) through which obesity results in lower wages. Using firm-level data that gives me unique access to proxies for productivity and discrimination against obese individuals, I find that inputs to productivity, particularly health, are important confounders of the wage-obesity relationship. I fail to find any evidence of discrimination against obese employees, but I do find that among females the negative relationship between wages and obesity exists only among mothers.

EXAMINING THE EFFECT OF HEALTH BEHAVIORS ON WAGES AND HEALTHCARE UTILIZATION IN MODELS WITH ENDOGENEITY

by

Matthew J. Trombley

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2014

Approved by

Christopher A. Swann
Committee Chair

To my parents, Larry and Jenny Trombley

APPROVAL PAGE

This dissertation, written by Matthew J. Trombley, has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair Christopher A. Swann

Committee Members Jeremy W. Bray

Joseph V. Terza

Stephen P. Holland

October 13, 2014
Date of Acceptance by Committee

August 28, 2014
Date of Final Oral Examination

ACKNOWLEDGMENTS

I would like to acknowledge the enormous effort, support, and patience of my dissertation advisor Christopher Swann. Even if he were not my advisor I likely would've graduated saying that I learned more from him about economics than from anyone else. I would like to thank my committee members for their hard work and dedication to my success: Dr. Jeremy Bray for his encouragement and advice, and for making Essay 3 possible with access to data; Dr. Joseph Terza for all that he taught me about econometrics, his guidance, and for his collaboration on essay 1; and Dr. Stephen Holland for his support, advice, and suggestions over the years.

I would like to acknowledge the contributions of Jean Rosales, who always made sure that all deadlines were met and all transitions went smoothly, and who also provided friendship and a listening ear. I would like to acknowledge the Work, Family, and Health Network for providing data for Essay 3, and acknowledge the contributions of Jesse Hinde and Michael Mills towards Essay 3. I would like to thank the remaining faculty, staff, and my friends in the program for all of their valuable assistance, with particular thanks to Matthew Rabbitt, who never hesitated to lend a hand or offer encouragement.

At Furman University I would like to thank Dr. Jeffrey Yankow for instilling in me a love of economics, Dr. Kenneth Peterson for all of his valuable advice, and Dr. Thomas Smythe for the opportunities he gave me and the belief he had in me.

Lastly, I would like to acknowledge all of the friends and family members who supported and encouraged me throughout the program. Special thanks to my wife for all

she sacrificed while I was in graduate school, and to my parents, who instilled in me everything I needed to complete this dissertation.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

CHAPTER I

ESSAY 1: EFFICIENT ESTIMATION OF PARAMETRIC COUNT MODELS WITH ENDOGENEITY: A MONTE CARLO ANALYSIS

1. Introduction

Count data regression models are a popular method of estimation when the outcome of interest assumes only nonnegative integer values. The standard specification for count data is the Poisson distribution. Efficiency of a Poisson estimator requires the restrictive assumption that the conditional mean equals the conditional variance: the assumption of equidispersion. This assumption is tenuous in most situations, and much of the count modeling literature consists of refinements to the basic Poisson in an effort to provide a more flexible, and therefore efficient, fit for count data that are not equidispersed.

One issue that has received less attention is the modeling of count processes in the presence of endogenous switching. Endogenous switching models are appropriate when (a) the outcome of interest occurs as one of two regimes, determined by a binary "switching" variable, and (b) the switching and outcome variables are correlated through unobserved confounders. The following study utilizes Terza's (2009) Extended Olsen Model (EOM) in order to account for endogenous switching in either of its two main incarnations: sample selection or endogenous treatment. We compare the performance of the standard Poisson model to several generalizations, including the negative binomial (NB), restricted generalized Poisson (RGP) and the Conway-Maxwell Poisson (CMP).

Unlike the popular NB model that can fit overdispersed data (variance exceeds the mean) but not underdispersed data (mean exceeds the variance), the RGP and CMP are able to fit data under either dispersion condition.

The RGP has featured sporadically in the economics literature (Price; 2009; Wang & Famoye, 1997) but failed to catch on as an alternative to the NB or Poisson specifications, likely due to the restricted range of underdispersion it is able to fit, and its similarity to the NB model in the case of overdispersion. To date, the CMP has not penetrated the econometric literature, although it is increasingly applied in other disciplines. Consequently, several papers have discussed the performance of the CMP relative to other models (Francis et al., 2012; Sellers & Shmueli, 2010; Zou, Lord, & Geedipally, 2011). However, no study has yet investigated the efficacy of the CMP or RGP in the context of endogenous switching. The purpose of this paper is to present the EOM framework, and compare the performance of the relatively unknown RGP and CMP to the widely used NB and Poisson models under conditions of endogeneity.

Results suggest that in the case of endogenous treatment, the new CMP model is most robust to misspecification, and, in general, provides both the best overall fit of the data and the minimum bias of the treatment effect estimates. However, the model typically performs poorly in accurately estimating coefficient values. The NB and RGP models both perform well in the case of endogenous sample selection, and either is preferable to the CMP in cases of sample selection with positive dependence. However, the CMP is the superior model in situations of negative dependence in both the endogenous treatment and endogenous sample selection cases. The standard Poisson is

the least robust of the four models, performing relatively poorly when the data are not generated by a pure Poisson process.

2. Parametric Count Models

2.1. Poisson Specification and Dependence

The standard econometric model for count data is the Poisson distribution. The Poisson probability mass function (PMF) can be expressed as:

$$P(Y = y|X) = \frac{e^{-\lambda}\lambda^{y}}{y!} \qquad y = 0, 1, 2 \ldots \qquad \text{(Eq. 1)}$$

where $E[Y \mid X] = V[Y \mid X] = \lambda = \exp(X\beta)$, X is a (n x k) vector of covariates, and β is a (k x 1) vector of parameters to be estimated. The primary shortcoming of the Poisson specification is the assumption of *equidispersion*, i.e., that the conditional variance is equal to the conditional mean. Provided the conditional mean is correctly specified, Poisson estimates retain their consistency if this assumption does not hold. However, the precision of the estimates is diminished, and making economic inferences may become problematic (Cameron & Trivedi, 2005).

The assumption of equidispersion is identical to the assumption of independence of the events. The independence assumption requires that the probability of successive events (e.g., physician visits) is conditionally independent of the number of events that have previously occurred (e.g., previous physician visits.) (This refers to the independence of events comprising a single count, rather than independence of successive counts over time.) If prior events make successive events more likely, then the counts are said to have positive occurrence dependence, which results in

overdispersion: a variance greater than the mean. If prior events make successive events less likely, then the counts are said to have negative occurrence dependence, which results in underdispersion: variance less than the mean (Winkelmann, 2008).[1] The Poisson is thus troubling from a theoretical standpoint since any individual heterogeneity that is not controlled for by X will introduce dependence between outcomes, even if the heterogeneity is exogenous. The potential problem with the Poisson then is not just that it is too restrictive of a statistical model to perfectly fit the data, but also too restrictive of a theoretical model to fit most economic events.

2.2. Individual Heterogeneity and Dispersion

One way to model overdispersion is to introduce a stochastic error term. As mentioned previously, count regression typically assumes that the expectation of Y conditional on X is an exponential function of a linear index: $E[Y|X] = \exp(X\beta)$. However, it is standard econometric practice to model the linear expectation of Y as having an additive error, such that: $E[Y|X] = \exp(X\beta+\varepsilon)$. The "unobserved heterogeneity" ε introduces positive dependence between outcomes since it increases the probability of all positive realizations of Y but is not conditioned out of the expectation by X. Moreover, exponential transformation of the error will manifest as a positive

[1] Another closely related statistical framework considers duration dependence rather than occurrence dependence. In this case duration refers to the amount of "time" elapsed between successful outcomes (i.e., a count of failures), where "time" is measured as the number of binary draws from a distribution that could result in success or failure. As discussed in Winkelmann (2008), the statistical consequences of this framework are essentially the same as occurrence dependence. We choose to use the occurrence framework for the remainder of the study with the opinion that dependence of occurrence rather than duration is a more intuitive way to understand the relationship between count outcomes.

value, resulting in a variance larger than the mean.[2] This helps demonstrate the relationship between positive dependence and overdispersion.

Models of continuous, unobserved heterogeneity are known as continuous mixture models, because they are generated as a mixture of a Poisson process and a second, continuous, positive distribution (typically Gamma or log-normal.) The marginal density of Y can be recovered by integrating out ε over its assumed distribution, in which case the independence assumption is restored.

Practitioners typically avoid this necessity by assuming that the unobserved heterogeneity follows a gamma distribution. Doing so allows for a closed-form solution to the marginal density of Y that may be expressed as

$$P(Y = y|X, \nu) = \frac{\lambda^y}{y!}\frac{\Gamma(\nu+y)}{\Gamma(\nu)\Gamma(y+1)}\left(\frac{\lambda}{\nu+\lambda}\right)^y \qquad \text{(Eq. 2)}$$

The result is the popular Negative Binomial (NB) specification, which includes an additional parameter ν that models overdispersion, even if the assumed data generating process does not involve continuous heterogeneity. ν is constrained to be ≥ 0, with the standard Poisson nested at ν = 0. The conditional mean remains identical to the Poisson specification, but the variance is now permitted to deviate from the mean and may be expressed as $Var[Y|X,\nu] = \lambda(1+\nu^{-1}\lambda)$. Although frequently appearing in the literature, the

[2] The error is sometimes modeled as a multiplicative error, with mean 1, expressed as $E[Y|X]=\exp(X\beta)u$. It may be more sensible to use this notation in the case of a positively distributed error such as the Gamma. However, since $\exp(X\beta)u = \exp(X\beta+\varepsilon)$ with $\varepsilon = \ln(u)$ and $E[\ln(u)] = 0$, the two notations are essentially equivalent. Rather than juggle two conventions I choose to use the additive linear error for parsimony with standard linear models.

NB is limited to modeling overdispersion, where the variance exceeds the mean. If data are underdispersed, the NB will collapse to a nested standard Poisson.

2.3. Count Models with Statistical Dependence

Another group of count models includes an additional parameter that allows models to account for the variance of Y without making an assumption about the distribution of the error. One example that is flexible enough to handle both over- and underdispersion is the restricted generalized Poisson (RGP) model (Famoye, 1993). The PMF of the generalized Poisson model can be expressed:

$$P(Y = y|X, \nu) = \left(\frac{\lambda}{1+\nu\lambda}\right)^{y} \frac{(1+\nu\lambda)^{y-1}}{y!} \exp\left(\frac{-\lambda(1+\nu y)}{1+\nu y}\right) \quad \text{(Eq. 3)}$$

with the dispersion parameter ν restricted to the range [min(-1/max(λ),-1/max(Y)),1), where λ is the conditional expectation of Y. $\nu > 0$ indicates the data are over-dispersed, while $\nu < 0$ indicates under-dispersion: the standard Poisson distribution is nested at $\nu = 0$. As in the case of the NB model, the conditional mean of Y remains the same as the Poisson specification and the variance is now a function of both λ and ν: $V[Y|X,\nu] = \lambda(1+\nu\lambda)^2$. Although the RGP can model underdispersed data, the range of underdispersion it can fit is limited to the value min(-1/max(λ),-1/max(Y)). This restriction is set because for values of $\nu < 0$, the CDF does not sum to one. So long as ν is constrained to fall within the restricted range, the truncation error will only be approximately 0.5%, and estimates will be consistent. However, if the true value of ν falls below the restriction, then the size of the error will increase to the point that the model fails to converge. This limitation means that the magnitude of underdispersion the

model can handle is inversely related to the maximum value of Y. This is often not problematic as higher underdispersion typically results in shorter right tails of the distribution. However, in cases where there is both overdispersion from unobserved heterogeneity, and negative statistical dependence, the RGP may have trouble fitting the data, an outcome that will be explored in Section 5.4.

Another model of statistical dependence is the Conway Maxwell Poisson (CMP). The CMP was formulated by Conway and Maxwell (1962) in the context of queuing systems for engineering applications. A reformulation of the CMP in the regression context was provided by Shmueli, Minka, Kadane, Borle, and Boatwright (2005), opening the door for additional applications of the model. Recent analyses include the modeling of automobile accidents (Lord, Geedipally, & Guikema, 2010; Lord, Guimeka, & Geedipally, 2008), risk analysis (Guikema & Goffelt, 2008), and customer behavior (Borle, Dholakia, Singh, & Westbrook, 2007), among others (Sellers, Borle, & Shmueli, 2011). Despite the increasing popularity of the model, it has yet to be utilized in the economics literature.

The CMP is a straightforward extension of the standard Poisson, with pmf

$$P(Y = y|X, \nu) = \frac{\lambda^y}{(y!)^\nu Z(\lambda,\nu)} \qquad \text{for } y = 0, 1, 2 \ldots \qquad \text{(Eq. 4)}$$

where $\nu \geq 0$, and $Z(\lambda,\nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$. Like the NB and RGP, the CMP nests the standard Poisson ($\nu = 1$). In addition, the CMP nests two other common specifications: the geometric ($\nu = 0$, $\lambda < 1$), and Bernoulli ($\nu \rightarrow \infty$ with probability $\lambda/1+\lambda$; Shmueli et al.,

2005). The nested Poisson allows for a simple statistical test of whether or not the specification varies significantly from the standard Poisson. Unlike the RGP, the CMP is theoretically unlimited in the range of underdispersion it is able to fit and is even capable of modeling binary outcomes, giving it an unmatched flexibility among fully parametric models. The CMP also differs from the NB and RGP models in that the dispersion parameter ν enters the expression for conditional variance, in addition to conditional mean.

According to Schemueli et al. (2005) the first two moments of the CMP are

$$E[Y|X,\nu] = \frac{\partial \ln Z}{\partial \ln \lambda} = \lambda \frac{\partial \ln Z}{\partial \lambda} \quad \text{(Eq. 5)}$$

and

$$V[Y|X,\nu] = \frac{\partial^2 \ln Z}{\partial^2 \ln \lambda} \approx \lambda \frac{\partial E[Y|X,\nu]}{\partial \ln \lambda}. \quad \text{(Eq. 6)}$$

An obvious concern with the CMP specification as presented is that $Z(\lambda,\nu)$ is an infinite summation with no closed form. There are two possible approaches to this issue. The first is to use an approximation for $Z(\lambda,\nu)$ provided by Minka, Shmeuli, Kadane, Borle, and Boatwright (2003) who demonstrate that

$$Z(\lambda,\nu) \approx \frac{\exp\left(\nu\lambda^{\frac{1}{\nu}}\right)}{\lambda^{(\nu-1)/2\nu}(2\pi)^{(\nu-1)/2}\sqrt{\nu}} \quad \text{(Eq. 7)}$$

Given 7, the first two moments may then also be approximated:

$$E[Y|X,\nu] = \lambda^{1/\nu} - \frac{\nu-1}{2\nu} \quad \text{(Eq. 8)}$$

and

$$V[Y|X,\nu] = \frac{\lambda^{1/\nu}}{\nu}. \quad \text{(Eq. 9)}$$

The approximations come at the cost of accuracy. Minka et al. (2003) warn that the approximated moments will not hold if $\nu > 1$ (negative dependence) or $\lambda < 10^{\nu}$ (small mean with marginal positive dependence). The approximation for $Z(\lambda,\nu)$ is also biased. The approximated value is generally within 5% of the true value. However, if λ falls below 1, or if the data are highly overdispersed ($\nu < 0.5$) the absolute percentage error increases rapidly, ranging from 6% to 100% (Shmueli et al., 2005). If this approximation is used in the calculation of the likelihood function, coefficient estimates will become biased, further undermining the estimated relationship between X and Y.

An alternative approach is to estimate $Z(\lambda,\nu)$ directly. Due to the factorial term in the denominator, the summation is convergent. This allows $Z(\lambda,\nu)$ to be calculated to within some acceptable truncation error. Furthermore, since the Z-function is convergent, the derivative of the summation can be expressed as the summation of the derivative, which allows $E[Y|X,\nu]$ to be computed to within an acceptable truncation error, as well, so that

$$E[Y|X,\nu] = \lambda\left(\frac{\sum_{j=1}^{\infty}\frac{j\lambda^{j-1}}{(j!)^{\nu}}}{\sum_{j=0}^{\infty}\frac{\lambda^{j}}{(j!)^{\nu}}}\right) \quad \text{(Eq. 10)}$$

2.4. Other Count Models

The literature contains several other parametric approaches that can model both over- and under-dispersed count data. Efron's (1986) double-Poisson model includes an additional parameter that can fit both types of dispersion. However, the double-Poisson is problematic in that the PMF is not defined when Y = 0. Additionally, the CDF of the model is not guaranteed to sum to one. Although the double-Poisson has seen a great deal of application in the statistical literature, the model has failed to catch on in the applied econometrics literature, and we omit this model as a comparison against the CMP.

Another branch of the parametric literature involves modeling dispersed Poisson data using a polynomial expansion around the standard Poisson distribution.[3] The primary benefit of this class of models is the relaxing of the assumption that the expected value of Y is an exponential function of a linear index, i.e., that $E[Y] = \exp(X\beta+\varepsilon)$. The converse is that the moments of the distribution become more complex, as does the computation of corresponding policy effects and standard errors. The two main drawbacks are the additional parameter that must be estimated for each increased order of polynomial, as well as the fact that the correct polynomial order is unknown to the researcher a priori, and must be determined using various fit statistics. Additionally,

[3] A similar, semi-parametric approach uses series expansion to approximate the density of the unobserved heterogeneity, maintaining the assumption of a linear exponential index. See Gurmu, Rilstone, and Stern (1999) for an example.

polynomial expansion models are not amenable to integration due to the higher order values of the expectation that must be computed. This makes adapting the model to account for unobserved heterogeneity problematic. Although a notable segment of the literature, polynomial expansion methods have not gained the traction enjoyed by the standard Poisson or continuous mixture models, particularly in the health economics/health services literature, and we also do not consider these models for comparison.

There are several special cases of count outcomes that appear frequently in the literature, that we do not consider in this study. The finite mixture model assumes that the data belong to one of K discrete classes with unknown probability π_k, each of which has a given count specification $f_k(Y|X)$ (Winkelmann, 2008). The Zero-Inflated model fits count data when zeros may occur due to one of two processes: a binary "participation" process, as well as zeros occurring naturally as part of the assumed count specification (Lambert, 1992; Mullahy, 1986). Endogenous sample selection is a special case of the Zero-Inflated model, where the errors of the participation equation and the count outcome remain correlated after controlling for observables. The Zero-Inflated model is closely related to the Hurdle model (Mullahy, 1986). The Hurdle model also assumes a binary determinant of participation: however, the Hurdle model assumes that all zeros are due to non-participation, while the count outcome is a truncated distribution that generates only non-zero values.

These special cases provide an extension of a given count distribution (e.g., Poisson, CMP) rather than a replacement for them. For this reason we do not compare

behavior of the various specifications under each of these frameworks (with the exception of the special Zero-Inflated case of endogenous selection). Although there may be some merit in such an analysis, comparison of each count model under every scenario in the literature lies outside the scope of this study.

3. Count Models in the Full-Information Extended Olsen Framework

A common difficulty faced in applied research is the presence of endogeneity in the model. No matter how robust, a count probability mass by itself is unable to correctly model the outcome of interest if the true data generating process suffers from endogeneity. A frequently encountered incarnation of the endogeneity problem is the presence of a binary endogenous switching variable (Terza, 2009). Switching refers to a situation where the outcome of interest, Y, occurs as one of two possible values depending on the value of a binary "switching" variable, X_s. For instance, researchers interested in estimating the effect of insurance on the number of physician visits would have to account for the fact that they are unable to observe the number of physician visits that would have occurred if individuals were in the opposite insurance state from the one in which they are observed. If the probability of obtaining insurance is related with the number of physician visits in ways unobservable to the researcher (e.g., underlying health) then the switching is said to be endogenous.

Switching models are a variant of the Roy Model, where the observed outcome Y depends on the value of a latent variable X_s^*. Two common manifestations of switching are sample selection and treatment. Sample selection occurs when X_s places Y into one of two regimes so that $Y_i = y$ if $X_{si} = 1$ and $Y_i = 0$ if $X_{si} = 0$. In the treatment case, Y

occurs as being influenced by X_s(treated) or uninfluenced by X_s (untreated). Then $Y_i = y_{1i}$ if $X_{si} = 1$, $Y_i = y_{0i}$ if $X_{si} = 0$. In both instances of switching, the value of Y that could have occurred but did not, remains unobservable.

Consistent estimation of Y requires accounting for the relationship between Y and X_s. In the endogenous switching case, Y and X_s are correlated through both observed and unobserved attributes, so that controls for observable characteristics are insufficient to consistently estimate Y. The following provides a framework for estimating count models in the presence of either endogenous sample selection or endogenous treatment.

Let X_u be an unobserved confounder through which X_s and Y are related. We formalize the relationship between X_s and X_u as:

$$Xs = I(W\alpha + Xu > 0) \qquad \text{(Eq. 11)}$$

Where $W = [X_o\ W^+]$, X_o is an observed vector of covariates, W^+ is a vector of identifying instrumental variables, and $(X_u|W)$ is a known binary distribution. The count PMFs remain as before, except now

$$\lambda = \exp(Xo\beta o + Xu\beta u). \qquad \text{(Eq. 12)}$$

Let $f(Y,X_s)$ refer to the joint distribution of the outcome of interest (Y), and the (possibly endogenous) switching variable (X_s). We can express the joint distribution of the two variables as

$$f(Y, X_s|W) = \int_{-\infty}^{\infty} f(Y, X_s, X_u|W)d(X_u|W) = \int_{-\infty}^{\infty} f(Y|X_s, W, X_u)f(X_s, X_u|W)d(X_u|W). \quad \text{(Eq. 13)}$$

Since $(X_u|W)$ is a known distribution, the form of $f(X_s,X_u|W)$ is also known:

$$f(X_s, X_u|W) = X_s g_{[X_u > -W\alpha]}(X_u|W) + (1 - X_s) g_{[X_u < -W\alpha]}(X_u|W), \qquad \text{(Eq. 14)}$$

where $g_{[q]}(X_u|W)$ is the conditional distribution of X_u over the interval q. Using (Eq. 13) and (Eq. 14) we obtain

$$\begin{aligned} f(Y, X_s|W) = {} & X_s \int_{-W\alpha}^{\infty} f(Y|X_s, W, X_u) g(X_u|W) d(X_u|W) \\ & + (1 - X_s) \int_{-\infty}^{-W\alpha} f(Y|X_s, W, X_u) g(X_u|W) d(X_u|W) \end{aligned} \qquad \text{(Eq. 15)}$$

in the case of ET, and

$$f(Y, X_s|W) = X_s \int_{-W\alpha}^{\infty} f(Y|X_s, W, X_u) g(X_u|W) d(X_u|W) + (1 - X_s) G(-W\alpha) \qquad \text{(Eq. 16)}$$

in the SS case, where $f(Y|X_s,W,X_u)$ is the PMF of the specified count distribution, with λ expressed as in (12). The likelihood function follows simply from (15) or (16):

$$L(\theta|Y, X_s, W) = \prod_{i=1}^{n} f(Y_i, X_{si}|W_i) \qquad \text{(Eq. 17)}$$

The integration of the error term out of the PMF of Y is analogous to the integration of the heterogeneity term in the continuous mixture model. However, Terza (2009) demonstrates that under the EOM framework, the exact form of the error of Y need not be specified so long as it is a linear function of X_u conditional on W. Our model thus fits dispersion along two dimensions: by modeling unobserved heterogeneity (X_u) that is jointly correlated with X_s and Y, as well as modeling exogenous dependence with

the parameter ν. Lastly, assuming the vector of instruments is valid, the estimated value of β_u allows for straightforward testing of the null hypothesis that X_s is exogenous.

Since X_u only appears in the λ term, the expectation of λ can be recovered by integrating out X_u, so that

$$\hat{\lambda} = \int_{-\infty}^{\infty} \exp(X_o\hat{\beta}_o + X_u\hat{\beta}_u)g(X_u|W)d(X_u|W) \qquad \text{(Eq. 18)}$$

If we assume that $g(X_u|W) \sim N(0,1)$ then we can rewrite (1.18) as

$$\hat{\lambda} = \exp(X_o\hat{\beta}_o)\exp\left(\frac{\hat{\beta}_u^2}{2}\right) \qquad \text{(Eq. 19)}$$

(We maintain the assumption of standard normally distributed errors for the remainder of the paper. Proof of (Eq. 19) is contained in Appendix A). Now $E[Y|X_o,X_u,\nu]$ can be expressed as in (Eq. 5) or (Eq. 8) where $\hat{\lambda}$ replaces λ in the formulation.

4. Estimating Policy Effects in the Extended Potential Outcomes Framework

Coefficient estimates of count data models are of limited usefulness, providing only the direction of the relationship between explanatory variables and the outcome of interest Y. Often, of primary consideration to researchers is estimating the effect that the change in a specific variable exerts on Y. Particularly, we assume that studies are focused on a certain variable (the *policy variable* - X_p) that is at present, or in the future will be, under the control of a policy-making entity. In standard econometric models, estimating the effect that a change in X_p would have on Y (the *policy effect* - PE) is a straightforward exercise. In the case of endogenous switching, outcomes occur in one of two regimes (e.g., treated or untreated) and directly observing outcomes in the opposite

regime from that which occurred is not possible. Furthermore, assignment to regimes is considered to depend on both observed (X_o) and unobserved (X_u) confounders. Correcting for endogenous assignment is necessary to produce unbiased estimates of policy effects.

In Terza's (2012) Extended Potential Outcomes Framework, the expected value of Y is said to be "mean causal" if we can assume that the vector $V = [X_o, X_u]$ is comprehensive: i.e., that V comprises all possible confounders for X_p and Y. In this case, conditional on V, differences in the mean observed value of Y can be exclusively attributed to differences in the value of X_p. In the extended POF, the policy effect of interest can be broadly stated as the difference between the distributions of $Y_{X_{p1}}$ and, $Y_{X_{p2}}$ where X_{p1} and X_{p2} represent well-defined and distinct counterfactually imposed pre- and post-intervention versions of the policy variable, respectively.[4] Without loss of generality, we represent the policy increment and pre- and post-policy scenarios as Δ, $X_{p1} = X_p^*$ and $X_{p2} = X_p^* + \Delta$, respectively, where X_p^* is a counterfactual version of X_p representing a (possibly degenerate) random variable. For the remainder of the discussion we will focus on the following average incremental effect (AIE) as the policy effect of interest

$$\mathrm{AIE}(\Delta) = \mathrm{E}\left[Y_{X_p^*+\Delta}\right] - \mathrm{E}\left[Y_{X_p^*}\right] \qquad \text{(Eq. 20)}$$

[4] Simply put, the framework involves a thought experiment where we begin by assuming everyone in the population has $X_p = X_{p1}$ (a specific value or distribution of the treatment variable), with corresponding outcome Y_{Xp1}. We then assume that everyone is shifted to $X_p = X_{p2}$ by some exogenous force, and note the corresponding outcome Y_{Xp2}. With the assumption that V is comprehensive, the difference between Y_{Xp2} and Y_{Xp1} can be casually interpreted as the effect of the "treatment" (i.e., the shifting of X_{p1} to X_{p2}). This reflects the total effect of both "prevention" (all individuals in the sample reporting X_{p2} remain at X_{p2}) and "treatment" (all individuals in the sample reporting Xp1 are shifted to X_{p2}.)

Expression (20) can be generalized to include binary or continuous changes in X_{p1}. For example, when the policy variable is binary, if we set $X_p^* = 0$ and $\Delta = 1$ then (20) measures the average treatment effect (ATE). When the policy variable is continuous and Δ approaches 0 then $\lim_{\Delta \to 0} (\text{AIE}(\Delta) / \Delta)$ represents the average marginal effect (AME) of an infinitesimal change in the policy variable

In the case of endogenous treatment, the endogenous variable X_s is also the policy variable of interest, X_p. As a binary variable, the appropriate policy effect is the average treatment effect (ATE) that would occur if X_p was exogenously shifted from 0 to 1. Let $\lambda_1 = \exp(\beta_p + X_o\beta_o)\exp\left(\frac{\beta_u^2}{2}\right)$ and $\lambda_0 = \exp(X_o\beta_o)\exp\left(\frac{\beta_u^2}{2}\right)$. Then, given (Eq. 11) and (Eq. 19) we can write

$$\text{ATE} = E_{X_o}[\lambda_1 - \lambda_0] \qquad \text{(Eq. 21)}$$

and

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\lambda}_{1i} - \hat{\lambda}_{0i} \qquad \text{(Eq. 22)}$$

Where $\hat{\lambda}$ is the estimate of λ constructed using $[\hat{\beta}_p\ \hat{\beta}_o\ \hat{\beta}_u]$, consistent estimators of $[\beta_p\ \beta_o\ \beta_u]$.

In the SS case, we assume that X_p is exogenous. Since the policy variable is not constrained to be binary as in the case of ET, all three policy effects discussed previously are feasible. The estimator for the ATE remains as in (1.22). Considering a non-binary

X_p, we can write $\lambda_{p2} = \exp(X_{p2}\beta_{p2} + X_o\beta_o)\exp\left(\frac{\beta_u^2}{2}\right)$ and $\lambda_{p1} = \exp(X_{p1}\beta_{p1} + X_o\beta_o)\exp\left(\frac{\beta_u^2}{2}\right)$. The AIE can now be expressed as:

$$\text{AIE} = E_{X_o, X_{p1}}[\lambda_{p2} - \lambda_{p1}] \quad \text{(Eq. 23)}$$

which can be estimated by

$$\widehat{\text{AIE}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\lambda}_{p2i} - \hat{\lambda}_{p1i}. \quad \text{(Eq. 24)}$$

If X_p is continuous, and the researcher has no policy-relevant increment in mind, then the policy effect of interest is the AME.

$$\text{AME} = E_{X_o, X_p}[\beta_p \lambda] \quad \text{(Eq. 25)}$$

which can be consistently estimated by:

$$\widehat{\text{AME}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_p \hat{\lambda}_i \, . \quad \text{(Eq. 26)}$$

The previous formulations for the policy effects do not hold with the CMP distribution since the expected value of Y does not have a closed-form solution. With our assumption that X_u is standard normally distributed we can express (Eq. 18) as

$$\lambda = \int_{-\infty}^{\infty} \exp(X_p\beta_p + X_o\beta_o + X_u\beta_u)\varphi(X_u)dX_u. \quad \text{(Eq. 27)}$$

Appendix A demonstrates that the endogenous version of λ has a closed-form solution. Unfortunately, (Eq. 10) does not hold if the closed-form solution is used when calculating the summation terms. Instead, (Eq. 10) must be re-expressed as

$$E[Y|X,\nu] = \frac{\sum_{j=1}^{\infty} \frac{j\left[\int_{-\infty}^{\infty} \exp(X_p\beta_p + X_o\beta_o + X_u\beta_u)\varphi(X_u)dX_u\right]^j}{(j!)^\nu}}{\sum_{j=0}^{\infty} \frac{\left[\int_{-\infty}^{\infty} \exp(X_p\beta_p + X_o\beta_o + X_u\beta_u)\varphi(X_u)dX_u\right]^j}{(j!)^\nu}} \quad \text{(Eq. 28)}$$

(*i* subscripts here, and the in the remainder of the section, have been dropped for convenience.)

If the distribution is assumed to be standard normal (28) can be computed using standard Gauss-Hermite quadrature. Equation (28) forms the basis for the "true" CMP estimated policy effects.

As discussed above, the ATE is equal to the difference between $Y_{X_{p2}^*}$ and $Y_{X_{p1}^*}$. Let

$E[Y_{X_{p2}^*}|X,\nu]$ equal (27) with $\lambda_2 = \int_{-\infty}^{\infty} \exp(\beta_p + X_o\beta_o + X_u\beta_u)\varphi(X_u)dX_u$ and $E[Y_{X_{p1}^*}|X,\nu]$ equal to (27) with $\lambda_1 = \int_{-\infty}^{\infty} \exp(X_o\beta_o + X_u\beta_u)\varphi(X_u)dX_u$. Then

$$ATE = E[Y_{X_{p2}^*}|X,\nu] - E[Y_{X_{p1}^*}|X,\nu]. \quad \text{(Eq. 29)}$$

The AIE can be expressed identically to the ATE, except in this case

$\lambda_2 = \int_{-\infty}^{\infty} \exp(\beta_p(X_p + \Delta) + X_o\beta_o + X_u\beta_u)\varphi(X_u)dX_u$ and

$\lambda_1 = \int_{-\infty}^{\infty} \exp(\beta_p X_p + X_o\beta_o + X_u\beta_u)\varphi(X_u)dX_u$. Calculation of the AIE using λ_2 and λ_1 remains the same in (Eq. 29).

The average marginal effect is the derivative of 10 with respect to β_p, which is equal to

$$\mathrm{AME} = \frac{\partial \mathrm{E}[\mathrm{Y}|\mathrm{X},\nu]}{\partial \beta_p} = \mathrm{X}_p \frac{\frac{\sum_{j=0}^{\infty} \lambda^j \sum_{j=0}^{\infty} j^2 \lambda^{j-1}}{\sum_{j=0}^{\infty} (j!)^{\nu} \sum_{j=0}^{\infty} (j!)^{\nu}} - \frac{\sum_{j=0}^{\infty} j\lambda^{j-1} \sum_{j=0}^{\infty} j\lambda^{j-1}}{\sum_{j=0}^{\infty} (j!)^{\nu} \sum_{j=0}^{\infty} (j!)^{\nu}}}{\frac{\sum_{j=0}^{\infty} \lambda^j}{\sum_{j=0}^{\infty} (j!)^{\nu}}} \tag{30}$$

where λ is the same as (Eq. 27).

5. Evaluation and Comparison of the Estimators

5.1. Count Models with Endogenous Heterogeneity

In practice, dispersion is treated as a single statistical phenomenon to be fit by the selected specification. Dependence models, rather than continuous mixture models, are the preferred method of fitting overdispersion since they do not require integration, nor do they require specifying a distribution for the dispersion.[5,6] It is possible, however, that variance in the model could arise from both unobserved heterogeneity and other forms of dependence. Individual-level heterogeneity (i.e., stochastic error) should be taken as given in any econometric model. If theory also suggests another form of dependence, then dependence models alone may not provide the best fit for the data. Moreover, if dependence does not follow the same distribution as the heterogeneity, then heterogeneity

[5] Although technically the NB is a continuous mixture model, the closed form makes it functionally similar to a dependence model. Chapter 2 of Winkelmann (2008) describes how the pure negative binomial model (as opposed to the Gamma-Poisson mixture expressed as a negative binomial) arises from a positive linear dependence process. Although NB regression implicitly assumes that the heterogeneity is Gamma distributed, the simulations show that ν is functionally without an assumed distribution.

[6] Of the papers used to guide the simulation study (see Appendix C), virtually every parametric specification involves a negative binomial specification, while only two utilize a different continuous mixture model (log normal). No studies utilize the RGP or CMP models.

models will misspecify the density of the dispersion. In fact, it is possible for heterogeneity to increase the variance of the model, while negative dependence simultaneously decreases the variance of the model, an outcome that will be explored in Section 5.4. As discussed in Section 3, one advantage of dependence distributions specified under Terza's EOM is that the models account separately for both dependence and individual heterogeneity without the need to exactly specify the distribution of the heterogeneity. This added flexibility should give the models more robustness than either mixture or dependence models in isolation, even allowing them to account for processes that pull the variance in opposite directions. We explore this flexibility under various combinations of dependence in Section 5.4, and explore the ability of a heterogeneity model to properly account for multiple sources of dispersion.

