2015

Three Essays in Health Economics

Lizhong Peng
Lehigh University

Three Essays in Health Economics

by

Lizhong Peng

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Doctor of Philosophy

in

Business and Economics

Lehigh University

April, 2015

Approved and recommendation for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Date: April 10, 2015

Dissertation Advisor: Chad D. Meyerhoefer

Committee Members:

James A. Dearden

Shin-Yi Chou

Samuel H. Zuvekas

Acknowledgements

During my doctoral studies I have benefited from the encouragement and support of many people. I am especially grateful to my advisor, Professor Chad Meyerhoefer, for his patient guidance and for being a constant source of support and inspiration. More importantly, Chad has taught me how to be a researcher. It has been a great pleasure working with him both as his student and colleague. I also would like to thank Professor James Dearden, Professor Shin-Yi Chou, and Doctor Samuel Zuvekas for serving as members of my dissertation committee. Their suggestions and insights are invaluable to this thesis. My gratitude also goes to those who had helped and inspired me over the past several years. Lastly, I am greatly indebted to my family for their love and support.

Contents

List of Tables

List of Figures

Abstract

This dissertation consists of three essays. The first essay examines the unintended consequence of Medicare pay-for-performance programs. I find evidence that the CMS case mix adjustment formula for patient experience measures in the Hospital Value-based Purchasing Program (HVBP) over-corrects (under-corrects) for the effect of patient health status on favorable survey responses for surgical (obstetric) patients, which creates scope for hospital to risk select patients on the basis of health status. Using inpatient discharge data from Pennsylvania and Maryland, I find that average patient severity increased among surgical patients and decreased among obstetric patients after the HVBP took effect. In addition, I find weak evidence of an increase in patient experience measures as a result of the HVBP, but no such effect is found for clinical process measures.

In the second essay, I estimate the short-term effect of depression on labor market outcomes using data from the 2004-2009 Medical Expenditure Panel Survey. After accounting for the endogeneity of depression through a correlated random effects panel data specification, I find that depression reduces the contemporaneous probability of employment by 2.6 percentage points. I do not find evidence of a causal relationship between depression and hourly wages or weekly hours worked. In addition, I examine the effect of depression on work impairment and found that depression increases annual work loss days by about 1.4 days (33 percent), which implies that the annual aggregate productivity loses due to depression range from $700 million to 1.4 billion in 2009 USD.

In the third essay, I investigate the health impacts of unconventional natural gas development of Marcellus shale in Pennsylvania between 2001 and 2013. Through a multivariate regression analysis that compares changes in hospitalization rates over time for air pollution-sensitive disease in counties with unconventional gas wells to changes in hospitalization rates in non-well counties, I find significant associations between shale gas development and hospitalizations for acute myocardial infarction (AMI), chronic obstructive pulmonary disease

(COPD), asthma, pneumonia, and upper respiratory infections (URI). These adverse effects on health are consistent with higher levels of air pollution resulting from unconventional natural gas development.

Chapter 1

Introduction

The Affordable Care Act (ACA) mandates that the Centers for Medicare & Medicaid Services (CMS) incorporate pay-for-performance (P4P) adjustments into Medicare payments. In response, CMS has begun to implement several P4P programs that are aimed at promoting quality and efficiency in the delivery of inpatient care. To ensure fair and accurate comparison of providers' quality, proper "risk adjustment" of the metrics used to determine P4P payments plays a critical role. The second chapter of this dissertation examines the unintended consequences of P4P programs when measures of providers' quality are not fully adjusted for the differences in patient characteristics. Using data from various sources, I study how the Hospital Value-based Purchasing Program (HVBP), Medicare's flagship P4P program, affects the patient mix treated at participating hospitals. I find consistent evidence that hospitals engaged in patient selection through exploiting the incomplete case mix adjustment used by CMS to set payment under the HVBP. In light of recent health care reforms, the findings in this paper have important policy implications. A good case mix adjustment (risk adjustment) method is critical to implementing large-scale public P4P initiatives like the HVBP. Since patient mix varies greatly among different types of hospitals in the U.S., hospitals treating certain patient populations will be unfairly penalized if the performance measures are not fully risk-adjusted. By implication, if safety net hospitals that operate in low-income areas consistently under-perform on quality measures mainly due to the patient mix, they may avoid certain types of patients that are likely to lower their performance scores, which could in turn widen the disparities in access to care for vulnerable populations.

In addition to P4P, I am also interested in the impact of the ACA on other aspects of the health care system. Under the individual and small group exchange provisions of the ACA, a large number of previously uninsured individuals with severe mental disorders will gain coverage. Thus, a correct understanding of economic costs associated with mental disorders is important when evaluating the welfare effects of such initiatives. In the third chapter of this dissertation, I investigate how depression, one of the most common mental health disorders among U.S. adults, affects individuals' short-term job market opportunities. Using panel data techniques, I find that in the short-run, depression has limited adverse effects on key labor market indicators such as employment, wages, or weekly hours worked, but a large effect on workplace productivity loss. These results imply that from a social standpoint part of the costs of coverage expansions under the ACA may be offset through less productivity loss.

Although most of my research focuses on the health care system, I have also conducted research on other topics of broad policy interest. For example, while the boom of unconventional natural gas development (i.e. "fracking") has contributed significantly to energy security in the U.S., it has also raised public health concerns among policy makers and the general public. However, to date most evidence on adverse health effects of unconventional natural gas development is anecdotal. This lack of rigorous scientific evidence has resulted in a debate on the safety of unconventional natural gas development that is founded on speculations, making it difficult for regulators to establish appropriate policies. In the last chapter, I examine the health impacts of the unconventional natural gas development of Marcellus shale in Pennsylvania between 2001 and 2011. I find significant associations between shale gas development and hospitalizations for air-pollution-sensitive conditions among adult residents in Pennsylvania. This is the first study that establishes a consistent link between unconventional natural gas development and higher rates of disease.

Chapter 2

Patient Selection Under Incomplete Case Mix Adjustment: Evidence from the Hospital Value-based Purchasing Program

2.1 Introduction

Historically, hospitals treating Medicare patients are reimbursed on the basis of the quantity of services provided. In the early years after its establishment in 1965, Medicare used a fee-for-service payment model that fully reimburses health care providers for the cost of treatment. Under such a system, hospitals have incentives to oversupply medical services to patients because they do not bear the marginal cost of treatment. In October 1983, Medicare implemented the prospective payment system under which hospitals are paid a flat rate for treating patients with similar conditions. This payment scheme also incentivizes hospitals to strategically choose the extent of treatment, since sicker patients are more costly to treat. Ellis (1998) shows that prospectively paid health care providers have incentives to oversupply medical services to low cost patients and undersupply services to high cost patients. Moreover, providers have no incentives to improve the quality of care if payments are solely based on volume.

In recent years, public policy makers and private insurers have become increasingly interested in the pay-for-performance (P4P) schemes, which reward providers that demonstrate their services are high quality. With the passage of the Affordable Care Act (ACA), the focus has now been shifted toward containing cost growth and promoting better quality of care. In response, the Centers for Medicare & Medicaid Services (CMS) inaugurated the Hospital Value-Based Purchasing Program (HVBP) as an effort to improve the quality of inpatient care. The HVBP is the first large-scale public P4P program that affects acute care hospitals nationwide. Due to the increasing popularity of P4P schemes among the public and private payers, there is a growing literature on the economic evaluation of the effectiveness and efficiency of such quality initiatives. In general, there is mixed evidence regarding the effectiveness of P4P on improving the quality of care (for example, see Mullen, Frank and Rosenthal, 2010; Rosenthal et al., 2005; Li et al., 2014; Ryan, Blustein and Casalino, 2012; Werner et al., 2011).