5.2. Simulation Background

The estimation of treatment effects for count data under the EOM and EPOF can be executed with many parametric count specifications. The following section characterizes the performance of the CMP, RGP, NB, and Poisson specifications under several distributions of data. To the degree possible, our simulations are guided by the health economics literature. Adapting the literature into a representative simulation scheme was conducted in an organized but unscientific manner. Studies from the health economics literature ranging as far back as 20 years were collected. Those that did not report a mean of the count variable, or that utilized a method other than parametric maximum likelihood (e.g., nonlinear least squares with an exponential mean) were not considered, leaving a total of 44 papers. In general, count models are utilized within the

field to estimate healthcare demand (e.g., physician visits) or substance abuse (e.g., alcoholic drinks per week). A majority of the count data literature is focused on the former, with physician/general practitioner (GP) visits being the primary variable of interest, and specialist visits, ER visits, prescription drug use, and inpatient hospital nights/weeks generally accounting for the rest of the healthcare demand literature.

Mean averages of GP visits typically fell in a range between 1 and 6, while non-GP visits generally had a mean less than 1. Means from the substance abuse literature ranged anywhere between <1 and 99 (excepting a single outlier on each end, all other values were between 4 and 17). (See Appendix C for details.) Since GP visits comprise the bulk of the literature, the values drawn from these studies were selected to guide our simulation design.[7] The data considered were drawn from myriad datasets, and were often separated by gender; some studies pooled multiple years and others only reported an annual mean across multiple years. Rather than attempting to approximate an unweighted "mean of the means," we simply chose a mean that we felt was representative of the data in general. The mean selected is 3.

In order to keep the study focused, we limit the policy effect estimated to the average treatment effect (ATE) rather than the average marginal (AME) or incremental effects (AIE) discussed earlier. Current parametric methods of estimating nonlinear endogenous policy effects are intended to fit binary, rather than continuous (or multi-

[7] Although the substance abuse literature has higher means, we believe it is reasonable to assume that the behavior of the count models considered will not differ much between two means that are both relatively small. Of more interest are the extremely small values of the healthcare demand studies. Such small values are almost certainly created by dual data generating processes: a binary variable of requiring treatment, and a count of demand conditional on seeking treatment. In general, studies did not report conditional means, but we feel safe assuming the mean conditional on requiring specialized treatment is roughly in line with that of the unconditional mean of GP visits.

level discrete) variables. Although the AIE and AME can be computed in the endogenous sample selection case (under the assumption of exogeneity for the policy variables), estimating the ATE in the sample selection case is congruent with the binary policy effect estimated in the endogenous treatment case.

The literature provides much less guidance regarding estimated policy effects. With few exceptions, results were reported as coefficient values rather than marginal or treatment effects. We thus selected three "true" treatment levels for our simulations. Section 5.3 considers performance of the models estimating a "small" treatment effect (10% of the mean), and a "large" treatment effect (100% of the mean), generated using a standard Poisson distribution with log-normally distributed heterogeneity, and no dependence. Ten percent is likely nearing the lower bound of what is economically significant in the case of a binary variable. Although effect sizes greater than 100% of the mean are possible, it is likely that the magnitude of the 100% effect is sufficient to serve as a "large" treatment effect, and the ability of the various estimators to estimate the ATE would likely be similar for all values greater than 100%. Section 5.4 considers data generated with both heterogeneity and dependence. These data sets have a "moderate" treatment effect of approximately 25% of the mean.

The four models are compared according to four criteria. The relative accuracy of the models is determined by computing an absolute percent bias for the coefficients, as well as the expected value of Y and the ATE, where

$$\mathrm{ABP}(\beta) = \frac{1}{k}\sum_{i=1}^{k}\left|\frac{\hat{\beta}_i - \beta}{\beta}\right| \qquad \text{(Eq. 31)}$$

and k is the number of repetitions.

Relative efficiency of estimated treatment effects is compared using the Mean Squared Error (MSE)

$$\mathrm{MSE} = \mathrm{Var}\left(\widehat{\mathrm{ATE}}\right) + \left(\widehat{\mathrm{ATE}} - \mathrm{ATE}\right)^2. \qquad \text{(Eq. 32)}$$

Goodness of fit is determined using the Akaike Information Criterion, a measure that penalizes for additional parameters, and then commonly appears in the literature when comparing count models.[8]

$$\mathrm{AIC} = 2j - 2\ln L, \qquad \text{(Eq. 33)}$$

where j is the number of parameters in the model and lnL is the opmitized value of the log-likelihood function. Although there is no test statistics to determine what a "good" fit is, the AIC provides a measure of relative fit, where smaller values indicate a superior fit of the data.[9]

All of the simulations have several components in common. Recall that X_s is the binary selection/treatment variable, X_p represents the policy variable of interest (which may be endogenous), X_o is the remaining observed data, X_u is an unobserved (scalar) confounder that enters the equations for both Y and X_s, and W^+ is an instrumental

[8] For some examples of AIC in count model selection in health economics, see Deb and Trivedi (1997, 2002), Gerdtham and Trivedi (2001), Liu and Gupta (2011), and Schmitz (2012).

[9] We calculated BIC for each model as well: however, the relative AIC and BIC values between models were virtually identical, and we do not report the BIC values.

variable for the binary selection/treatment equation. The data are generated according to the following:

$X_o \sim U(0.5,1)$, $W^+ \sim U[0,1]$, $X_u \sim N(0,1)$

$X_p = 1(u > 0.45)$ if exogenous (where u is standard uniformly distributed)

$X_p = 1(\alpha_o X_o + \alpha_w W^+ + \alpha_c + X_u > 0)$ if endogenous and

$X_s = 1(\alpha_p X_p + \alpha_o X_o + \alpha_w W^+ + \alpha_c + X_u > 0)$ for sample selection.

In the case of endogenous selection, $[\alpha_p\ \alpha_o\ \alpha_c\ \alpha_W] = [-0.5\ 1\ 1\ 0.5]$, resulting in a roughly 64% probability of selection. Endogenous treatment has participation coefficients $\alpha = [-0.47\ 1\ 0.5]$ resulting in a 55% probability of treatment.

As demonstrated in 1.19, the expected influence of β_u on the expected value of Y has a closed-form solution when X_u is standard normally distributed. β_u was selected so that unobserved heterogeneity served as "multiplier" of roughly 10%: i.e., $\exp\left(\frac{\beta_u^2}{2}\right) = 1.1$. For all simulations, $\beta_u = 0.437$ in the Poisson, RGP, and NB cases, and $\beta_u = 0.437\nu$ in the CMP scenario. Where possible, the constant β_c was selected to account for roughly 1/3 of the effect of observables on the expected value of Y. $\beta_c = 0.334$ in the Poisson, RGP, and NB cases (except for the case of 100% treatment effect, where βc is adjusted to hold Y constant despite the larger β_p value.)[10]

All CMP data is generated based on 4, with the true Z calculated to within a truncation error of $1e^{-5}$. Estimation using the CMP model also computes the true Z to within $1e^{-5}$ of the "true" value, and post-estimation computation of the expectation of Y

[10] Data generated according to a CMP process do not strictly follow this outline, since the model does not have a reliable closed-form solution to calculate the coefficient values. The values assigned correspond as closely as possible to the plan of assignment discussed above.

and the ATE are calculated using 10 and 28 respectively. (For comparison, estimations of Y and the ATE computed from 8 are reported in Appendix B.)

Each simulation is run 500 times with $n = 5000$. All simulations are executed using Gauss-Legendre quadrature with ten points of support. Since the focus of the analysis is on the performance of the count specification with regard to the β coefficients and ATE, we do not report the FIML estimations of the α parameters or the predicted probability of selection/treatment.

5.3. Simulation 1—Estimating "Small" and "Large" Treatment Effects

The first set of simulations compares the effectiveness of the models under endogenous sample selection or endogenous treatment with both small (10%) and large (100%) binary treatment effects (TE). The data are generated according to a Poisson distribution with endogenous heterogeneity in the second stage equation as described in 12. Although the standard Poisson is the true data generating process, the CMP, RGP, and NB all nest the Poisson, and therefore should not be at a disadvantage.[11] Estimated values are presented in Tables 1-4, with absolute percent bias in parentheses.

The CMP produces the best estimate in the small TE case. Both the CMP and Poisson have virtually identical bias and MSE values, but the CMP produces the most accurate point estimate of the four models. Despite this, the log likelihood values generated by the CMP estimation do not reject the null hypothesis of a standard Poisson

[11] The Poisson being described here is not technically a pure Poisson process, but a linear function of a log-normal mixture of a Poisson specification (whereas the other three models are linear functions of a log-normal mixture combined with a dependence specification). This Poisson mixture is technically a model of overdispersion and is not subject to the same assumptions as a pure Poisson. Therefore the standard errors used to calculate the MSE are not subject to the corrections necessary for a pure Poisson in the absence of equidispersion.

specification according to the likelihood ratio test. The NB and RGP models both reject the standard Poisson model in favor of a dispersion model. The RGP edges out the NB according to the information criterion, but the NB model has a slightly lower MSE. Both distributions pay a small penalty in bias and MSE relative to the CMP and Poisson specifications.

Table 1

10% Treatment Effect with Endogenous Treatment

	β_p = .111	β_o = .720	β_c = .334	β_u = .437	Y = 3.00	ATE = .330	MSE	AIC	lnL
Poisson	.105 (19.07)	.718 (1.30)	.340 (5.27)	.442 (2.70)	2.99 (1.02)	.314 (19.01)	.062	23066.84	11529.42
NB	.119 (22.17)	.719 (1.37)	.335 (5.31)	.429 (3.74)	2.97 (1.36)	.353 (21.86)	.079	23027.99	11509.00
RGP	.121 (21.79)	.720 (1.27)	.334 (5.14)	.429 (3.37)	2.99 (1.00)	.362 (21.82)	.082	23002.19	11496.10
CMP	.112 (17.80)	.701 (3.55)	.312 (11.36)	.421 (5.75)	2.89 (4.00)	.334 (18.82)	.065	23071.87	11530.94

In the large treatment effect case all four models perform substantially better, although the Poisson provides the best fit according to every metric. None of the three dispersion models reject the null hypothesis that the standard Poisson is the best specification, and the Poisson also has the least bias and most accurate point estimate. The CMP performs slightly better than the NB and RGP in terms of bias and MSE, and has similar IC values to the NB. The RGP, which had the best fit according to the AIC in the small TE case now has the worst.

In both sample selection models, the Poisson has the worst fit statistics, and is rejected in favor of all three dispersion models by a likelihood ratio test. However, the Poisson also has the lowest MSE in both cases. Although it has the lowest bias of the ATE estimate in the small TE case, it has the highest bias of the ATE estimate in the large TE case. The CMP performs well in terms of bias, but has a higher MSE than the other three specifications.

Table 2

100% Treatment Effect with Endogenous Treatment

	β_p = 1.331	β_o =.755	β_c =-.621	β_u = .437	Y = 3.00	ATE = 3.15	MSE	AIC	lnL
Poisson	1.33 (1.82)	.754 (1.24)	-.615 (3.58)	.440 (3.03)	2.68 (10.83)	3.15 (1.80)	5.103	21526.67	10759.34
NB	1.351 (2.41)	.757 (1.32)	-.625 (3.62)	.419 (4.91)	2.66 (11.47)	3.19 (2.33)	5.254	21701.92	10845.96
RGP	1.351 (2.29)	.757 (1.25)	-.625 (3.56)	.418 (4.89)	2.69 (10.71)	3.21 (2.55)	5.331	21724.52	10857.26
CMP	1.290 (3.77)	7.25 (4.53)	-.624 (3.55)	.411 (7.10)	2.63 (12.70)	3.16 (1.90)	5.223	21705.50	10847.75

Table 3

10% Treatment Effect with Sample Selection

	β_p = .111	β_o = .720	β_c = .334	β_u = .437	Y = 3.00	ATE = .330	MSE	AIC	lnL
Poisson	.109 (18.34)	.719 (1.84)	.365 (10.59)	.390 (10.74)	3.04 (1.45)	.329 (18.35)	.065	17963.88	8977.88
NB	.117 (18.02)	.703 (2.53)	.427 (27.63)	.358 (19.18)	3.09 (3.07)	.358 (19.18)	.073	17901.78	8945.89
RGP	.117 (18.08)	.703 (2.54)	.428 (28.03)	.300 (31.21)	3.09 (3.15)	.359 (19.33)	.074	17901.46	8945.73
CMP	.121 (20.40)	.756 (5.15)	.280 (16.55)	.349 (20.07)	2.89 (3.66)	.348 (18.52)	.076	17923.16	8956.58

In all four models the CMP tends to do the worst job of estimating coefficient values. This is to be expected since the CMP coefficient estimates must be rescaled as $\hat{\beta}/\hat{\nu}$ in order to make a comparison to models with an exponential conditional mean. The rescale is an approximation involving two estimated values, and is therefore subject to greater bias than the other models. The CMP also performs the worst in estimating the true value of Y in three out of the four cases. The NB, RGP, and Poisson produce very similar coefficient estimates in the endogenous treatment models, and similar Y estimates in all four cases. The NB and RGP models do both fair slightly worse estimating the constant β_c and the heterogeneity term β_u in the sample selection case. However, this does not affect the estimates of the coefficient of interest β_p or the estimated ATE, which are fairly similar to the CMP and Poisson in the sample selection case.

Table 4

100% Treatment Effect with Sample Selection

	β_p = 1.218	β_o = -.527	β_c = .334	β_u = .437	Y = 3.12	ATE = 3.22	MSE	AIC	lnL
Poisson	1.220 (2.04)	-.523 (2.50)	.330 (7.40)	.440 (3.36)	3.07 (2.73)	3.17 (3.36)	2.383	15902.92	7947.46
NB	1.225 (2.06)	-.531 (2.48)	.355 (10.05)	.406 (7.78)	3.17 (2.15)	3.28 (2.97)	2.930	15693.86	7851.93
RGP	1.224 (2.04)	-.529 (2.20)	.349 (8.39)	.415 (5.71)	3.16 (1.76)	3.26 (2.57)	2.877	15695.90	7842.95
CMP	1.257 (3.63)	-.541 (3.46)	.300 (11.86)	.423 (4.39)	3.10 (1.51)	3.27 (2.53)	3.013	15695.91	7842.96

None of the first set of simulations provides a clear recommendation for any of the four specifications. However, all four specifications nest the true data generating

process, and none of the simulations introduces dependence. Section 5.4 explores the performance of the specifications when they do not correspond exactly to the true data generating process, and when both heterogeneity and dependence cause dispersion. The robustness of the various specifications to incorrect assumptions will provide a better demonstration of which model is preferable when the true distribution of the data is unknown.

5.4. Simulations 2 & 3—Positive and Negative Dependence

The second and third set of simulations compare the performance of the four models under conditions of positive and negative dependence in both the endogenous sample selection and endogenous treatment cases. Data with positive dependence are generated using all three dependence models, while data with negative dependence are generated using the CMP specification. All four count models are used to produce estimates from each of the three data configurations with positive dependence, in order to determine which model performs best when it is not the true data generating distribution. The NB is not considered in analyzing data with negative dependence since it collapses to the Poisson in this instance.

The ATE is selected to be a "moderate" effect size of approximately 25% of the mean, and the mean remains at 3. The dispersion is selected so that the conditional variance is approximately 7.3. We target this variance with all three data generating processes in order to keep the comparison data sets as similar as possible. Results are presented in Tables 5-10.

Table 5

CMP Generated Data with Endogenous Treatment and Positive Dependence ($\nu = 0.281$)[12]

	β_p = .141	β_o = -.005	β_c = .042	β_u = .123	Y = 3.13	ATE = .917	MSE	AIC	lnL
Poisson	-.097 (168.95)	-.025 (457.71)	.326 (677.34)	.202 (65.44)	3.80 (21.44)	-1.460 (259.21)	5.741	27906.29	13948.15
NB	.094 (33.14)	-.002 (76.67)	.246 (486.87)	.065 (47.06)	3.36 (8.15)	1.187 (29.95)	.116	27207.99	13599.00
RGP	.097 (31.03)	-.002 (81.70)	.248 (491.27)	.065 (47.27)	3.17 (3.91)	1.09 (20.63)	.128	26742.13	13366.07
CMP	.141 (11.05)	-.005 (103.79)	.039 (53.43)	.122 (9.89)	3.05 (2.33)	.921 (10.95)	.024	26435.09	13212.55

Table 6

NB Generated Data with Endogenous Treatment and Positive Dependence ($\nu = 25.00$)

	β_p = .281	β_o = .629	β_c = .334	β_u = .437	Y = 3.00	ATE = .825	MSE	AIC	lnL
Poisson	.214 (23.71)	.620 (1.92)	.365 (9.55)	.487 (11.63)	2.99 (1.14)	.637 (22.88)	.198	23676.38	11833.19
NB	.285 (9.47)	.629 (1.57)	.336 (5.70)	.432 (3.68)	2.97 (1.31)	.842 (9.36)	.285	23656.49	11823.25
RGP	.285 (9.49)	.629 (1.57)	.337 (5.70)	.433 (3.68)	2.97 (1.31)	.841 (9.38)	.285	23656.40	11823.20
CMP	.281 (7.48)	.632 (3.36)	.299 (12.17)	.466 (6.82)	2.53 (15.65)	.661 (19.95)	.092	23008.46	11499.23

One takeaway from all of the simulations with positive dependence is that in the case of endogenous treatment the standard Poisson with heterogeneity performs poorly when there is additional dependence. The Poisson has the worst bias in estimating

[12] Estimates of CMP coefficients produced by the NB, RGP, and Poisson models are multiplied by 0.281 post-estimation in order to rescale them for comparison with the CMP. This cannot be done in reverse since the "true" dispersion parameter from the CMP is unknown if the data are not generated by a CMP process.

treatment effects, even predicting a large effect in the wrong direction in the case of the CMP. In all three cases the Poisson has the worst fit statistics, and the LR test correctly rejects it in favor of any of the three dependence models. The one advantage the Poisson possesses is that it has the lowest variance of the estimates. However, this cannot overcome its primary weakness of large and potentially catastrophic bias.

Table 7

RGP Generated Data with Endogenous Treatment and Positive Dependence ($\nu = .025$)

	β_p = .281	β_o = .629	β_c = .334	β_u = .437	Y = 3.00	ATE = .826	MSE	AIC	lnL
Poisson	.193 (31.36)	.617 (2.17)	.374 (11.93)	.503 (15.15)	2.99 (1.11)	.574 (30.46)	.198	23764.23	11877.12
NB	.282 (9.75)	.628 (1.55)	.338 (5.91)	.434 (3.74)	2.97 (1.32)	.832 (9.58)	.280	23734.25	11862.13
RGP	.282 (9.84)	.628 (1.55)	.338 (5.93)	.437 (3.77)	2.97 (1.31)	.832 (9.67)	.280	23734.06	11862.03
CMP	.267 (10.64)	.654 (4.15)	.262 (21.93)	.473 (8.37)	2.97 (1.30)	.758 (11.31)	.249	23746.73	11868.37

Table 8

CMP Generated Data with Endogenous Sample Selection and Positive Dependence ($\nu = 0.281$)

	β_p = .141	β_o = -.005	β_c = .042	β_u = .123	Y = 3.13	ATE = .917	MSE	AIC	lnL
Poisson	.075 (46.67)	.029 (745.99)	.153 (264.42)	.193 (57.07)	2.70 (13.59)	.709 (22.93)	.066	18931.99	9461.00
NB	.090 (35.97)	-.005 (91.27)	.261 (523.59)	.057 (53.45)	3.10 (1.83)	.971 (9.59)	.023	18545.38	9267.69
RGP	.091 (35.63)	-.006 (94.15)	.262 (526.40)	.054 (55.73)	3.11 (1.86)	.979 (10.21)	.025	18544.00	9267.00
CMP	.140 (8.39)	-.004 (138.29)	.037 (46.80)	.113 (14.21)	3.05 (2.70)	.925 (8.05)	.015	18529.36	9259.68

Table 9

NB Generated Data with Endogenous Sample Selection and Positive Dependence ($\nu = 25.00$)

	β_p = .281	β_o = .629	β_c = .334	β_u = .437	Y = 3.00	ATE = .826	MSE	AIC	lnL
Poisson	.277 (7.86)	.635 (2.41)	.335 (7.54)	.429 (3.97)	3.01 (1.14)	.817 (7.78)	.265	18214.33	9102.65
NB	.286 (7.67)	.616 (2.63)	.414 (23.94)	.322 (26.22)	3.09 (2.93)	.865 (8.67)	.282	18121.29	9055.65
RGP	.287 (7.73)	.616 (2.66)	.417 (24.74)	.317 (27.30)	3.09 (3.06)	.868 (8.86)	.284	18120.47	9055.24
CMP	.308 (11.55)	.805 (8.60)	.199 (40.55)	.376 (13.95)	3.49 (14.06)	.964 (17.22)	.557	18651.22	9320.61

Table 10

RGP Generated Data with Endogenous Sample Selection and Positive Dependence ($\nu = 0.025$)

	β_p = .281	β_o = .629	β_c = .334	β_u = .437	Y = 3.00	ATE = .826	MSE	AIC	lnL
Poisson	.275 (8.95)	.641 (2.93)	.321 (8.09)	.442 (4.03)	3.02 (1.10)	.814 (8.85)	.271	18368.79	9179.40
NB	.286 (7.77)	.617 (2.59)	.411 (23.02)	.327 (25.11)	3.09 (2.85)	.862 (8.66)	.281	18175.59	9082.80
RGP	.287 (7.85)	.617 (2.60)	.414 (23.90)	.322 (.263)	3.09 (2.98)	.866 (8.88)	.284	18174.89	9082.45
CMP	.310 (12.09)	.814 (9.80)	.172 (48.46)	.387 (11.40)	3.50 (13.95)	.959 (16.82)	.560	18717.12	9353.56

The CMP performs well in all three cases. The CMP has the best fit statistics when it is the correctly specified distribution, and also when the data follow an NB distribution. In the case of RGP data, the CMP has fit statistics comparable to the true RGP model. The CMP also has a lower variance of the estimator than the RGP and NB models, resulting in a smaller MSE even when the data follow a NB or RGP

specification. Although the CMP has somewhat high bias with NB distributed data, the bias of the NB model is even worse with CMP data. The CMP also performs better on RGP data than the RGP performs with CMP data, although when both models are pitted against NB data, the RGP has the lowest bias of the two, landing in a virtual tie with the true NB model. As in the case of simulations 1-4, the CMP does a relatively poor job fitting the coefficients compared to the NB and RGP models, although the Poisson fares the worst in this regard.

Although the CMP is arguably the superior model in the case of endogenous treatment, this advantage does not hold in the case of sample selection. The CMP performs best across the board when it is the true data generating process, but the bias of the treatment effect estimates, and the mean squared error are the worst of all four models when the data are NB or RGP distributed. Moreover, the CMP does substantially worse fitting the data than any of the other three models. Contrary to the endogenous treatment case, the NB and RGP models perform fairly well (and nearly identically) when they do not reflect the true data generating process.

The Poisson performs reasonably well in the CMP case, but is still the worst by every metric, indicating that it may have trouble handling dependence when it re-enters the conditional mean as it does in CMP distributed data. The Poisson fares better with RGP and NB data, actually having the least bias and smallest MSE in the NB case, and fitting the data better than even the CMP in both cases. Overall, the NB model edges the RGP slightly in fit and bias over the three models, although the two models are virtually interchangeable.

In case of negative dependence the mean remains at 3 and the ATE remains at 25% of the mean. However, the presence of the heterogeneity presents a problem, since it overdisperses the data relative to a basic specification. In the case of the CMP, since the dependence parameter enters the conditional mean (including the heterogeneity term), it takes an extreme level of negative dependence (approaching a binary outcome) to make the conditional variance less than the conditional mean. Rather, the model presented for simulation represents a case where individual heterogeneity increases the variance of the model, but negative dependence decreases the variance, leaving the model with net overdispersion.

One interesting finding of the present study is the failure of the RGP model in the case of negative dependence with overdispersion induced by individual heterogeneity. The restrictions on the RGP force the dispersion parameter ν to be such that $\nu > \min(-1/\max(\lambda),-1/\max(Y))$ when the dependence is negative. This is typically not a problem, but the mix of net overdispersion with negative dependence results in such large values of Y and/or λ that ν is constrained to be virtually 0. Although in theory this should collapse the model to the Poisson, it rather prevents the model from achieving concavity and it is unable to converge consistently in this case. As such, we do not consider the RGP for comparison with the CMP or Poisson models in the case of negative dependence.

In the ET case, the Poisson performed fairly poorly in terms of bias of the ATE estimate and the MSE. As expected the CMP performed much better, but still had a fairly large bias and MSE in comparison to data with positive dependence. Both models

performed better in the sample selection case, and the Poisson was actually more accurate in estimating the coefficient values. However, the CMP still fit the data better, rejecting the null hypothesis of a Poisson specification, in addition to having both lower bias and MSE of the ATE. (See Tables 11 and 12.)

Table 11

CMP Generated Data with Endogenous Treatment and Negative Dependence ($\nu = 4.50$)

	β_p = 1.023	β_o = 2.506	β_c = 2.540	β_u = 1.965	Y = 3.03	ATE = .766	MSE	AIC	lnL
Poisson	2.323 (127.10)	3.000 (22.47)	0.822 (67.63)	1.301 (33.79)	2.89 (2.98)	1.463 (93.54)	1.384	21250.64	10620.32
CMP	0.942 (8.18)	1.84 (25.02)	1.700 (33.07)	1.343 (31.63)	2.90 (2.57)	.939 (24.21)	.292	20398.54	10194.27

Table 12

CMP Generated Data with Endogenous Sample Selection and Negative Dependence ($\nu = 4.50$)

	β_p = 1.023	β_o = 2.506	β_c = 2.540	β_u = 1.965	Y = 3.03	ATE = .766	MSE	AIC	lnL
Poisson	1.213 (18.82)	2.599 (6.06)	2.224 (12.43)	1.184 (39.75)	3.12 (4.78)	.825 (10.18)	.228	16689.19	8339.60
CMP	.466 (54.40)	1.050 (57.17)	1.103 (56.57)	.634 (67.75)	3.03 (1.98)	.793 (7.60)	.184	16484.67	8237.34

6. Discussion

The results of the simulations did not overwhelmingly favor one model over the others. All four models performed comparably well at estimating both “large” and “small” treatment effects when the data follow a Poisson distribution. This result is not surprising given that all three of the dispersion models nest the standard Poisson.

The strongest conclusion to be drawn from the simulations is that the Poisson specification does not fare well when there is both unobserved heterogeneity and occurrence dependence (i.e., the Poisson is not robust to misspecification). The three dependence models perform substantially better in terms of fit and bias of the treatment effect in virtually every case. Although the Poisson does have a low variance of the estimate, and generally does a comparable job estimating coefficient values, it is also the only specification to fail catastrophically and predict a treatment effect that was both large and in the wrong direction.

The simulations show that in the case of endogenous treatment, the CMP is the most robust to being the incorrect specification in terms of fit of the data and accuracy of the estimated ATE. However, the CMP performs worse than the other two dispersion models in the sample selection case by virtually every metric. Additionally, the CMP produces the worst coefficient estimates in nearly every model, in both the ET and SS case, and the approximation that re-scales the CMP to be comparable to the other models becomes less reliable the more dispersed the data become.

These results suggest that the CMP is the preferred specification if the data suffer from endogenous treatment, and the researcher is primarily interested in estimating treatment effects. Under positive occurrence dependence, either the NB or RGP models are appropriate when the data involve endogenous sample selection, or when the researcher is primarily interested in coefficient estimates. However, if there is suspected negative dependence, both the NB and RGP models will be inappropriate, and the CMP is preferred no matter what the objective.

The primary shortcoming of the CMP model is the failure of the approximation for Z during integration. Data that deviate significantly from a standard Poisson distribution result in unreliable estimates when the approximation is used. This requires that Z be computed directly to within an acceptable truncation error, resulting in a massive penalty in computing time. Running on Stata 12 MP with two processors, the Poisson, NB, and RGP endogenous treatment models each took less than ten seconds per repetition. The CMP took roughly three minutes. In the sample selection case, the Poisson, NB, and RGP models still took under ten seconds, whereas the time necessary for the CMP to execute rose to nearly 35 minutes. These discrepancies arose with simulated data, and it's likely that real data would exacerbate the differences in speed between the two models. Despite this issue, the simulation results suggest that the CMP is a preferable model when the research goal is estimation of an endogenous treatment effect, or estimation of an effect with both unobserved heterogeneity and negative dependence.

There are several potential approaches for increasing the scope of this study that we elected not to pursue. In the case of endogeneity it may be helpful to test the response of the model to various levels of selection or treatment. While characterizing the performance of the EOM under several proportions of selection/treatment may be beneficial, the relative performance of the various count specifications should remain constant under different conditions of endogeneity. A more relevant extension would be to characterize the performance of the count models under a different parametric framework of endogeneity. For instance, a growing literature has begun to approach

endogenous count models using a Copula specification, whereby the errors of X_s and Y are modeled as marginal distributions linked by a dependence parameter, rather than belonging to a joint distribution (Van Ophem, 2000; Zimmer & Trivedi, 2006). This approach is advantageous in that it requires limited assumptions regarding the relationship between the errors of Y and X_s. Moreover, Copulas do not force variance-increasing heterogeneity on the model in order to account for endogeneity: omitting heterogeneity from the model may allow better comparisons of the specifications, particularly the CMP and RGP.

However, Copula estimation of variables with discrete margins may encounter problems in convergence. Methods to transform discrete marginals into a continuous distribution are available, but will introduce a small amount of additional bias to the model (Trivedi & Zimmer, 2005.) Additionally, if there is reason to believe that there is both unobserved heterogeneity and latent dispersion in the model, the Copula approach will not provide a computational benefit as an integration of the PMF will still be required. Regardless, the merits of the Copula framework relative to the EOM remain a topic for future review.

CHAPTER II

ESSAY 2: ESTIMATING THE RELATIONSHIP BETWEEN MARIJUANA USE DISORDER AND EMERGENCY ROOM UTILIZATION AMONG MEDICAID RECIPIENTS

1. Abstract

The study uses data from the National Survey of Drug Use and Health (NSDUH) across seven years (spanning 2005–2011) to investigate the relationship between marijuana use disorder and total emergency room (ER) visits in the past year among adult (age 18+) Medicaid recipients. No economic study of demand for medical care has ever looked at marijuana use separate from other drugs to disentangle the effect of marijuana versus so-called "hard" drugs: this distinction is of growing policy-relevance as the national debate grows over the future of marijuana regulation, and states begin to consider the ramifications of relaxing current restrictions. Consistent with previous literature I find a positive relationship between hard drug use disorder and ER visits. However, I fail to find a positive relationship between marijuana use disorder and ER visits, conditional on hard drug use disorder Using a two-stage nonlinear least squares approach, I estimate an average treatment effect ranging from roughly -0.62 to -1.23 visits among females, and from -0.61 to -1.03 visits among males. Although these estimates are generally not significantly different from zero, the preferred specification rejects an ATE greater than 0.24 for females and 0.23 for males at the 95% significance level, suggesting that if a positive relationship exists, it is fairly small.

2. Introduction

Marijuana is a topic of growing national interest: 20 states and the District of Columbia have legalized medical marijuana, while seventeen have decriminalized marijuana, and both Colorado and Washington recently legalized the limited sale and distribution of marijuana (Office of National Drug Control Policy, 2013). As debate continues regarding the future of marijuana policy, it is important for policymakers to understand the potential ramifications of relaxing marijuana regulations. In this study I use data from the National Survey of Drug Use and Health (NSDUH) to estimate the relationship between marijuana use disorder and total annual ER visits among Medicaid recipients. Marijuana use disorder (MUD) refers to either dependence on or abuse of marijuana, as defined by the Diagnostic and Statistical Manual of Mental Disorders (Appendix D). Given the heterogeneity of consumption among marijuana users, MUD provides a clinically significant definition of use that reflects severe levels of consumption. While previous studies have investigated the effect of marijuana consumption on education or employment (e.g., Bray, Zarkin, Ringwalt, & Qi, 2000; McCaffrey, Pacula, Han, & Ellikson, 2010; Yamada, Kendix, & Yamada, 1996; Popovici & French, 2014) no study has investigated the effect of marijuana consumption on acute healthcare utilization. Estimates of the relationship between MUD and ER visits will be of use to state and federal policymakers considering the potential costs of relaxed marijuana regulation.

Although marijuana is typically considered less dangerous than so-called hard drugs (e.g., cocaine, heroin, methamphetamines, prescription narcotics), its consumption

is correlated with numerous health issues such as respiratory problems, liver disease, cardiovascular disease, sexually transmitted disease, and increased mortality among AIDS patients (Brook, Stimmel, & Brook, 2008; Gordon, Conley, & Gordon, 2013; Mittleman, Lewis, Maclure, Sherwood, & Mueller, 2001; Schuster, Crane, & Gonzalez, 2012; Sidney, Beck, Tekawa, Quesenberry, & Friedman, 1997; Smith & Crespo, 2001). Marijuana consumption may also trigger or exacerbate schizophrenic episodes (Rey & Tennent, 2002), and daily smokers of marijuana are more likely than non-users to visit the physician for respiratory problems or injury (Polen, Sidney, Tekawa, Sadler, & Friedman, 1993).

While several previous studies have investigated the potential causal impact of drug use on demand for acute medical care no study has specifically tried to isolate the effect of marijuana use. This has left a need for marijuana-specific estimates that may inform the current policy debate regarding the relaxation of marijuana regulations at the state and federal level. Moreover, no previous study has considered the effect of drug use (marijuana or otherwise) among Medicaid recipients. Care covered by Medicaid derives from Federal and state discretionary budgets. With the Affordable Care Act expanding Medicaid coverage to include all individuals under 65 earning less than 133% of the federal poverty limit beginning in 2014 (among participating states), information regarding demand for health services among Medicaid recipients may be especially

valuable to policymakers, particularly those from states considering Medicaid expansion.[13]

This study is also the first in the literature to utilize an instrumental variables count data model to estimate the relationship between marijuana dependence and ER visits, making estimates robust to both time-invariant and time-variant endogeneity. This study is therefore the most econometrically rigorous study of the relationship between drug use and ER visits to date, and demonstrates an approach that may be used to estimate other healthcare outcomes among drug users.

3. Prior Research

Several individual-level economic studies have investigated the direct effect of marijuana use on physical or mental health. Williams and Skeels (2006) estimate the joint impact of tobacco and marijuana use on self-reported health in Australia and find that weekly use of marijuana is associated with a 6-18 percentage point reduction in the probability of reporting excellent or very good health, conditional on smoking status. Marijuana users were also 22 percentage points more likely to have seen a doctor in the past twelve months. A similar study by van Ours and Williams (2012) finds that marijuana use among citizens of Amsterdam is correlated with diminished physical functioning among males and diminished mental functioning among both males and females.

[13] Estimates from current Medicaid recipients are not perfectly generalizable to individuals who are not currently covered by Medicaid. However, since the influx of new beneficiaries will disproportionately consist of males and individuals without children, it is likely that estimates in this study will serve as a lower bound on the relationship between drugs and healthcare use.

Several previous studies have attempted to link the health effects of drug use to demand for medical care. French, McGeary, Chitwood, and McCoy (2000) study the use of outpatient services, ER visits, and total hospital admissions among residents of Dade County Florida who use illicit drugs weekly or more. They find that use of illicit drugs is significantly associated with fewer outpatient visits, but increased ER and hospital admissions. However, the authors assume exogeneity of drug use, which may result in biased estimates if drug use is endogenous to health care utilization. McGeary and French (2000) use the National Household Survey on Drug Abuse (NHSDA—predecessor to the NSDUH) to investigate the relationship between weekly drug use and the probability of any ER services used in the last year, using a bivariate probit with instrumental variables to account for the potential endogeneity of drug use. The authors find that the probability of ER utilization is significantly higher among both male and female drug users. They also find that the error terms in the drug use and acute medical care equations are significantly negatively correlated, and results of a Hausman test reject the null hypothesis that drug use is exogenous to acute medical care. French, Fang, and Balsa (2011) utilize a fixed effects regression in lieu of an instrumental variables approach to investigate the effect of "casual" (less than sample median) or "heavy" (greater than sample median) drug use frequency on the probability of any ER visit or hospital admission, and counts of ER visits and hospital admissions. Heavy drug use was correlated with a significantly higher probability of any ER use among both males and females, but only hospital stays for women. Both heavy drug using males and females were admitted to the hospital more frequently, but drug use intensity had no

effect on the number of ER visits. Drug use of any intensity was also significantly correlated with increased probability of and number of injuries among both males and females. However, these studies fail to disentangle the effect of marijuana from other hard drugs, and the implications for states considering new marijuana policy is unclear.

4. Conceptual Model

Acute medical care is one service of many on which people are able to spend their time or money, and they will select a utility maximizing number of ER visits, subject to a budget constraint. I assume that ER visits do not provide any utility in and of themselves, but rather, going to the ER restores utility by alleviating or eliminating disutility that is caused by an injury or illness. To consume an ER visit, individuals incur a cost (e.g., time and money), and so only visit the ER when the utility restored by doing so exceeds the utility that could be purchased by using the same resources (e.g., time and money) on other goods or services.[14] Therefore, demand for ER visits depends on the amount of utility that would be gained by going to the ER and the amount of utility that would be foregone by going.

Consistent with French et al. (2011), I assume that individuals have a latent probability of negative health shocks determined by $H(X_o, ,M(X_o,X_u),D(X_o,X_u), X_u , \mu)$, where X_o is a vector of observable individual demographic and socioeconomic characteristics, M is a binary variable representing MUD, D is set of binary variables representing alcohol, nicotine, and/or hard drugs use disorders, X_u represents unobserved heterogeneity that may be jointly correlated with MUD and the probability of a negative

[14] The monetary cost of visiting the ER is capped at $3.90 for Medicaid recipients, so time costs and "psychic" costs of going to the ER will be the primary costs associated with ER visits.

health shock (e.g., future discount rate, health knowledge), and μ is a stochastic error term capturing factors that may affect health but are assumed to be exogenous to substance use disorders.

The probability of negative health shocks may be increased by MUD through illness (e.g., respiratory issues among marijuana smokers, or reduced time and money for investment in health stock), or through an increased probability of injury (e.g., driving while under the influence). On the other hand, the probability of health events may be decreased if, for example, individuals with MUD are less likely to participate in activities that can result in injury (e.g., less time driving, less time pursuing outdoor activities). If those with MUD are more likely to suffer adverse health events, then the expected benefit of going to the ER should be higher than among non-dependents, and *ceteris paribus* marijuana dependents should visit the ER more often. Conversely, if those with MUD are less likely to suffer adverse health events then their expected benefit of going to the ER should be lower, and we should see fewer visits to the ER.

While the demand for ER visits (ER) will partly (or even primarily) be determined by H, other inputs such as M, D, factors contained in X_o (e.g., relative income) or X_u (e.g., future discount rate) will also influence demand by determining the tradeoff in utility between the consumption of ER visits and the consumption of other goods (including marijuana), conditional on H.[15] The full demand equation for acute medical care may thus be expressed as

[15] Since data are not available on the price of acute medical care or marijuana, these factors are also contained in X_u.

$$ER = f(H(X_o, X_u, M, D, \mu), X_o, X_u, M, D). \quad \text{(Eq. 1)}$$

Since H is unobserved, the goal of the empirical model is to estimate the reduced-form equation

$$ER = f(X_o, X_u, M , D). \quad \text{(Eq. 2)}$$

The effect of MUD is considered independently of other illicit drugs for several reasons. First, marijuana is the most commonly used illicit drug in the United States: in 2012 roughly 7% of Americans twelve and older had used marijuana at least once in the past month (National Institute on Drug Abuse, 2014). Although the risk of dependence is estimated to be less than for other substances such as nicotine, alcohol, or other drugs, roughly one in ten individuals who ever try marijuana will at some point develop dependence: a risk that rises to one in two among daily users (Copeland & Swift, 2009). Second, marijuana is policy-relevant in isolation given the current political climate where the need to assess the potential costs of expanded marijuana use, and the subsequent potential for increased levels of marijuana use disorder, is a pressing issue at the state and federal level. Finally, there is reason to believe that the effect of marijuana on demand for acute medical care may differ to that of other illicit drugs. In general, marijuana is substantially cheaper in monetary terms than many other illicit drugs (Fries, Anthony, Cseko, Gaither, & Sculman, 2008) and due to the relative prevalence of marijuana is also likely cheaper in time costs to obtain. Moreover, consumption methods for marijuana are generally less prone to directly cause disease (e.g., infection), or to result in direct health

shocks from accidental overdose or substance adulteration (Ashton, 2001; British Medical Association, 2013).

One remaining identification issue is that X_u is not observed. Since M is a function of X_u, failure to condition on these unobservable factors will produce estimates that are confounded by the joint relationship of Y and M to X_u. To account for this, I utilize a nonlinear instrumental variables method introduced by Terza (1998, 2009). By assuming that X_u is continuous and standard normally distributed conditional on observable covariates (and with at least one valid instrument that is correlated with M and independent of Y conditional on X_o), I am able to condition on X_u, which renders M exogenous to Y. This approach will be addressed further in the empirical section.

5. Data and Variables

5.1. NSDUH

The National Survey on Drug Use and Health (NSDUH) is a nationally representative sample of non-institutionalized Americans age twelve and older, funded by the Substance Abuse and Mental Health Services Administration. The survey is a repeated cross-section conducted annually and includes approximately 70,000 observations per year, of which roughly 55,000 are available in the public use data set. The observations not in the public use file are dropped at random to help ensure that responses are not identifiable, and the sampling weights are adjusted to account for their deletion. Therefore, the available observations should retain their representativeness, and the randomness of the process should alleviate concern about the possibility of endogeneity being introduced into the sampling procedure.

The NSDUH is uniquely suited to the present study since it is administered with the primary intent of obtaining information regarding substance use, including marijuana and other illicit drugs, as well as licit substances like alcohol and nicotine. In addition to information regarding annual ER visits, the data contain a rich collection of demographic and socioeconomic characteristics. The data also provide information regarding respondent health, as well as personality traits that may be correlated with substance use or healthcare demand such as preferences for risk-taking or dangerous behavior.

My analysis sample is limited to Medicaid recipients between 18 and 64 years of age. From an initial sample of 30,893, roughly 6% of observations (1,575) report an annual family income of over $75,000. There are several plausible explanations for receipt of Medicaid with such high income. Financial eligibility for Medicaid is adjusted according to the number of children an individual has. If a respondent has many children, then $75,000 may not exceed the threshold to surrender eligibility. Another possibility is that an individual who qualifies for supplemental security income (SSI) may qualify automatically for Medicaid. The family income may exceed $75,000 if one or more non-disabled family member(s) works. However, in the data individuals reporting an income over $75,000 do not have more children, on average, nor are they more likely to report disability. I therefore assume that individuals reporting such a high income are either incorrect about their income, or incorrect about qualifying for Medicaid, and anyone reporting over $75,000 in annual family income is dropped from the sample.

From the remaining sample of 29,318 respondents, I drop 1,392 observations due to missing responses.[16] This leaves a final sample of 27,841 (19,766 females and 8,075 males). Sampling weights are provided with the data to ensure national representativeness and also to account for survey non-response that is potentially endogenous. Unweighted data may also be heteroskedastic, in which case the sampling weights will improve the precision of the estimates (Cameron & Trivedi, 2005). Therefore, all regressions are weighted to account for both the sampling design, and the possibility of endogenous sampling. It is also common in the economic literature to split the analysis between males and females when drug use is an independent variable of interest (French, Roebuck, & Alexandre, 2001; MacDonald & Pudney, 2000; Popovici & French, 2014) particularly when the dependent variable relates to health or healthcare (French et al., 2011; McGeary & French, 2000; Van Ours & Williams, 2012). This approach has also recently been used when considering Medicaid recipients or potential Medicaid recipients (Brown & Finkelstein, 2008; Hamersma & Kim, 2013) and so I adopt this approach moving forward.

5.2. Variables and Summary Statistics

The variables used in this analysis may be divided into five broad groups: the dependent variable, independent variable of interest, instruments, and controls. Table 13 describes the distribution of ER visits, while Table 14 provides weighted summary statistics for all five groups of variables.

[16] This includes 849 missing responses for annual ER visits, 107 missing responses for the instrumental variable, and 436 missing responses for other independent variables. Observations dropped due to missing data are assumed to be missing at random.

Table 13

Weighted Sample Means by Gender

	Female	Male
Dependent Variable		
ER Visits	1.455 (2.754)	1.202 (2.550)
Independent Variable of Interest		
Marijuana Use Disorder	0.022 (0.146)	0.047 (0.212)
Dependence Only	0.016 (0.127)	0.032 (0.176)
Marijuana Abuse	0.005 (0.073)	0.015 (0.122)
Instrumental Variables		
Neither Approve Nor Disapprove	0.506 (0.500)	0.579 (0.494)
Somewhat Disapprove	0.102 (0.303)	0.111 (0.315)
Strongly Disapprove	0.392 (0.488)	0.310 (0.462)
Demographic and Household Characteristics		
White	0.441 (0.497)	0.480 (0.500)
Black	0.268 (0.443)	0.237 (0.426)
Asian	0.025 (0.155)	0.029 (0.169)
Other	0.039 (0.194)	0.032 (0.176)
Hispanic	0.228 (0.419)	0.222 (0.415)
Age 18-24	0.219 (0.414)	0.208 (0.406)
Age 25-29	0.189 (0.392)	0.123 (0.328)
Age 30-34	0.125 (0.331)	0.095 (0.293)
age 35-49	0.276 (0.447)	0.314 (0.464)
age 50-64	0.191 (0.393)	0.261 (0.439)
Married	0.248 (0.432)	0.305 (0.461)
Not Married	0.280 (0.449)	0.200 (0.400)
Never Married	0.472 (0.499)	0.495 (0.500)
Kids	1.338 (1.127)	0.899 (1.111)
Pregnant	0.070 (0.256)	-
No HH Member Over 65	0.917 (0.276)	0.885 (0.319)
1 HH Member Over 65	0.072 (0.258)	0.095 (0.293)
2+ HH Members Over 65	0.011 (0.105)	0.021 (0.142)
CBSA Pop. > 1 Million	0.501 (0.500)	0.489 (0.500)
CBSA Pop. < 1 Million	0.420 (0.494)	0.418 (0.493)
Does Not Live in CBSA	0.072 (0.258)	0.094 (0.291)
Human Capital and Financial Resources		
Less than High School	0.329 (0.470)	0.391 (0.488)
High School	0.388 (0.487)	0.385 (0.487)
Some College	0.229 (0.420)	0.166 (0.372)
College Graduate	0.054 (0.226)	0.059 (0.235)

Table 13

(Cont.)

	Female	Male
Human Capital and Financial Resources (cont.)		
Full Time	0.192 (0.394)	0.241 (0.428)
Part Time	0.142 (0.349)	0.115 (0.319)
Disabled – No SSI	0.088 (0.283)	0.140 (0.346)
Disabled and Collects SSI	0.155 (0.362)	0.214 (0.410)
Did Not Work Last Week	0.422 (0.494)	0.291 (0.454)
Family Income < $20,000	0.597 (0.491)	0.567 (0.495)
Family Income $20-50,000	0.346 (0.476)	0.366 (0.482)
Family Income > $50,000	0.058 (0.233)	0.067 (0.250)
Private Insurance	0.070 (0.255)	0.090 (0.286)
Family Collects Food Stamps	0.585 (0.493)	0.487 (0.500)
Family Collects Public Asst	0.190 (0.392)	0.124 (0.329)
Family Collects SSI	0.133 (0.339)	0.136 (0.343)
0 Phone Lines	0.324 (0.468)	0.320 (0.466)
1 Phone Line	0.644 (0.479)	0.646 (0.478)
2 Phone Lines	0.032 (0.175)	0.035 (0.183)
Other Substance Use Disorders		
Other Drug Use Disorder	0.031 (0.172)	0.048 (0.214)
Alcohol Use Disorder	0.064 (0.245)	0.123 (0.329)
Nicotine Use Disorder	0.276(0.447)	0.337 (0.473)
Supplementary Control Variables - Health		
Asthma	0.131 (0.338)	0.080 (0.272)
Bronchitis	0.068 (0.251)	0.044 (0.206)
Pneumonia	0.022 (0.147)	0.016 (0.127)
Sinusitis	0.050 (0.218)	0.017 (0.128)
Stroke	0.008 (0.088)	0.011 (0.105)
High Blood Pressure	0.141 (0.348)	0.153 (0.360)
Heart Disease	0.034 (0.181)	0.045 (0.208)
Diabetes	0.078 (0.268)	0.082 (0.274)
Hepatitis	0.010 (0.098)	0.025 (0.156)
STD	0.028 (0.164)	0.008 (0.089)
HIV	0.003 (0.053)	0.014 (0.118)
Ulcer	0.024 (0.054)	0.023 (0.149)
Other Disease	0.009 (0.095)	0.016 (0.127)
Depression	0.172 (0.378)	0.111 (0.314)

Table 13

(Cont.)

	Female	Male
Supplementary Control Variables – Attitudes and Beliefs		
Risk	0.106 (0.307)	0.195 (0.040)
Danger	0.123 (0.329)	0.238 (0.043)
Seatbelt	0.084 (0.277)	0.129 (0.396)
Religion Important	0.783 (0.412)	0.700 (0.426)
Service Attender	0.345 (0.475)	0.271 (0.444)
Religious Friends	0.373 (0.484)	0.375 (0.375)
Relig. Affects Decisions	0.727 (0.446)	0.657 (0.657)
Supplementary Control Variables – Illegal Activities		
Ever Arrested	0.206 (0.404)	0.421 (0.494)
Offered Drugs	0.120 (0.325)	0.194 (0.396)
Time Control Variables		
2005	0.139 (0.346)	0.122 (0.327)
2006	0.125 (0.331)	0.126 (0.332)
2007	0.140 (0.347)	0.138 (0.345)
2008	0.134 (0.340)	0.123 (0.329)
2009	0.189 (0.346)	0.147 (0.354)
2010	0.155 (0.362)	0.159 (0.366)
2011	0.168 (0.374)	0.184 (0.388)
Quarter 1	0.226 (0.418)	0.237 (0.426)
Quarter 2	0.260 (0.438)	0.268 (0.443)
Quarter 3	0.266 (0.442)	0.249 (0.432)
Quarter 4	0.248 (0.432)	0.245 (0.430)
N	19766 [9,719,097]	8,075 [4,966,773]

Notes. Standard deviation in parentheses. Brackets contain the population size represented by the weighted data. Risk refers to often or always preferring to participate in risky behavior. Danger refers to often or always preferring to participate in dangerous behavior. No seatbelt indicates an individual rarely or never wears a seatbelt when riding as a passenger in a car.

Table 14

Weighted Distribution of Annual ER Visits

	Female	**Male**
0	0.473	0.550
1	0.202	0.191
2	0.159	0.131
3	0.065	0.048
4	0.030	0.027
5-10	0.057	0.039
11-15	0.015	0.010
16-20	0.004	0.001
21-25	0.004	0.002
>25	0.002	0.002
N	19,766 [9,719,097]	8,075 [4,966,773]

Notes. Brackets contain the population represented by the weighted data. Values in the table are collapsed due to the rarity of high-valued outcomes, but the analysis considers the full, unaltered distribution of ER visits.

The dependent variable is the count of total ER visits in the past twelve months. Females visited the ER approximately 1.5 times in the past year while males attended 1.2 times. Roughly half of respondents did not visit the ER in the past year, while 20% went one time, 15% went twice, and roughly 5% went more than 4 times. Fewer than 2% of males or females went more than 10 times.

The primary independent variable of interest is a binary indicator for marijuana use disorder (MUD), which refers to either abuse or dependence upon marijuana in the past twelve months. Definitions of abuse and dependence are based on the Diagnostic and Statistical Manual of Mental Disorders Fourth Edition (DSM-IV). The DSM-IV was published by the American Psychiatric Association in 1994 and serves as the medical

standard for diagnosing mental health disorders.[17] According to the DSM-IV individuals are substance abusers if they meet 1 of 4 substance abuse measures but are not dependent on the substance. Substance dependence indicates that they meet 3 of 6 DSM dependence criteria, and therefore the two diagnoses are mutually exclusive, with dependence being a more severe level of use than abuse. Diagnostic criteria for abuse and dependence are listed in Appendix D.[18]

Roughly 2% of females and 5% of males are classified as having MUD. Rates of MUD among individuals who have used marijuana at least once in the past year are 15% and 21%, respectively. A brief descriptive analysis comparing respondents with MUD to those without is presented in section 4.3.

When allowing for MUD to be endogenous, identification of the coefficient for marijuana dependence requires one or more variables that are excluded from the equation for ER visits—that is, instrumental variables. Valid instruments must be both correlated with MUD conditional on X_o and also independent of acute medical care use conditional on X_o. In the current data, one measure that is plausibly independent of the process generating ER visits is the respondent's opinion of another adult trying marijuana once or twice. Respondents were asked "How do you feel about adults trying hashish once or twice?" Responses include "neither approve nor disapprove," "somewhat disapprove," and "strongly disapprove," as well as "don't know" and "refused." I create dummy

[17] The DSM-V was published in 2013, but the updated version has not yet been adopted by the NSDUH survey. Version V has eliminated the distinction between abuse and dependence, classifying both as "use disorder."

[18] Alcohol and certain "hard drugs" (e.g., heroin, cocaine, painkillers, etc.) also have a 7th dependence criterion that refers to withdrawal symptoms. This is not a determining factor for marijuana or nicotine dependence.

variables for both slight and strong disapproval, in reference to all other possible responses (roughly 1.3% of respondents did not know or refused). Roughly 1/2 of respondents neither approved nor disapproved of adults trying marijuana, while approximately 1/3 strongly disapprove, and the remainder disapprove somewhat.

The controls variables (X_o) are observable factors that may directly affect demand for medical care through their impact on health, or indirectly through their effect on the relative cost of ER visits. X_o is split into four subsets of variables. The organization of the variables is intended to group together measures that are in the same theoretical category (e.g., demographic controls vs. human capital controls). The groups are also generally intended to correspond with the risk each set presents of introducing endogenous regressors to the model.

The most basic subset of variables refers to controls for demographic and other household characteristics that are plausibly exogenous to ER visits, such as race/ethnicity (white, black, Hispanic, Asian, other). This also includes categorical measures of age (18–24; 25–29; 30–34; 35–49; 50–64), which may enter the model both indirectly, as health stock is expected to decline with age, and directly, since risky behavior is expected to be negatively correlated with age. Marital status, pregnancy status (for females), number of children under 18 living at home, and number of household members over age 65 will directly affect demand for acute medical care by reducing the amount of household resources available for individual *i*. These variables may also capture other types of heterogeneity, as married individuals and/or parents may behave differently to

single or childless individuals.[19] The size of the Core Based Statistical Area (> 1 million, < 1 million, or no CBSA) in which the respondent lives controls for the access to both acute medical care as well as illicit drugs.

Roughly 45% of the sample (both male and female) is white, with the other half being split between black and Hispanic.[20] The demographic distribution of the sample is roughly consistent with national averages reported by CMS. Nationwide in 2009, Medicaid beneficiaries were 59% female (41% male), and 41% white (23% black, and 22% Hispanic).[21] The sample is split roughly evenly among all age categories, with the exception of 30–34 year olds who comprise about 10%, and 35–49 year olds who comprise about 30% of the sample. Only 25–30% of respondents are currently married, and the vast majority (roughly 90%) does not have any household members over 65 living with them. Seven percent of females are pregnant at the time of the survey, and on average, females have 1.4 children, while males have less than 1.[22]

Another subset of control variables contains measures of human capital, employment, and financial resources. Education (less than a high school diploma, high school, some college, and college graduate) provides a measure of human capital, which is expected to increase the efficiency with which individuals "produce" health

[19] Whether these variables "cause" individual attributes to change or merely signal their presence (e.g., marriageability signals greater responsibility) is irrelevant to the current analysis.

[20] The NSDUH data are coded such that anyone who self-classifies as "white-Hispanic" or "black-Hispanic" etc. is categorized as Hispanic, not white or black.

[21] CMS: Medicaid Statistical Information System, 2014. Retrieved May 25, 2014, from http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MSIS-Tables.html

[22] The number of children in the public use data are top-coded at 2, which explains why the mean number of children in the sample may appear small.

(Grossman, 1972). Employment status (full time, part time, or did not work last week) accounts for the opportunity cost of time spent investing in health as well as time spent receiving acute medical care. Additional employment categories include individuals who reported that they did not work the previous week due to disability but did not collect supplemental security income (SSI), and those who reported that they did not work due to disability and do collect SSI. Categorical family income (<$20,000; $20,000–$50,000; $50,001–$75,000) proxies for financial resources available to invest in health or to directly purchase acute medical care or drugs (recall that observations reporting more than $75,000 in annual family income were dropped from the sample). Three additional proxies for financial resources include whether any member of the family collects food stamps, whether any member of the family collects welfare, job placement assistance, or childcare assistance, and whether anyone in the respondent's family besides the respondent collects supplemental security income.[23] I also include an indicator for possession of private insurance, which directly affects the price of visiting the ER.[24] The number of phone lines (0, 1, or 2) in the house is included as a final proxy for financial resources.

Over 60% of respondents reported that they did not work for pay in the previous week, and consequently about 60% of the sample reports an annual family income of less

[23] Information is not provided on whether the respondent collects SSI him or herself. Therefore, if an individual reports not working due to disability and also indicates that somebody in their family collects SSI, I code that individual as disabled and collecting SSI. If an individual reports not working due to disability but also indicates that nobody in the family collects SSI, they are coded as disabled without SSI. Therefore the indicator for "family collects SSI" indicates that the family receives SSI and the respondent does not self-report a disability.

[24] Possession of private insurance does not automatically disqualify one from qualifying for Medicaid. It is conceivable to envision low-income individuals (particularly those with children) receiving some form of employer-provided insurance but still falling below the income threshold for Medicaid.

than $20,000. Only about 5% of respondents are college graduates, while roughly one-third of the respondents have only a high school diploma, and another one-third of the respondents did not finish high school. Approximately half the sample has at least one family member on food stamps, and 20% of families have at least one member collecting welfare, job placement assistance, or childcare assistance. Over 60% of families have at least one phone line but nearly one-third of families have none. Less than 10% of respondents report access to private insurance. These results are consistent with the low-income status expected among Medicaid recipients, and support the possibility that ER visits may be generated differently among this subsample compared to the US population at large.

The set of controls most strongly suspected to be endogenous are indicators for alcohol use disorder, nicotine use disorder, and a use disorder for any illicit drug besides marijuana (so-called "hard" drugs). Three percent of females and 5% of males have a hard drug use disorder, while 6% of females and 12% of males have an alcohol use disorder, and roughly 30% of males and females have a nicotine use disorder. These other substance use disorders are expected to either reduce individual health stock, or to directly cause an adverse health event (e.g., alcohol poisoning or drug overdose), although they could have indirect effects on demand for ER visits. Overall, substance dependence is expected to be positively correlated with ER visits.

The fourth and final set of controls refers to additional variables used for robustness checks in Section 7.4. These include variables controlling for the

respondent's health, the respondent's attitudes and beliefs, and the respondent's participation in illegal activities.

Measures of health include binary indicators for whether the respondent has been diagnosed by a physician with any of the following in the past year: asthma, bronchitis, pneumonia, sinusitis, stroke, high blood pressure, heart disease, diabetes, hepatitis, STD, HIV, ulcer, lung cancer, liver cirrhosis, pancreatitis, and tuberculosis.[25] An indicator for depression is also included to account for potential mental health confounders.[26] Health is considered to be the primary factor that affects demand for ER visits. By controlling for as many diseases as possible, I am able to account for one of the key vectors through which MUD may influence demand for ER visits. If marijuana consumption is positively correlated with disease as predicted by the medical literature (Brook et al., 2008; Gordon et al., 2013; Mittleman et al., 2001; Polen et al., 1993; Rey & Tennent, 2002; Schuster et al., 2012; Sidney et al., 1997; Smith & Crespo, 2001), then inclusion of these controls should decrease the value of the coefficient on MUD. However, these variables may be endogenous to ER utilization since affirmative response requires diagnosis by a medical professional, which may occur as the result of an ER visits.

Measures of attitudes and beliefs refer to the respondent's preference for risky or dangerous behavior, as well as the importance they ascribe to their religious beliefs. Preference for risk-taking, preference for danger, and actual risky-taking (i.e., seat belt use) are measured as the response to the following questions: "How often do you like to

[25] Due to the relatively rarity of lung cancer, cirrhosis, pancreatitis, and tuberculosis, these four diseases are lumped together into a single "other disease" category.

[26] Although a more specific mental health measure is also available, the questions and scoring vary significantly between survey years, prohibiting use of a single measure across the multiple survey years.

test yourself by doing something a little risky?"; "How often do you get a kick out of doing things that are a little dangerous?"; and "How often do you wear a seatbelt when you ride in the front passenger seat of the car?"[27] Possible responses are "Never," "Seldom," "Often," and "Always." Binary indicators were created for individuals who "Often" and "Always" seek risk or danger, and who "Seldom" or "Never" wear a seatbelt.

Individual religiosity is determined by the response to four questions. Three questions ask for agreement/disagreement with the following statements: "My religious beliefs are very important," "My religious beliefs influence my decision making," and "It is important that my friends share my religious beliefs." Responses range from strongly disagree to strongly agree. The responses are transformed into binary variables reflecting agreement (1) or disagreement (0). The final measure of religiosity is the number of religious services attended in the past 12 months. Responses include six categories ranging from never to more than once per week. Individuals who attended at least 6-24 times in the past year are coded as attenders (1), while those attending fewer than 6 times are coded as non-attenders (0).[28]

Risk-taking or dangerous behavior is expected to increase the demand for ER visits by increasing the probability of an adverse health event. To the extent that

---

[27] A separate question asked how often a seatbelt was worn when driving a car. A possible response to this was "I do not drive." This was not a possible response to the passenger seat belt question. Non-response to the passenger seatbelt question was less than for other questions regarding risk/danger suggesting that few if any non-drivers refused to answer the passenger question.

[28] The range 6–24 is a full response category. Therefore the two alternatives were to say that attendance once every two months is the minimum for an attender or that twice in one month is the minimum. I opted for the former.

individuals who develop MUD have different preferences for risky behavior, controlling for these preferences should help to eliminate omitted variables bias. Religious beliefs may also affect behavior in ways that affect the demand for ER visits, and are expected to be negatively correlated with MUD since the moral code of many of the religions common in the United States prohibits the use of drugs.

The controls for illegal activities consist of binary indicators for affirmative response to the questions "Not counting minor traffic violations, have you ever been arrested and booked for breaking the law?" and "In the past 30 days has anyone approached you to sell you an illegal drug?" Both these variables may be endogenous to MUD. Presumably individuals with high levels of marijuana use are frequently offered drugs, while in many states the use of marijuana can lead to arrest. However, individuals with MUD may be involved in other risky or illegal activities besides drug use that could be correlated with demand for ER visits (whether or not they state a preference for risky activities). Therefore, omission of these controls may bias estimates of the relationship between MUD and ER visits.

Two control variables that are not available in the public use NSDUH data are state of residence or region of the country. Although drug regulations and Medicaid participation criteria vary from state to state, the unobserved effect of state-specific and regional attributes should be absorbed by X_u, which should in turn be accounted for by the instrumental variables approach as detailed in Section 6. Since the NSDUH data are intended to be nationally representative, the estimated relationship between marijuana

dependence and healthcare utilization should be consistent with the average relationship that would be observed if state indicators were available.

5.3. Descriptive Analysis

Table 15 presents the weighted mean of annual ER visits, rates of other substance use disorders (hard drugs, nicotine, alcohol), and opinion of other adults trying marijuana, by MUD. The comparison shows that females with MUD visit the ER significantly more times than those without. However, males and females with MUD are significantly more likely to be dependent on any of the three other substances, which may be correlated with negative health outcomes. Table 15 also suggests that the instrumental variables are correlated with MUD, as expected. Females with a marijuana use disorder are 30 percentage points less likely to strongly disapprove of adults trying marijuana, while males with MUD are 25 percentage points less likely. The difference is significant in both cases. Females with MUD are also significantly less likely to somewhat disapprove, a difference of roughly 4 percentage points.

Table 16 compares the full distribution of ER visits between those with and without MUD. Among both males and females those with MUD are significantly more likely to visit the ER. Females with MUD are significantly more likely to visit between 11 and 20 times, while males with MUD are significantly more likely to visit the ER one time.

Table 15

Comparison of Weighted Means by Marijuana Use Disorder

	Females		Males	
	No Marijuana Use Disorder	Marijuana Use Disorder	No Marijuana Use Disorder	Marijuana Use Disorder
Dependent Variable				
ER Visits	1.439	2.184**	1.189	1.478*
	(2.715)	(4.389)	(2.541)	(2.892)
Other Substance Dependence				
Other Drug Dependence	0.025	0.260***	0.037	0.270***
	(0.156)	(0.441)	(0.190)	(0.444)
Alcohol Dependence	0.058	0.334***	0.107	0.461***
	(0.234)	(0.472)	(0.309)	(0.499)
Nicotine Dependence	0.271	0.509***	0.327	0.548***
	(0.448)	(0.500)	(0.469)	(0.498)
Instrumental Variables				
Somewhat Disapprove	0.103	0.066***	0.110	0.136
	(0.302)	(0.248)	(0.313)	(0.343)
Strongly Disapprove	0.399	0.092***	0.322	0.051***
	(0.488)	(0.358)	(0.467)	(0.219)
N	19,101	665	7,458	617
	[9,506,713]	[212,384]	[4,733,374]	[233,399]

Notes. Standard deviation in parentheses. Brackets contain the population represented by the weighted data. Asterisks denote significant difference in weighted means between those with and without MUD within a gender. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Overall, descriptive analysis suggests that those of either gender with MUD are generally more likely to use the ER. However, both males and females with MUD are more likely to have other substance use disorders, which may be confounding the results. Marijuana use disorder and ER utilization are both complex processes and a more thorough analytic approach is necessary to disentangle the causal impact of MUD on ER utilization. The empirical model used to estimate this causal effect is detailed in the next section.

Table 16

Weighted Distribution of Annual ER Visits by Marijuana Use Disorder

	Females With No Marijuana Use Disorder	Females With Marijuana Use Disorder	Males With No Marijuana Use Disorder	Males With Marijuana Use Disorder
0	0.477	0.327***	0.559	0.426***
1	0.201	0.223	0.188	0.270**
2	0.158	0.194	0.129	0.162
3	0.064	0.111	0.047	0.061
4	0.030	0.045	0.027	0.019
5–10	0.058	0.052	0.039	0.049
11–15	0.014	0.043**	0.010	0.004
16–20	0.004	0.011***	0.001	0.002
21–25	0.004	0.002	0.002	0.005
>25	0.002	0.004	0.002	0.003
N	19,101 [9,506,713]	665 [212,384]	7,458 [4,733,374]	617 [233,399]

Notes. Brackets contain the population represented by the weighted data. Asterisks denote significant difference between marijuana dependents and non-dependents within a given gender. Significance is determined by a survey-adjusted χ^2 statistic. $^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.10$. Values in the table are collapsed due to the rarity of high-valued outcomes, but the analysis considers the full, unaltered distribution of ER visits.

6. Empirical Approach

The estimation objective is the average effect of M on ER where ER is the reduced form demand for emergency room visits as a function of X_o, X_u, M, and D. This average "treatment" effect (ATE) of marijuana use disorder measures the partial change in demand for ER visits that occurs (or would occur) if an average individual in the population switched from a non-dependent to a dependent state. The ATE can be defined as:

$$ATE = E[E[ER_1] - E[ER_0]] = E[E[ER|X, M = 1] - E[ER|X, M = 0]] \quad \text{(Eq. 3)}$$

where ER_1 is the count of ER visits that did (or would) occur if an individual suffers from MUD, and ER_0 is the count that did (or would) occur if the same individual does not have a marijuana use disorder.

Estimation is complicated by the fact that for a given individual only one state is observed: individual *i* either suffers from MUD or not. The observed value of M is not randomly assigned, but rather depends on characteristics of an individual or her environment (X) that may also affect demand for acute healthcare, some of which are observed (X_o) and some of which are not (X_u). Failure to account for this "selection" into MUD will lead to biased and inconsistent estimates of the effect of MUD on healthcare utilization.

Consistent estimation of the ATE is possible given a vector of control variables X that contains all variables correlated with ER and M. Following Terza (2009), I express these variables as $X = [X_o\ D\ X_u]$ where X_o is a vector of observable characteristics, D is a vector of other possible substance use disorders (alcohol, nicotine, or hard drugs) and X_u is a continuous scalar representing the combined effect of unobservable characteristics (including measurement error). As discussed in Section 4, X_o, D, and X_u may affect ER directly but may also operate indirectly through their effect on H.

I allow for M to be endogenous to health (and therefore to utilization of medical care) since the consumption of marijuana and medical care may depend, in part, on common unobservable factors (X_u). For example, potential marijuana users may discount the future higher than non-potential users, and consequently invest less in their health stock. Use of illicit drugs may also reflect a low stock of knowledge, both about the

danger of illicit drug use and of other health behaviors that may change one's stock of health (e.g., proper exercise or nutrition).[29,30]

Following the approach discussed in Terza (2009), the conditional expectation of ER can be expressed as

$$E[ER|M,X] = \frac{1}{n}\frac{\sum_i exp(M_i\beta m + Xoi\beta o + D_i\beta D + Xui\beta u)W_i}{\sum_i W_i} \quad \text{(Eq. 4)}$$

where W_i refers to the individual-level sampling weights.

Marijuana use disorder can be represented by a probit specification such that $M = 1(M^* > 0)$ where:

$$M^* = W\delta + X_u, \quad \text{(Eq. 5)}$$

$W = [X_o\ D\ W^+]$, W^+ is a vector of identifying instruments and X_u is standard normally distributed conditional on W and independent of W. Unobserved correlation between M and ER occurs due to a joint dependence upon X_u, and therefore M and ER are independent conditional on X_o, D, and X_u.

X_u can be conditioned out of the ATE by integrating over the assumed density. Since X_u is assumed to be standard normally distributed conditional on W the estimated ATE can be expressed as

[29] For example, Smith and Crespo (2001) find that marijuana users consume more sodium, pork, and cheese, but less fruit than non-users.