One potential unintended consequence inherent with this type of quality initiative is that providers may avoid patients that will lower their performance (Shen, 2003). Studies have found a correlation between providers' performance scores under P4P and patient characteristics. In particular, Chien et al. (2012) find that physician groups that serve in low-income areas of California received lower performance scores in a major P4P program. Jha, Orav and Epstein (2011) find that hospitals that care for greater number of elderly black and Medicaid patients are less likely to have high quality ratings. If the measures used to evaluate providers' quality of care are not fully adjusted for the differences in patient population of each provider, they may strategically engage in patient selection in order to increase their performance ratings.[1]

Ideally, if one is to eliminate the scope for patient selection in P4P programs, the performance measures should be selected in a way such that they are not affected by patient characteristics. One example of such a measure would be the percentage of cases where

[1]The quality measures commonly used in P4P programs fall into 4 categories (James, 2012): (1) process measures of the performance of clinically-validated activities that contribute to good health outcomes; (2) outcome measures of the effects of health care on patient's health status; (3) experience measures of patients' perceptions of the health care they received; and (4) structural measures of the features of health care organizations that are related to the ability to provide high-quality health care.

surgical patients are given proper antibiotics to prevent infection. Clinical activities like this are under the control of hospitals and are not dependent on patient characteristics or illness severity. However, in order to attain a comprehensive evaluation of quality of care, the use of measures that are likely to be correlated with patient characteristics is sometimes necessary (for example, outcome-based measures like the mortality rate for pneumonia patients). Under those circumstances, proper adjustment for differences in patient characteristics is critical to ensure fair comparisons across health care providers serving different patient populations. For example, safety net hospitals that treat a large share of disadvantaged patients may rank low in *measured* quality if outcome-based measures such as mortality rates are not fully adjusted for patient illness severity (Jha, Orav and Epstein, 2011).

A large number of statistical methods for adjusting patient characteristics have been developed in the health care industry for such purposes.[2] Public payers often use "risk adjustment"[3] to properly set premiums for private insurers under contract to provide services so that there is no incentive for them to select healthier enrollees or avoid sicker enrollees. For example, CMS uses risk adjustment models in setting payments to private insurers in Medicare Advantage (MA) plans. Also, in recent years, both public policy makers and private payers began publishing "report cards" that publicly report information on health care providers and their quality of health care, which necessitates detailed and credible adjustment for patient characteristics (Werner and Asch, 2005). The newly created HVBP requires that payments are risk adjusted because some of the measures used in the HVBP, such as patients' perceptions of their hospital stay, are closely related to their characteristics, while others are directly based on patient health outcomes (for example, 30-day mortality rate of heart failure patients). However, almost all current risk-adjusters are criticized for low explanatory power. With a large portion of the variance in patient outcomes/costs/responses unexplained, health care providers might still have the ability to "game" the system and

[2]See Schone and Brown (2013) for a detailed description of these methods.

[3]"Risk adjustment" refers to the methods for determining whether patient with certain characteristics have higher utilization of medical services.

benefit from doing so.

As noted by Dranove et al. (2003), health care providers may have superior information on patient conditions than the econometricians. As a result, statistical modeling may not be able to fully account for the effect of patient characteristics on quality measures that are based on or closely related to these same factors. Moreover, risk-averse providers may still want to engage in patient selection even if risk adjustment is accurate on average. Researchers have found evidence of patient selection in various settings. For example, Dranove et al. (2003) find that publication of health care report cards induce providers to select healthier patients for heart surgery. In the context of private insurance, Brown et al. (2014*b*) find that after CMS introduced risk adjustment in its Medicare Advantage Plans, private insurers responded by increasing selection of enrollees along dimensions that are excluded from the CMS risk adjustment formula. In general, incomplete adjustment for patient characteristics creates scope for patient selection, and economic agents tend to respond to this type of incentives quickly.

In this paper, I investigate hospital behavior under quality incentives with incomplete adjustment for patient characteristics. The HVBP provides an ideal setting for this type of study. First, the HVBP employs a set of process measures that are designed to assess the appropriateness of care provided by hospitals. A clinical process refers to the activities performed for or by a patient, and process measures assess whether these activities are properly carried out by health care providers. There is no need to adjust these measures for patient characteristics since hospitals are not able to improve their performance by avoiding or attracting particular types of patients. Second, in addition to the process measures, CMS also includes experience measures (used to assess patient satisfaction) and outcome measures in the program. Since there is a well-documented correlation between patients' perceptions for health care and demographic characteristics, CMS uses coefficients from regression-based models to explicitly adjust these measures for patient characteristics such as age, educational attainment, self-perceived health, and a set of clinical risk factors in order to "level the playing field for all hospitals" (O'Malley et al., 2005). This is called the patient case mix adjustment.[4] However, if such adjustments are incomplete, then hospitals will still have an incentive to select patients who are most likely to supply high ratings. Using data from a large hospital network in Pennsylvania, I find that the CMS patient case mix adjustment[5] formula for patient experience measures have failed to account for the interactive effects between patient health status and service line,[6] thereby failing to account for differences in patients that affect these experience ratings. My results suggest that relative to medical patients, obstetric patients who perceive themselves to be sicker tend to evaluate their hospital experiences more negatively, whereas surgical patients who perceive themselves to be sicker are more likely to give favorable responses, after conditioning on the same variables CMS uses in its case mix adjustment formula. As a result, the current CMS patient case mix adjusters might have created scope for hospitals to systematically select patients that can increase their scores on experience measures.

Based on these findings, I test whether the implementation of the HVBP results in selection on the basis of patient illness severity using inpatient discharge data from the states of Pennsylvania and Maryland from 2009 to 2013. Since acute care hospitals in Maryland are exempt from the HVBP, I employ a difference-in-differences strategy for identification using patients and hospitals in Maryland as the comparison group. I find that relative to patients in Maryland, there is a decrease in average illness severity for obstetric patients at Pennsylvania hospitals after HVBP became effective, whereas the average illness severity increased among surgical patients in Pennsylvania. This provides evidence that the HVBP induced hospitals in Pennsylvania to engage in patient selection in both directions.

[4] See Appendix 2.7.2 for a detailed description of the CMS patient case mix adjustment method for experience measures.

[5] Case mix refers to the type or mix of patients treated by a hospital. Later in this paper, the phrase "case-mix adjustment" is used interchangeably with "risk adjustment" because they both refer to the practice of adjusting for differences in patient characteristics.

[6] O'Malley et al. (2005) find that the most important case-mix adjustment factors are hospital service line, age, race, level of education, self-perceived health status, language spoken at home, and having a circulatory disorder. They also find strong interactive effects between service and age , but do not find the interactions between service line and self-perceived health to have any significant effect on patients responses on HCAHPS, the survey CMS uses for experience measures in HVBP.

Finally, I look at changes in HVBP performance measures over time. Relative to hospitals in Maryland, there is no evidence of improvement in process measures for hospitals in Pennsylvania. However, I do find that evidence that experience measures increased in hospitals in Pennsylvania after the implementation of the HVBP, which is consistent with HVBP-induced changes in hospitals' patient selection incentives.

The rest of the paper is organized as follows: Section 2.2 provides an overview of HVBP; Section 2.3 describes the data; Sections 2.4 and 2.5 presents the empirical strategies and results, respectively; Section 2.6 concludes.