[30] Individuals with a high propensity to consume marijuana may also have a less reliable 12-month recall, and therefore measurement error in Y may also be a function of X_u.

$$\widehat{ATE} = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} E\left[\widehat{ER_{1i}}\right] - E[\widehat{ER_{0i}}]]\varphi(X_u)dX_u \qquad \text{(Eq. 6)}$$

where $\varphi(\cdot)$is a standard normal density, $E\left[\widehat{Y_{1i}}\right]$ is equivalent to (4) with $M_i = 1$, and $E[\widehat{Y_{0i}}]$ is equivalent to (4) with $M_i = 0$. The ATE can be computed from (6) by obtaining consistent estimates for the parameters β_m, β_o, and β_u.

To obtain parameter estimates, an estimator must be selected that appropriately accounts for the data generating process that produces ER visits. Consistent with a traditional count measure, the number of ER visits is the sum of a number of binary outcomes. Under the assumptions of the conceptual model, a severe enough health event will result in a visit to the ER. While a single underlying condition may trigger repeated ER visits, a single ER visit cannot directly cause another one. (That is, the condition may be severe enough that the patient will remain in the hospital but the severity of the shock cannot result in a repeat admission to the ER during the same trip to the hospital.) ER visits may thus be modeled by a single count distribution.

I model ER visits using the two-stage nonlinear least squares (2SNLS) approach introduced by Terza (1998). Although less efficient than the full information maximum likelihood (FIML) approach proposed by Terza (1998, 2009), 2SNLS is more robust to misspecification since only the conditional mean of ER visits needs to be specified rather than the entire conditional distribution (although both approaches requires that the distribution of X_u is correctly specified). As discussed previously, I assume X_u is standard normally distributed conditional on W. Then, as presented in Terza (1998) the

exponential conditional expectation of *ER* when X_u is standard normally distributed can be expressed as:

$$E[ER|W,M]=M\ exp(M\beta_M\ +\ X_o\beta_o^+ + D\beta_D)\frac{\Phi(W\hat{\delta}+\beta_u)}{\Phi(W\hat{\delta})} + (1-M)exp(\ X_o\beta_o^+ + D\beta_D)\frac{\Phi(-W\hat{\delta}-\beta_u)}{\Phi(-W\hat{\delta})} \quad (7)$$

where β_o^+ contains a constant term that is shifted upward by $\left(\frac{\beta_u^2}{2}\right)$ and $\hat{\delta}$ refers to estimates obtained from probit regression of M. Applying NLS to (7) allows for consistent estimates of β_M, β_o, and β_u, which may be substituted into (6) to produce estimates of the ATE with respect to ER visits. The ATE thus captures the expected change in the number of ER visits that would occur if all individuals in the population switched from a state of no MUD to MUD. Unfortunately, cluster identifiers are unavailable in the public-use data so standard errors are computed using the Huber-White sandwich estimator.[31] Additionally, the covariance matrix must be adjusted for the first stage estimation of δ, an adjustment discussed further in Appendix E.

Specification of the model is complicated by the fact that elements of X_o may be endogenous to ER visits and/or marijuana use disorder (i.e., elements of X_o may also be functions of X_u). For instance, educational and employment decisions may be correlated with unobservable factors (e.g., future discount rate) that are expected to influence the decision to consume drugs, or the decision to invest in one's health stock. Other

[31] Estimates of the standard error should be clustered within sampling units. However, sampling unit identifiers are not provided in the public use data file and estimation of clustered standard errors is thus impossible.

measures of substance use disorder (hard drugs, alcohol, and nicotine) are particularly problematic, since all forms of substance use are likely generated by similar processes. (This is especially true for hard drugs since in most states consumption of nicotine and alcohol is legal, while consumption of hard drugs or marijuana is prohibited.)

Estimates of parameters for endogenous variables will not be consistent. Moreover, any endogenous variable in the model for ER visits that is correlated with M will also render estimates of the coefficient for M inconsistent (and by extension, estimates of the ATE as well.) Additionally, variables endogenously correlated with the probability of marijuana use disorder may lead to inconsistent estimates of δ that in turn affect second stage estimation of the model for ER visits.

The solution is not as simple as omitting any variables that are considered to be endogenous. Although the use of instrumental variables in the second stage should render the coefficient estimate for M consistent despite the omission of variables that belong in the second stage equation, the omission of variables that belong in the first-stage probit equation for M will render estimates of δ inconsistent, and by extension affect second stage estimates. However, in order for the instrumental variables strategy to be feasible, all variables included in the first stage that do not meet the criteria to be exclusion restrictions must also be included in the second stage, and so variables endogenous to ER visits that belong in the equation for M cannot simply be omitted from both stages without risking additional problems. Moreover, it is neither known with certainty which variables in particular are endogenous, nor the severity of the endogeneity problem.

The approach moving forward is intended to assess the tradeoffs between potential omitted variables bias on the one hand and inconsistencies potentially introduced by endogenous variables on the other. Ideally, a specification that minimizes the likelihood of endogenous regressors while maximizing the possibility for omitted variables bias will produce estimates very similar to one that minimizes the risk of omitted variables bias but maximizes the number of potentially endogenous variables. If this is the case, then it bolsters the credibility of the results and suggests that the estimates are reasonably close to the "true" parameter values. However, if the estimates are highly unstable as tradeoffs are made between endogenous regressors and omitted variables bias, it may suggest that the empirical model is not a good match for the data. The analytic strategy is described in more detail in the following section.

7. Results

In this section I present results for four sets of regression analyses. These include NLS estimates of the demand for ER visits; 2SNLS estimates of the demand for ER visits, which account for the potential endogeneity of MUD; first-stage probit estimates for MUD; and follow-up 2SNLS estimates, which further investigate the results from the first set of 2SNLS estimates.

For the first three sets of regressions I run three separate specifications for both males and females to determine the stability of the estimates under various combinations of potential omitted variables bias and potentially endogenous regressors. This will inform not only on the robustness of the results, but also provide insight into which variables may be more damaging to include than exclude. The analytic strategy consists

of moving along the spectrum from a large risk of omitted variables bias and small risk of endogenous regressors, to a larger risk of endogenous regressors and reduced risk of omitted variables bias. The three specifications correspond to the initial three sets of control variables introduced in Section 5. In each specification a new set of variables is added so that the third specification contains all three sets of variables.

The three sets of variables include controls for demographic and household characteristics; education, employment, and financial resources; and whether the respondent has a substance use disorder for hard drugs, alcohol, or nicotine. The variables are grouped in such a way as to balance theoretical considerations (i.e., what types of variables control for roughly the same type of characteristics) with the goal of moving from the most plausibly exogenous set of variables to the one most suspected to be endogenous. Although a handful of variables could reasonably be assigned to other groups, the assigned categories and order of analysis are expected to fulfill the analytic objective.

In the fourth set of regressions, I conduct robustness checks of the initial 2SNLS results by running 2SNLS models with additional controls for health, attitudes/beliefs, and participation in illegal activities, corresponding to the fourth set of controls described in Section 5. I also conduct 2SNLS regression without the sampling weights. The results are discussed below.

7.1. Exogenous Model Development

Exogenous results for females and males are presented in Tables 17 and 18, respectively. Among females, MUD is positively and significantly correlated with ER

visits in the basic demographic model. On average, respondents with MUD visit the ER 0.7 more times than those without, a change of roughly 1/4 of a standard deviation. These results are robust to the inclusion of controls for human capital. However, inclusion of controls for other substance use disorders attenuates the estimated ATE by nearly 50%, suggesting that the previous estimates were upward biased due to the high prevalence of other substance use disorders among those with a marijuana use disorder. This is consistent with the findings of William and Skeels (2006) who show that the positive correlation between tobacco and marijuana use confounds the predicted effect of marijuana on health when tobacco is not controlled for.

Table 17

NLS Estimates of ER Visits for Females

	Demographic	**Human Capital**	**Other Substances**
Variable of Interest			
MUD	0.399**	0.394**	0.250*
	(0.121)	(0.129)	(0.132)
ATE	0.701**	0.690**	0.407
	(0.256)	(0.270)	(0.242)
Demographics			
Black	0.127*	0.081	0.104
	(0.069)	(0.072)	(0.070)
Asian	-1.028***	-0.816***	-0.763***
	(0.164)	(0.158)	(0.156)
Other	0.025	-0.011	0.005
	(0.112)	(0.113)	(0.112)
Hispanic	-0.373***	-0.407***	-0.373***
	(0.064)	(0.071)	(0.073)
Age 18–24	0.036	0.475***	0.467***
	(0.091)	(0.095)	(0.094)
Age 25–29	0.085	0.416***	0.375***
	(0.092)	(0.101)	(0.100)

Table 17

(Cont.)

	Demographic	Human Capital	Other Substances
Demographics (cont.)			
Age 30–34	0.051	0.297**	0.265**
	(0.100)	(0.112)	(0.115)
Age 35–49	0.149	0.260**	0.232**
	(0.094)	(0.094)	(0.095)
Not Married	0.016	-0.113	-0.114
	(0.083)	(0.086)	(0.086)
Never Married	-0.145*	-0.252**	-0.262**
	(0.083)	(0.083)	(0.083)
Kids	-0.164***	-0.118***	-0.109***
	(0.027)	(0.029)	(0.028)
Pregnant	-0.152**	-0.110*	-0.102
	(0.071)	(0.064)	(0.064)
1 HH Member > 65	-0.173	-0.155	-0.160
	(0.107)	(0.106)	(0.105)
2+ HH Members > 65	0.328	0.298	0.277
	(0.446)	(0.368)	(0.357)
Small CBSA	0.163**	0.130**	0.130**
	(0.059)	(0.063)	(0.064)
No CBSA	0.107	0.001	0.003
	(0.102)	(0.114)	(0.111)
Constant	0.525***	-0.074	-0.102
	(0.123)	(0.281)	(0.282)
Human Capital			
Less than HS		0.049	0.028
		(0.257)	(0.266)
High School		-0.159	-0.173
		(0.257)	(0.265)
Some College		-0.091	-0.110
		(0.260)	(0.268)
Full Time		-0.039	-0.034
		(0.072)	(0.071)
Part Time		-0.223***	-0.211**
		(0.066)	(0.065)
Disable No SSI		0.690***	0.671***
		(0.096)	(0.094)

Table 17

(Cont.)

	Demographic	Human Capital	Other Substances
Human Capital (cont.)			
Disable Collects SSI		-0.066	-0.048
		(0.101)	(0.104)
Income < $20,000		0.231**	0.228**
		(0.111)	(0.112)
Income $20-50,000		0.223**	0.226**
		(0.108)	(0.109)
Insurance		-0.123	-0.103
		(0.085)	(0.084)
Food Stamp		0.187**	0.172**
		(0.075)	(0.074)
Public Assistance		0.065	0.059
		(0.079)	(0.080)
Family SSI		0.146**	0.154**
		(0.063)	(0.063)
1 Phone		-0.032	-0.022
		(0.059)	(0.059)
2+ Phones		0.186	0.181
		(0.174)	(0.172)
Other Substances			
Alcohol			0.114
			(0.111)
Nicotine			0.086
			(0.065)
Hard Drugs			0.267**
			(0.117)
N	19,101	19,101	19,101
	[9,506,713]	[9,506,713]	[9,506,713]

Notes. The dependent variable is annual ER visit. All models control for year and quarter of interview. Omitted categories include white, age 50-64, married, no household members over 65, respondent lives in CBSA with population over 1 million, college education, did not work last week, family income $50,000-$75,000, and no phone lines in household. Huber/White robust standard errors reported in parentheses. Standard errors for ATE computed using Delta Method. Brackets contain "true" number of observations represented by weighted estimates. $^{*}p < 0.1$ $^{**}p < 0.05$ $^{***}p < 0.01$.

Table 18

NLS Estimates of ER Visits for Males

	Demographic	Human Capital	Other Substances
Variable of Interest			
MUD	0.201	0.026	-0.253
	(0.139)	(0.144)	(0.278)
ATE	0.265	0.030	-0.229
	(0.199)	(0.173)	(0.224)
Demographics			
Black	0.073	0.101	0.081
	(0.124)	(0.136)	(0.152)
Asian	-0.858***	-0.661**	-0.375
	(0.245)	(0.255)	(0.274)
Other	0.254*	0.170	0.173
	(0.148)	(0.150)	(0.166)
Hispanic	-0.149	-0.143	0.339**
	(0.202)	(0.210)	(0.153)
Age 18–24	-0.259*	0.187	-0.071
	(0.139)	(0.150)	(0.190)
Age 25–29	-0.114	0.164	0.015
	(0.142)	(0.169)	(0.176)
Age 30–34	-0.128	0.058	-0.121
	(0.160)	(0.176)	(0.259)
Age 35–49	-0.018	0.017	0.219*
	(0.135)	(0.137)	(0.128)
Not Married	0.387**	0.291*	0.068
	(0.135)	(0.158)	(0.162)
Never Married	0.103	-0.071	-0.002
	(0.125)	(0.136)	(0.160)
Kids	-0.083*	0.009	-0.047
	(0.043)	(0.049)	(0.059)
1 HH Member > 65	0.094	0.013	0.444***
	(0.193)	(0.197)	(0.133)
2+ HH Members > 65	-0.127	-0.082	-0.153
	(0.217)	(0.226)	(0.225)
Small CBSA	0.376***	0.357**	0.537***
	(0.112)	(0.119)	(0.115)
No CBSA	0.109	-0.017	0.101
	(0.127)	(0.133)	(0.164)
Constant	0.233	-0.780**	-1.689**
	(0.209)	(0.368)	(0.598)

Table 18

(Cont.)

	Demographic	Human Capital	Other Substances
Human Capital			
Less than HS		0.165	0.483
		(0.265)	(0.442)
High School		0.027	0.673
		(0.270)	(0.446)
Some College		0.144	0.664
		(0.281)	(0.464)
Full Time		-0.157	-0.279**
		(0.118)	(0.123)
Part Time		-0.141	-0.112
		(0.127)	(0.142)
Disable No SSI		0.676***	0.783***
		(0.142)	(0.154)
Disable Collects SSI		0.719***	0.822***
		(0.127)	(0.136)
Income < $20,000		0.353**	0.730**
		(0.161)	(0.274)
Income $20–50,000		0.306*	0.879**
		(0.159)	(0.270)
Insurance		-0.368*	-0.304*
		(0.220)	(0.172)
Food Stamp		0.071	0.047
		(0.119)	(0.110)
Public Assistance		0.250	0.053
		(0.177)	(0.181)
Family SSI		0.435***	0.540***
		(0.101)	(0.123)
1 Phone		0.061	-0.265**
		(0.124)	(0.116)
2+ Phones		-0.060	-0.386*
		(0.247)	(0.222)
Other Substances			
Alcohol			0.727***
			(0.108)
Nicotine			-0.202
			(0.127)
Hard Drugs			0.268
			(0.195)

Table 18

(Cont.)

	Demographic	Human Capital	Other Substances
N	8,075 [4,966,773]	8,075 [4,966,773]	8,075 [4,966,773]

Notes. The dependent variable is annual ER visit. All models control for year and quarter of interview. Omitted categories include white, age 50-64, married, no household members over 65, respondent lives in CBSA with population over 1 million, college education, did not work last week, family income $50,000-$75,000, and no phone lines in household. Huber/White robust standard errors reported in parentheses. Standard errors for ATE computed using Delta Method. Brackets contain "true" number of observations represented by weighted estimates. $^{*}p < 0.1$ $^{**}p < 0.05$ $^{***}p < 0.01$.

Among males, MUD is positively but insignificantly correlated with ER visits in the basic demographic model, resulting in roughly 0.25 additional ER visits per year. Inclusion of controls for human capital reduces the effect to essentially zero. However, inclusion of controls for other substance use disorders renders the ATE negative such that respondents with MUD visit the ER roughly 0.25 fewer times per year, on average. Although this is not a significant result, it is consistent with the finding among females that the correlation between MUD and other substance use disorders may be upward biasing the relationship between MUD and ER visits.

In general, the control variables are estimated to have the relationship with ER visits that would be expected *a priori*. Among both males and females, respondents reporting a disability are significantly more likely to use the ER, while those with private insurance are significantly less likely. Lower income is associated with more ER visits among both males and females, while employment is associated with fewer ER visits. Age is negatively correlated with ER visits among both males and females, although this

result is consistent with previous studies regarding drug use and ER visits (French et al., 2000; McGeary & French, 2000).

Among females hard drug use disorder is positively and significantly correlated with ER visits, and the coefficient is similar in magnitude to the coefficient for MUD (β_M). Among males, only alcohol use disorder is significantly correlated with ER visits, although hard drug use disorder is also positively correlated with ER visits, in contrast to the negative estimated correlation between MUD and ER visits.

Progressing from the demographic model to the model controlling for other substance use disorders appears to reduce omitted variables bias, although it remains uncertain whether or not the estimates are biased by potential endogeneity of the control variables. However, the relative stability of the estimates, particularly among females, suggests that the NLS model is robust to the tradeoffs between potential omitted variables bias and bias from potentially endogenous regressors. At this stage it appears that MUD may have a small, positive correlation with ER visits, although the most comprehensive model suggests there may be a null, or even negative relationship among males. As mentioned previously there is reason to believe that these estimates may be biased if there is correlation between MUD and the error term for ER visits that is not accounted for by the control variables. In the next section I repeat the preceding set of specifications using a 2SNLS model, allowing me to control for the potential endogeneity of MUD.

7.2. Endogenous Model Development

The direction and magnitude of the estimates from the 2-Stage Nonlinear Least Squares (2SNLS) model differ markedly from those produced by the exogenous model.

These results are reported in Tables 19 (females) and 20 (males). Among females MUD is predicted to have a large and negative, but statistically insignificant relationship to ER visits. The ATE ranges from -0.85 in the baseline demographic specification to -0.79 in the specification controlling for other forms of substance use disorder. The estimate of β_M is only significant in the specification controlling for other substance use disorders, and the ATE is not significant in any of the three models. However, the coefficient for the unobserved heterogeneity, β_u, is significant and positive in all three specifications. This indicates that the relationship between MUD and ER visits is significantly confounded by unobserved factors correlated with both MUD and the demand for ER visits. As with the exogenous model, the demographic and human capital specifications produce similar estimates to each other, with controls for other substance use disorders slightly attenuating the estimates of β_M and β_u.

Among males, marijuana use disorder is also estimated to have a negative relationship with ER visits, although the relationship is only significant in the substance use disorder specification. The estimated ATE ranges from -0.61 in the demographic model to -0.92 in the substance use disorder specification but is not significant in any of the three specifications. β_u is only significant in the substance use disorder specification, making it less clear than among females whether MUD is endogenous to ER visits. However, the sign and magnitude of the estimates for the ATE and β_u are roughly similar to those for females, suggesting that the lack of significance may be partly due to the reduced sample size among males. The pattern of 2SNLS coefficient estimates across specifications is similar to that from the exogenous model, with controls for human

capital and financial resources somewhat reducing the value of β_M, and controls for substance use disorder substantially reducing the value of β_M.

Table 19

2SNLS Estimates of ER Visits for Females

	Demographic	**Human Capital**	**Other Substances**
Variable of Interest			
MUD	-0.841	-0.847	-0.745*
	(0.590)	(0.515)	(0.447)
ATE	-0.858	-0.862	-0.787
	(0.711)	(0.609)	(0.522)
Demographics			
Black	0.132*	0.088	0.126*
	(0.068)	(0.071)	(0.069)
Asian	-1.027***	-0.814***	-0.738***
	(0.164)	(0.158)	(0.156)
Other	0.028	-0.009	0.016
	(0.114)	(0.114)	(0.113)
Hispanic	-0.387***	-0.425***	-0.368***
	(0.066)	(0.072)	(0.073)
Age 18–24	0.131	0.597***	0.554***
	(0.111)	(0.116)	(0.103)
Age 25–29	0.155	0.515***	0.441***
	(0.104)	(0.111)	(0.101)
Age 30–34	0.101	0.373**	0.309**
	(0.106)	(0.115)	(0.114)
Age 35–49	0.168*	0.291**	0.237**
	(0.093)	(0.094)	(0.094)
Not Married	0.043	-0.090	-0.100
	(0.087)	(0.088)	(0.088)
Never Married	-0.112	-0.225**	-0.248**
	(0.086)	(0.084)	(0.083)
Kids	-0.178***	-0.136***	-0.118***
	(0.029)	(0.031)	(0.029)
Pregnant	-0.158**	-0.116*	-0.107*
	(0.071)	(0.064)	(0.064)

Table 19

(Cont.)

	Demographic	Human Capital	Other Substances
Demographics (cont.)			
1 HH Member > 65	-0.180*	-0.160	-0.149
	(0.107)	(0.106)	(0.105)
2+ HH Members > 65	0.321	0.290	0.264
	(0.442)	(0.362)	(0.350)
Small CBSA	0.170**	0.135**	0.134**
	(0.059)	(0.063)	(0.064)
No CBSA	0.102	-0.006	0.003
	(0.102)	(0.113)	(0.110)
Constant	0.536***	-0.098	-0.141
	(0.124)	(0.284)	(0.283)
Human Capital			
Less than HS		0.062	0.022
		(0.261)	(0.273)
High School		-0.162	-0.190
		(0.264)	(0.275)
Some College		-0.095	-0.123
		(0.266)	(0.278)
Full Time		-0.042	-0.029
		(0.071)	(0.070)
Part Time		-0.215**	-0.202**
		(0.067)	(0.065)
Disable No SSI		0.709***	0.684***
		(0.099)	(0.097)
Disable Collects SSI		-0.069	-0.050
		(0.104)	(0.106)
Income < $20,000		0.219**	0.215*
		(0.111)	(0.113)
Income $20–50,000		0.210*	0.213*
		(0.108)	(0.109)
Insurance		-0.107	-0.085
		(0.087)	(0.087)
Food Stamps		0.205**	0.177**
		(0.076)	(0.074)
Public Assistance		0.079	0.074
		(0.082)	(0.083)

Table 19

(Cont.)

	Demographic	Human Capital	Other Substances
Human Capital (cont.)			
Family SSI		0.149**	0.160**
		(0.064)	(0.063)
1 Phone		-0.011	0.003
		(0.061)	(0.061)
2+ Phones		0.191	0.192
		(0.170)	(0.169)
Other Substances			
Alcohol			0.237*
			(0.132)
Nicotine			0.120
			(0.074)
Hard Drugs			0.450***
			(0.131)
β_u	0.617*	0.619**	0.523**
	(0.328)	(0.279)	(0.233)
N	19,101	19,101	19,101
	[9,506,713]	[9,506,713]	[9,506,713]

Notes. The dependent variable is annual ER visit. All models control for year and quarter of interview. Omitted categories include white, age 50-64, married, no household members over 65, respondent lives in CBSA with population over 1 Million, college education, did not work last week, family income $50,000-$75,000, and no phone lines in household. Huber/White robust standard errors reported in parentheses and corrected for 2-stage estimation. Standard errors for ATE computed using Delta Method. Brackets contain "true" number of observations represented by weighted estimates. $^{*}p < 0.1$ $^{**}p < 0.05$ $^{***}p < 0.01$.

Table 20

2SNLS Estimates of ER Visits for Males

	Demographic	Human Capital	Other Substances
Variable of Interest			
MUD	-0.659	-0.889	-1.490**
	(1.158)	(0.994)	(0.595)
ATE	-0.610	-0.742	-0.920
	(1.239)	(1.017)	(0.586)
Demographics			
Black	0.079	0.108	0.049
	(0.126)	(0.139)	(0.153)
Asian	-0.900***	-0.716**	-0.428
	(0.257)	(0.263)	(0.268)
Other	0.263*	0.177	0.153
	(0.151)	(0.155)	(0.174)
Hispanic	-0.157	-0.165	0.315**
	(0.203)	(0.207)	(0.147)
Age 18–24	-0.168	0.270	0.087
	(0.209)	(0.199)	(0.192)
Age 25–29	-0.036	0.242	0.156
	(0.198)	(0.211)	(0.190)
Age 30–34	-0.088	0.100	-0.169
	(0.180)	(0.192)	(0.255)
Age 35–49	0.015	0.049	0.255**
	(0.144)	(0.142)	(0.129)
Not Married	0.425**	0.339*	0.133
	(0.158)	(0.174)	(0.170)
Never Married	0.134	-0.045	0.026
	(0.136)	(0.138)	(0.162)
Kids	-0.086**	0.003	-0.057
	(0.043)	(0.050)	(0.059)
1 HH Member > 65	0.077	-0.011	0.382**
	(0.199)	(0.197)	(0.135)
2+ HH Members > 65	-0.147	-0.086	-0.155
	(0.221)	(0.229)	(0.224)
Small CBSA	0.371**	0.340**	0.536***
	(0.114)	(0.121)	(0.118)
No CBSA	0.084	-0.053	0.025
	(0.141)	(0.150)	(0.170)

Table 20

(Cont.)

	Demographic	Human Capital	Other Substances
Demographics (cont.)			
Constant	0.250	-0.783**	-1.780**
	(0.208)	(0.364)	(0.602)
Human Capital			
Less than HS		0.228	0.666
		(0.279)	(0.485)
High School		0.073	0.845*
		(0.277)	(0.487)
Some College		0.193	0.799
		(0.288)	(0.501)
Full Time		-0.194	-0.370**
		(0.131)	(0.140)
Part Time		-0.169	-0.166
		(0.133)	(0.148)
Disable No SSI		0.618***	0.702***
		(0.158)	(0.158)
Disable Collects SSI		0.684***	0.783***
		(0.131)	(0.138)
Income < $20,000		0.356**	0.711**
		(0.158)	(0.269)
Income $20-50,000		0.302*	0.870**
		(0.156)	(0.269)
Insurance		-0.385*	-0.416**
		(0.221)	(0.186)
Food Stamp		0.098	0.090
		(0.121)	(0.106)
Public Assistance		0.224	-0.018
		(0.182)	(0.197)
Family SSI		0.439***	0.564***
		(0.100)	(0.120)
1 Phone		0.064	-0.275**
		(0.120)	(0.114)
2+ Phones		-0.062	-0.399*
		(0.245)	(0.227)

Table 20

(Cont.)

	Demographic	Human Capital	Other Substances
Other Substances			
Alcohol			0.943***
			(0.174)
Nicotine			-0.198
			(0.127)
Hard Drugs			0.684**
			(0.267)
β_u	0.473	0.508	0.849**
	(0.685)	(0.585)	(0.395)
N	8,075	8,075	8,075
	[4,966,773]	[4,966,773]	[4,966,773]

Notes. The dependent variable is annual ER visit. All models control for year and quarter of interview. Omitted categories include white, age 50-64, married, no household members over 65, respondent lives in CBSA with population over 1 million, college education, did not work last week, family income $50,000-$75,000, and no phone lines in household. Huber/White robust standard errors reported in parentheses and corrected for 2-stage estimation. Standard errors for ATE computed using Delta Method. Brackets contain "true" number of observations represented by weighted estimates. $^{*}p < 0.1$ $^{**}p < 0.05$ $^{***}p < 0.01$.

In general, the coefficients on the control variables remain similar to the exogenous model. However, the coefficients for the other substance use disorder variables shift substantially. Among females the coefficient for alcohol nearly doubles and becomes significant, while the coefficient for hard drugs increases from 0.27 to 0.45 and remains significant. Among males the magnitude of the alcohol coefficient also increases (and remains significant), while the coefficient for hard drugs increases from 0.27 to 0.68 and becomes significant. The 2SNLS results also indicate that the average effect of MUD on total ER visits differs from the average effect of hard drug use disorders since the hard drug coefficient is large and positive, while the marijuana coefficient is large and negative.

Given the evidence that MUD is endogenously correlated with ER visits due to unobserved factors, at least among females, the 2SNLS model is preferred to NLS. I select the final specification, which controls for demographics, human capital, and other substance use disorders, as the preferred specification for additional analysis. In addition to being the most comprehensive model, the results from this specification among males warrant further attention. Although the ATE is within the range of estimates from the female specifications, the estimate of β_M seems implausibly large, which may indicate that the ATE is overestimated. Therefore, this specification invites additional checks to determine whether this is a spurious result, or whether it reflects the true relationship between MUD and ER visits.

7.3. First-Stage Estimation

Although the endogeneity-corrected estimates differ from the exogenous ones, they are no more likely to be correct if the probability of MUD is not appropriately estimated in the first stage. To address this concern I survey the general predictions of the first stage before turning to examine the validity of the instrumental variables. Tables 21 and 22 show the first-stage estimation results for females and males, respectively.

Asian males are less likely than whites to suffer from MUD, while black females are more likely. Among both males and females age is negatively correlated with the probability of having MUD, while individuals who are not currently married or never have been married are significantly more likely to have MUD. Living outside of a core-based statistical area (i.e., rural residence) is also negatively correlated with MUD among both genders. For males, college graduates and full time workers are significantly less

likely to have MUD. For females, having children is negatively associated with MUD. However, among females education is uncorrelated with MUD, and income is uncorrelated with MUD among both genders.[32] The only control for financial resources that is significant among females is whether the respondent's current household has a single phone line, an outcome that is positively correlated with MUD. Whether or not the respondent reports a disability is uncorrelated with MUD, which helps to alleviate concerns of reverse-causality (i.e., that chronic poor health causes marijuana use that leads to MUD).

Table 21

First-Stage Probit Estimates of Marijuana Use Disorder for Females

	Demographic	**Human Capital**	**Other Substances**
Instrumental Variables			
Slightly Disapprove	-0.426***	-0.434***	-0.408***
	(0.079)	(0.079)	(0.086)
Strongly Disapprove	-0.710***	-0.716***	-0.623***
	(0.117)	(0.118)	(0.121)
IV Significance	57.93***	58.24***	41.22***
1st Stage	{0.000}	{0.000}	{0.000}
IV Significance	2.67	1.940	1.890
2nd Stage	{0.263}	{0.379}	{0.389}
Demographics			
Black	0.134*	0.120	0.282***
	(0.076)	(0.077)	(0.085)
Asian	0.034	0.068	0.272
	(0.232)	(0.231)	(0.251)

[32] Although it may seem odd that human capital is uncorrelated with MUD among females, these results are consistent with McGeary and French (2000) who find that education is uncorrelated with chronic drug use among both males and females, and that neither income nor employment status is correlated with chronic drug use among females. The negative correlation between employment and MUD among males is also consistent with McGeary and French (2000).

Table 21

(Cont.)

	Demographic	Human Capital	Other Substances
Demographics (cont.)			
Other	0.063	0.074	0.187
	(0.120)	(0.116)	(0.128)
Hispanic	-0.041	-0.046	0.105
	(0.091)	(0.089)	(0.095)
Age 18–24	0.847***	0.985***	1.035***
	(0.197)	(0.188)	(0.188)
Age 25–29	0.721***	0.860***	0.850***
	(0.204)	(0.195)	(0.190)
Age 30–34	0.661**	0.783***	0.806***
	(0.212)	(0.201)	(0.202)
Age 35–49	0.282	0.345*	0.319*
	(0.186)	(0.180)	(0.184)
Not Married	0.302**	0.274**	0.247*
	(0.128)	(0.125)	(0.129)
Never Married	0.294**	0.272**	0.213*
	(0.111)	(0.108)	(0.111)
Kids	-0.141***	-0.153***	-0.102**
	(0.033)	(0.037)	(0.038)
Pregnant	-0.037	-0.020	0.035
	(0.082)	(0.083)	(0.087)
1 HH Member > 65	-0.093	-0.097	-0.047
	(0.130)	(0.132)	(0.143)
2+ HH Members > 65	-0.023	-0.039	-0.035
	(0.189)	(0.191)	(0.216)
Small CBSA	0.033	0.025	0.016
	(0.066)	(0.065)	(0.071)
No CBSA	-0.089	-0.107	-0.116
	(0.109)	(0.112)	(0.121)
Constant	-2.461***	-2.878***	-3.199***
	(0.204)	(0.275)	(0.308)
Human Capital			
Less than HS		0.249	0.119
		(0.152)	(0.171)
High School		0.124	0.030
		(0.148)	(0.170)

Table 21

(Cont.)

	Demographic	Human Capital	Other Substances
Human Capital (cont.)			
Some College		0.111	0.017
		(0.158)	(0.177)
Full Time		-0.055	-0.034
		(0.083)	(0.089)
Part Time		0.084	0.141*
		(0.074)	(0.081)
Disable No SSI		0.172	0.103
		(0.150)	(0.164)
Disable Collects SSI		-0.008	0.059
		(0.176)	(0.186)
Income < $20,000		-0.012	-0.048
		(0.103)	(0.113)
Income $20–50,000		-0.048	-0.068
		(0.107)	(0.116)
Insurance		0.113	0.201*
		(0.109)	(0.118)
Food Stamp		0.088	0.039
		(0.074)	(0.081)
Public Assistance		0.074	0.057
		(0.071)	(0.074)
Family SSI		0.022	0.059
		(0.080)	(0.087)
1 Phone		0.143**	0.167**
		(0.069)	(0.073)
2+ Phones		-0.070	-0.101
		(0.173)	(0.203)
Other Substances			
Alcohol			0.605***
			(0.082)
Nicotine			0.314***
			(0.076)
Hard Drugs			0.956***
			(0.106)

Table 21

(Cont.)

	Demographic	Human Capital	Other Substances
N	19,101 [9,506,713]	19,101 [9,506,713]	19,101 [9,506,713]

Notes. The dependent variable is a binary indicator for marijuana use disorder. All models control for year and quarter of interview. Omitted categories include white, age 50-64, married, no household members over 65, respondent lives in CBSA with population over 1 million, college education, did not work last week, family income $50,000-$75,000, and no phone lines in household. Omitted IV category refers to "Neither agree nor disagree." { } Indicates p-value for joint test of variable significance. 2nd-Stage Significance refers to the results of a heuristic test for excludability of the IVs from the 2nd stage, in which the IVs are included in an NLS regression of ER visits and tested for joint significance. Brackets contain "true" number of observations represented by weighted estimates. Standard errors are in parentheses. $^{*}p < 0.1$ $^{**}p < 0.05$ $^{***}p < 0.01$.

Table 22

First-Stage Probit Estimates of Marijuana Use Disorder for Males

	Demographic	Human Capital	Other Substances
Instrumental Variables			
Somewhat Disapprove	-0.090	-0.092	-0.055
	(0.108)	(0.109)	(0.109)
Strongly Disapprove	-0.896***	-0.887***	-0.826***
	(0.106)	(0.105)	(0.105)
IV Significance	72.36***	71.26***	63.57***
1st Stage	[0.000]	[0.000]	[0.000]
IV Significance	2.49*	7.98***	6.90***
2nd Stage	[0.083]	[0.000]	[0.001]
Demographics			
Black	0.163	0.116	0.157
	(0.100)	(0.097)	(0.102)
Asian	-0.651**	-0.621**	-0.561**
	(0.240)	(0.250)	(0.257)
Other	0.149	0.126	0.165
	(0.161)	(0.150)	(0.171)
Hispanic	0.065	0.054	0.138
	(0.128)	(0.125)	(0.124)
Age 18–24	0.900***	0.876***	1.008***
	(0.194)	(0.187)	(0.189)

Table 22

(Cont.)

	Demographic	Human Capital	Other Substances
Demographics (cont.)			
Age 25–29	0.818***	0.825***	0.820***
	(0.197)	(0.188)	(0.198)
Age 30–34	0.641**	0.653**	0.629**
	(0.208)	(0.199)	(0.205)
Age 35–49	0.331*	0.349**	0.283
	(0.189)	(0.178)	(0.181)
Not Married	0.485**	0.451**	0.365**
	(0.158)	(0.158)	(0.161)
Never Married	0.256**	0.233*	0.131
	(0.118)	(0.120)	(0.127)
Kids	-0.043	-0.050	-0.031
	(0.035)	(0.037)	(0.038)
1 HH Member > 65	-0.196	-0.198	-0.234
	(0.133)	(0.140)	(0.156)
2+ HH Members > 65	-0.015	0.019	0.076
	(0.274)	(0.283)	(0.329)
Small CBSA	-0.003	-0.027	-0.047
	(0.092)	(0.088)	(0.092)
No CBSA	-0.251**	-0.275**	-0.318**
	(0.119)	(0.121)	(0.135)
Constant	-2.526***	-3.068***	-3.581***
	(0.187)	(0.406)	(0.478)
Human Capital			
Less than HS		0.928**	0.876**
		(0.303)	(0.361)
High School		0.812**	0.790**
		(0.304)	(0.364)
Some College		0.834**	0.857**
		(0.311)	(0.374)
Full Time		-0.234**	-0.241**
		(0.098)	(0.102)
Part Time		-0.085	-0.008
		(0.099)	(0.111)
Disable No SSI		-0.240	-0.229
		(0.150)	(0.181)

Table 22

(Cont.)

	Demographic	Human Capital	Other Substances
Human Capital (cont.)			
Disable Collects SSI		-0.039	-0.041
		(0.139)	(0.138)
Income < $20,000		-0.035	-0.008
		(0.129)	(0.138)
Income $20–50,000		-0.093	-0.090
		(0.119)	(0.127)
Insurance		-0.082	-0.024
		(0.109)	(0.125)
Food Stamp		0.036	-0.022
		(0.089)	(0.092)
Public Assistance		-0.057	-0.095
		(0.105)	(0.111)
Family SSI		-0.052	-0.094
		(0.098)	(0.107)
1 Phone		-0.111	-0.107
		(0.078)	(0.082)
2+ Phones		-0.024	0.054
		(0.196)	(0.246)
Other Substances			
Alcohol			0.737***
			(0.091)
Nicotine			0.336***
			(0.081)
Hard Drugs			0.960***
			(0.147)
N	8,075 [4,966,773]	8,075 [4,966,773]	8,075 [4,966,773]

Notes. The dependent variable is a binary indicator for marijuana use disorder. All models control for year and quarter of interview. Omitted categories include white, age 50–64, married, no household members over 65, respondent lives in CBSA with population over 1 Million, college education, did not work last week, family income $50,000-$75,000, and no phone lines in household. Omitted IV category refers to "Neither agree nor disagree." { } Indicates *p*-value for joint test of variable significance. 2nd-Stage Significance refers to the results of a heuristic test for excludability of the IVs from the 2nd stage, in which the IVs are included in an NLS regression of ER visits and tested for joint significance. Brackets contain "true" number of observations represented by weighted estimates. Standard errors are in parentheses.
$^{*}p < 0.1$ $^{**}p < 0.05$ $^{***}p < 0.01$.

All three other forms of substance use disorder are strongly and positively correlated with MUD, which is consistent with the positive omitted variables bias on β_M observed in the second stage. Given the large magnitude of these three coefficients relative to the other controls, it is likely that inclusion of the substance use disorder controls improves the accuracy of estimates of the linear index for MUD at the individual level, which should in turn improve the second stage estimates. This bolsters the case for the comprehensive substance use disorder specification being preferred to the human capital and demographic specifications.

Overall, the coefficients for the controls in the first stage do not seem problematic. However, consistency of the 2SNLS estimates requires that the instrumental variables (IVs) are valid. Validity of the IVs requires both that they strongly correlated with the probability of MUD, conditional on all other control variables, and that conditional on all other control variables, they are uncorrelated with the error term for ER visits. Recall from section 4 that the single instrument (measured by two variables) is the respondent's opinion of another adult trying marijuana once or twice. Responses include "neither approve nor disapprove," "somewhat disapprove," and "strongly disapprove." Indicators for "somewhat" and "strong" disapproval are included in the first stage, with "neither approve nor disapprove" as the omitted category.

The first stage results suggest that the IVs are jointly significant among males and females. The chi-squared value for females is roughly 58 for both the demographic and human capital models, and diminishes to 41 in the substance use disorder specification. For males the IVs have even more predictive power, with a chi-squared value of 71–72 in

the demographic and human capital specifications and 64 in the substance use disorder specification. These results suggest that the IVs are sufficiently strong.

The theoretical justification for exclusion of the IVs from the equation for ER visits rests on the assertion that conditional on all of the available controls it is reasonable to assume that one's opinion of adults trying marijuana is independent of health or of other behavior that may lead to ER visits. Unfortunately, when the model is just-identified, there is no formal empirical test that the instrumental variables are excludable from the equation for ER visits. One heuristic test used previously in the substance use literature is to include the instruments directly in an exogenous version of the second stage NLS model and test whether the IVs are jointly significant (see e.g., Kenkel & Terza, 2001). Among females the IVs are jointly insignificant, with a p-value ranging from 0.26 to 0.39. However, among males the IVs are jointly significant in all three models. Interestingly, the IVs become more strongly correlated with ER visits after controls for human capital are added, and the controls for other forms of substance use disorder barely diminish this joint significance, such that the statistical case for excludability is greatest in the least comprehensive model.

These results provide only circumstantial rather than direct evidence in favor of the instrument (for females) or against them (for males). However, the response of the IVs to the inclusion of additional controls among males suggests that the instrument is capturing statistical noise rather than any true underlying information about individual demand for ER visits. Moreover, the evidence suggests that the IVs are excludable among females, and the female estimates are broadly similar to those for males (with the

exception of β_M in the final specification.) Therefore, I continue under the assumption that the IVs are valid. To provide more support for the 2SNLS estimates, I run several additional specifications. Results of these robustness checks are presented in the next section.

7.4. Robustness Checks

The purpose of the analyses in this section is to not only demonstrate the robustness of the results from the preferred specification, but also to try to determine if there are any previously omitted factors that may explain why marijuana is negatively correlated with ER visits (and for males, why the estimated coefficient is improbably large). The first of the extended specifications includes disease indicators that proxy for individual health. The second extended specification includes measures of attitudes/beliefs, consisting of measures of preference for risky and dangerous behavior, as well as religiosity. The third extended specification includes controls for illegal behavior. Unlike the first set of specifications, each additional set of controls enters in isolation, rather than building on the previous set of controls (i.e., the "illegal activities" controls are introduced separately from the "attitudes" controls). As a final robustness check, I re-run the substance use disorder specification without weighting the first or second stage. Results are presented in Table 23 for females and Table 24 for males.

Among females, inclusion of controls for health slightly increases the magnitude of β_M and β_u although β_M becomes insignificant. Four diseases, including asthma, bronchitis, ulcers, and "other diseases" are positively correlated with ER visits, while none are significantly negatively correlated with ER visits. The increasing coefficient

magnitude for β_M and β_u is consistent with the assumed positive correlation between negative health events (i.e., disease) and MUD, although the effect does not seem to be particularly large.

Table 23

2SNLS Robustness Checks for Females

	Other Substances	Health	Attitudes	Illegal Activities	Unweighted
Variable of Interest					
MUD	-0.745*	-0.789	-0.528	-1.467**	-0.625**
	(0.447)	(0.497)	(0.410)	(0.619)	(0.204)
ATE	-0.780	-0.867	-0.615	-1.245*	-0.629**
	(0.561)	(0.607)	(0.530)	(0.721)	(0.273)
Demographics					
Black	0.126*	0.123	0.101	0.133*	0.098**
	(0.069)	(0.098)	(0.075)	(0.073)	(0.031)
Asian	-0.738***	-0.655***	-0.759***	-0.691***	-0.672**
	(0.156)	(0.173)	(0.165)	(0.157)	(0.220)
Other	0.016	-0.019	-0.020	0.015	0.001
	(0.113)	(0.125)	(0.120)	(0.113)	(0.047)
Hispanic	-0.368***	-0.347***	-0.384***	-0.358***	-0.248***
	(0.073)	(0.100)	(0.076)	(0.074)	(0.042)
Age 18-24	0.554***	0.576***	0.504***	0.540***	0.480***
	(0.103)	(0.119)	(0.101)	(0.110)	(0.060)
Age 25-29	0.441***	0.500***	0.396***	0.405***	0.387***
	(0.101)	(0.117)	(0.105)	(0.100)	(0.060)
Age 30-34	0.309**	0.300**	0.253**	0.252**	0.315***
	(0.114)	(0.121)	(0.121)	(0.115)	(0.067)
Age 35-49	0.237**	0.261**	0.217**	0.150	0.113**
	(0.094)	(0.104)	(0.099)	(0.101)	(0.053)
Not Married	-0.100	-0.070	-0.090	-0.125	-0.012
	(0.088)	(0.095)	(0.090)	(0.094)	(0.039)
Never Married	-0.248**	-0.113	-0.245**	-0.275**	-0.197***
	(0.083)	(0.109)	(0.083)	(0.088)	(0.034)
Kids	-0.118***	-0.075*	-0.112***	-0.134***	-0.060***
	(0.029)	(0.039)	(0.029)	(0.033)	(0.013)
1 HH Member > 65	-0.107*	-0.129*	-0.103	-0.096	-0.129**
	(0.064)	(0.072)	(0.066)	(0.065)	(0.043)
2+ HH Members > 65	0.264	0.197	0.272	0.154	-0.028
	(0.350)	(0.309)	(0.337)	(0.343)	(0.129)
Small CBSA	0.134**	0.089	0.147**	0.164**	0.141***
	(0.064)	(0.072)	(0.066)	(0.072)	(0.027)

Table 23

(Cont.)

	Other Substances	Health	Attitudes	Illegal Activities	Unweighted
Demographics (cont.)					
No CBSA	0.003	0.019	0.021	0.043	0.013
	(0.110)	(0.111)	(0.110)	(0.108)	(0.045)
Constant	-0.141	-0.249	-0.142	-0.079	-0.306**
	(0.283)	(0.322)	(0.286)	(0.273)	(0.123)
Less than HS	0.022	0.087	-0.032	-0.095	0.219**
	(0.273)	(0.279)	(0.289)	(0.295)	(0.080)
High School	-0.190	-0.160	-0.234	-0.294	0.081
	(0.275)	(0.277)	(0.291)	(0.301)	(0.080)
Some College	-0.123	-0.159	-0.167	-0.231	0.082
	(0.278)	(0.283)	(0.295)	(0.311)	(0.081)
Full Time	-0.029	-0.035	-0.024	0.016	-0.075**
	(0.070)	(0.061)	(0.071)	(0.084)	(0.038)
Part Time	-0.202**	-0.214***	-0.209**	-0.192**	-0.179***
	(0.065)	(0.064)	(0.067)	(0.066)	(0.043)
Disable No SSI	0.684***	0.483***	0.701***	0.700***	0.653***
	(0.097)	(0.105)	(0.099)	(0.109)	(0.043)
Disable Collects SSI	-0.050	0.078	-0.072	-0.103	-0.007
	(0.106)	(0.122)	(0.111)	(0.120)	(0.043)
Income < $20,000	0.215*	0.098	0.252**	0.207*	0.096
	(0.113)	(0.155)	(0.109)	(0.119)	(0.066)
Income $20–50,000	0.213*	0.074	0.254**	0.228*	0.096
	(0.109)	(0.147)	(0.107)	(0.118)	(0.066)
Insurance	-0.085	-0.151	-0.107	-0.098	-0.011
	(0.087)	(0.114)	(0.088)	(0.094)	(0.053)
Food Stamp	0.177**	0.093	0.182**	0.197**	0.178***
	(0.074)	(0.107)	(0.075)	(0.075)	(0.030)
Public Assistance	0.074	0.113	0.079	0.108	0.067**
	(0.083)	(0.101)	(0.091)	(0.095)	(0.028)
Family SSI	0.160**	0.179**	0.149**	0.159**	0.191***
	(0.063)	(0.069)	(0.062)	(0.066)	(0.036)
1 Phone	0.003	0.056	0.004	0.051	-0.022
	(0.061)	(0.066)	(0.062)	(0.074)	(0.026)
2+ Phones	0.192	0.167	0.201	0.294	0.090
	(0.169)	(0.144)	(0.176)	(0.209)	(0.074)

Table 23

(Cont.)

	Other Substances	Health	Attitudes	Illegal Activities	Unweighted
Other Substances					
Alcohol	0.237*	0.263**	0.163	0.255*	0.174***
	(0.132)	(0.116)	(0.118)	(0.131)	(0.051)
Nicotine	0.120	0.049	0.138*	0.112	0.154***
	(0.074)	(0.091)	(0.079)	(0.090)	(0.028)
Hard Drugs	0.450***	0.435**	0.381**	0.421**	0.451***
	(0.131)	(0.142)	(0.130)	(0.141)	(0.058)
Health Controls					
Asthma		0.292***			
		(0.078)			
Bronchitis		0.457***			
		(0.125)			
Pneumonia		-0.078			
		(0.135)			
Sinusitis		0.007			
		(0.122)			
Stroke		0.248			
		(0.220)			
High BP		0.033			
		(0.089)			
Heart Disease		0.170			
		(0.123)			
Diabetes		-0.104			
		(0.112)			
Hepatitis		-0.097			
		(0.232)			
STD		0.075			
		(0.148)			
HIV		-0.072			
		(0.237)			
Ulcer		0.624***			
		(0.152)			
Other Disease		0.560**			
		(0.236)			

Table 23

(Cont.)

	Other Substances	Health	Attitudes	Illegal Activities	Unweighted
Health Controls (cont.)					
Depression		0.061			
		(0.101)			
Attitude Controls					
Risk			0.008		
			(0.089)		
Danger			0.102		
			(0.076)		
Seatbelt			0.112		
			(0.082)		
Religion Important			0.030		
			(0.073)		
Service Attender			0.094		
			(0.068)		
Religious Friends			0.181**		
			(0.068)		
Religious Decisions			-0.165**		
			(0.070)		
Illegal Activity Controls					
Ever Arrested				0.193**	
				(0.068)	
Offered Drugs				0.439**	
				(0.180)	
β_u	0.523**	0.553**	0.409**	0.898**	0.396***
	(0.233)	(0.259)	(0.203)	(0.335)	(0.106)
N	19,101 [9,506,713]	19,101 [9,506,713]	19,101 [9,506,713]	19,101 [9,506,713]	19,101 [9,506,713]

Notes. Dependent variable is annual ER visits. All models control for year and quarter of interview. Omitted categories include white, age 50-64, married, no household members over 65, respondent lives in CBSA with population over 1 Million, college education, did not work last week, family income $50,000-$75,000, and no phone lines in household. Huber/White robust standard errors reported in parentheses and corrected for 2-stage estimation. Standard errors for ATE computed using Delta Method. $^{*}p < 0.1$ $^{**}p < 0.05$ $^{***}p < 0.01$.

Table 24

2SNLS Robustness Checks for Males

	Other Substances	Health	Attitudes	Illegal Activities	Unweighted
Variable of Interest					
MUD	-1.490**	-1.678***	-1.304**	-1.533**	0.041
	(0.595)	(0.451)	(0.591)	(0.612)	(0.284)
ATE	-0.920	-1.032	-0.853	-0.956	0.047
	(0.586)	(0.444)	(0.564)	(0.582)	(0.327)
Demographics					
Black	0.049	0.050	-0.023	-0.010	0.113**
	(0.153)	(0.126)	(0.156)	(0.139)	(0.054)
Asian	-0.428	-0.359	-0.528**	-0.225	-0.358
	(0.268)	(0.239)	(0.265)	(0.240)	(0.280)
Other	0.153	0.188	0.021	0.123	0.064
	(0.174)	(0.151)	(0.193)	(0.174)	(0.073)
Hispanic	0.315**	0.281**	0.168	0.202	-0.093
	(0.147)	(0.142)	(0.149)	(0.135)	(0.072)
Age 18-24	0.087	0.367**	0.130	0.095	0.262**
	(0.192)	(0.182)	(0.193)	(0.190)	(0.095)
Age 25-29	0.156	0.400**	0.154	0.189	0.276**
	(0.190)	(0.190)	(0.185)	(0.194)	(0.090)
Age 30-34	-0.169	0.258	-0.148	-0.160	0.003
	(0.255)	(0.276)	(0.230)	(0.237)	(0.113)
Age 35-49	0.255**	0.344**	0.202	0.190	0.074
	(0.129)	(0.119)	(0.133)	(0.125)	(0.078)
Not Married	0.133	0.114	0.164	0.089	0.024
	(0.170)	(0.135)	(0.175)	(0.162)	(0.072)
Never Married	0.026	0.051	0.023	-0.043	-0.196**
	(0.162)	(0.126)	(0.158)	(0.149)	(0.062)
Kids	-0.057	-0.025	-0.065	-0.048	0.009
	(0.059)	(0.049)	(0.056)	(0.060)	(0.023)
1 HH Member > 65	0.382**	0.339**	0.306**	0.367**	0.053
	(0.135)	(0.126)	(0.130)	(0.129)	(0.079)
2+ HH Members > 65	-0.155	-0.201	-0.393*	-0.129	-0.138
	(0.224)	(0.277)	(0.233)	(0.234)	(0.212)
Small CBSA	0.536***	0.472***	0.516***	0.526***	0.198***
	(0.118)	(0.113)	(0.123)	(0.116)	(0.050)

Table 24

(Cont.)

	Other Substances	Health	Attitudes	Illegal Activities	Unweighted
Demographics (cont.)					
No CBSA	0.025	-0.116	-0.061	0.030	0.048
	(0.170)	(0.183)	(0.178)	(0.170)	(0.078)
Constant	-1.780**	-1.675***	-1.648**	-1.668***	-0.491**
	(0.602)	(0.434)	(0.510)	(0.497)	(0.221)
Human Capital					
Less than HS	0.666	0.583*	0.438	0.425	0.137
	(0.485)	(0.303)	(0.399)	(0.360)	(0.160)
High School	0.845*	0.672**	0.609	0.628*	0.228
	(0.487)	(0.300)	(0.398)	(0.355)	(0.158)
Some College	0.799	0.730**	0.479	0.649*	0.219
	(0.501)	(0.325)	(0.417)	(0.376)	(0.164)
Full Time	-0.370**	-0.340**	-0.289**	-0.356**	-0.085
	(0.140)	(0.127)	(0.131)	(0.133)	(0.074)
Part Time	-0.166	-0.160	-0.208	-0.170	-0.015
	(0.148)	(0.130)	(0.151)	(0.158)	(0.079)
Disable No SSI	0.702***	0.586***	0.722***	0.718***	0.845***
	(0.158)	(0.143)	(0.152)	(0.150)	(0.074)
Disable Collects SSI	0.783***	0.596***	0.775***	0.761***	0.731***
	(0.138)	(0.132)	(0.131)	(0.140)	(0.070)
Income < $20,000	0.711**	0.575**	0.559**	0.703**	0.106
	(0.269)	(0.206)	(0.227)	(0.271)	(0.105)
Income $20–50,000	0.870**	0.529**	0.724***	0.811**	0.013
	(0.269)	(0.185)	(0.218)	(0.264)	(0.107)
Insurance	-0.416**	-0.457**	-0.524**	-0.337*	0.000
	(0.186)	(0.188)	(0.187)	(0.197)	(0.087)
Food Stamp	0.090	0.057	0.103	0.095	0.145**
	(0.106)	(0.103)	(0.101)	(0.102)	(0.048)
Public Assistance	-0.018	0.130	0.058	0.064	0.003
	(0.197)	(0.163)	(0.196)	(0.182)	(0.057)
Family SSI	0.564***	0.446***	0.596***	0.508***	0.445***
	(0.120)	(0.105)	(0.113)	(0.119)	(0.062)
1 Phone	-0.275**	-0.194*	-0.201*	-0.156	-0.106**
	(0.114)	(0.111)	(0.104)	(0.105)	(0.045)

Table 24

(Cont.)

	Other Substances	Health	Attitudes	Illegal Activities	Unweighted
Human Capital (cont.)					
2+ Phones	-0.399*	-0.259	-0.293	-0.326	-0.131
	(0.227)	(0.245)	(0.202)	(0.236)	(0.136)
Other Substances					
Alcohol	0.943***	0.908***	0.836***	0.738***	0.318***
	(0.174)	(0.124)	(0.156)	(0.151)	(0.063)
Nicotine	-0.198	-0.115	-0.213*	-0.320**	0.097**
	(0.127)	(0.118)	(0.127)	(0.125)	(0.048)
Hard Drugs	0.684**	0.767***	0.519**	0.451**	0.394***
	(0.267)	(0.165)	(0.239)	(0.221)	(0.090)
Health Controls					
Asthma		0.243			
		(0.168)			
Bronchitis		-0.451*			
		(0.231)			
Pneumonia		0.461**			
		(0.221)			
Sinusitis		0.541			
		(0.330)			
Stroke		0.078			
		(0.289)			
High BP		0.210*			
		(0.114)			
Heart Disease		0.403**			
		(0.136)			
Diabetes		0.525***			
		(0.119)			
Hepatitis		-0.237			
		(0.172)			
STD		-0.636**			
		(0.286)			
HIV		-0.191			
		(0.202)			
Ulcer		0.147			
		(0.230)			

Table 24

(Cont.)

	Other Substances	Health	Attitudes	Illegal Activities	Unweighted
Health Controls (cont.)					
Other Disease		0.216			
		(0.244)			
Depression		-0.124			
		(0.155)			
Attitude Controls					
Risk			-0.044		
			(0.127)		
Danger			0.424***		
			(0.117)		
Seatbelt			-0.142		
			(0.161)		
Religion Important			-0.178		
			(0.127)		
Service Attender			-0.027		
			(0.136)		
Religious Friends			0.300**		
			(0.112)		
Religious Decisions			0.253*		
			(0.133)		
Illegal Activity Controls					
Ever Arrested				0.379***	
				(0.114)	
Offered Drugs				0.648***	
				(0.150)	
β_u	0.849**	0.970***	0.729*	0.822**	-0.038
	(0.395)	(0.248)	(0.383)	(0.377)	(0.154)
N	8,075 [4,966,773]	8,075 [4,966,773]	8,075 [4,966,773]	8,075 [4,966,773]	8,075 [4,966,773]

Notes. Dependent variable is annual ER visits. All models control for year and quarter of interview. Omitted categories include white, age 50-64, married, no household members over 65, respondent lives in CBSA with population over 1 million, college education, did not work last week, family income $50,000-$75,000, and no phone lines in household. Huber/White robust standard errors reported in parentheses and corrected for 2-stage estimation. Standard errors for ATE computed using Delta Method.
*$p < 0.1$ **$p < 0.05$ ***$p < 0.01$.

Inclusion of the attitudes/beliefs controls among females attenuates β_M and β_u by roughly 20–25% relative to the substance use disorder specification, indicating an (insignificant) ATE of -0.62, compared to the -0.78 predicted by the preferred specification. Controls for risk and danger preference, as well as seatbelt use are insignificantly correlated with ER visits. Preferring one's friends to share one's religion is positively correlated with ER visits, while allowing one's religion to affect one's decisions is negatively correlated with ER visits. This suggests that among females, religious beliefs, but not attitudes towards risk-taking, are a significant confounder of the relationship between MUD and ER visits. However, it is also possible that the potential endogeneity of beliefs to MUD or health is downward biasing the estimates.

Including controls for illegal activities nearly doubles the magnitude of both β_M and β_u, rendering both coefficients roughly equivalent to the estimates from the preferred specification among males. β_M and the ATE are both significant, although the coefficient estimate is an implausibly large -1.47 (the next largest coefficient in magnitude is 0.700 for individuals who report disability but do not collect SSI). Being arrested and being offered drugs are both positively and significantly correlated with ER visits. Although both variables are likely confounders of the relationship between MUD and ER visits, the magnitude of the new coefficient estimate for MUD suggests that these variables may be endogenous to the model and therefore biasing the coefficient estimates. Regardless of the source of the error, it seems implausible that the inflated estimates reflect the true relationship between MUD and ER visits.

Lastly, in the unweighted model, β_M and β_u are both attenuated. However, dropping the weights also decreases the standard errors so that β_M and the ATE are both significant. The unweighted model predicts that individuals with MUD utilize the ER -0.69 times per year on average, compared to the estimated ATE of -0.79 from the weighted model. In general, the other coefficients from the unweighted model are fairly consistent with those from the weighted model, suggesting that the weights do not have a dramatic effect on the results among females.

Among males inclusion of the health variables increases the magnitude of both β_M and β_u, consistent with the female results. Pneumonia, high blood pressure, heart disease, and diabetes are all positively and significantly correlated with ER visits, although bronchitis and sexually transmitted diseases are both significantly negatively correlated with ER visits. The downward shift in the value of β_M and β_u supports the assumed positive relationship between MUD and negative health events, but as with females the magnitude of the shift remains comparatively small. Considering that health should be one of the key, if not primary, determinants of ER visits, this suggests that the relationship between MUD and health events (as I am able to measure them) is not particular large.

Inclusion of controls for attitudes/beliefs attenuates β_M and β_u, which is also consistent with the female results. However, β_M remains an implausibly large -1.30, indicating that this set of variables cannot explain the inflation of the coefficient caused by inclusion of the substance use disorder controls. A preference for dangerous behavior is significantly positively correlated with ER visits, as is the preference that one's friends

share one's religious beliefs, and whether one allows religion to influence decision making. This, coupled with the attenuation of β_M and β_u, indicates that attitudes toward risky behavior, as well as religious beliefs, are significant confounders of the relationship between MUD and ER visits (although preferring one's friends to share one's religious beliefs has the opposite sign as in the results for females). However, these confounders cannot account for the inflated estimate of β_M and β_u.

In contrast to the female specification, including controls for illegal activities has virtually no effect on β_M or β_u. Although being arrested or offered drugs are both positively and significantly correlated with ER visits, it appears that these variables are not significant confounders of the model among males, and provide no new information regarding the accuracy of the coefficient estimates among males.

In another contrast to females, the unweighted model differs drastically from the weighted model among males. β_M and β_u both become essentially 0, while many of the other coefficients shift substantially (although no coefficients that are significant in the weighted model become significant with an opposing sign). It is still the case that the weighted results for males are considered closer to the "true" result since the weighting accounts for potentially endogenous sampling, and also makes the sample more closely correspond to the true population. However, the unweighted results may reflect the possibility that the male subsample is more heterogeneous than the female subsample, such that the weights have a greater influence on the results.

In general, the extended specifications support the finding that MUD and ER visits are negatively correlated, although the results are not consistently significant for

females, and questionably large among males. However, with the exception of the controls for illegal activities among females, no set of variables drastically affects the results. Even this inflated estimate, though potentially inconsistent, is larger in magnitude than the preferred model, and thus cannot account for the negative estimate of β_M. Although β_M is not consistently significant for females, β_u remains significant in all models, supporting the hypothesis that MUD and ER visits are endogenously correlated through unobservable factors.

8. Discussion

In this study I find evidence that marijuana use disorder is not positively correlated with ER visits, and may actually be negatively correlated with ER visits. The precise effect is uncertain as estimates of the ATE are generally insignificant, but weighted 2SNLS estimates range from roughly -0.62 to -1.23 visits among females, and from -0.61 to -1.03 visits among males (although the upper bound estimates are considered to be implausibly large, a result I will return to shortly). Although MUD and ER visits are predicted to have a positive relationship among females in the exogenous specifications, all 2SNLS models for females reject the null hypothesis that MUD is exogenous to ER visits, suggesting that the exogenous estimates are biased. Among males exogeneity of MUD is rejected in the substance use disorder specification (and all subsequent weighted robustness checks) although this may be attributable to an inflated coefficient estimate for the unobserved heterogeneity rather than a true effect. However, the coefficient estimates for unobserved heterogeneity among males are roughly similar to that among females, and the lack of significance may be partly due to the fact that the

male subsample is less than half the size of the female subsample. This result, along with the theoretical considerations discussed in Section 4, suggests that unobservable factors jointly correlated with MUD and ER visits are significantly confounding the estimated relationship between these two outcomes. This result is consistent with McGeary and French (2000) who find that chronic drug use (marijuana and/or hard drugs) is endogenous to the probability of any ER visit, and French et al. (2011) who find that time-invariant endogeneity may be biasing estimates of the effect of heavy drug use on total ER visits and hospital admissions. In all weighted models, for both genders, unobserved heterogeneity is positively correlated with ER visits, suggesting an upward omitted variables bias for the relationship between MUD and ER visits. Thus, while the true relationship between MUD and ER visits may be null, the evidence suggests that even the most comprehensive NLS specification considered is positively biased by omitted variables.

The validity of the 2SNLS results, and subsequent rejection of exogeneity, rests on the validity of the instrumental variables (IVs) utilized in the first stage. In all specifications the instrumental variables are significantly correlated with the probability of MUD with a chi-squared statistic over 40 for females and over 60 for males. This indicates that the IVs are sufficiently strong. Although I cannot directly test for the excludability of the IVs, I perform a heuristic test by including the IVs in an NLS model for ER visits. Among females the IVs are insignificantly correlated with ER visits. Among males the IVs are jointly significant with ER visits, although the largest F-statistic is less than 8.0 in magnitude. The IVs become more significant when controls

for human capital and financial resources are added, and controls for other substance use disorders barely attenuates these significant results. This suggests that among males the IVs may be capturing statistical noise rather than a true relationship between the instrument and ER visits, particularly given the theoretical argument in support of the IVs.[33] Moreover, the similarity of the estimates for males and females suggests that the estimates for males are not being substantially biased by potentially invalid instruments. (Put another way, it would seem odd for the instruments to bias the estimates so similarly for both males and females when the IVs seem to have a different relationship to ER visits among males compared to females. On the other hand, the true relationship between MUD and ER visits being similar for males and females is not particularly odd).

Another result of interest is the difference in effect between hard drugs and marijuana. In all 2SNLS models hard drug use disorder is positively and significantly correlated with ER visits, a result consistent with the previous literature. However, the predicted relationship between MUD and ER visits is negative in all of the 2SNLS models (except for the unweighted male specification). In the NLS model for males, the coefficient for hard drug use is positive and significant, while the coefficient for MUD is negative and insignificant. This suggests that even if the exogenous model is correct, the effect of MUD still differs from that of other drugs among males.

The most troubling result is the large discrepancy between the weighted and unweighted model for males. The most likely explanation is simply that the male

[33] In an unreported specification I repeat the test with all control variables simultaneously, including the controls for health, attitudes/beliefs, and illegal activities. The instruments remain jointly significant at the 0.001 level. This provides further evidence that these results may be spurious.

subsample is more heterogeneous than the female subsample. Although the unweighted model differs from the weighted model among males, the weighted models for males produce broadly similar results to the weighted models among females. This suggests that, rather than the heterogeneity of the male subsample rendering the weights invalid, the heterogeneity of the male subsample makes the weighs even more important, since the heterogeneity is exacerbated by the failure to appropriately weight the estimates. Even in the "worst case" scenario that the unweighted model is the correct one, this would still indicate a null result for MUD, which would remain in contrast both to the positive and significant coefficient for hard drugs indicated by the unweighted model, and to the previous literature.

An additional pair of other results stands out as requiring additional discussion. The first is the inflated estimates of β_M and β_u in the substance use disorder specification of the 2SNLS model among males. In this specification, inclusion of controls for alcohol, nicotine, and hard drug use disorders increases β_M by more than 50%. Moreover, the magnitude of the inflated coefficient is nearly twice the size of the next largest control variable coefficient, indicating that it is likely inaccurate. Additional controls for health, attitudes/beliefs, and illegal activities fail to reduce the estimates of β_M or β_u to the more reasonable (though still large) values observed for males in the demographic and human capital specifications. This suggests that the substance use disorder model among males may be suffering bias from endogenous regressors. This possibility is supported by the fact that numerous coefficients were inflated compared to the demographic and human capital specifications (see e.g., Age 35-49, 1 household

member over 65, all education controls, full time work, both income variables, and the constant). This across-the-board inflation did not occur among females, so whatever caused these other coefficients to be inflated is likely the same factor that inflated β_M and β_u. One additional possibility is that the model is not over-inflated relative to the other models. The estimate of ATE, -0.92, is nearly identical to the estimate among females in the model which only controlled for demographic and human capital control. It is possible that the inflation of the other coefficients (some negative some positive) offset the inflation of β_M, such that the predicted ATE remained roughly consistent.

A similar inflation of β_M and β_u occurs among females with the inclusion of controls for ever being arrested, and being offered drugs in the past 30 days. Unlike with males, the ATE estimate of -1.23 for this "inflated" model is nearly 1/3 larger than estimates from any of the other specifications among either gender, and is therefore unlikely to be correct. Also in contrast to males, this model did not experience across-the-board inflation. However, several of the coefficients that inflated among males (e.g., education controls) shift upward noticeably compared to other weighted specifications among the robustness checks. The control for age 35-49 deflates noticeably, which is in contrast to the results among males but consistent with overall bias, which may lend more credence to the possibility that the "illegal activities" variables are endogenous to ER visits and therefore biasing the estimates.

It is uncertain exactly why β_M and β_u are inflated to implausibly large values in certain specifications, although bias from endogenous regressors is a likely culprit. In both cases estimates become substantially more negative, rather than flipping signs and

becoming positive. Moreover, among males, the "inflated" estimate is not substantially attenuated by controls for health, attitudes/beliefs, or illegal activities. Therefore, it seems reasonable to assert that these models provide evidence against a positive relationship between MUD and ER visits, even if they cannot conclusively demonstrate a negative relationship.

To summarize, in the preceding analyses I find that marijuana use disorder is not positively correlated with ER visits and may be negatively correlated. Among females, hard drug use disorder is significantly positively correlated with ER visits in the NLS specification, while among both males and females hard drug use disorders are significantly and positively correlated with ER visits in 2SNLS models that correct for the endogeneity of MUD. This result is consistent with previous literature regarding drug use and acute healthcare utilization (McGeary & French, 2000; French et al., 2000; French et al., 2011). This suggests that the effect of marijuana use disorder on ER visits does, in fact, differ meaningfully from the effect of hard drug use disorders.

I also find evidence that the relationship between MUD and ER visits is confounded by unobserved heterogeneity, particularly for females. The results suggest a positive relationship between unobserved heterogeneity and ER visits that is significant in all specifications for females. Although insignificant for males, the magnitude of the coefficients remains large and roughly consistent with the estimates for females, and it's possible that the insignificance is due to the small size of the male subsample relative to the female subsample. The positive coefficient for unobserved heterogeneity suggests that β_M is being upward biased in the exogenous model, which supports the argument that

a null effect is the lower bound for the relationship between MUD and ER visits. The endogeneity of MUD, like the negative relationship between MUD and ER visits, depends on the validity of the instrumental variables used in the first stage. Although the validity of the instruments remains unprovable, an inconsistent 2SNLS would not undermine the evidence suggesting that MUD is uncorrelated with ER visits among males, a finding at odds with the previous literature.