2.2 Institutional background

In order to improve the quality of care and contain cost growth, some private insurers employ the pay-for-performance (P4P)[7] schemes adjust payments to health care providers based on measured quality. A typical P4P program rewards health care providers if they meet certain quality standard or score highly on agreed-upon performance measures. The underlying rationale is that if providers' behavior can be influenced by how they are reimbursed, then tying financial incentives to the delivery of medical care will lead providers to behave in a way that results in improved service delivery. The Affordable Care Act (ACA) mandates that CMS adopt a pay-for-performance program for hospitals nationwide. On April 29, 2011, CMS issued a rule that finalized the Hospital Value-based Purchasing Program (HVBP) under the Inpatient Prospective Payment System (IPPS) in federal fiscal year (FY) 2013, the first year of HVBP implementation. According to CMS, HVBP is aimed at incorporating P4P into the current Medicare payment system with the intent of promoting better clinical outcomes for hospital patients as well as improving their experience during hospital stays.

The HVBP program became effective for all Medicare inpatient discharges nationwide (except for Maryland) occurring after October 1, 2012 (FY 2013) in all eligible acute care hospitals (CMS, 2012).[8] The HVBP is budget-neutral by design: it is funded by first reducing

[7]Pay-for-performance (P4P) is often used interchangeably with value-based purchasing (VBP).

[8]Hospitals are not eligible for HVBP if (1) they are excluded from the Inpatient Prospective Payment

the base operating diagnosis-related group (DRG) payment for all inpatient discharges and then redistributing these funds among the participating hospitals based on their performance scores. Under the program, each eligible hospital's performance is evaluated both in a *performance period* and in a *baseline period*. CMS collects data on an established set of quality measures in both periods. Hospitals are then scored on each measure for "achievement points" by comparing their performance to *all* participating hospitals nationwide during the performance period. They are also scored for "improvement points" by comparing the data collected in the performance period to those collected (on themselves) in the baseline period. For each measure, the final score is the greater of the achievement points or the improvement points. The combined scores across all measures (total performance score, or TPS) is then translated to a payment adjustment factor (also called as incentive payment percentage) that applies to all inpatient discharges during a fiscal year.

One notable feature of the HVBP is that payment adjustment factor in future time periods will be determined by performance measures in the current period. For example, for the FY 2013 program, the performance period is July 1, 2011 to March 31, 2012, and the baseline period is July 1, 2009 to March 31, 2009. For each hospital, performance measures based on data collected during these two periods are used to determine the payment adjuster for Medicare inpatient discharges between Oct 1, 2012 and September 30, 2013. I discuss this feature and its implication in greater detail in the data section.

Beginning Oct 1, 2012, all Medicare inpatient discharges at hospitals that participate in the HVBP are paid with a certain percentage withholding of the base DGR payment plus the incentive payment percentage.[9] For FY 2013 HVBP, hospitals are evaluated along 12 process measures (70% of TPS) and 8 experience measures (30% of TPS). In the finalized

System (IPPS) (for example, psychiatric, rehabilitation, long-term care, children's, and cancer hospitals) ; (2) they did not participate in Hospital Inpatient Quality Reporting (IQR) during the HVBP performance period; (3) they are cited by the Secretary of Health and Human Services for deficiencies during the performance period that pose an immediate jeopardy to patients health or safety; (4) they do not meet the minimum number of cases, measures, or surveys required by HVBP (CMS, 2012).

[9]For example, the amount of withholding is 1 percent of base DRG payments for FY 2013, so if a hospital's payment adjustment factor is 1.2 percent, then it is effectively receiving a 0.2 percent bonus for each inpatient discharge.

provisions for HVBP in FY 2014 (beginning October 1, 2013) and FY 2015 (beginning October 1, 2014), CMS sequentially added outcome measures and efficiency measures, and also changed the relative weights of each domain in the total performance score.[10] In addition to the increase in the number of measures, the amount of withholding was also increased to 1.25% of the base DRG payment for FY 2014 and 1.5% of the base payment for FY 2015. According to the CMS plan, more measures will be added to the program and the amount of withholding will be gradually increased to 2 percent of the base DRG payment in FY 2017 (CMS, 2012).

In 2008, the Maryland Health Services Cost Review Commission (Maryland HSCRC) began implementing the Quality Based Reimbursement Initiative (QBR), which allocated rewards and penalties for hospitals based on their performance in process of care measures for heart attack, heart failure, pneumonia, and surgical infection prevention. In FY 2012, the Maryland HSCRC added experience measures (the same as those in HVBP) to the QBR program. Due to the consistent design between the Maryland QBR and HVBP, CMS exempted the acute care hospitals in the state of Maryland from HVBP. Maryland hospitals began receiving QBR-adjusted payments in FY 2010, and the payment rates were determined by their performance in calender year (CY) 2008 (Maryland HSCRC, 2011).[11] In this paper, the patients and hospitals in Maryland constitute the comparison group.

2.3 Data

The empirical analysis in this paper relies on data from several sources. First, in order to evaluate how the HVBP program affects hospital measured quality, I collected hospital performance data between 2009Q1 and 2013Q2 from *Hospital Compare*, a website administered by CMS that publishes information about the quality of care at hospitals across the United States. As part of the Hospital Quality Initiative announced by the Department of Health and Human Services (HHS) in 2001, Hospital Compare began to collect and publish

[10] See Appendix Table 2.7 for relative domain weights for the first three years of the HVBP.
[11] The baseline period for the first year of Maryland QBR is CY 2007.

information on quality of care, patient satisfaction, and pricing and cost of medical services at hospitals nationwide. Under the HVBP, hospitals are rewarded (or penalized) for their relative standings in process measures, experience measures, outcome measures (added for FY 2014), and efficiency measures (added for FY 2015).

In this paper I focus on process measures and experience measures. The main reason for excluding outcome measures and efficiency measures is that the data collection periods are inconsistent between public reporting (the data available on Hospital Compare) and the actual data CMS uses to calculate outcome and efficiency measures).[12] Also, because of the way Hospital Compare archives historical data, only annualized performance data can be retrieved before 2013, although CMS calculates hospitals performance scores using quarterly data. The most recently available data are through June 30.

For the FY 2013 HVBP, CMS used the 12 process measures hospitals reported to Medicare via the Hospital Inpatient Quality Reporting (IQR) program (see Appendix Table 2.9 for these measures).[13] All 12 measures are the proportion of patients that received appropriate care. For example, HF-1 is the proportion of heart failure patients that were given discharge information. I exclude four measures (AMI-7a, AMI-8a, SCIP-VTE-2, and SCIP-Inf-4) from the analysis because more than half of the hospitals in both Maryland and Pennsylvania had missing data on these measures. In addition, information were lacking for all but two Maryland hospitals in 2011 on 5 measures (SCIP-Card-2, SCIP-VTE-2, SCIP-Inf-1, SCIP-Inf-2, and SCIP-Inf-2). Therefore, I exclude 2011 for the analysis on these measures. For each process measure, I only include in the estimation sample hospitals with complete data for all years. As a result, the final estimation sample contains 135 Pennsylvania hospitals and 41 Maryland hospitals.[14]

[12]For example, the performance period for the 30-day mortality rate for AMI patients in FY 2014 HVBP is from July 1, 2011 to June 30, 2012. However, the same measure reported in the publicly available 2012 Hospital Compare archived file is based on data collected between July 1, 2009 and June 30, 2012.