These findings do, however, come with several caveats and shortcomings. As discussed in Section 5, the lack of state-level controls introduces omitted variables bias that may not be completely accounted for by the instrumental variables. The study also lacks controls for the sampling units, which prevents me from estimating clustered standard errors. Thus it is possible that the estimated standard errors are too small. However, even using potentially deflated standard errors, all exogenous models among males fail to reject the null hypothesis that MUD and ER visits are uncorrelated, while the most comprehensive exogenous model among females is only significant at the 10% level. Therefore, it is unlikely that the use of clustered standard errors would alter the evidence that the lower bound of the relationship between MUD and ER visits is zero.

Lastly, despite the evidence that MUD is uncorrelated with ER visits, this result only refers to the partial relationship between MUD and ER visits. It cannot, for instance, account for the effect of MUD on education, income, or probability of dependence upon other substances, which all may in turn affect demand for ER visits. I also cannot control for the duration of the marijuana use disorder. While it appears that MUD is positively correlated with negative health events, reduction in the health stock

may be gradual, and therefore those with longer spells of MUD may have a stronger demand for medical care. A corollary to this is that I cannot control for heterogeneous effects of marijuana dependence within the population. Although the estimation goal is an average for the entire population of Medicaid recipients, the magnitude of the relationship between marijuana dependence and acute medical care may differ between young and old, white and nonwhite, etc. Such differences may be relevant to policy considerations.

Ideally, future research will not only incorporate state-level controls, but also leverage state-level policy regarding marijuana, alcohol, or tobacco, into viable instruments that would allow the model to be overidentified, and shore up the 2SNLS results. Future research should also focus on examining whether the results differ between individuals by age or race. These effects will help policymakers to have a clearer understanding of the expected long-term ramifications of marijuana policy.

CHAPTER III

ESSAY 3: INVESTIGATING THE NEGATIVE RELATIONSHIP BETWEEN WAGES AND OBESITY: NEW EVIDENCE FROM THE WORK, FAMILY, AND HEALTH NETWORK

1. Abstract

A substantial literature has established that obesity is negatively associated with wages, particularly among females. However, prior research has found limited evidence in support of the factors hypothesized to underlie this relationship. Utilizing data from a single U.S. telecommunications firm I add to the literature by exploring the influence of productivity and discrimination on wages for workers who are and who are not obese, using control variables that are typically unavailable in national-level datasets. Consistent with previous research, I find that obesity is negatively associated with wages among females. Results suggest that differences in productivity attributable to human capital accumulation and health account for approximately half of this wage penalty. I find no evidence of coworker or manager discrimination against obese employees among males or females. However, I find evidence that the wage-obesity penalty among females occurs only among obese mothers, a result that may suggest differences in unobserved productivity between obese and non-obese mothers.

2. Introduction

Obesity is a substantial health issue in the United States, affecting 35 percent of adults in 2010—a proportion projected to increase to 51% by 2030 (Ogden, Carroll, Kit, & Flegal, 2012; Finkelstein et al., 2012).[34] Finkelstein, Trogdon, Cohen, and Dietz (2009) estimate that in 2008 obesity accounted for 10 percent of medical spending in the US at a cost of $147 billion. In addition to direct medical costs, obesity may incur significant economic costs through lost productivity. Numerous studies link obesity to greater rates of both absenteeism and presenteeism (i.e., reduced work effort while present on the job), and it is estimated that in 2008 obesity-related absenteeism and presenteeism cost employers $42.8 billion in lost productivity (Finkelstein, DiBonaventura, Burgess, & Hale, 2010).[35] Perhaps unsurprisingly, obesity has been linked to lower wages across the demographic spectrum in the United States, Europe, and even China.[36] Subsequent studies using instrumental variables suggest that the relationship between obesity and wages is not attributable to reverse causality.[37] This has

[34] Body Mass Index (BMI) is the standard scale for establishing a healthy weight to height ratio. BMI is defined as the ratio of kilograms of weight to squared-meters of height. Healthy BMI is classified within the range [18.5, 25), while [25, 30) is classified as overweight, and people with BMI $\geq$ 30 are classified as obese.

[35] For examples of studies regarding obesity and absenteeism/presenteeism, see Howard and Potter (2012); Goetzel et al. (2010); Ricci and Chee (2005); Tsai, Ahmed, Wendt, Bhojani, and Donnelly (2008); Burton et al. (1998); Tucker and Friedman (1998); Schmier, Jones, and Halpern (2006); Finkelstein et al. (2010); Gates, Succop, Brehm, Gillespie, and Sommers (2008); and Pronk et al. (2004).

[36] See for example Register and Williams (1990); Loh (1993); Pagan and Davila (1997); Averett and Korenmann (1999); Cawley (2004); Garcia and Quintana-Domeque (2007); Lundborg, Bolin, Hojgard, and Lindgren (2007); Greve (2008); and Shimokawa (2008).

[37] For a thorough review of the wage-BMI instrumental variables literature, see Kortt and Leigh (2010).

led researchers to attempt to uncover the factors that could be underlying or confounding the wage-obesity relationship. There are two primary hypotheses.

The first hypothesis is that obese workers are less productive than non-obese workers in some way that is not captured in standard wage equations. These differences are typically attributed to differences in human capital accumulation or differences in underlying health. The second primary hypothesis is that the residual wage penalty for obesity that remains unexplained after controlling for typical human capital and demographic variables may be attributable to discrimination against obese employees. While the potential presence of systematic discrimination may seem more of a legal or ethical dilemma than an economic one, the socio-economic implications are significant. Nonwhites, less-educated individuals, and low-income females are more likely to be obese than others (Flegal, Carroll, Kit, & Ogden, 2012; Ogden, Carroll, Kit, & Flegal, 2010). To the extent that vulnerable populations are more likely to be obese, the wage effects of discrimination may have repercussions on income equality or equity of opportunity. Additionally, discrimination may exacerbate productivity losses, for instance if it limits beneficial cooperation in the workplace. Discrimination against obese individuals has also been shown to increase the negative health effects of obesity (Schafer & Ferraro, 2011) which may increase absenteeism or presenteeism costs to employers.

The current study considers both the productivity and discrimination hypotheses. The study utilizes employee-level data from a US IT firm that provide measures of health and human capital accumulation unavailable in previous studies. Information about the

organization of the firm also allows us to directly test for the possibility of discrimination against obese employees. In addition to these two hypotheses, I investigate the previously unconsidered possibility that children are confounding the wage-obesity relationship. While children are a standard inclusion in previous wage equations, our study is the first to interact children with obesity to test whether the wage penalty varies between parents and non-parents, a possibility that could explain why the estimated penalty for obesity is consistently larger among females than among males.

The data provide other advantages over previous datasets. All employees in the dataset are high-income earners relative to the US population at large (the median annual salary is $85,000), which greatly alleviates concerns about possible reverse causality between low income and obesity. Measures of annual salary used to construct hourly wages are retrieved from administrative records, and BMI measures are directly gathered by trained data collectors. By eliminating self-reports of BMI and salary, the dependent variable and independent variable of interest should be less subject to measurement error than in previous studies.[38]

Consistent with previous research I find that obese females earn significantly less than normal-weight females: a penalty of nearly 7%. Obese males earn a roughly 4 percent wage premium compared to normal-weight males, although this result is not statistically significant. Results indicate that human capital accumulation explains a

[38] As acknowledged by previous studies (e.g., Burkhauser & Cawley, 2008), BMI is an imperfect measure of health, as ratios of weight to height do not perfectly account for body frame or muscle mass. To circumvent this issue, several recent studies have substituted percent body fat or waist size for BMI in the regression equation (Bozoyan & Wolbring, 2011; Johansson, Bockerman, Kiiskinen, & Heliovaara, 2009; Wada & Tekin, 2010). Unfortunately the available data do not contain measures of body fat or waist circumference. Since the majority of the literature uses the BMI measures previously described rather than waist size, using BMI is consistent with providing the best comparison possible.

significant amount of the difference in wages between obese and normal-weight employees among both males and females. Contrary to previous research, I find evidence that health is an important confounding factor among males, which suggests that efforts to improve employee health may yield significant gains in productivity. I do not find evidence of peer or supervisor discrimination against obese employees, but I do find that the wage penalty for obese females only occurs among obese mothers, and that obese females with no children actually earn wage premiums similar to obese males.

The remainder of the paper is as follows. Section 3 discusses previous attempts to explain the well-documented negative relationship between wages and obesity. Section 4 introduces the econometric model, while Section 5 describes the data in detail. Section 6 reports the results of tests for differences in productivity and for the possibility of discrimination. Lastly, Section 7 explores the possibility that the wage-obesity relationship is confounded by unobserved factors related to parenthood.

3. Background and Previous Literature

As evidence has mounted in support of the negative wage-obesity relationship, a subset of the literature has shifted focus from estimating the relationship between obesity and wages to identifying the possible factors(s) that may be underlying this relationship (hereafter referred to as the "wage penalty"). There are two broad hypotheses for the underlying source of the wage penalty. The first hypothesis is that obese employees are less productive than normal-weight ones for reasons that are not captured by standard wage equations. These differences are typically attributed to either differences in human capital accumulation or differences in health. This is an important distinction.

Differences in human capital are typically attributed to differences in unobserved individual attributes that could be correlated with both human capital investment and the probability of becoming obese (such as time-preference or intrinsic motivation). In this case obesity is a confounding rather than a causal mechanism: it is not causing lower wages but is merely the physical manifestation of other attributes that lead to lower wages. Moreover, these attributes are considered immutable. One's time preference or ambition is not likely to respond to intervention at the public or private level. On the other hand, differences in underlying health may reflect a causal relationship. Even if obesity and reduced health occur simultaneously, reduced wages attributed to reduced health would signal the possibility of restoring lost productivity through appropriate health intervention at the public or private (i.e., firm) level.

Support for the human capital hypothesis is offered by Baum and Ford (2004), who find that obese U.S. workers earn significantly lower returns to tenure on the job, a result which they attribute to differences in investment in the on-the-job training. Atella, Pace, and Vuri (2008) attempt to recreate this result among European employees but find no difference in returns to participation in training programs between obese and normal-weight workers.

The literature provides less support for the health hypothesis. Brunello and D'Hombres (2007) and Baum and Ford (2004) control for health using a binary indicator for "poor health" or "health limitations," respectively, while Atella et al. (2008) use a measure of absenteeism (days of work missed in the last four work weeks due to illness). All three studies fail to find evidence that obese workers are less productive due to lower

average health. Johansson et al. (2009) do find that a binary indicator for "good health" is significantly correlated with wages among Finnish workers, but inclusion of the measure only slightly attenuates the observed wage penalty for obesity. Similarly, Lundborg et al. (2007) find that controlling for chronic conditions, mobility, and self-reported health status partially attenuates the wage penalty among European workers over 50 years of age, but leaves a significant wage penalty unaccounted for. Gregory and Ruhm (2009) point out that medical expenditures do not begin to increase with increasing BMI until well after the point at which BMI begins to adversely affect wages, casting further doubt on the health hypothesis. Ultimately, no supplemental control for productivity has been able to fully account for the observed wage penalty, leading researchers to consider alterative explanations.

The second primary hypothesis for the wage penalty is that obese workers are subject to workplace discrimination.[39] Two recent wage studies have found evidence in support of this hypothesis. Han, Norton, and Stearns (2009) find that the wage penalty for obese employees is higher in jobs that require higher levels of interpersonal skill. A similar study by Johar and Katayama (2012) finds that obese employees in socially-oriented jobs (those requiring "authority" or "nurturance") also face a higher wage penalty than those in non-social jobs. However, these studies cannot differentiate between possible customer and employer discrimination (although Baum and Ford, 2004

[39] A third hypothesis posits that total compensation for obese employees is consistent with normal-weight employees, but that wages decrease as employers shift compensation to cover higher insurance premiums incurred by obese workers but borne by the firm. Bhattacharya and Bundorf (2009) find strong evidence for this effect in the US, but Baum and Ford (2004) find that obese American workers with employer-provided insurance actually earn more than obese workers without it. Atella et al. (2008) also fail to find evidence for the insurance hypothesis among European workers.

find that obese workers in customer-oriented occupations do not face a higher wage penalty than those outside of such occupations suggesting that the results may be driven by employer discrimination). Moreover, in both of these studies, obese employees in the less-social jobs still face a wage penalty. This suggests that even if discrimination is occurring, it cannot fully account for the observed wage penalty.

In the current study I consider two types of employer discrimination: direct and indirect. Direct discrimination refers to obese employees being paid less simply because they are disliked by their employers. Standard economic theory posits that competitive markets should eliminate this type of behavior, although evidence for labor market discrimination according to gender, race, or sexual orientation remains, suggesting that this type of discrimination cannot be ruled out (see e.g., Biddle & Hamermesh, 2013; Laurent & Mihoubi, 2012).

Indirect discrimination is more subtle. In this case, non-obese employees may mistreat their obese coworkers, or refuse to cooperate with them to the same degree they would their non-obese peers. Due to this negative behavior, obese employees will be less effective at their jobs, therefore "earning" their reduced wages, even though the effect is still attributable to discrimination. Several studies have found qualitative evidence for this type of discrimination. Carr and Friedman (2005) report that BMI is positively correlated with *perceived* workplace discrimination (e.g., rudeness, being treated as less intelligent, etc.), and that such workplace discrimination is more prevalent among white collar workers. Obese workers are also more likely to be perceived by their coworkers as

lazy, lacking in self-control, of lower ability, and less likely to get along with and be accepted by coworkers and subordinates (Rudolph, Wells, Weller, & Baltes, 2009).

In addition to the previously considered hypotheses, I also consider one possibility that has not been previously addressed in the literature: that having children is a factor that may both increase BMI and decrease wages. This would help to explain the discrepancy in results between obese males and females sometimes encountered in the literature (see e.g., Cawley, 2004; Greve, 2008; Johansson et al., 2009; Hildebrande & Van Kern, 2010). Women have larger biological roles in pregnancy and childbirth that may affect body mass or disrupt human capital accumulation, and in the US women still handle the majority of child care responsibilities (Craig, 2006). Children are a standard control variable in the literature and therefore not the source of any omitted variables bias. What has not been accounted for is the possibility that the relationship between children and wages differs between obese and normal-weight women, or that the relationship between wages and obesity differs between mothers and non-mothers.

Research has suggested that women who are obese in early adulthood are less likely to ever have children and that those who do will have fewer children than mothers who were not obese in early adulthood (Frisco & Weden, 2013; Frisco, Weden, Lippert, & Burnett, 2012). Therefore the presence of children may provide information on the timing of obesity. The timing of obesity may influence the wage-obesity relationship in ways that can be captured by this variation, something that I consider in Section 7.

4. Econometric Model

Following the classic Mincer (1975) approach, wages may be modeled as

$$W = \exp(X'\beta + \delta_1 OV + \delta_2 OB + \varepsilon) \quad \text{(Eq. 1)}$$

where OV and OB refer to indicators for overweight and obese, with BMI < 25 (normal weight) serving as the reference category, and X is a set of control variables available in national-level datasets, including demographic and occupational controls, as well as basic measures of human capital accumulation (e.g., experience, education).[40]

In general, the aim of regressing wages on obesity is to estimate the average percentage difference in wages between obese and normal-weight (or non-obese) individuals conditional on X (hereafter referred to as the average "treatment" effect of obesity [ATE_{OB}]). This ATE may be defined as

$$ATE_{OB} = \exp(\delta_2) - 1.^{41} \quad \text{(Eq. 2)}$$

Practitioners generally make several additional assumptions about the model. The first is that

$$\varepsilon = Z'\gamma + \nu \quad \text{(Eq. 3)}$$

[40] I omit the "underweight" category in the analysis as only 5 respondents meet this definition. Therefore the normal-weight category refers to anyone with a BMI of less than 25, rather than the strict definition of individuals with a BMI between 18.5 and 25.

[41] In the literature ATE_{OB} is often approximated by δ_2 itself. This distinction is unimportant for the present analysis.

where Z is a vector of explanatory variables that are not contained in X. The second is that corr(OB,Z|X) $\neq 0$, so that estimates of δ_2 are biased due to the omission of Z. Finally, it is assumed that ν is exogenous and mean zero conditional on both X and Z, so that estimates of δ_2 would be consistent if Z was included in the model.

Previous studies (e.g., Atella et al., 2008; Baum & Ford, 2004; Cawley, 2004) have utilized a mixture of instrumental variables and individual fixed effects in order to condition out Z and retrieve consistent estimates of δ_2. As discussed in Section 3, a segment of the literature has recently shifted from consistently estimating the wage-obesity relationship (δ_2) towards uncovering the elements of Z. However, previous studies have not utilized a rigorous approach to achieving this goal. Typically, if inclusion of a new variable set (Z_k) appears to substantially attenuate the estimate of δ_2 or renders the estimate insignificant, then the new set is deemed to be an element of Z. In this study I take this approach a step further by using a generalized Hausman test to test the null hypothesis that the estimate of δ_2 from the regression equation excluding Z_k is identical to that produced by the equation that includes Z_k (δ_2'). Although motivated by Baron and Kenny's (1986) test for mediating effects, our approach is virtually identical to previous efforts in the literature to classify Z, except I rely on a statistical rather than "eyeball" test.[42]

Let W_{ij} be the hourly wage rate of individual *i* in work group *j*. In our sample a work group refers to a collection of employees who all report to the same manager.

[42] This approach is similar to the test for collapsibility proposed by Clogg, Petkova, and Shihadeh (1992). Their approach requires the assumption of normally distributed and homoscedastic errors. I assume that errors are clustered, and therefore the assumptions necessary for that test are not met.

Virtually all collaboration at the firm occurs within a work group and I assume that both observed and unobserved work group attributes may influence the wages of employee *i*. I express the true wage equation for our sample as:

$$W_{ij} = \exp(X'_{ij}\beta + Z'_{ij}\gamma + \lambda_1 OV_i + \lambda_2 OB_i + u_j + \xi_i) \quad \text{(Eq. 4)}$$

where X_{ij} are individual and group-level control variables consistent with those drawn from national-level datasets; Z_{ij} are individual and group-level variables that remain jointly correlated with W_{ij} and OB_i conditional X_{ij}; λ_2 is the true value of the regression parameter for obesity; u_j is a work-group level random effect and ξ_i is an individual-level stochastic shock, both of which are exogenous and mean zero conditional on X_{ij} and Z_{ij}.

I define a "potentially consistent" estimator of λ_2 as one that *would* consistently estimate λ_2 if the regression contained the entire vector Z (i.e., every possible variable jointly correlated with wages and obesity not contained in X) and that would consistently estimate the appropriate standard error σ_{λ_2}. Furthermore, estimates of λ_2 produced by this estimator would consistently estimate the true population ATE_{OB} if $\hat{\lambda}_2$ was inserted into equation (2). I define the estimator as "potentially" consistent since I do not assume that the set of *k* variables Z_K available in our data represents the full set of potentially omitted variables jointly correlated with wages and obesity (that is, I do not assume that I ever successfully estimate the "true" relationship between wages and obesity).

The previous literature almost universally utilizes OLS on a log-transformed model of wages. However, I deviate from this approach since it may not fulfill the assumptions of the potentially consistent estimator. As discussed by Blackburn (2007)

and Manning and Mullahy (1998), estimates of λ_2 from a log-wage model may not recover the true ATE_{OB} if there is heteroskedasticity in the error term. Moreover, due to the presence of u_j, a log-wage model will not produce the correct residuals to construct cluster-robust standard errors since

$$E[y_j - \exp(X'_{ij}\beta + Z'_{ij}\gamma + \lambda_1 OV_i + \lambda_2 OB_i)] \neq E[\ln(y_j) - X'_{ij}\beta + Z'_{ij}\gamma + \lambda_1 OV_i + \lambda_2 OB_i].$$

Therefore, I utilize a generalized linear model with a log link function as a potentially consistent estimator in place of OLS.[43]

If Z is jointly correlated with wages and obesity, and Z_k is some subset of Z, then including Z_k in the regression should reduce (or potentially eliminate) omitted variables bias that is present, such that δ'_2 is different in magnitude to δ_2, and $\lambda_2 - \delta'_2 < \lambda_2 - \delta_2$ (i.e. δ'_2 is closer to the "true" parameter value than is δ_2). Our estimation goal then is neither λ_2 nor ATE_{OB} but Δ, where $\Delta = \delta'_2 - \delta_2$. I assert that the greater the magnitude of Δ, the greater the confounding effect of Z_k. By imposing a "potentially" consistent estimator of ATE_{OB}, I ensure that Δ is a consistent estimate of the confounding effect of Z_k, and that the cluster robust standard errors are not underestimated.

I use a generalized Hausman test to determine whether Δ is statistically significant.[44] If so, then I can reject the null hypothesis that $\delta'_2 = \delta_2$. Although rejecting

[43] When dealing with clustered as opposed to panel data with nonlinear outcomes, random effects will only improve efficiency and will not affect the consistency of the estimates (Cameron & Trivedi, 2005). On the other hand, integration of the random effect requires the assumptions that the errors are normally distributed and homoscedastic within a cluster. Failure of these assumptions will render the model inconsistent. This, coupled with the practical shortcomings of the nonlinear random effects model makes us opt to construct clustered standard errors rather than cluster random effects.

the null of a statistical test cannot prove that the omission of Z_k was causing omitted variables bias, such a rejection when coupled with valid theoretically consideration allows us to make a compelling case that Z_k is an explanatory factor underlying the wage-obesity relationship. Elements of Z_k include controls for employee productivity (human capital accumulation and health) as well as controls for peer or supervisor discrimination. These elements will be detailed in Sections 5.3 and 6.3.

5. Data

5.1. Work, Family, and Health Network

The data were obtained from the Work, Family, and Health Network (WFHN). The WFHN was created by the National Institutes of Health (NIH) and Centers for Disease Control and Prevention (CDC) to study the relationship between work, family life, and health outcomes. The WFHN is comprised of four research centers, a translational coordinating center (TCC), and a data and methods coordinating center (DCC). The four research centers are the University of Minnesota, Penn State University, Harvard University, and Portland State in conjunction with Purdue University. The Kaiser Permanente Center for Health Research serves as the TCC, and RTI International serves as the DCC. Members of the WFHN were tasked with determining the health effects of an intervention intended to: (a) increase employees' control over their work time, and (b) improve supervisor and coworker support for employees' family and personal lives. Two firms were selected for intervention, one of

[44] Tests are 1 tailed with $p \leq 0.1$ as the cutoff for "significant" difference. The residuals used to estimate $\hat{\sigma}^2$ are clustered at the work group level.

which is an American telecommunications firm. A single cross-section of data from this firm, collected prior to the intervention, is utilized in the present study.

The sample is limited to full-time, permanent (non-contractor) employees, each of whom belongs to one of 106 "work groups."[45] The work groups are collections of employees who report to the same manager and may collaborate with each other frequently. Employees operate at one of thirteen sites, and all sites are located in one of two urban locations in two separate states.[46] The mean work group size in the sample is approximately 12, and the average work site hosts about 58 employees.

Employees are classified by the firm's human resources (HR) department as either support personnel (e.g., network administrators, administrative assistants) or core personnel (those directly involved in the firm's core business). Within the firm there are also four broad occupational classes based on primary job function. Each work group emphasizes or is entirely devoted to one of the four functions. Each employee in the sample is designated by HR as belonging to one of four job categories based on the function emphasized by her or his work group (that is, occupational controls are at the work group rather than individual level). The support/core and occupational categories are utilized as additional controls in the wage equation as discussed in Section 5.2.[47]

---

[45] The data do not contain any "blue collar" support personnel (e.g., custodians, security).

[46] In addition to work site indicators, all models include an indicator for those who did not report a work site.

[47] The distinction between core/support and occupation assignment are defined for the benefit of researchers and are not official administrative divisions. However, these divisions represent real and significant differences between tasks performed and potentially the compensation that accompanies each task. Disclosure of details regarding the four job functions is prohibited.

5.2. Standard Variables

The dependent variable is a constructed measure of hourly wages. Annual income is obtained from administrative records. Annual hours are constructed by multiplying self-reported average weekly hours by 52 and subtracting out the number of vacation hours taken by the employee that year (which is also recorded in the administrative data). Hourly wages (W_{ij}) are the ratio of annual income to annual hours.[48] Controls for the hourly wage, indicated by the vector X_{ij}, include age, squared age, tenure with the firm, tenure squared, education, race, nativity, married/cohabitating status, number of children, occupation, an indicator to differentiate between support and core employees, and indicators for state and worksite.[49,50] Following the literature, separate models are run for males and females.

[48] Using an hourly wage measure as the dependent variable raises the concern that low wages are a consequence of long hours rather than reduced compensation, and that long hours could be correlated with obesity. However, constructed hourly wage is an issue with all data. For instance, in the 2010 wave of the NLSY79, 56 percent of employed respondents reported a time-unit of compensation other than hourly (Bureau of Labor Statistics, 2013). Thirty-seven percent of employed respondents reported compensation as an annual value. Moreover, a relationship between longer hours and obesity is consistent with a productivity hypothesis. Salaried compensation is offered with the expectation of a certain level of output per unit of time (e.g. every week). Individuals who take more hours to produce that level of output are, by definition, less marginally productive. Therefore long hours "causing" obesity over time still indicates that reduced productivity is confounding the wage-obesity relationship. One other possibility is that individuals put in longer hours than necessary in order to signal commitment with the hope of higher future salary. If this is the case I should see that obese employees have higher returns to tenure than normal-weight employees. However, as reported in Appendix G, this is not the case.

[49] The binary division between support and core workers is a qualitative distinction in job type not captured by the four job categories. The difference between staff and senior workers may reflect differences in ability or human capital accumulation and therefore this distinction is not considered until later models in Section 6.2.

[50] Race enters the equation as a binary white/nonwhite variable. Roughly half of non-whites are Indian-Asians, while about 20% are "Other Asian," 10% are African American, and the remainder self-identify as Pacific Islander, Native American, or "Other."

Observations with missing data for BMI or salary are dropped from the sample.[51] All missing observations are assumed to be missing at random (Rubin, 1976), and the observed subsample is thus representative of the full population of non-executive, white-collar employees within the firm. Due to the small sample size, extra attention is paid to outliers that may have an undue influence on the parameter estimates, and I drop any observation that fulfills both of two conditions: belonging to either the top or bottom 1-percentile of BMI, and belonging to either the top or bottom 1-percentile of hourly wages. This resulted in two additional males and one additional female being dropped, leaving a final sample of 452 men and 295 women. Summary statistics are provided in Tables 25 and 26 for females and males, respectively.

Table 25

Summary Statistics (Female)

		Full Sample (n = 295)		Obese (n = 92)		Normal Weight (n = 111)	
		M	*SD*	*M*	*SD*	*M*	*SD*
Standard Variables	BMI	28.28	6.61	36.05***	5.68	22.44	1.75
	Overweight	31.18%	0.46				
	Obese	31.18%	0.46				
	Hourly Wage	38.76	7.85	37.00***	7.96	40.35	7.61
	Married or Cohabitating	70.85%	0.46	54.34%***	0.50	72.97%	0.45
	Total Number of Children	1.63	1.24	1.73*	1.34	1.48	1.19

[51] See Appendix F for information regarding missing data.

Table 25

(Cont.)

		Full Sample (n = 295)		Obese (n = 92)		Normal Weight (n = 111)	
		M	*SD*	*M*	*SD*	*M*	*SD*
	Nonwhite	24.07%	0.43	17.39%*	0.38	27.03%	0.45
	Born Abroad	19.32%	0.40	8.70%***	0.28	27.03%	0.45
	College Graduate (4-year)	66.44%	0.47	55.43%***	0.50	75.68%	0.43
	Age	46.88	8.38	48.49**	7.68	45.11	9.14
	Tenure with Firm (years)	16.04	9.94	17.66**	10.19	15.01	9.96
	Support Personnel	5.76%	0.23	8.70%	0.28	5.41%	0.23
Productivity Controls (Z)	Staff	36.27%	0.48	38.04%	0.49	32.43%	0.47
	Senior	57.97%	0.49	53.26%**	0.50	62.16%	0.49
	Physical Function	92.63%	13.38	87.14***	17.69	96.00	7.61
	Loud Snoring	25.76%	0.44	40.22%***	0.49	10.82%	0.31
	C-Reactive Protein[a]	3.06	4.34	5.74***	6.24	1.53	2.25
	Cholesterol Ratio[b]	3.87	1.09	4.17***	1.18	3.61	0.94
	Hypertension	31.18%	0.46	43.48%***	0.50	18.92%	0.39
	Heart Rate	72.35	11.10	73.66*	12.39	71.37	10.58

Notes: a: C-Reactive Protein concentration is measured in mg/L. For reference, CRP levels below 1.0 are considered low-risk for heart disease, 1.0-2.99 is considered average risk, and greater than 3.0 is high risk. b: A cholesterol ratio below 3.5 is considered optimal, 3.5-5 is normal, and >5 is considered high.
* Indicates obese employees are significantly different from normal-weight employees at the 0.1 level. ** $p < 0.05$ *** $p < 0.01$

Table 26

Summary Statistics (Male)

		Full Sample (n = 452)		Obese (n = 135)		Normal Weight (n = 123)	
		M	*SD*	*M*	*SD*	*M*	*SD*
Standard Variables	BMI	28.10	4.90	33.74***	4.38	23.04	1.55
	Overweight	42.92%	0.50				
	Obese	29.87%	0.46				
	Hourly Wage	41.02	8.23	42.20*	9.23	40.24	8.51
	Married or Cohabitating	84.74%	0.36	85.93%	0.35	81.30%	0.39
	Total Number of Children	1.64	1.43	1.83**	1.49	1.32	1.25
	Nonwhite	29.20%	0.46	17.04%***	0.38	45.53%	0.50
	Born Abroad	31.89%	0.44	16.30%***	0.37	47.97%	0.50
	College Graduate (4-year)	83.63%	0.37	73.33%***	0.44	91.87%	0.27
	Age	45.12	8.70	46.96***	9.04	43.24	9.38
	Tenure with Firm (years)	12.07	8.16	12.60**	8.11	10.82	7.35
	Support Personnel	6.86%	0.25	4.44%*	0.21	10.47%	0.31
Productivity Controls (Z)	Staff	33.41%	0.47	34.07%	0.48	33.33%	0.47
	Senior	59.73%	0.49	61.48%	0.49	56.10%	0.50
	Physical Function	95.28	10.21	92.96***	12.30	97.02	6.57
	Loud Snoring	32.74%	0.47	40.00%**	0.49	23.58%	0.43
	C-Reactive Protein[a]	1.87	2.32	2.53***	2.86	1.04	1.08
	Cholesterol Ratio[b]	4.47	1.35	4.60**	1.11	4.31	1.68
	Hypertension	45.35%	0.50	55.56%***	0.50	29.27%	0.45
	Heart Rate	69.75	11.33	73.40***	11.96	68.26	10.73

Notes: a: C-Reactive Protein concentration is measured in mg/L. For reference, CRP levels below 1.0 are considered low-risk for heart disease, 1.0-2.99 is considered average risk, and greater than 3.0 is high risk.
b: A cholesterol ratio below 3.5 is considered optimal, 3.5-5 is normal, and >5 is considered high.
* Indicates obese employees are significantly different from normal-weight employees at the 0.1 level.
** $p < 0.05$ *** $p < 0.01$

Respondents' annual salaries average about $82,000 for females and $88,000 for males: equivalent to hourly wages of roughly $39 for females and $41 for males. The majority of the sample has at least a four-year degree with virtually the entire remainder having some college education and only a few holding just a high school diploma.[52] Employees are generally mid-career with a mean age of about 46 (minimum 26, maximum 64) and 12–15 years of experience with the current company. Roughly 25 percent of the sample is nonwhite, while about 32 percent of males and 19 percent of females were born abroad. Approximately 85 percent of males and 71 percent of the females are married or cohabitating, and both subsamples have slightly less than two children, on average.

Both men and women have an average BMI of 28, about 1.5 units higher than the national average, with females having a slightly higher variance. Approximately 32 percent of females and 42 percent of males are clinically overweight (≤ 25 BMI < 30), while about 30 percent of males and 32 percent of females are clinically obese (BMI $\geq$ 30). The proportions overweight and obese are roughly consistent with unadjusted national averages (except for the slightly high proportion of overweight males).[53]

---

[52] Testing found no significant difference in wages between those with just a high school diploma and those with some college but no bachelor's degree. Individuals without a bachelor's degree do not appear to have diminished standing with the firm. Employees without a degree are more likely to be classified as senior than staff, and are represented across all four job classes. The primary difference is that employees without a degree have been with the company longer, suggesting a longer path to promotion, or the possibility that older employees were "grandfathered" in before more rigorous selection was enacted.

[53] Parametric studies typically model the relationship between BMI and wages by utilizing a linear measure of BMI, possibly paired with a quadratic term, or by creating a set of categorical BMI indicator variables defined as underweight (BMI < 18.5), overweight ($25 \leq$ BMI < 30), or obese (BMI ≥ 30), with the normal range ($18.5 \leq$ BMI < 25) as the reference group. Testing suggests a nonlinear relationship between wages and BMI that is best captured by the categorical measures.

Comparisons between obese and normal-weight employees indicate significant differences between the two subsamples. Relative to their normal-weight coworkers, obese females are significantly less likely to be married, nonwhite, or born abroad. The average obese female is also about three years older than her normal-weight peers, with an extra two years of tenure with the present firm. Despite the additional experience, obese females earn nearly $3 less per hour than normal-weight women, a significant difference equivalent to nearly one-half of a standard deviation. However, obese females are nearly 20 percent less likely to have completed a 4-year degree and 10 percentage points less likely to be "senior" core personnel. Among males, obese workers actually earn significantly more than their normal-weight peers, a difference of $1.50, or about one-fifth of a standard deviation. Obese males are significantly less likely to be nonwhite or born abroad, and also significantly less likely to have finished college. Like females, obese males in the sample are roughly three years older with an additional two years of experience with the firm.

5.3. Additional Controls for Productivity

The control variables (X_{ij}) in Section 5.2 are those that are included as standard in previous wage models in the wage-obesity literature. However, these controls may be insufficient to fully account for differences in productivity between obese and normal-weight workers (as evidenced by the significant wage-obesity relationship typically estimated in the literature). The WFHN data include additional controls (Z_{ij}) for employee productivity that allow me to more precisely account for differences in

productivity that may remain after accounting for the standard controls, either due to differences in human capital accumulation or health.[54]

The first set of variables controls for possible differences in accumulation of occupation-specific human capital, i.e., skills and abilities that directly translate to the job (and which should entail a mixture of workplace experience and natural ability). Core personnel in the data are differentiated between "staff" and "senior" level workers. Senior level workers do not possess administrative authority over staff workers. Rather, senior status indicates a meaningful accumulation of human capital that has resulted in one or more promotions.[55] Indicator variables accounting for this division should capture potential differences in human capital accumulation between obese and non-obese employees. Among both males and females roughly 60 percent of employees have reached senior status, and approximately 35 percent are staff level (with the remainder belonging to support positions). Obese females are significantly less likely to be senior-level employees, a difference of nearly 10 percentage points, although this discrepancy does not hold among males.

The second set of variables contains proxies for facets of health that may affect productivity, including physical function, sleep, and cardiovascular health. The first health proxy is a comprehensive measure of physical function, rated on a scale from 0-100. This variable is constructed from nine separate questions, which assess self-reported

[54] Tests for discrimination rely on organizational data rather than specific variables. The approach for these methods will be outlined in Section 6.3.