[13]CMS added one process measure to the FY 2014 program. and excluded one measure for FY 2015 program. See Appendix Table 2.9 for details.

[14]Hospitals can have missing data for a measure when 1) they do not report that measure to CMS, or 2) there are two few cases so that they become ineligible for that measure.

The patient experience measures are derived from the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS).[15] The HCAHPS was developed jointly by the Agency for Healthcare Research and Quality (AHRQ) and CMS. It was first implemented in October 2006 and survey results became publicly available in March 2008. HCAHPS was the first standardized survey allowing comparisons of patient perceptions for inpatient care across hospitals nationally (Issac et al., 2010). Eligible adult patients with various medical conditions are randomly selected by HCAHPS between 48 hours and 6 weeks after discharge.[16] The survey includes 33 questions which ask about patient experiences in the hospital and collect other health and demographic information such as self-assessed health status, race and ethnicity, and level of education. Based on the responses to a selected set of questions, CMS calculates and publishes 8 experience of care measures after adjustment for survey mode,[17] non-response, and patient case mix.[18] Each measure is the proportion of patients who gave the "top-box" responses (most positive) on HCAHPS.[19] For example, the measure "pain management" is the proportion of patients who answered that "pain was always well controlled during the hospital stay".

Instead of using the total performance scores for clinical process and patient experience calculated by CMS,[20] I use as outcome variables the underlying proportions (rates) used to derive these scores. The reason for doing this is twofold. First, the CMS composite scores reflect a hospital's relative standing among all hospitals nationwide as well as its improvement over time. For example, a poor performer can earn high composite scores by substantially improving their ratings over time. Therefore, looking at the composite scores obscures comparisons across hospitals. Second, using the underlying rates for each individual

[15]See Appendix 2.7.3 for the survey instrument.

[16]To be eligible for HCAHPS, patients must be at least 18 years old at admission; have non-psychiatric principal diagnosis; stay at least one night in the hospital; and be discharged alive.

[17]The HCAHPS can be administered via mail, telephone, or a combination of the two.

[18]See Appendix Table 2.10 for a complete list of experience measures.

[19]To receive an experience of care domain score under HVBP, hospitals must have at least 100 completed HCAHPS surveys.

[20]In addition to total performance scores, CMS also publishes separate domain scores for process measures and experience measures.

measures allows us to examine the effect of the HVBP on specific and meaningful aspects of hospital care rather than a simple overall rating.

In addition to collecting aggregate HCAHPS scores for all hospitals in Pennsylvania, I obtained the HCAHPS survey responses from a large hospital network in Pennsylvania for 2012 and 2013. I then linked the survey results to the inpatient database obtained from the same hospital network. The combined dataset contains patients' responses on HCAHPS as well as their demographic and diagnostic information. I then subset to eligible patients with complete surveys, thereby excluding 366 patients from the full sample of 3,729 patients.[21] Since the experience of care measures are computed using the top-box responses for a subset of questions in the survey, I constructed a set of dichotomous indicators for whether the patient answered the top-box for the questions related to a particular measure and use these as dependent variables. Following the CMS method for patient case mix adjustment (Elliott et al., 2009; O'Malley et al., 2005), I assigned patients to three different service lines (obstetric, surgical, and medical) based on the Diagnostic Related Group (DRG). Self-assessed health status is a categorical variable that ranges from 1 to 5 (in decreasing order: excellent, very good, good, fair, poor). Similarly, educational attainment is an ordinal variable ranging from 1 to 6 in increasing order (8th grade or less, some high school but did not graduate, high school graduate, some college or 2-year degree, 4-year college graduate, more than 4-year college degree). The final estimation sample is comprised of 3,363 patients, among which 6.3 percent are obstetric patients, 33.2 are surgical patients, and 58.5 percent are medical patients.[22]

My third data source is the 2009-2013 inpatient discharge data for the states of Pennsylvania and Maryland. The Pennsylvania inpatient discharge data were obtained from the Pennsylvania Healthcare Cost Containment Council (PHC4), while the Maryland inpatient

[21] A survey is deemed complete if the patient has valid responses more than half of the questions. In addition, patients who do not respond to the question that asks about the overall health status are excluded.

[22] Around 2 percent of the patients cannot be assigned to any service line mainly because they do not have a valid DRG on the record. However, I do not exclude them from the estimation sample because they can contribute to the identification of the correlation between self-assessed health status and top-box responses.

discharge files came from two different sources: the 2009-2012 discharge data were obtained from the State Inpatient Databases of the Healthcare Cost and Utilization Project at AHRQ (HCUP SID, 2014);[23] and the 2013 data were obtained directly from Maryland HSCRC.

Collectively, these three datasets constitute the universe of inpatient discharges in the two states during the sample period. They contain information on patient demographics such as age, gender, race and ethnicity, as well as principle and secondary diagnoses and procedures performed during the inpatient admission. After excluding patients that are discharged from ineligible hospitals in Pennsylvania, and those below 18 years of age because they are not eligible for HCAHPS, the final estimation sample contains 7,050,841 discharges in 137 Pennsylvania hospitals and 3,030,898 discharges in 47 Maryland hospitals.

My main measure of patient illness severity is a "risk score" computed from the principle and secondary ICD-9 codes[24] reported on the inpatient discharge record. I use the hierarchical condition categories (HCC) risk adjustment model developed by CMS, which was first created in 2004 to adjust capitation payment to private health insurers for beneficiaries enrolled in Medicare Advantage (MA) plans (Pope et al., 2004). One concern is whether the HCC risk adjustment method is a good measure of patient illness severity in itself. Despite its relatively low explanatory power and poor performance in predicting expenditures for high-cost patients (Schone and Brown, 2013),[25] it is still one of the state-of-the-art risk adjustment methods that have been developed so far. Compared to other measures for patient illness severity commonly used in the literature such as the Charlson comorbidity index (Charlson et al., 1987) and Elixhauser index (Elixhauser et al., 1998), the HCC risk adjustment model is considerably more comprehensive and includes more conditions and complications that

[23]For an overview of HCUP SID, see `http://www.hcup-us.ahrq.gov/sidoverview.jsp`.

[24]The Maryland inpatient data report up to 29 secondary ICD-9 codes for each discharge. However, the Pennsylvania inpatient data only report 8 secondary ICD-9 codes before 2011 (The number was increased to 17 after 2011). As a result, I only use the first 7 secondary ICD-9 codes across all years for both states when computing the risk scores to make sure that the changes in the risk scores are not a artifact of how hospitals document and report patient diagnoses.

[25]Using Medicare claims and enrollment data, Pope et al. (2011) report that the R^2 (the proportion of variance in Medicare expenditures that is explained by the variables in the model) of CMS HCC V12 model ranges between 8% and 11%.

can better capture patient risk.[26] Using a sample of Medicare beneficiaries, Li, Kim and Doshi (2010) find that the HCC model outperforms the Charlson and Elixhauser indexes in predicting in-hospital and 6-month mortality rates among AMI, congestive heart failure, diabetes mellitus, and stroke patients. This provides evidence that the HCC model can be used as a patient risk adjuster (therefore a measure of illness severity) although it was originally developed to predict cost for Medicare patients.

I use the CMS-HCC Version 12 model to compute the risk scores. Each patient's diagnoses are first assigned to condition categories.[27] Hierarchies are then imposed so that a patient is coded for only the most severe manifestation among related diseases. A set of coefficients (weights) are then applied to each HCC and the interactions among them. In addition to the diagnostic information, the HCC risk scores also account for patients' gender, age, Medicaid eligibility, and disability status.[28] Finally, for each patient, the HCC risk score equals the sum of coefficients across all HCCs (Pope et al., 2011).[29] As an alternative measure of patient illness severity, I use in-hospital mortality. Figure 2.1 contains the trends in in-hospital mortality and risk scores over time by state. While there is no noticeable differences in both outcomes between the two states before the HVBP was announced, risk scores in Pennsylvania hospitals increased at a slightly faster rate than for Maryland hospitals after the announcement of the HVBP. Table 2.1 contains descriptive statistics of patient demographics and risk scores by state. In general, patients in Pennsylvania are older and have higher risk scores.

[26]For example, the Charlson index only includes 17 conditions (comorbidities) and does not explicitly account for patient demographics such as age and gender.

[27]The Maryland inpatient discharge data do not contain an unique patient identifier. As a result, it is impossible to identify a patient over time. Therefore, a "patient" is technically an inpatient discharge record.

[28]The inpatient discharge data do not contain information on patients' Medicaid eligibility or disability status. Therefore, the HCC risk scores used in this paper do not account for the risk associated with these two factors.

[29]See Appendix 2.7.1 for a detailed description of the computation of HCC risk scores.

2.4 Empirical strategy

In this section I describe the empirical strategies used to test whether the HVBP induces hospitals to engage in patient risk selection.

2.4.1 Patient illness severity and survey responses

First, I consider the association between patient illness severity and HCAHPS responses. For each of the experience measures, I estimate the following linear probability model using exactly the same variables CMS uses in its patient case mix adjustment:

$$y_{ij} = X_i'\beta + \delta\, health_i + \varepsilon_{ij}, \tag{2.1}$$

where y_{ij} is a binary indicator set to 1 if patient i answers the "top-box" response on questions related to measure j; $health_i$ is patient i's self-assessed overall health status; X_i includes controls for patient age (10-year age categories), gender, survey language (a binary indicator for English), level of education (8th grade or less, some high school but did not graduate, high school graduate, some college or 2-year degree, 4-year college graduate, more than 4-year college degree), the type of services received at the hospital (obstetric, surgical, and medical), and the interactions between service line and a categorical variable for patient age.[30] The parameter of interest is δ, which is the correlation between patient self-assessed health and "top-box" responses. If $\delta < 0$, then "sicker" patients are less likely to give top-box responses on HCAHPS.[31] If the HCAHPS scores are *not* risk-adjusted, this negative correlation will give hospitals an incentive to risk select healthier patients. In theory, perfect case mix adjustment will eliminate the scope for any patient selection (either negative or

[30]Following the CMS case mix adjustment formula, I assign patients into 8 age categories: 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, and above 85. The service line by age interaction variables are created as the product of a binary indicator for service line and a *single* categorical variable ranging from 1 to 8, corresponding to the each of 8 age categories.

[31]We use patients' self-assessed (subjective) health status as a global measure of health given its consistent and strong association with actual (objective) health status (such as prevalence of diseases and mortality risk) reported in the epidemiological literature (for example, see Wu et al., 2013; Miilunpalo et al., 1997; Mossey and Shapiro, 1982).

positive). However, if conditioning on the variables in equation (2.1) is not sufficient, then hospital will still have an incentive to identify patients who give more (or less) favorable responses *conditional on* variables used in the CMS risk adjustment formula.

To explore this possibility, I further estimate

$$y_{ij} = X_i'\beta + \delta\, health_i + \alpha_1\, obstetric_i \times health_i + \alpha_2\, surgical_i \times health_i + \varepsilon_{ij}, \tag{2.2}$$

where $obstetric_i$ and $surgical_i$ are binary indicators for whether the patient received obstetric or surgical care at the hospital (the reference group is medical patients).[32] The motivation for doing this is that patients who receive different types of services may have totally different experience with the hospital. In the methodology study for the current CMS case mix adjustment model, O'Malley et al. (2005) report a strong relationship between service line and patients' rating of the hospital services, and suggest that the case mix adjustment model should include interactions of each included variable with service if a single model is to be used for patients across all service lines. If α_1 and α_2 are statistically different from 0, then hospitals will have an incentive to risk select patients along a particular service line. For example, if $\alpha_1 < 0$ ($\alpha_1 > 0$), then conditional on self-assessed health status, sicker obstetric patients are less (more) likely to give the most positive survey responses, which raise hospitals' scores under the HVBP. In other words, the CMS case mix adjustment formula under-corrects (over-corrects) for the effect of patient severity on top-box responses among obstetric patients. Hospitals would have an incentive to risk select obstetric patients (the same argument applies to surgical patients). It should be noted that the actual power of the selection incentive will depend on the extent to which case mix adjustment changes their experience scores.

[32]Patients are assigned to one of these three mutually exclusive service lines (maternity care, medical, and surgical) based on MS-DRG codes. A detailed list of codes for each service line can be found at `http://www.hcahpsonline.org/Files/MS_DRG_V32_Table.pdf`.

2.4.2 Patient-level analysis

Next, I test the effect of the HVBP on patient illness severity using a difference-in-differences (DID) strategy. The baseline econometric model is

$$y_{ijt} = \alpha_0 \, PA_{ijt} + \sum_{k=1}^{3} \alpha_k \, post_{ijt}^{k} + \sum_{k=1}^{3} \delta_k (PA_{ijt} \cdot post_{ijt}^{k}) + X_{ijt}' \beta + \phi_j + \xi_t + \varepsilon_{ijt}, \tag{2.3}$$

where y_{ijt} is a measure of illness severity for patient i discharged from hospital j in year t, X_{ijt} is a vector of patient characteristics such as age (5-year categories), gender, race and ethnicity (white, black, Asian, Hispanic, and other race), type of admission (emergent, urgent, elective, and other type of admission), and type of insurance (private insurance, Medicare, Medicaid, self-pay, other insurance); and ϕ_j and ξ_t are hospital fixed-effects and time effects, respectively. I use the binary indicator PA_{ijt} to denote patients discharged from hospitals in Pennsylvania. The performance period for FY 2013 program is July 1, 2011 to March 31, 2012. For FY 2014 program, the performance period is designated as April 1, 2012 to December 12, 2012. For FY 2015 program, the performance period is January 1, 2013 to December 31, 2013.[33] Hospitals began to receive adjusted payments after October 1, 2012, when the FY 2013 HVBP went into effect. Since hospitals' payment adjustment factors are determined by the data collected before the program became effective, they might respond to the HVBP strategically even before they feel real financial impact. To capture this feature of program implementation, I create 3 *mutually exclusive* binary indicators that correspond to the HVBP regime: $post_{ijt}^{1}$ is set to 1 for patients that were discharged between 2011Q3 and 2012Q1 (the performance period for FY 2013 program); $post_{ijt}^{2}$ is set to 1 for inpatient discharges that occurred between 2012Q2 and 2012Q3 (the performance period for FY 2014 program);[34] $post_{ijt}^{3}$ equals 1 for inpatient discharges occurred after 2012Q4 (the

[33] See Appendix Table 2.8 for a complete description of baseline and performance periods for HVBP in FY 2013-2015.

[34] Technically speaking, the performance period for FY 2014 is 2012Q2 to 2012Q4. In order to capture the effect of actual payment adjustment on hospital behaviors, I assign 2012Q4 to the third phase of HVBP.

performance period for FY 2015 program/initial post payment period). The coefficients of interest are δ_k ($k = 1, 2, 3$), which are the DID estimates of the effect of the HVBP program on patient illness severity over time. The rejection of null hypothesis $\delta_k = 0$ is consistent with the hypothesis that the implementation of the HVBP caused patient selection.