[55] Staff may be either "Staff I" or "Staff II" and seniors as "Senior I" or "Senior II." Therefore someone joining the firm as a Staff I would need two promotions to reach Senior I, while someone hired as a Senior I could be promoted and still appear in the data as a "Senior." The distinction between sub-levels I and II is unavailable in the data.

limitations on everyday activities such as walking up stairs, carrying groceries, and bending/stooping, among others. Responses may be either "Yes, limited a lot"; "Yes, limited some" or "No, not limited at all." The responses are transformed by the DCC (RTI, International) into a 0-100 scale based on scoring devised by the RAND Corporation. [56] A score of 100 reflects full functionality (able to run and play sports) and 0 is barely functional (health severely limits all everyday activities). Although software development is not a physically strenuous job, it is still reasonable to assume that diminished physical function may affect on-the-job performance. In general, the sample is highly functional, with average ratings over 90 percent. However, obese employees are significantly limited compared to their normal-weight coworkers, a difference of 10 percentage points (roughly one standard deviation) among females and 5 percentage points (roughly one-half a standard deviation) among males.

Another way in which obesity may result in reduced productivity is if obesity substantially disrupts sleep. Obesity is significantly correlated with obstructive sleep disorder (aka "sleep apnea"), which in turn has been associated with decreased cognitive function (Engleman & Douglas, 2004; Ulfberg, Carter, Talback, & Edling, 1996; Vgontzas et al., 1994). Although the data lack an exact measure of sleep apnea, they do contain a self-reported indicator variable for "loud snoring." The indicator denotes positive response to the question "During the past month, have you ever snored loudly, or

[56] The questions comprising the physical limitations measure are derived from the Medical Outcomes Study 36-Item Short-Form Survey (Ware & Sherbourne, 1992). Scoring from 0-100 is based on the RAND 36-Item Health Survey 1.0 (www.rand.org/health/surveys_tools/mos/mos_core_36item_scoring.html).

been told you were snoring loudly?"[57] Severity of snoring has been identified as one of the primary predictors of sleep apnea among obese patients (Vgontzas et al., 1994). Therefore, controlling for loud snoring may serve as a proxy for the effect of obesity on sleep. Approximately one-fourth of females and one-third of males report loud snoring. Obese females are more likely to snore heavily by 30 percentage points, while for males the differences is 15 percentage points.

The final set of health variables contains four measures of cardiovascular health including blood serum levels of C-reactive protein (CRP), a biomarker for inflammation; blood serum levels of cholesterol; blood pressure; and heart rate.[58] All four of these measures capture elements of risk for cardiovascular disease, as well as overall cardiovascular fitness. Measures of cholesterol are transformed into a ratio of total to HDL ("good") cholesterol, a measure that is more strongly correlated with heart disease than total cholesterol or LDL ("bad") cholesterol and that may better capture the negative lifestyle behaviors associated with BMI (Kinosian et al., 1994). Measures of both CRP and the cholesterol ratio appear to follow a lognormal distribution (and to be heteroskedastic in relation to log wages), and so these variables are transformed by the natural log before entering the model. Measures of blood pressure and heart rate refer to

[57] This question distinguishes people who have reported snoring (but not loudly) from people who have reported "loud" snoring. Nested within the "loud snoring" indicator are respondents who have "snorted/gasped" or "stopped breathing/struggled for breath." These sub-measures are not considered separately due to the small number of respondents (particularly females) who suffer from these conditions. Moreover, analysis by Maislin et al. (1995) suggests that loud snoring is nearly as correlated with apnea as snorting/gasping, and more correlated with apnea than stopped breathing/struggled for breath.

[58] Data for C-Reactive protein and cholesterol are missing for 7% and 13% of the sample, respectively. Several observations were also missing data on heart rate. Since these are not the variables of interest and are assumed to be missing at random, missing values were imputed using a modified regression-based EM algorithm. See Appendix F for details.

the mean value of three measures obtained on three separate days. The average blood pressure measure is used to construct an indicator for high blood pressure (hypertension), which refers to systolic pressure greater than 140 mmHg or diastolic pressure greater than 90 mmHg.

Among both males and females, obese employees have significantly worse measures of cardiovascular health compared to their normal-weight coworkers. Obese females measure nearly one standard deviation higher in their cholesterol ratio and CRP levels, and are approximately 25 percentage points more likely to have clinically high blood pressure. Females also have significantly higher heart rates, although the difference is not substantial (about 2.5 beats per minute [BPM]: less than one-fifth a standard deviation). Obese males also have higher measures of inflammation and cholesterol ratio compared to normal-weight males, equal to approximately one-half of a standard deviation. Obese males are also 25 percentage points more likely to have high blood pressure, and have a heart rate approximately 5 BPM (one-half standard deviation) higher than normal-weight employees. Taken together, the set of health variables suggests that obese workers are less healthy than normal-weight employees in physical function, cardiovascular health, and potentially in quality of sleep. Incorporating these measures should capture differences in productivity that are attributable to health. Tests for differences in productivity between obese and normal-weight workers attributable to differences in health and human capital are provided in Section 6.2.

6. Results

6.1. Baseline Results

Results from the initial model containing only standard control variables (equation 1) are reported in Table 27. Among females, overweight employees earn approximately 5.5 percent less than those in the normal-weight category, while obese employees earn roughly 6.8 percent less, on average. Neither overweight nor obese males earn significantly different wages compared to their normal-weight coworkers. The lack of a significant wage-obesity relationship for males is consistent with the literature, which has found mixed results regarding the wage-obesity relationship among men. The simplest explanation is that obesity is a more imperfect measure of physical health for males, since body mass index (BMI) does not distinguish muscle from other body mass. It may also be the case that the wage-obesity pathway operates differently for males versus females, a possibility that will be explored further in the remainder of Section 6.

Table 27

The Relationship between Wages and Obesity with Standard Control Variables (Model 1)

	Female (n = 295)	Male (n = 452)
Obese	-0.068**	0.039
	(0.033)	(0.025)
Overweight	-0.055*	-0.003
	(0.032)	(0.022)
Married or Cohabitating	0.026	0.018
	(0.032)	(0.026)

Table 27

(Cont.)

	Female (n = 295)	Male (n = 452)
Total Number of Children	-0.002 (0.012)	0.008 (0.008)
Nonwhite	-0.046 (.029)	-0.026 (0.025)
Born Abroad	0.025 (0.034)	0.039 (0.027)
College Graduate (4-year)	0.060** (0.025)	0.057* (0.033)
Age (10 years)	0.285** (0.125)	0.342*** (0.092)
Age Squared (100 squared years)	-0.027** (0.013)	-0.032*** (0.010)
Tenure with Firm (10 years)	-0.157*** (0.062)	-0.020 (.045)
Tenure Squared (100 squared years)	0.042*** (0.015)	0.011 (0.011)
Core (Staff and Senior)	0.185** (0.076)	0.231*** (0.038)

Note: The dependent variable is hourly wages. The overweight and obese coefficients report wages relative to normal-weight employees (BMI < 25). Estimates are obtained using a generalized linear model with a log link function. The model includes indicators for worksite, state, and job category. Standard errors are clustered at the work group level and reported in parentheses.

Table 28 reports the results of tests for the hypothesized productivity mechanisms (human capital accumulation and health). Results examining possible peer or supervisor discrimination are reported in Table 29, while Table 30 shows results for models that test whether parenthood is a confounding factor of the wage-obesity relationship among females. For the sake of brevity, only results for obese workers relative to normal-weight

workers will be reported in results tables moving forward. All subsequent models include indicators for overweight, and all interaction models include an overweight interaction term in addition to the reported obese interaction term. Results for overweight employees are qualitatively similar to those for obese ones.

Table 28

Productivity Differences as a Potential Factor Underlying the Wage-Obesity Relationship

	Baseline	**Human Capital**	**Health**	**Combined**
Female				
Obese	-0.068** (0.033)	-0.054* (0.033)	-0.049 (0.035)	-0.039 (0.032)
Δ		0.014 # [0.065]	0.019 [0.157]	0.029# [0.066]
Male				
Obese	0.039 (0.025)	0.016 (0.024)	0.79*** (0.027)	0.045* (0.026)
Δ		-0.023 # [0.010]	0.040 # [0.000]	0.006 [0.318]

Note: The dependent variable is hourly wage. Estimates are obtained using a generalized linear model with a log link function. Models control for employee's state, site, age, age^2, race, nativity, marital status, number of children, tenure with the firm (in years), tenure^2, and job category, unless reported otherwise. All models also contain an indicator for overweight. The normal-weight category therefore serves as the point of reference for all obesity and obesity-interaction coefficients. Standard errors are clustered at the work group level and reported in parentheses. The p-value for the chi-squared statistic from a generalized Hausman test is reported in brackets.

Signifies that the obesity coefficient is significantly different from the Baseline coefficient estimate.

* $p < 0.1$, ** $p < 0.5$, *** $p < 0.1$

Table 29

Discrimination as a Potential Factor Underlying the Wage-Obesity Relationship

		Health and Human Capital	Opposite Gender	Obese Coworkers	Workgroup Fixed Effects
Females	Obese	-0.039 (0.032)	-0.045 (0.078)	-0.005 (0.048)	-0.060* (0.033)
	Proportion Opposite Gender		-0.088 (0.98)		
	Obese X Proportion OG		0.005 (0.123)		
	Proportion Obese			0.129 (0.102)	
	Obese x Proportion Obese			-0.124 (0.114)	
	Δ		-0.006 [0.434]	0.034 [0.176]	-0.021 [0.390]
Males	Obese	0.045* (0.026)	0.032 (0.039)	0.121*** (0.034)	0.039 (0.025)
	Proportion Opposite Gender		-0.050 (0.047)		
	Obese X Proportion OG		0.030 (0.085)		
	Proportion Obese			0.177** (0.079)	
	Obese x Proportion Obese			-0.259*** (0.105)	
	Δ		-0.013 [0.318]	0.076# [0.009]	-0.006 [0.202]

Note: The dependent variable is hourly wage. Estimates are obtained using a generalized linear model with a log link function. Models control for employees' state, site, age, race, nativity, marital status, number of children, tenure with the firm (in years), job category, indicators for staff and senior, an indicator for loud snoring, as well as log cholesterol ratio, log CRP plasma concentration, hypertension, and heart rate. All models also contain an indicator for overweight, and overweight is interacted with proportion opposite gender in model (9) and proportion obese in model (10). The normal-weight category therefore serves as the point of reference for all obesity and obesity-interaction coefficients. Standard errors clustered at work group level and reported in parentheses. The p-value for the chi-squared statistic from a generalized Hausman test is reported in brackets. # Signifies that the obesity coefficient is significantly different from the obesity coefficient in model (5), the preferred productivity model.

* $p < 0.1$, ** $p < 0.5$, *** $p < 0.1$

Table 30

Parenthood as a Potential Factor Underlying the Wage-Obesity Relationship

		Health and Human Capital	Child-Interaction	Single-Parent
Females	Obese	-0.039 (0.032)	0.051 (0.041)	0.041 (0.053)
	Children	-0.007 (0.012)	0.009 (0.014)	0.012 (0.013)
	Obese x Children		-0.053*** (0.016)	-0.070*** (0.024)
	Obese x Married x Children			0.027 (0.024)
	Δ		0.090# [0.001]	0.080# [0.007]
Males	Obese	0.045* (0.026)	0.046 (0.031)	0.044 (0.032)
	Children	0.008 (0.006)	0.018 (0.013)	0.018 (0.013)
	Obese x Children		-0.002 (0.012)	0.019 (0.037)
	Obese x Married x Children			-0.022 (0.035)
	Δ		0.001 [0.492]	-0.001 [0.458]

Note: The dependent variable is hourly wages. Estimates are obtained using a generalized linear model with a log link function. Models control for employee's state, site, age, race, nativity, marital status, number of children, tenure with the firm (in years), job category, indicators for staff and senior, physical function, an indicator for loud snoring, as well as log cholesterol ratio, log CRP plasma concentration, hypertension, and heart rate. All models also contain an indicator for overweight, and overweight is interacted with children in models (12) and (13). The normal-weight category therefore serves as the point of reference for all obesity and obesity-interaction coefficients. Standard errors are clustered at the work group level and reported in parentheses. The p-value for the chi-squared statistic from a generalized Hausman test is reported in brackets. # signifies that the obesity coefficient differs significantly from model (11) – which is identical to the preferred productivity model (model 5). Models 12 and 13 are re-runs of Model (5) – the preferred productivity model - with and without a control for children.
* $p < 0.1$, ** $p < 0.5$, *** $p < 0.1$

6.2. Productivity

To control for differences in underlying human capital I replace the indicator for "core" jobs with two indicators differentiating between staff and senior positions. Since transition from staff to senior requires one or more promotions, achieving senior level indicates a demonstrable level of human capital accumulation that has been recognized by the firm. This may provide a more precise measure of overall human capital accumulation than simple years of tenure with the firm. If obese employees accumulate human capital at a different rate than normal-weight employees (or if they are discriminated against in promotion), they may be disproportionately represented at the staff level. If this is the case, controlling for the staff/senior distinction should attenuate the coefficients for the overweight and obese indicators.

As shown in the second column of Table 27, accounting for the distinction between staff and senior employees significantly reduces the obesity coefficient among males, suggesting that obese males actually have greater accumulated human capital relative to normal-weight males. Inclusion of these variables decreases the magnitude of the estimated wage penalty among females from -6.8 percent to -5.4 percent, a statistically significant reduction. This result suggests that differences in human capital accumulation are a significant confounder of the wage-obesity relationship, a finding consistent with Baum and Ford (2004).[59]

[59] However, as seen in Appendix G, neither obese males nor females have significantly lower returns to years of tenure, despite not controlling for the staff/senior distinction. This may suggest that the staff/senior distinction is a better measure of human capital accumulation than are years of tenure.

In the health model in Table 28, I test whether differences in productivity attributable to underlying health may be biasing the estimated relationship between wages and obesity. This model omits the controls for staff and senior and introduces a measure of physical function, a proxy for sleep quality, and proxies for cardiovascular fitness. Proxies for cardiovascular health include log C-Reactive Protein concentration, log cholesterol ratio, resting heart rate, and an indicator for hypertension. Contrary to previous research, controlling for employee health reveals a large negative bias in the obesity coefficient among males, suggesting that health is a key omitted variable. The wage penalty for females is attenuated from -6.8 percent to -4.9 percent while the male wage premium from obesity increases significantly from 3.9 percent to 7.9 percent.

The combined model in Table 28 contains both the health controls and the staff/senior indicators to determine if a joint model of productivity is able to explain the entire wage-obesity relationship. Among males the health and human capital variables essentially cancel each other out. Controlling for human capital accumulation reduces the wage premium significantly, whereas including controls for cardiovascular health significantly increases the wage premium. The net effect is a small positive increase that leaves a large and significant wage premium of 4.5 percent. For females the combined model reduces the obesity penalty from -6.8 percent to -3.9 percent, a large and significant reduction. The magnitude of the coefficient shift in the combined model of health and human capital is roughly consistent with the sum of the coefficient shifts from the individual health and human capital models. This suggests that the two sets of proxy variables are capturing different elements of the wage function rather than jointly

measuring the same element. For instance, health stock could be correlated with the same unobservable attributes (e.g., time-preference or motivation) that are intended to be captured by the proxy for human capital. If this were the case then lost wages attributable to lower health would not be recoverable through health intervention, since the underlying individual traits causing the health disparities would remain. However, the health variables do not seem to be capturing the same unobservable attributes supposed to be captured by the human capital model, suggesting that employee productivity could be enhanced if employee health was improved.

Although the remaining female wage penalty and male wage premium are statistically insignificant, the sign and magnitude of the estimates are not inconsistent with the previous literature, suggesting that meaningful unexplained wage differences may still exist between obese and normal-weight employees. Consistent with prior research, I next consider the possibility that the remaining differences in wages may be attributable to discrimination.

6.3 Discrimination

This section utilizes three approaches to test for two possible sources of discrimination. The first two approaches use work group-level measures to test for coworker discrimination. Work group-level measures include the proportion of obese coworkers in an employee's work group, and the proportion of the work group of opposite gender in each employee's work group.[60,61] Research has shown that

[60] The sample mean of the work group characteristics is consistent with the sample mean of the characteristics themselves (i.e., the mean proportion of obese coworkers across work groups equals the proportion of the overall sample that is obese).

individuals are more likely to discriminate publicly when it is considered socially acceptable to do so (Crandall, Eshelman, & O'Brien, 2002). Presumably as the proportion of obese workers in a work group increases, the acceptability of discrimination towards obese members of that work group will diminish. The proportion of work group members of the opposite gender is also hypothesized to affect the probability of discrimination again obese employees, as social norms of body size may be enforced more rigidly by members of the same sex, or members of the opposite sex. The peer discrimination models include an interaction between the indicators for overweight/obese and the measures of proportion opposite gender and proportion obese. This provides a test of whether the wage penalty varies based on either work group characteristic. The third discrimination model utilizes work group fixed effects to test whether unobserved work group attributes are driving the wage penalty among obese workers. Recall that each work group reports to a single manager. Therefore key unobserved work group characteristics include the attributes of the manager and her or his potential attitude towards obese subordinates. The first two discrimination models should account for the primary channels through which indirect, coworker discrimination would manifest. Therefore the work group fixed effects model can be thought of as a manager fixed effects model that is able to account for direct supervisor discrimination.[62]

---

[61] An attempt was made to generate gender-specific obesity proportions, but there were several work groups that had only one male or female, making this approach infeasible.

[62] In order to compute work group-level fixed effects males and females are combined into a single regression. This avoids dropping observations for individuals who do not have another member of the same gender in their work group (all observations belong to work groups of at least two individuals, but within gender-specific regressions several individuals do not have any other within-group observations to provide variation). This strategy also boosts the number of observations per workgroup helping to alleviate

All models condition on the controls for human capital and health from the combined productivity model.

The results in Table 29 show that among females there is no evidence for coworker discrimination. The proportion of male coworkers in a work group is insignificantly correlated with wages, and the coefficients on the interaction terms do not indicate that the relationship between work group gender composition and wages varies by BMI category. Controlling for the proportion of obese coworkers in one's work group reduces the observed wage penalty to zero, although this is not a significant change from the combined productivity specification. The proportion of obese coworkers is insignificantly correlated with wages, and the coefficient for the obesity interaction is also insignificant. Moreover, the sign of the interaction coefficient for both overweight and obese employees is the opposite of the hypothesized direction. The wages for both obese and overweight females diminishes as the proportion of obese coworkers increase.

Among males the proportion of females in the work group has no effect on the wage of overweight or obese males. However, as the proportion of obese workers in a group increases, the wages of overweight and obese males actually fall significantly. An obese male with no other obese group members earns a significant wage premium of approximately 12 percent, while obese males in a group comprised entirely of obese coworkers face a nearly 14 percent wage penalty.

---

the potential incidental parameters problem that arises when using least squares dummy variables to estimate a nonlinear fixed effects model. Heckman (1981) suggests that having 8 observations per unit is sufficient to mitigate the problem. In the combined sample the average group size is 12, which should be sufficient to allow for consistent estimation of the parameter of interest (δ_2). Based on differences between coefficient estimates by gender in the previous regression analyses, gender-interactions are included for age, age^2, tenure, tenure^2, log C-reactive protein, log-cholesterol ratio, hypertension, and tiers 2 and 3, in addition to interactions for overweight and obese.

It is unclear what is causing this result. If work groups comprised almost entirely of obese workers are less productive as a group, or reflect over-representation of obese employees in lower-paying positions, then higher proportions of obesity should be negatively correlated with wages among normal-weight workers as well, which is not what the model predicts. An alternative explanation is that normal-weight workers are able to outperform more of their immediate peers if they are in a work group with a high proportion of obese coworkers. If this were the case, then overweight workers should also perform better in comparison to obese employees. However, the overweight interaction term is actually more negative than the obesity interaction for both males and females, suggesting that whatever factor is diminishing wages for heavier employees has a greater effect for overweight compared to obese employees. While the underlying explanation for these results is uncertain, the direction of the effect does not suggest that indirect, peer-level discrimination is an issue within the sample. I turn next to the possibility of supervisor discrimination.

Controlling for unobserved work group (manager) attributes decreases the wage premium among obese males from 4.5 percent to 3.9 percent. The difference is statistically insignificant and the change in coefficient is the opposite direction to what would be expected if unobserved discrimination was occurring. Controlling for work group fixed effects among female employees increases the magnitude of the wage penalty roughly 2.1 percentage points for obese employees. This result is also insignificant and suggests that the coefficients from the pooled work group models are positively biased among females, which is the opposite of what would be expected if manager

discrimination was a dominant factor. Ultimately, though the possibility of discrimination cannot be completely ruled out, the results from the previous analyses do not uncover any evidence in support of this hypothesis.

7. Parenthood

The final set of analyses investigates the possibility that the female wage penalty and male wage premium that remains unaccounted for may be attributable to parenthood. Although this is expected to affect females more than males, males are included in the analysis for comparison of the results. To test this possibility I augment the health and human capital model with interactions between children and the overweight/obesity indicators.

Including an interaction between children and obesity does not produce any significant results among males. However, the results show a significant difference in the wage-obesity relationship between obese mothers and non-mothers. Obese females appear to pay a significant penalty of over 5 percent for each additional child. Obese females without children no longer face any wage penalty, and may in fact earn a wage premium. The results appear to be nonlinear. Obese mothers with one child face no penalty, while those with two children face a penalty of roughly 17%, and those with three or more children face a penalty of roughly 15% compared to normal weight mothers. These results do not hold for overweight females, who face no additional wage penalty for fertility. Moreover, overweight females without children continue to incur a wage penalty of approximately 4 percent, although the estimate remains insignificant.

The difference in wages between obese mothers and non-mothers likely has a straightforward explanation. Non-mothers may have a preference for market production over non-market production. Or perhaps the absence of children allows obese non-mothers to prioritize their careers, regardless of preference. However, the discrepancy in wages between normal-weight and obese mothers is more unusual.

The results indicate that there are important underlying differences between mothers who are obese and those who are not since there is no wage penalty associated with children among normal-weight mothers. This may suggest underlying differences in productivity that are not captured by other control variables. For example, females who are more efficient at work may also receive greater returns to investment in health and thus be less likely to become obese. It is also possible that the underlying differences between obese and normal-weight mothers is a manifestation of differences in human capital accumulation, although accounting for the distinction between staff and senior employees cannot account for the wage penalty attributed to obese mothers. It may also be the case that obese mothers are more likely to be single mothers, which may influence both body mass and productivity. However, controlling for single-motherhood does not significantly change the coefficient on the child-obesity interaction, suggesting that having a spouse present is not the primary factor that differentiates obese mothers from normal-weight mothers. The result is also not attributable to a wage penalty that is increasing in BMI. Obese Mothers with two children have the same mean BMI as those with one child. Although obese mothers with three or more children have a slightly

higher BMI (1.5 units) compared to those with only one child, all obese mothers have a lower mean BMI than non-mothers (Appendix H).

The one outstanding characteristic of obese mothers who are penalized is their age. On average, obese mothers with two children are five years older than normal-weight mothers with two children (48 vs. 43). The same pattern holds among mothers with three or more children (52 vs. 47), and in both cases the difference is statistically significant. Obese mothers with one child (who face no wage penalty) are not significantly older. Although data are not available for all children, the age of the youngest child among obese mothers is, on average, five years older than that for normal-weight mothers. This suggests that obese mothers have older children at home rather than the possibility that obese females delay having children. Findings from the labor literature regarding the motherhood wage penalty indicate that the penalty may fade with time (Baum, 2003; Buligescu, de Crombrughe, Mentesoglu, & Montizaan, 2009), which suggests that normal weight mothers, with younger children, should have a higher penalty compared to older (obese) mothers rather than no penalty. It is possible that the negative health associated with obesity does not begin to have adverse effects on productivity until a more advanced age. Yet, obese non-mothers also have an average age of 47, roughly equivalent to obese mothers with two children. It also does not appear that having children significantly reduces health among obese females (Appendix H), although it is possible that the dual burden of childrearing and low average health are what cause productivity to be reduced. However, this cannot be conclusively demonstrated in the

current data, and the exact mechanism underlying this result remains an issue for future research.

8. Conclusion

Consistent with previous literature I find a significant negative relationship between wages and obesity among females. I also find that obese males earn a wage premium, although this result is not significant. To explain these results, I test for differences in productivity between obese and normal-weight employees. I find evidence that human capital accumulation and health are both confounding the wage-obesity relationship, although neither factor can account for the full difference in wages between obese and non-obese employees. Results suggest that differences in measured health are not attributable to unobserved individual attributes that are correlated with wages (e.g., time-preference, ambition), which suggests that a substantial portion of productivity losses attributable to obese workers may be recoverable through appropriate health intervention.

I find no evidence for peer or supervisor discrimination among males or females. Subsequent analyses indicate that obese females with children are the primary recipients of the wage-penalty for obesity in the sample. Obese mothers incur a significant wage penalty of roughly 5 percent per additional child, on average. Estimates also suggest that obese females without children may actually earn more than normal-weight females, although this result is not statistically significant.

The wage difference between obese mothers and normal-weight mothers may be attributable to unobserved differences in productivity. However, this discrepancy cannot

be ascribed to human capital accumulation, delayed child-bearing, or single-parenthood.[63] Uncovering the exact mechanism driving the productivity difference between obese and normal-weight females is a promising avenue for future research: a better understanding of this mechanism could yield valuable information for employers or policymakers wishing to improve maternal health and well-being.

These results may also inform research on women's wages in other branches of the literature. For instance, the motherhood wage penalty established in the labor literature (see Budig & England, 2001) may fall more heavily on obese than normal-weight women. Lastly, these results may provide guidance for future wage-obesity analyses at the national level. If the frequently observed wage penalty for obese women is primarily attributable to obese mothers across region and industry, this may provide justification for public policy aimed at promoting maternal health. Such policy could help to recover some of the lost productivity and medical spending currently attributable to obesity.

[63] Although without a full panel of adulthood it is impossible to account for past spells of single-motherhood that may have affected investments in health or interruptions in work experience.

REFERENCES

Aitkin, M., & Aitkin, I. (1996). A Hybrid EM/Gauss-Newton Algorithm for maximum likelihood in mixture distributions. *Statistics and Computing, 6,* 127–130.

Ashton, H. C. (2001). Pharmacology and effects of cannabis: A brief review. *British Journal of Psychiatry, 178,* 101–106.

Atella, V., Pace, N., & Vuri, D. (2008). Are employers discriminating with respect to weight? European evidence using quantile regression. *Economics and Human Biology, 6,* 305–329.

Averett, S., & Korenman, S. (1999). Black-White differences in social and economic consequences of obesity. *International Journal of Obesity, 23,* 166–173.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Baum, C. (2003). The effect of work interruptions on women's wages. *Labour, 16*(1), 1–36.

Baum, C. L. II, & Ford, W. F. (2004). The wage effects of obesity: A longitudinal study. *Health Economics, 13,* 885–899.

Bhattacharya, J., & Bundorf, M. K. (2009). The incidence of the healthcare costs of obesity. *Journal of Health Economics, 28*(3), 649–658.

Biddle, J., & Hamermesh, D. (2013). Wage discrimination over the business cycle. *IZA Journal of Labor Policy, 2*(7), 1–19.

Blackburn, M. L. (2007). Estimating wage differentials without logarithms. *Labour Economics, 14,* 73–98.

Borle, S., Dholakia, U. M., Singh, S. S., & Westbrook, R. A. (2007). The impact of survey participation on subsequent customer behavior: An empirical investigation. *Marketing Science, 26*(5), 711–726.

Bozoyan, C., & Wolbring, T. (2011). Fat, muscles, and wages. *Economics and Human Biology, 9,* 356–363.

Bray, J. W., Zarkin, G. A., Ringwalt, C., & Qi, J. (2000). The relationship between marijuana initiation and dropping out of high school. *Health Economics, 9,* 9–18.

British Medical Association Board of Science. (2013). *Drugs of dependence: The role of medical professionals.* Retrieved from http://bma.org.uk/news-views-analysis/in-depth-drugs-of-dependence/full-report

Brook, J. S., Stimmel, M. A., & Brook, D. W. (2008). Association between early marijuana use and subsequence academic achievement and health problems: A longitudinal study. *The American Journal on Addictions, 17*(2), 155–160.

Brown, J., & Finkelstein, A. (2008). The interaction of public and private insurance: Medicaid and the long-term care insurance market. *American Economic Review, 98*(3), 1083–1102.

Brunello, G., & D'Hombres, B. (2007). Does body weight affect wages? Evidence from Europe. *Economics and Human Biology, 5*(1), 1–19.

Budig, M. J., & England, P. (2001). The wage penalty for motherhood. *American Sociological Review, 66*(2), 204–225.

Buligescu, B., de Crombrughe, D., Mentesoglu, G., & Montizaan, R. (2009). Panel estimates of the wage penalty of maternal leave. *Oxford Economic Papers, 61,* 35–55.

Bureau of Labor Statistics. (2013). *National Longitudinal Survey of Youth 1979. Survey Year 2010*. Retrieved December, 2013, from https://www.nlsinfo.org/investigator/pages/search.jsp#

Burkhauser, R. V., & Cawley, J. (2008). Beyond BMI: The value of more accurate measures of fatness and obesity in social science research. *Journal of Health Economics, 27*(2), 519–552.

Burton, W. N., Chen, C. Y., Schultz, A. B., & Edington, D. W. (1998). The economic costs associated with body mass index in a workplace. *Journal of Occupational and Environmental Medicine, 40*(9), 786-792.

Cameron, C., & Trivedi, P. (2005). *Microeconomics: Methods and applications.* Cambridge, MA: Cambridge University Press.

Carr, D., & Friedman, M. A. (2005). Is obesity stigmatizing? Body weight, perceived discrimination, and psychological well-being in the United States. *Journal of Health and Social Behavior, 46,* 244–259.

Cawley, J. (2004). The impact of obesity on wages. *The Journal of Human Resources, 39*(2), 451–474.

Centers for Disease Control and Prevention. (n.d.). *National Hospital Ambulatory Medical Care Survey: 2010 Emergency Department Summary Tables.* Retrieved June 23, 2014, from http://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2010_ed_web_tables.pdf

Centers for Medicare and Medicaid Services. (2012). *Medicaid Statistical Information System.* Retrieved May 25, 2014, from http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MSIS-Tables.html

Clogg, C. C., Petkova, E., & Shihadeh, E. S. (1992). Statistical methods for analyzing collapsibility in regression models. *Journal of Educational and Behavioral Statistics, 17*(1), 51–74.

Conway, R. W., & Maxwell, W. L. (1962). A queuing model with state dependent rates. *Journal of Industrial Engineering, 12,* 132–136.

Copeland, J., & Swift, W. (2009). Cannabis use disorder: Epidemiology and management. *International Review of Psychiatry, 21*(2), 96–103.

Craig, L. (2006). Does father care mean fathers share? A comparison of how mothers and fathers in intact families spend time with children. *Gender and Society, 20,* 259–281.

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology, 82*(3), 359–378.

Deb, P., & Trivedi, P. (1997). Demand for medical care by the elderly: A finite mixture approach. *Journal of Applied Econometrics, 12,* 313–336.

Deb, P., & Trivedi, P. (2002). The structure of demand for health care: Latent class versus two-part models. *Journal of Health Economics, 21,* 601–625.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association, 81*(395), 709–721. doi: 10.1080/01621459.1986.10478327

Engleman, H. M., & Douglas, N. J. (2004). Sleep 4: Sleepiness, cognitive function, and quality of life in obstructive sleep apnoea/hypopnoea syndrome. *Thorax, 59*(7), 618–622. doi: 10.1136/thx.2003.015867

Famoye, F. (1993). Restricted Generalized Poisson Regression Model. *Communications in Statistics—Theory and Methods, 22*(5), 1335–1354.

Finkelstein, E. A., Khavjoi, O. A., Thompson, H., Trogdon, J. G., Pan, L., Sherry, B., & Dietz, W. (2012). Obesity and severe obesity forecasts through 2030. *American Journal of Preventive Medicine, 42*(6), 563–570.

Finkelstein, E., DiBonaventura, M. d., Burgess, S., & Hale, B. (2010). The costs of obesity in the workplace. *Journal of Occupational and Environmental Medicine, 52*(10), 971–976.

Finkelstein, E., Trogdon, J., Cohen, J. W., & Dietz, W. (2009). Annual medical spending attributable to obesity: Payer-and service-specific estimates. *Health Affairs, 28*(5), 822–832.

Flegal, K., Carroll, M., Kit, B., & Ogden, C. (2012). Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. *Journal of the American Medical Association, 307*(5), 491–497.

Francis, R. A., Geedipally, S. R., Guikema, S. D., Dhavala, S. S., Lord, D., & LaRocca, S. (2012). Characterizing the performance of the Conway-Maxwell Poisson Generalized Linear Model. *Risk Analysis, 32*(1), 167–183.

French, M. T., Fang, H., & Balsa, A. I. (2011). Longitudinal analysis of changes in drug use and health services utilization. *Health Services Research, 46*(3), 877–899.

French, M. T., McGeary, K. A., Chitwood, D. D., & McCoy, C. B. (2000). Chronic illicit drug use, health services utilization and the cost of medical care. *Social Science and Medicine, 50,* 1703–1713.

French, M. T., Roebuck, M. C., & Alexandre, P. K. (2001). Illicit drug use, employment, and labor force participation. *Southern Economic Journal, 68*(2), 349–368.

Fries, A., Anthony, R. W., Cseko, Jr., A., Gaither, C. C., & Sculman, E. (2008). *The price and purity of illicit drugs 1981–2007.* Technical Report: Contract DASW01-04-C-0003, Task ES-55-2448, "Drug Use Indicators" for the Office of National Drug Control Policy. Institute for Defense Analyses.

Frisco, M. L., & Weden, M. (2013). Early adult obesity and U.S. women's lifetime childbearing experiences. *Journal of Marriage and Family, 75,* 920–932.

Frisco, M. L., Weden, M. M., Lippert, A. M., & Burnett, K. D. (2012). The multidimensional relationship between early adult body weight and women's childbearing experiences. *Social Science and Medicine, 74,* 1703–1711.

Garcia, J., & Quintana-Domeque, C. (2007). Obesity, employment, and wages in Europe. *The Economics of Obesity – Advances in Health Economics and Health Services Research, 17,* 187–217.

Gates, D., Succop, P., Brehm, B., Gillespie, G., & Sommers, B. (2008). Obesity and presenteeism: The impact of body mass index on workplace productivity. *Journal of Occupational and Environmental Medicine, 50*(1), 39–45.

Gerdtham, U.-G., & Trivedi, P. (2001). Equity in Swedish healthcare revisited: New results based on the finite mixture model. *Health Economics Letters, 10,* 565–572.

Goetzel, R. Z., Gibson, T. B., Short, M. E., Chu, B.-C., Waddell, J., Bowen, J., . . . DeJoy, D. M. (2010). A multi-worksite analysis of the relationships among body mass index, medical utilization, and worker productivity. *Journal of Occupational and Environmental Medicine, 52*(1), 52–58.

Gordon, A. J., Conley, J. W., & Gordon, J. M. (2013). Medical consequences of marijuana use: A review of current literature. *Current Psychiatry Reports, 15,* 419.

Gregory, C. A., & Ruhm, C. J. (2009). Where does the wage penalty bite? In M. Grossman & N. Mojan (Ed.), *Economic aspects of obesity* (pp. 315–348). Chicago, IL: University of Chicago Press.