The validity of the DID strategy hinges on the assumption that the two states would have had a common trend in patient illness severity in the absence of the HVBP. However, it is possible that patients in Maryland might not be a perfect comparison group because Maryland implemented similar quality initiatives earlier in 2008. To address this concern, I estimate a model of the following specification:

$$y_{ijt} = \alpha_0 \, PA_{ijt} + \sum_{k=1}^{3} \alpha_k \, post^k_{ijt} + \sum_{k=1}^{3} \delta_k (PA_{ijt} \cdot post^k_{ijt}) + X'_{ijt}\beta + \lambda \, D_t \cdot PA_{ijt} + \phi_j + \xi_t + \varepsilon_{ijt}, \quad (2.4)$$

which allows the two states to have different pre-HVBP trends. In this specification, D_t is a single linear trend over the sample period. The estimates of δ_k are robust to pre-existing differential trends in outcome y_{ijt}.

I estimate both equations (2.3) and (2.4) on HCC risk scores and a binary indicator of in-hospital mortality using all patients aged 18 or older at the time of admission. To see whether selection occurs in different service lines, I re-run the models on the samples of obstetric patients, surgical patients, and medical patients. If the estimates of δ_k are statistically significant in any of the subsamples, then it suggests that patient selection has occurred in that particular patient group (service line). Despite the right-skewness of the HCC risk scores,[35] I report the estimates from models that use raw scale as the dependent variable due to the fact that transforming the dependent variable in the DID framework changes the identifying assumption. Specifically, when using logged risk scores as the dependent variable, identification requires that the trend of risk scores on the *log* scale is the same in the absence of the HVBP. Since the risk scores are computed as a weighted

[35]See Appendix Figure 2.4 for a histogram of these risk scores

sum across a set of conditions, using the raw scale is more appropriate.[36]

2.4.3 Hospital-level analysis

In the final component of my empirical analysis, I investigate the effect of the HVBP on process and experience measures. Since all measures are proportions, they are bounded between 0 and 1. One can estimate the following transformed model when the dependent variable is *strictly* bounded in the unit interval (Papke and Wooldridge, 1996):

$$\Lambda^{-1}(y_{jt}) = \alpha_1 post_{jt} + \alpha_2 PA_{jt} + \delta(PA_{jt} \cdot post_{jt}) + \phi_j + \xi_t + \varepsilon_{jt}, \tag{2.5}$$

where $\Lambda^{-1}(z) = \ln \frac{z}{1-z}$; y_{jt} is hospital j's performance at time t; ϕ_j and ξ_t are hospital fixed-effects and time effects, respectively. Since only annualized hospital performance data are available, $post_{jt}$ is set to 1 for 2013 (post payment).

When y can take on values of 0 and 1 with positive probability, the above transformation fails. But, if the sample size from which the proportion is derived is known, we can modify (2.5) to account for the mass points at 0 and 1 (Maddala, 1983):

$$\tilde{y}_{jt} = \alpha_1 post_{jt} + \alpha_2 PA_{jt} + \delta(PA_{jt} \cdot post_{jt}) + \phi_j + \xi_t + \varepsilon_{jt}, \tag{2.6}$$

where

$$\tilde{y}_{jt} = \ln \frac{y_{jt} + (2n_{jt})^{-1}}{1 - y_{jt} + (2n_{jt})^{-1}}. \tag{2.7}$$

In (2.7) n_{jt} is the sample from which proportion y_{jt} is derived. The parameter of interest δ is the DID estimate of the effect of the HVBP. In fact, one can just estimate the DID model on the raw data without transforming the dependent variable. However, ignoring the bounded nature of proportional dependent variable is problematic. Generally speaking, the validity of the DID approach hinges on the assumption that the trend in the dependent variable is the

[36] When I estimate models using logged risk scores as the dependent variable, the results are both qualitatively and quantitatively similar.

same across the treatment and comparison group, regardless of the level of the dependent variable in the pre-period. This point becomes more subtle when we use proportions as dependent variables. If initially a proportion differs in level in the treatment and control group, then an identical change in the underlying variable from which the proportion is derived can have drastically different effects on the proportion (Mullen, Frank and Rosenthal, 2010). Estimating the effect of the HVBP on the *log odds-ratio* ($\Lambda^{-1}(\cdot)$) can account for this non-linearity. Since the experience measures are strictly bounded between 0 and 1, I estimate them on the untransformed model (2.5). In contrast, almost all of the process measures have a mass point at 1. Therefore, I estimate them on the transformed model (2.6).

2.5 Results

In this section I present the main results from the empirical analysis.

2.5.1 Patient illness severity and HCAHPS survey responses

I report the estimates from equations (2.1) and (2.2) in Appendix Table 2.11. In general, the estimated coefficient on self-assessed health status is negative and statistically significant across all experience measures with the exception of discharge information. This is consistent with the findings from the methodological study of the HCAHPS case mix adjustment: Sicker patients are less likely to give positive evaluation of health care (O'Malley et al., 2005). After adding the interaction terms between service line and self-assessed health status, two interesting findings arise. First, the coefficient on the interaction term *health* $\times$ *obstetric* is negative and statistically significant in the regressions on "cleanliness and quietness of hospital environment" and "overall rating of hospital" (columns 12 and 16 of Table 2.11). This suggests that relative to medical patients, sicker obstetric patients are less likely to give top-box responses even after conditioning on self-assessed health status. In other words, the CMS case mix adjustment formula under-corrects for the tendency of sicker obstetric patients to give less favorable responses on HCAHPS. Second, for surgical patients, although none of coefficient estimates of the interaction terms *health* $\times$ *surgical* are statistically significant at 10 percent level, most of them (5 out of 8) are positive. This indicates that the CMS case mix adjuster may over-corrects for the tendency of sicker surgical patients to give less favorable responses. Taken together, these results suggest that hospitals in Pennsylvania have an incentive to avoid sicker obstetric patients and attract sicker surgical patients under the HVBP.

2.5.2 Patient-level analysis

I estimate equations (2.3) and (2.4) on in-hospital mortality and risk scores using the full sample of inpatient discharges from Pennsylvania and Maryland as well as three different subsamples: obstetric patients, surgical patients, and medical patients. Due to the fact that Maryland implemented similar quality initiatives in 2008, patients discharged from hospitals in Maryland might not constitute a perfect comparison group. For that reason, my preferred specification is the model that controls for a linear state-specific trend. For in-hospital mortality, the estimated coefficients on the interaction terms *PA* $\times$ *performance period* 1, *PA* $\times$ *performance period* 2, and *PA* $\times$ *post payment* are all negative and statistically significant in the regression that controls for a single linear trend and its squared (column 1 of Table 2.3) in the full sample. This suggests that after the implementation of the HVBP, patients in Pennsylvania were less likely to die in the hospital relative to patients in Maryland (a 15 percent decrease relative to the state average in-hospital mortality rate). Under the same specification, I find a decrease of similar magnitude in the mortality rate among medical patients (column 7 of Table 2.3). However, no such decrease is found in obstetric patients and surgical patients. After adding a state-specific linear trend in the model (columns 2, 4, 6 and 8 of Table 2.3), the estimated coefficients on the interaction terms remain significant in the full sample and the subsample of medical patients (the magnitude of the decreases remains the same). For obstetric and surgical patients, none of the coefficients on the interaction terms are statistically significant. Taken together, these results imply that the decrease in mortality is mostly driven by medical patients. However, we cannot conclude that patient illness severity decreased as a result of the HVBP based on the changes in mortality because an increase in hospitals' clinical quality can also account for that.