Greve, J. (2008). Obesity and labor market outcomes in Denmark. *Economics and Human Biology, 6,* 350–362.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy, 80,* 223–255.

Guikema, S., & Goffelt, J. (2008). A flexible count data regression model for risk analysis. *Risk Analysis, 28*(1), 213–223.

Gurmu, S., Rilstone, P., & Stern, S. (1999). Semiparametric Estimation of Count Regression Models. *Journal of Econometrics, 88,* 123–150.

Hamersma, S., & Kim, M. (2013). Participation and crowd out: Assessing the effects of parental medicaid expansions. *Journal of Health Economics, 32,* 160–171.

Han, E., Norton, E., & Stearns, S. (2009). Weight and wages: Fat versus lean paychecks. *Health Economics, 18,* 535–548.

Heckman, J. J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In C.

Manski & D. McFadden (Eds.), *Structural analysis of discrete data with econometric applications*. Boston, MA: MIT Press.

Hildebrand, V., & Van Kerm, P. (2010). *Body size and wages in Europe: A semi parametric analysis.* CEPS/INSTEAD Working Paper 2010–09.

Howard, J., & Potter, L. (2012). An assessment of the relationships between overweight, obesity, related chronic conditions, and worker absenteeism. *Obesity Research and Clinical Practice, 8*(1), e1–e15. doi: 10.1016/j.orcp.2012.09.002

Johansson, E., Bockerman, P., Kiiskinen, U., & Heliovaara, M. (2009). Obesity and labour market success in Finland: The difference between having a high BMI and being fat. *Economics and Human Biology, 7,* 36–45.

Johar, M., & Katayama, H. (2012). Quantile regression analysis of body mass and wages. *Health Economics, 21,* 597–611.

Kenkel, D., & Terza, J. (2001). The effect of physician advice on alcohol consumption: Count regression with an endogenous treatment effect. *Journal of Applied Econometrics, 16*(2), 165–184.

Kortt, M., & Leigh, A. (2010). Does size matter in Australia? *Economics Record, 86*(272), 71–83.

Lambert, D. (1992). Zero-inflated Poisson regression with application to defects in manufacturing. *Technometrics, 34*(1), 1–14.

Laurent, T., & Mihoubi, F. (2012). Sexual orientation and wage discrimination in France: The hidden side of the rainbox. *Journal of Labor Research, 33*(4), 487–527.

Liu, Q., & Gupta, S. (2011). The impact of direct-to-consumer advertising of prescription drugs on physician visits and drug requests: Empirical findings and public policy implications. *Internal Journal of Research in Marketing, 28,* 205–217.

Loh, E. S. (1993). The economic effects of physical appearance. *Social Science Quarterly, 74*(2), 420–438.

Lord, D., Geedipally, S. R., & Guikema, S. (2010). Extension of the application of Conway-Maxwell-Poisson models: Analyzing traffic crash data exhibiting underdispersion. *Risk Analysis, 30*(8), 1268–1276.

Lord, D., Guimeka, S., & Geedipally, S. R. (2008). Application of the Conway-Maxwell-Poisson generalized linear model for analyzing motor vehicle crashes. *Accident Analysis and Prevention, 40,* 1123–1134.

Lundborg, P., Bolin, K., Hojgard, S., & Lindgren, B. (2007). Obesity and occupational attainment among the 50+ of Europe. *The Economics of Obesity–Advances in Health Economics and Health Services Research, 17*, 219–251.

MacDonald, Z., & Pudney, S. (2000). Illicit drug use, unemployment, and occupational attainment. *Journal of Health Economics, 19,* 1089–1115.

Maislin, G., Pack, A. I., Kribbs, N. B., Smith, P. L., Schwartz, A. R., Kline, L. R., . . . Dinges, D. F. (1995). A survey screen for prediction of apnea. *Sleep, 18*(3), 158–166.

Manning, W. G., & Mullahy, J. (2001). Estimating log models: To transform or not to transform? *Journal of Health Economics, 20,* 461–494.

McCaffrey, D. F., Pacula, R. L., Han, B., & Ellikson, P. (2010). Marijuana use and high school dropout: The influence of unobservables. *Health Economics, 19*(11), 1281–1299. doi: 10.1002/hec.1561

McGeary, K. A., & French, M. T. (2000). Illict drug use and emergency room utilization. *Health Services Research, 35*(1), 153–169.

Mincer, J. A. (1974). Schooling, experience, and earnings. *Journal of Political Economy, 83*(2), 444–446.

Minka, T. P., Shmeuli, G., Kadane, J. B., Borle, S., & Boatwright, P. (2003). Computing with the COM-Poisson Distribution. *Technical Report 775*. Pittsburgh, PA: Carnegie Mellon University Department of Statistics.

Mittleman, M. A., Lewis, R. A., Maclure, M., Sherwood, J. B., & Mueller, J. E. (2001). Triggering MI by marijuana. *Circulation, 103,* 2805–2809.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics, 33*(3), 341–365.

National Institute on Drug Abuse. (2014). *DrugFacts: Nationwide trends.* Retrieved June 23, 2014, from http://www.drugabuse.gov/publications/drugfacts/nationwide-trends

Office of National Drug Control Policy. (2013). *Marijuana Resource Center: State laws related to marijuana.* Retrieved September 6, 2013, from http://www.whitehouse.gov/ondcp/state-laws-related-to-marijuana

Ogden, C., Carroll, M., Kit, B., & Flegal, K. (2010). *Obesity and socioeconomic status in adults: United States, 2005–2008.* NCHS Data Brief No. 50, December 2010.

Retrieved February 16, 2013, from http://www.cdc.gov/nchs/data/databriefs/db50.pdf

Ogden, C., Carroll, M., Kit, B., & Flegal, K. (2012). *Prevalence of obesity in the United States, 2009–2010.* NCHS Data Brief No. 82, January 2012. Retrieved February 16, 2013, from http://www.cdc.gov/nchs/data/databriefs/db82.pdf

Pagan, J., & Davila, A. (1997). Obesity, occupational attainment, and earnings. *Social Science Quarterly, 78*(3), 756–770.

Polen, M. R., Sidney, S., Tekawa, I. S., Sadler, M., & Friedman, G. D. (1993). Health care use by frequent marijuana smokers who do not smoke tobacco. *The Western Journal of Medicine, 158*(6), 597–601.

Popovici, I., & French, M. T. (2014). Cannabis use, employment, and income: Fixed effects analysis of panel data. *Journal of Behavioral Health Services Research, 41*(2), 185–202. doi: 10.1007/s11414-013-9349-8

Price, G. (2009). The problem of the 21st century: Economics faculty and the color line. *Journal of Socio-Economics, 38*(2), 331–343.

Pronk, N., Martinson, B., Kessler, R., Beck, A., Simon, G., & Wang, P. (2004). The association between work performance and physical activity, cardiorespiratory fitness, and obesity. *Journal of Occupational and Environmental Medicine, 46*(1), 19–25.

Register, C., & Williams, D. (1990). Wage effects of obesity among young workers. *Social Science Quarterly, 71*(1), 130–141.

Rey, J. M., & Tennant, C. C. (2002). More evidence establishes clear link between use of cannabis and psychiatric illness. *BMJ, 325,* 1183–1184.

Ricci, J., & Chee, E. (2005). Lost productive time associated with excess weight in the US workforce. *Journal of Occupational and Environmental Medicine, 47*(12), 1227–1234.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63,* 581–592.

Rudolph, C. W., Wells, C. L. Weller, M. D., & Baltes, B. B. (2009). A meta-analysis of empirical studies of weight-based bias in the workplace. *Journal of Vocational Behavior, 74,* 1–10.

Schafer, M. H., & Ferraro, K. F. (2011). The stigma of obesity: Does perceived weight discrimination affect identity and physical health? *Social Psychology Quarterly, 74*(1), 76–97.

Schmier, J., Jones, M., & Halpern, M. (2006). Cost of obesity in the workplace. *Scandinavian Journal of Work, Environment, and Health, 32*(1), 5–11.

Schmitz, H. (2012). More health care utilization with more insurance coverage? Evidence from a latent class model with German data. *Applied Economics, 44,* 4455–4468.

Schuster, R. M., Crane, N. A., & Gonzalez, R. (2012). The influence of inhibitory control and episodic memory on the risky sexual behavior of young adult cannabis users. *Journal of International Neuropsychology Society, 18*(5), 827–833.

Sellers, K., Borle, S., & Shmueli, G. (2011). The COM-Poisson model for count data: A survey of methods and applications. *Applied Stochastic Models in Business and Industry, 28,* 104–116.

Sellers, K., & Shmeuli, G. (2010). A flexible regression for count data. *The Annals of Applied Statistics, 4*(2), 943–961.

Shimokawa, S. (2008). The labour market impact of body weight in China: A semiparametric analysis. *Applied Economics, 40,* 949–968.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., & Boatwright, P. (2005). A useful distribution for modeling discrete data: Revival of the Conway-Maxwell-Poisson Distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 54*(1), 127–142.

Sidney, S., Beck, J. E., Tekawa, I. S., Quesenberry, C. P., & Friedman, G. J. (1997). Marijuana use and mortality. *American Journal of Public Health, 87*(4), 585–590.

Smith, E., & Crespo, C. (2001). Dietary intake and nutritional status of US adult marijuana users: Results from the third national health and nutrition examination survey." *Public Health Nutrition, 4*(3), 781–786.

Terza, J. V. (1998). Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics, 84,* 129–154.

Terza, J. V. (2009). Parametric nonlinear regression with endogenous switching. *Econometric Reviews, 28*(6), 555–580.

Terza, J. V. (2012a). *Correct standard errors for multistage regression-based estimators: A practitioner's guide with illustrations.* Unpublished Manuscript.

Terza, J. V. (2012b, November). Smoking during pregnancy and birth weight: A practitioner's guide to health policy analysis from a potential outcomes perspective. *Working Paper*.

Trivedi, P., & Zimmer, D. (2005). Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics, 1*(1), 1–111.

Tsai, S. P., Ahmed, F. S., Wendt, J. K., Bhojani, F., & Donnelly, R. P. (2008). The impact of obesity on illness absence and productivity in an industrial population of petrochemical workers. *Annals of Epidemiology, 18,* 8–14.

Tucker, L. A., & Friedman, G. M. (1998). Obesity and absenteeism: An epidemiologic study of 10,825 employed adults. *American Journal of Health Promotion, 12*(3), 202–207.

Ulfberg, J., Carter, N., Talback, M., & Edling, C. (1996). Excessive daytime sleepiness at work and subjective work performance in the general population and among heavy snorers and patients with obstructive sleep apnea. *Chest, 110*(3), 659–663.

Van Ophem, H. (2000). Modeling selectivity in count-data models. *Journal of Business and Economic Statistics, 18*(4), 503–511.

Van Ours, J. C., & Williams, J. (2012). The effects of cannabis use on physical and mental health. *Journal of Health Economics, 31,* 564–577.

Vgontzas, A. N., Tan, T. L., Bixler, E. O., Martin, L. F., Shubet, D., & Kales, A. (1994). Sleep apnea and sleep disruption in obese patients. *Archives of Internal Medicine,* 154(15), 1705–1711.

Wada, R., & Tekin, E. (2010). Body composition and wages. *Economics and Human Biology, 8,* 242–254.

Wang, W., & Famoye, F. (1997). Modeling household fertility decisions with generalized Poisson regression. *Journal of Population Economics, 10*(3), 273–283.

Ware, Jr., J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care, 30*(6), 473–483.

Williams, J., & Skeels, C. (2006). The impact of cannabis use on health. *De Economist, 154*(4), 517–546.

Winkelmann, R. (2008). *Econometric analysis of count data.* Springer.

Yamada, T., Kendix, M., & Yamada, T. (1996). The impact of alcohol consumption and marijuana use on high school graduation. *Health Economics, 5,* 77–92.

Zimmer, D., & Trivedi, P. (2006). Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *Journal of Business and Economic Statistics, 24*(1), 63–76.

Zou, Y., Lord, D., & Geedipally, S. R. (2011, November 14). *Over- and under-dispersed count data: Comparing the Conway-Maxwell-Poisson and Double-Poisson distributions.* Presentation at the 91st Annual Meeting of the Transportation Research Board.

APPENDIX A

EXPECTED VALUE OF UNOBSERVED CONFOUNDER IN CONWAY-MAXWELL POISSON DISTRIBUTION

$$
\begin{aligned}
\int_a^b \exp(X\beta)\varphi(X)dX &= \frac{1}{\sqrt{2\pi}}\int_a^b \exp(X\beta)\exp\left(-\frac{1}{2}X^2\right)dX \\
&= \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{1}{2}X^2 + X\beta\right)dX \\
&= \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{1}{2}X^2 + X\beta - \frac{1}{2}\beta^2 + \frac{1}{2}\beta^2\right)dX \\
&= \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{1}{2}(X^2 - 2X\beta + \beta^2) + \frac{1}{2}\beta^2\right)dX \\
&= \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{1}{2}(X-\beta)^2 + \frac{1}{2}\beta^2\right)dX \\
&= \exp\left(\frac{1}{2}\beta^2\right)\frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{1}{2}(X-\beta)^2\right)dX \\
&= \exp\left(\frac{1}{2}\beta^2\right)\frac{1}{\sqrt{2\pi}}\int_{a+\beta}^{b+\beta} \exp\left(-\frac{1}{2}u^2\right)du \\
&= \exp\left(\frac{1}{2}\beta^2\right)\int_{a+\beta}^{b+\beta} \varphi(u)\,du \\
&= \exp\left(\frac{1}{2}\beta^2\right)[\Phi(b+\beta) - \Phi(a+\beta)]
\end{aligned}
$$

When a and b are negative and positive infinity (respectively), the second multiplicand is 1, proving the result.

EXPECTED VALUE OF UNOBSERVED CONFOUNDER IN (APPROXIMATED) CONWAY-MAXWELL POISSON DISTRIBUTION

$$
\begin{aligned}
\int_a^b \exp(X\beta)\varphi(X)dX &= \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(X\frac{\beta}{\nu}\right)\exp\left(-\frac{1}{2}X^2\right)dX \\
&= \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{1}{2}X^2 + X\frac{\beta}{\nu}\right)dX \\
&= \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{1}{2}X^2 + X\frac{\beta}{\nu} - \frac{1}{2}\left(\frac{\beta}{\nu}\right)^2 + \frac{1}{2}\left(\frac{\beta}{\nu}\right)^2\right)dX \\
&= \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{1}{2}\left(X^2 - 2X\frac{\beta}{\nu} + \left(\frac{\beta}{\nu}\right)^2\right) + \frac{1}{2}\left(\frac{\beta}{\nu}\right)^2\right)dX \\
&= \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{1}{2}\left(X - \frac{\beta}{\nu}\right)^2 + \frac{1}{2}\left(\frac{\beta}{\nu}\right)^2\right)dX \\
&= \exp\left(\frac{1}{2}\left(\frac{\beta}{\nu}\right)^2\right)\frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{1}{2}\left(X - \frac{\beta}{\nu}\right)^2\right)dX \\
&= \exp\left(\frac{1}{2}\left(\frac{\beta}{\nu}\right)^2\right)\frac{1}{\sqrt{2\pi}}\int_{a+\frac{\beta}{\nu}}^{b+\frac{\beta}{\nu}} \exp\left(-\frac{1}{2}u^2\right)du \\
&= \exp\left(\frac{1}{2}\left(\frac{\beta}{\nu}\right)^2\right)\int_{a+\frac{\beta}{\nu}}^{b+\frac{\beta}{\nu}} \varphi(u)\,du \\
&= \exp\left(\frac{1}{2}\left(\frac{\beta}{\nu}\right)^2\right)\left[\Phi\left(b + \frac{\beta}{\nu}\right) - \Phi\left(a + \frac{\beta}{\nu}\right)\right]
\end{aligned}
$$

When a and b are negative and positive infinity (respectively), the second multiplicand is 1, proving the result.

APPENDIX B

COMPARISON OF THE APPROXIMATED AND "TRUE" TREATMENT EFFECTS FOR THE CMP

As discussed in Section 1 there are two expressions for the conditional mean of Y in the CMP specification. The approximated value is expressed as

$$E[Y|X,\nu] = \lambda^{1/\nu} - \frac{\nu - 1}{2\nu}$$

whereas the "true" conditional mean can be expressed as

$$E[Y|X,\nu] = \lambda\left(\frac{\sum_{j=1}^{\infty}\frac{j\lambda^{j-1}}{(j!)^{\nu}}}{\sum_{j=0}^{\infty}\frac{\lambda^{j}}{(j!)^{\nu}}}\right).$$

Using the approximation typically results in additional bias of both the estimated treatment effect derived from the conditional expectation, and of the expectation itself. Comparison of the two methods is presented in the following tables. Absolute percentage bias is reported in parentheses. Mean squared error is shown in brackets for the estimated treatment effects.

Comparison of Estimated Y with Positive Dependence				
	Model	**Actual Value**	**"True" estimate**	**Approximation**
Endogenous Treatment	CMP	3.13	3.05 (2.33)	2.91 (6.84)
	NB	3.00	2.53 (15.65)	2.50 (16.65)
	RGP	3.00	2.97 (1.30)	2.91 (3.07)
Endogenous Sample Selection	CMP	3.13	3.05 (2.70)	2.93 (6.42)
	NB	3.00	3.49 (14.06)	3.40 (10.75)
	RGP	3.00	3.50 (13.95)	3.38 (10.31)

Comparison of Estimated Y with Negative Dependence				
	Model	**Actual Value**	**"True" estimate**	**Approximation**
Endogenous Treatment	CMP	3.03	2.90 (2.57)	3.62 (21.52)
Endogenous Sample Selection	CMP	3.03	3.03 (1.98)	3.23 (8.30)

Comparison of Estimated ATE with Positive Dependence				
	Model	**Actual Value**	**"True" estimate**	**Approximation**
Endogenous Treatment	CMP	.917	.921 (10.95) [.024]	.770 (16.97) [.042]
	NB	.826	.661 (19.95) [.092]	0.655 (20.69) [.037]
	RGP	.826	.758 (11.31) [.249]	0.743 (12.23) [.230]
Endogenous Sample Selection	CMP	.917	.925 (8.05) [.015]	.774 (15.90) [.033]
	NB	.826	.964 (17.22) [.557]	.936 (14.34) [.026]
	RGP	.826	.793 (7.60) [.184]	.923 (13.65) [.025]

Comparison of Estimated ATE with Negative Dependence				
	Model	**Actual Value**	**"True" estimate**	**Approximation**
Endogenous Treatment	CMP	0.766	.939 (24.21) [.292]	1.13 (48.92) [.143]
Endogenous Sample Selection	CMP	0.766	.793 (7.60) [.184]	0.829 (10.70) [.013]

While the comparison tables suggest that the approximation performs worst when the data are CMP distributed, a closer look at the data generating processes paints a slightly different picture. The CMP data with positive dependence were generated using $\nu = 0.281$. The estimated dispersion parameter values for the NB and RGP models were roughly 0.9 and 0.8, respectively. This suggests that the approximation breaks down when the positive dependence becomes too great, rather than being specific to a particular specification. The results from the tables of negative dependence also suggest the approximation does not perform well when the counts are negatively dependent. It is unclear at what level of positive dependence the approximations begin to increase substantially in bias, but it is recommended that the "true" conditional mean be used to produce estimates for any data that reject a standard Poisson distribution as the baseline specification (i.e., $\hat{\nu}$ is significant).

APPENDIX C

MEANS FROM THE SUBSTANCE ABUSE LITERATURE

Study	Outcome Measure	Means
Atella and Deb (2008)	Primary care visits last 4 weeks	0.235
	Public specialist visits last 4 weeks	0.096
	Private specialist visits last 4 weeks	0.136
Ayyagari et al. (2011)	Drinks per day	0.68
Bauer, Gohlmann, and Sinning (2007)	Cigarettes per Day	3.63–8.10
Birch, Eyles, and Newbold (1993)	Physician visits last year	2.79
Brown, Scheffler, Seo, and Reed (2006)	Cigarettes per month	99.38
Davalos, Fang, and French (2012)	Days of binge drinking last year	12.8
Deb and Holmes (2000)	Mental health visits last year	3.88
Deb and Trivedi (2002)	Physician visits last year	2.861
	Outpatient visits last year	3.546
Deb et al. (2006)	Physician visits last year	2.83–3.56
	Non-physician visits last year	0.18–1.87
	ER visits last year	0.13–0.28
Deb and Trivedi (2008)	Physician visits last year	3.52
	Non-physician visits last year	0.26
	Surgery last year	0.18
	ER visits last year	0.26

Study	Outcome Measure	Means
	Inpatient hospital visits last year	0.11
Decker et al. (2011)	Inpatient hospital visits last year	0.18–0.28
Farbmacher (2012)	Physician visits last three months	1.325
Gerdtham (1997)	Physician visits last year	2.06
	Weeks in hospital	0.3
Gerdtham and Trivedi (2001)	Physician visits last year	2.06
	Weeks in hospital last year	0.3
Greene (2009)	Physician visits last three months	3.18
	Hospital visits last year	0.14
Grootendorst (1995)	Prescription drugs in last month	2.02–2.11
Gupta and Greeve (2011)	Physician visits last year	5.19
Gustavsen, Nagya, and Wu (2010)	Physician visits last year	4.03
Hyppolite and Trivedi (2012)	Physician visits last year	2.35–2.66
Jimenez-Martin, Labeaga, and Martinez-Granado (2002)	Physician visits last year	3.39–3.53
	Specialist visits last year	1.07
Jochman and Leon-Gonzalez (2004)	Physician visits last four months	4.12
Lee and Kobayashi (2001)	Physician visits last year	3.34–3.64
	Hospital visits last year	0.60–0.65

Study	Outcome Measure	Means
Lourenco and Ferreira (2005)	Physician visits last year	5.52
Madden, Nolan, and Nolan (2005)	Physician visits last year	2.30–6.50
McLeod (2011)	Physician visits last year	3.26
Meyerhoeffer and Zuvekas (2009)	Physician visits last year	2.76
	Mental health visits last year	0.23
Moreira and Barros (2010)	Physician visits last three months	1.01
Mullahy (1998)	Physician visits last year	4.91
Nolan (2007)	Physician visits last year	3.4
Ovrum (2004)	Physician visits last year	4.94
Saez et al. (2006)	Physician visits last year	1.00–1.33
	Specialist visits last year	1.13–1.31
Sari (2009)	Physician visits last year	3.03–3.68
	Specialist visits last year	0.68–0.85
	Inpatient hospital visits last year	0.40–0.82
Sarma and Simpson (2006)	Physician visits last year	3.49
	Specialist visits last year	0.72
	Inpatient hospital visits last year	0.62
Schellhorn (2001)	Physician visits last year	2.15–2.77
	Specialist visits last year	1.12–2.19
Shafrin (2010)	Inpatient surgery last year	0.50
	Outpatient surgery last year	1.37

Study	Outcome Measure	Means
Sheu et al. (2004)	Cigarettes per day among daily smokers	16.10
Windmeijer and Silva (1997)	Physician visits last month	0.40
Winkelmann (2004)	Physician visits last year	2.39–2.69
Winkelmann (2006)	Physician visits last three months	2.46–2.96
Van Ourti (2004)	Physician visits last year	4.94
Yen and Jones (1996)	Cigarettes per day among daily smokers	9.38
Yen, Tang, and Su (2001)	Traditional medicine visits per month	0.22
Zhong (2010)	Physician visits last month	0.40
	Inpatient days last year	0.06
Zimmer and Trivedi (2006)	Physician visits last year	2.33–3.92
	Specialist visits last year	1.03–1.70
	ER visits last year	0.10–0.11

APPENDIX D

CRITERIA FOR MARIJUANA DEPENDENCE OR ABUSE

Dependence:

1. Spent a great deal of time over a period of a month getting, using, or getting over the effects of substance
2. Unable to keep set limits on substance use or used more often than intended.
3. Needed to use substance more than before to get desired effects or noticed that using the same amount had less effect than before.
4. Unable to cut down or stop using the substance every time he or she tried or wanted to.
5. Continued to use substance even though it was causing problems with emotions, nerves, mental health, or physical problems.
6. Reduced or gave up participation in important activities due to substance.

Abuse:

1. Respondent reported having serious problems due to substance use at home, work, or school.
2. Respondent reported using substance regularly and then did something where substance use might have put them in physical danger.
3. Respondent reporting substance use causing actions that repeatedly got them in trouble with the law.
4. Respondent reported having problems caused by substance use with family or friends and continued to use substance even though it was thought to be causing problems with family and friends.

APPENDIX E

CORRECTION OF SECOND-STAGE STANDARD ERRORS TO ACCOUNT FOR FIRST-STAGE ESTIMATION OF RELEVANT PARAMETERS (TERZA, 2012)

Let q represent the function that is to be optimized in the second-stage (the square of the residuals in the case of NLS). Then define the following:

$\nabla_{\delta}q$ = gradient of the second-stage optimization function with respect to the first-stage estimates
$\nabla_{\beta}q$ = gradient of the second-stage optimization function with respect to the second-stage estimates
$\nabla_{\delta\beta}q$ = cross-partial derivative of second-stage optimization function with respect to the first and second-stage estimates
$\nabla_{\beta\beta}q$ = Hessian matrix of the second-stage optimization function
Ω_1 = variance-covariance matrix from estimation of first-stage
Ω_2 = variance-covariance matrix from estimation of second-stage

When the second stage is estimated via NLS, as in the present case, the corrected variance-covariance matrix Ω^* can be computed as:

$$\Omega^* = E[\nabla_{\beta\beta}q]^{-1}E[\nabla_{\delta\beta}q]'\Omega_1E[\nabla_{\delta\beta}q]\ E[\nabla_{\beta\beta}q]^{-1} + \Omega_2$$

APPENDIX F

MISSING AND IMPUTED DATA

Omitted Observations		
Reason Omitted	Females	Males
Salary nonresponse	19 (5.9%)	36 (7.2%)
BMI nonresponse	6 (2.0%)	5 (1.1%)
TOTAL	25 (7.9%)	41 (8.4%)

Note. Values in parentheses refer to percentage of original (full) sample dropped for each reason

Observations were dropped from the analysis sample if respondents did not provide their salary or a BMI value. Nineteen females and 36 males failed to provide salary data. This represents approximately 6 percent of all females and 7 percent of all males. Six females and five males failed to provide their BMI. This represents 2 percent of females and 1 percent of males. Among the remainder of the control variables in Model (1) no nonresponses are recorded. The following table reports differences in means between non-respondents and respondents. On average, non-responding females have less tenure, and are more likely to be nonwhite or foreign. Male non-respondents also have less tenure, and more likely to be nonwhite or foreign. They are also significantly younger, on average. These marginal differences do not suggest that weight or salary-level are correlated with item non-response.

Summary Statistics for Observations with Omitted Data				
	Females		Males	
	Omitted	Retained	Omitted	Retained
Married or Cohabitating	60.00% (0.50)	70.85% (0.46)	85.37% (0.34)	84.74% (0.36)
Total Number of Children	1.52 (1.45)	1.63 (1.24)	1.61 (1.28)	1.64 (1.43)
Nonwhite	36.00%* (0.49)	24.07% (0.43)	48.78%* (0.49)	29.20% (0.46)
Born Abroad	32.00%* (0.48)	19.32% (0.40)	53.65%* (0.53)	31.89% (0.44)
College Graduate (4-year)	76.00% (0.44)	66.44% (0.47)	88.80% (0.32)	83.63% (0.37)
Age	46.52 (12.56)	46.88 (8.38)	41.95* (9.20)	45.12 (8.70)
Tenure with Firm (years)	11.12* (10.87)	16.04 (9.94)	8.80* (8.60)	12.07 (8.16)
Support Personnel	12.00% (0.33)	5.76% (0.23)	9.75% (0.30)	6.86% (0.25)
Staff	24.00% (0.44)	36.27% (0.48)	29.27% (0.46)	33.41% (0.47)
Senior	64.00% (0.50)	57.97% (0.49)	60.98% (0.49)	59.73% (0.49)

* Indicates significant difference in means between omitted and retained observations ($p < 0.1$).
Omitted observations were left out of the sample due to missing salary or missing BMI.

Comparison of Means Between Imputed and Reported Health Variables				
	Females		Males	
	Imputed	Non-imputed	Imputed	Non-imputed
C-Reactive Protein	1.65* (2.10) [15]	3.14 (4.42) [280]	1.88 (0.84) [34]	1.87 (2.41) [418]
Cholesterol Ratio	3.76 (0.25) [30]	3.88 (1.15) [265]	4.48 (0.26) [58]	4.47 (1.44) [394]
Heart Rate	73.40 (0.43) [2]	72.35 (11.14) [293]	68.85 (3.88) [4]	69.76 (11.38) [448]

Standard errors in parenthesis. *N* in brackets.
* Indicates significant difference between imputed and non-imputed means ($p < 0.1$).

Among control variables the only item non-response occurred among biological measures that proxy for cardiovascular health. Due to the relatively high number of non-responses values were imputed for missing observations rather than dropping the entire observation from the sample. Imputation followed the modified regression-based EM algorithm detailed in Cameron and Trivedi (2005) Section 27.5 (pp. 931–932). X_1 refers to control variables for observations with no missing responses. X_2 indicates control variables for observations with a missing response for one of the above biological variables. There are N_1 complete observations and N_2 incomplete observations. The algorithm is as follows:

1. Estimate $\hat{\beta}$ using the N_1 complete observations.
2. Estimate s^2 using only N_1 complete observations.
3. Generate N_2 estimates of the missing values:$\hat{y}_{MIS} = X_2\hat{\beta}$
4. Estimate $\hat{V}[\hat{y}_{MIS}] = s^2(I_{N2} + X_2[X_1'X_1]^{-1}X_2')$
5. Generate adjusted values $\widehat{y_{MIS}^{a}} = (\hat{V}^{-1/2}\hat{y}_{MIS}) \circ u_m$ where u_m is a Monte Carlo draw from the $N(0, s^2)$ distribution and $\circ$ denotes element by element multiplication.
6. Using the augment sample (y_1 and $\widehat{y_{MIS}^{a}}$)obtain a revised estimate of $\hat{\beta}$.
7. Repeat steps 1-6 using the revised $\hat{\beta}$ in step 1 for each iteration.

Consistent with Aitkin and Aitkin (1996), this algorithm is cycled until the difference between successive log likelihood values is less than 10^{-5}

The Relationship Between Wages and Obesity with Standard Control Variables Weighted for Nonresponse

	Female (n = 295)	Male (n = 452)
Obese	-0.065**	0.039
	(0.033)	(0.025)
Overweight	-0.051	-0.004
	(0.034)	(0.022)
Married or Cohabitating	0.027	0.017
	(0.032)	(0.026)
Total Number of Children	-0.002	0.008
	(0.012)	(0.008)
Nonwhite	-0.055*	-0.025
	(0.030)	(0.025)
Born Abroad	0.030	0.037
	(0.034)	(0.026)
College Graduate (4-year)	0.054**	0.060*
	(0.025)	(0.033)
Age (10 years)	0.305**	0.337***
	(0.124)	(0.093)
Age Squared (100 squared years)	-0.029**	-0.032***
	(0.014)	(0.011)
Tenure with Firm (10 years)	-0.148***	-0.027
	(0.062)	(.043)
Tenure Squared (100 squared years)	0.039***	0.012
	(0.015)	(0.011)
Core (Staff and Senior)	0.208**	0.229***
	(0.078)	(0.038)

Note. Dependent variable is hourly wages. The overweight and obese coefficients report wages relative to normal-weight employees (BMI < 25). Estimates are obtained using a generalized linear model with a log link function. The model includes indicators for worksite, state, and job category. Standard errors are clustered at the work group level and reported in parentheses.

* $p < 0.1$, ** $p < 0.5$, *** $p < 0.1$

APPENDIX G

ROBUSTNESS CHECKS

Returns to Job Tenure Among Obese Employees				
	Female (n = 295)		Male (n = 452)	
	Linear	Quadratic	Linear	Quadratic
Obese * Tenure (Years)	0.005 (0.030)	-0.001 (0.005)	-0.009 (0.027)	-0.005 (0.005)
Obese * Tenure^2 (10 years)		0.008 (0.094)		0.081 (0.092)

Note. Dependent variable is hourly wage. Estimates are obtained using a generalized linear model with a log link function. Models control for employee's state, site, age, age^2, race, nativity, marital status, number of children, job category, and an indicator for core vs. support employee, unless reported otherwise. All models also contain an indicator for overweight and all interactions are interacted with overweight. The normal-weight category therefore serves as the point of reference for all obesity and obesity-interaction coefficients. Standard errors are clustered at the work group level and reported in parentheses.
* p < 0.1, ** p < 0.5, *** p < 0.1

Nonlinear Wage Penalty of Children		
	Female (n = 295)	Male (n = 452)
Obese	0.062 (0.051)	0.036 (0.049)
Obese * 1 Child	0.007 (0.068)	0.010 (0.059)
Obese * 2 Children	-0.171*** (0.060)	0.026 (0.0960)
Obese * 3+ Children	-0.150** (0.064)	-0.050 (0.061)

Note. Dependent variable is hourly wage. Estimates are obtained using a generalized linear model with a log link function . Models control for employee's state, site, age, age^2, race, nativity, marital status, number of children, and job category, and an indicator for core vs. support employee, unless reported otherwise. All models also contain an indicator for overweight and all interactions are interacted with overweight. The normal-weight category therefore serves as the point of reference for all obesity and obesity-interaction coefficients. Standard errors are clustered at the work group level and reported in parentheses.
* p < 0.1, ** p < 0.5, *** p < 0.1

APPENDIX H

ADDITIONAL FEMALE SUMMARY STATISTICS

BMI of Obese Females by Number of Children	
No Children (n = 24)	37.71 (5.93)
1 Child (n = 12)	34.99 (4.12)
2 Children (n = 33)	34.95 (4.02)
3+ Children (n = 56)	35.56 (5.80)

Note. Standard errors in parentheses

Age of Obese Mothers Compared to Obese Non-Mothers and Normal-weight Mothers		
	Obese	Normal Weight
No Children	47.04 (6.36)	45.52 (10.51)
1 Child	44.75 (6.45)	45.76 (9.31)
2 Children	48.06** (8.75)	43.25 (7.91)
3+ Children	51.38* (7.01)	47.09 (9.71)

Note. Standard errors in parentheses
* Indicates obese mothers significantly older than normal-weight mothers at 10% significance level
** Indicates obese mothers significantly older than normal-weight mothers at 5% significance level

Health Status of Obese Females by Number of Children				
	No Children	1 Child	2 Children	3+ Children
C-Reactive Protein	6.57 (5.43)	7.85 (4.59)	3.93 (3.03)*	6.32 (9.91)
Cholesterol Ratio	4.17 (0.65)	3.97 (1.25)	4.17 (1.50)	4.18 (1.03)
Hypertension	0.33 (0.48)	0.42 (0.52)	0.36 (0.49)	0.65 (0.49)*
Heart Rate (BPM)	71.86 (12.85)	78.25 (10.40)	73.31 (13.43)	73.66 (11.43)
Physical Function	88.43 (13.40)	93.51 (11.56)	84.51 (22.81)	86.23 (15.70)
Heavy Snoring	0.33 (0.48)	0.42 (0.51)	0.39 (0.50)	0.48 (0.51)

Note. Standard errors in parentheses
* Indicates significant difference in health measure between mothers and non-mothers at 5% level