When patient risk scores are the outcome, I do not find any statistically significant effects in the full sample and subsample of medical patients. However, for obstetric patients, the estimated coefficient on $PA \times$ *post payment* is negative and statistically significant (columns 3 of Table 2.4). This indicates that relative to patients in Maryland, the average illness severity decreased for obstetric patients in Pennsylvania after the HVBP went into effect (a 53 percent decrease relative to the state mean). In contrast, the estimated coefficient on $PA \times$ *post payment* is positive and statistically significant for surgical patients, indicating an increase in average patient illness severity in this case (a 13 percent increase relative to the state mean). The estimated effects on risk scores are robust to adding a state-specific linear trend (columns 4 and 6 of Table 2.4). Interestingly, in my preferred specification, the estimated effects of the HVBP are both *positive* in the first two performance periods for obstetric patients while the two effects are both *negative* for surgical patients. One plausible explanation is that hospitals in Pennsylvania were first selecting patients based on their prior beliefs. After they gradually learned how the incompleteness of the CMS case mix adjustment formula could affect patient experience ratings, they began to systematically select healthier obstetric patients and sicker surgical patients.

2.5.3 Hospital-level analysis

Tables 2.5 and 2.6 report estimates from equations (2.5) and (2.6). For each of the measures, I estimate models with four different specifications for the time effects. The first two columns of Tables 2.5 and 2.6 present estimates from regressions that include a single linear trend and its squared, and the next two columns contain estimates from models that additionally control for a state-specific linear trend and its squared. For experience measures, all of the coefficient estimates are positive. Specifically, I find statistically significant increase for "communication with doctor", "pain management", "cleanliness and quietness of hospital environment", and "overall rating" in some of the specifications. In contrast, no effect is found for any of the process measures. Collectively, these estimates provide evidence that the HVBP led to an increase in Pennsylvania hospitals' measured performance on the HCAHPS, but has no detectable effect on the process of care measures.

2.6 Discussion and conclusion

In this paper, I investigate health care providers' selection behavior under financial incentives designed to improve the quality of care. Specifically, I examine how the Hospital Value-Based Purchasing Program introduced by CMS in mid-2011 affects the mix of patient treated at participating hospitals using a difference-in-differences strategy. Two findings arise from the empirical analysis: (1) The patient case mix adjustment formula used by CMS does not account for interactive effects of patient health status and service line (obstetric, surgical, and medical) on HCAHPS responses. Relative to medical patients, sicker obstetric (surgical) patients are less (more) likely to give positive evaluation of their hospital experiences conditional on their self-assessed health status. By implication, CMS over-adjusts (under-adjusts) the effect of patient health status on the probability of giving the most positive responses for surgical (obstetric) patients. (2) I also find that hospitals exploited the opportunity to select patient created by incomplete case mix adjustment, and that the average illness severity decreased by about 53 percent for obstetric patients and increased by about 13 percent for surgical patients after the HVBP went into effect. This indicates that hospitals engaged in patient selection in a way that boosts their performance scores under the HVBP.

Furthermore, I do not find any evidence of improvement in process measures after the HVBP became effective, but I do find weak evidence that the program was associated with increases in experience measures. Based on my models of patient illness severity, it is likely that at least part of this increase can be attributed to patient selection.

This study is subject to several limitations. First, the evolving nature of the HVBP poses significant difficulty in studying its effect. Overtime CMS has added several new measures to the HVBP to capture more aspects of hospital care. This means hospital responses to the HVBP will also evolve over time. For example, after adding outcome-based measures such as 30-day mortality rate for AMI, heart failure, and pneumonia patients, hospitals face a new selection incentive, depending on how well risk adjustment works for these measures. Therefore, it is not possible to fully isolate and quantify the extent of patient selection due to patient experience measures. If selection incentives for outcome measures work in the same direction as those for patient experience measures, then the estimated effect on changes in patient illness severity cannot be entirely attributed to the experience measures. Nonetheless, over the time period I examine, I find no effect of the HVBP on process measures.

A second limitation is that the analysis on the inpatient discharge data is at the "visit-level" instead of the patient-level. Ideally, the risk of a patient (illness severity) should be measured using information before the actual hospital admission (index admission) occurs.[37] The main reason for doing this is that most of the time the actual "patient selection" process effectively occurs before patients are admitted. Unfortunately, the Maryland inpatient discharge data do not contain a unique patient identifier, which makes it impossible to identify individual patients over time. As a result, the HCC risk scores are computed using contemporaneous diagnoses in both states. The main concern with this approach is the HCC risk scores may not fully capture patient illness severity if discharge records do not contain ICD-9 codes of all pre-existing conditions. However, under the current Medicare DRG system, hospitals have an incentive to code all pre-existing conditions because this can potentially lead to higher reimbursement. For that reason, using contemporaneous diagnosis codes to determine patient illness severity should be less of a concern.

Another limitation of this paper is that there are a number of other CMS initiatives that are concurrent with the HVBP, which could also result in patient selection. Among these new efforts, the most notable two programs are the the Hospital Readmissions Reduction

[37] A common measure of patient illness severity used in the literature is medical expenditure (based on payments) in the year prior to the index admission. For example, see Dranove et al. (2003).

Program (HRRP) and the Hospital-acquired Condition Reduction Program (HAC). The HRRP became effective on October 1, 2012, the same day as the HVBP. Under the program, hospitals with excess 30-day readmission rates for AMI, heart failure, and pneumonia patients are penalized by up to 1 percent of the base DRG payment.[38] This raises the concern that the main results in the paper could be mainly driven by the HRRP instead of the HVBP. In order to mitigate this concern, I estimated visit-level models using the subsamples of AMI, heart failure, and pneumonia patients. I do not find changes in mortality or HCC risk scores among these patients.[39] The HAC, on the other hand, will not become effective until FY 2015 (beginning October 1, 2014).[40] Since the data collection period for FY 2015 HAC ends on December 31, 2013, it is possible that the HAC affected hospitals' selection behavior prior to the time period corresponding to payment determination. However, the HAC payment reduction only affects hospitals that rank among the lowest-performing quartile in HAC measures. Therefore, the effect of the HAC on the results in this paper is limited.

Despite these limitations, the findings in this paper have important policy implications. Accurate case mix adjustment methods are critical when implementing large-scale public P4P initiatives like the HVBP, HRRP, and HAC to avoid unintended consequences like patient risk selection. Since patient mix varies widely among different types of hospitals in the U.S, hospitals treating certain types of patient populations will be unfairly penalized if their performance scores are not properly adjusted for the differences in patient characteristics. The results in this paper provide evidence that hospitals do strategically respond to the selection incentives created by insufficient risk adjustment. If safety net hospitals that operate in low-income areas consistently under-perform mainly due to the character-

[38]For FY 2015 HRRP, CMS added 30-day readmission rate for (1) patients admitted for an acute exacerbation of chronic obstructive pulmonary disease (COPD); and (2) patients admitted for elective total hip arthroplasty (THA) and total knee arthroplasty (TKA). Also, the amount of penalty can be as much as 3 percent of the base DRG payment in FY 2015.

[39]These results are not shown, but are available from the author upon request.

[40]The HAC uses AHRQ Patient Safety Indicators (PSI) and Centers for Disease Control and Prevention National Healthcare Safety Network (CDC NHSN) measures. The performance period for AHRQ PSIs is July 1, 2011 to June 30, 2013, and the performance period for CDC NHSN measures is January 1, 2012 to December 31, 2013.

istics of the patient they treat, they may avoid certain types of patients that are likely to lower their performance scores, which could in turn widen the disparities in access to care for vulnerable populations. Thus, without evidence showing the improvement of the quality of care, P4P programs like the HVBP might end up welfare-reducing from a social standpoint. To minimize the risk of unintended consequences in public P4P programs, policy makers should consider carefully when selecting new measures in the HVBP and develop more comprehensive risk adjustment methods for existing measures.

Tables

Table 2.1: Descriptive statistics: inpatient discharge data

	PA (N=7,050,841)		MD (N=3,030,898)	
Variables	Mean	Std. Dev.	Mean	Std. Dev.
Patient characteristics				
Age 18-24	0.057	0.232	0.065	0.247
Age 25-29	0.053	0.223	0.064	0.244
Age 30-34	0.053	0.223	0.065	0.246
Age 35-39	0.042	0.201	0.054	0.225
Age 40-44	0.045	0.208	0.057	0.231
Age 45-49	0.058	0.233	0.071	0.256
Age 50-54	0.072	0.259	0.082	0.274
Age 55-59	0.078	0.269	0.082	0.274
Age 60-64	0.083	0.275	0.082	0.274
Age 65-69	0.083	0.277	0.079	0.270
Age 70-74	0.081	0.273	0.073	0.260
Age 75-79	0.085	0.279	0.070	0.256
Age 80-84	0.091	0.287	0.069	0.253
Age 85-89	0.074	0.262	0.055	0.229
Age 90-100	0.045	0.207	0.034	0.181
Female	0.582	0.493	0.593	0.491
White	0.810	0.392	0.583	0.493
Black	0.130	0.336	0.330	0.470
Asian	0.007	0.085	0.018	0.133
Other race	0.052	0.223	0.069	0.254
Hispanic	0.025	0.157	0.033	0.178
Admission type-emergency	0.571	0.495	0.666	0.472
Admission type-urgent	0.170	0.376	0.127	0.332
Admission type-elective	0.253	0.435	0.167	0.373
Admission type-other	0.006	0.077	0.041	0.198
Insurance-private	0.309	0.462	0.319	0.466
Insurance-selfpay	0.020	0.140	0.055	0.229
Insuracne-Medicare	0.507	0.500	0.430	0.495
Insurance-Medicaid	0.151	0.358	0.167	0.373
Insurance-other	0.013	0.113	0.028	0.164
Year 2009	0.209	0.407	0.216	0.411
Year 2010	0.205	0.403	0.209	0.407
Year 2011	0.202	0.402	0.201	0.401
Year 2012	0.195	0.396	0.188	0.391
Year 2013	0.189	0.391	0.186	0.389
Performance period 1 (0/1)	0.149	0.357	0.147	0.354
Performance period 2 (0/1)	0.097	0.296	0.094	0.292

Post payment (0/1)	0.237	0.425	0.232	0.422
Patient outcomes				
HCC risk score	1.124	0.884	1.106	0.895
In-hospital mortality	0.020	0.142	0.020	0.141

Notes: *HCC risk scores are based on 3,030,611 inpatient discharges for Maryland and 7,049,364 inpatient discharges for Pennsylvania with valid and non-missing ICD-9-CM codes.

Table 2.2: Descriptive statistics: hospital performance measures

	Pennsylvania					Maryland				
Measures	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
*Experience of care** (%)										
Communication with nurses	685	60.134	5.004	46	80	205	59.495	4.284	50	70
Communication with doctor	685	76.661	3.987	58	87	205	73.683	4.743	60	82
Responsiveness of staff	685	78.549	3.711	61	93	205	77.249	3.218	69	87
Pain management	685	63.400	6.423	37	83	205	57.527	6.553	40	72
Cleanliness and quietness	685	68.772	3.840	54	82	205	66.449	3.924	56	75
Communication about medicines	685	59.730	4.230	44	78	205	57.361	4.731	46	66
Discharge information	685	83.244	3.662	66	93	205	82.039	4.175	70	91
Overall rating	685	65.340	7.366	35	94	205	63.639	7.433	42	82
*Process of care*** (%)										
HF-1	675	91.040	11.269	6	100	205	90.780	8.055	64	100
Pn-3b	675	96.745	3.714	59	100	205	94.805	4.247	78	100
Pn-6	675	94.268	4.805	73	100	205	94.215	5.754	50	100
SCIP-Card-2	516	94.771	6.445	50	100	152	94.211	5.557	73	100
SCIP-VTE-2	540	95.344	6.071	53	100	164	92.890	8.435	56	100
SCIP-Inf-1	532	97.075	5.336	31	100	164	96.506	3.597	79	100
SCIP-Inf-2	532	97.468	3.735	69	100	164	97.293	3.151	76	100
SCIP-Inf-3	532	95.333	6.076	44	100	164	94.622	5.587	61	100

Notes: *Means for experience measures are derived from hospitals with complete data for each measure for all 5 years (2009-2013).
**Means for process measures are derived from hospitals with complete data for each measure in all 4 years excluding 2011.

Table 2.5: Estimates from difference-in-differences models: HCAHPS experience of care measures

Measures	(1)	(2)	(3)	(4)
Communication with nurses	0.022	0.022	0.032	0.007
	(0.022)	(0.022)	(0.028)	(0.044)
Communication with doctor	0.036**	0.036**	0.022	0.037
	(0.018)	(0.018)	(0.025)	(0.045)
Responsiveness of staff	0.006	0.006	0.002	-0.002
	(0.018)	(0.018)	(0.026)	(0.045)
Pain management	0.044*	0.044*	0.037	0.021
	(0.024)	(0.024)	(0.031)	(0.050)
Cleanliness and quietness	0.035	0.035	0.067**	0.126***
	(0.021)	(0.021)	(0.028)	(0.042)
Communication about medicines	0.032	0.032	0.032	0.019
	(0.022)	(0.022)	(0.031)	(0.049)
Discharge information	0.020	0.020	0.036	0.064
	(0.031)	(0.031)	(0.036)	(0.062)
Overall rating	0.036	0.036	0.075**	0.050
	(0.028)	(0.028)	(0.035)	(0.051)
Linear trend	**X**	**X**	**X**	**X**
Linear trend sq		**X**	**X**	**X**
State-specific linear trend			**X**	**X**
State-specific linear trend sq				**X**
Hospital fixed-effects	**X**	**X**	**X**	**X**

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors are clustered at hospital level. Each cell represents a model.

Table 2.3: Estimates from difference-in-differences models: in-hospital mortality

	Full sample		Obstetric		Surgical		Medical	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PA × performance period 1	-0.001**	-0.002***	-0.000	-0.001**	0.000	-0.001	-0.002**	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
PA × performance period 2	-0.003***	-0.004***	-0.000	-0.001**	0.000	-0.001	-0.005***	-0.006***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)
PA × post payment	-0.003***	-0.004***	-0.001	-0.001	0.004**	0.001	-0.004***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Observations	10,081,739	10,081,739	1,008,497	1,008,497	2,451,094	2,451,094	5,445,658	5,445,658
Number of hospitals	184	184	150	150	184	184	184	184

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors are clustered at hospital level. Each column represents a model. Other control variables include 5-year age categories, gender, race and ethnicity (black, Hispanic, Asian, and other race with white as the base group), type of admission (urgent, trauma, elective, and other with emergency as the base group), insurance type (self-pay, Medicare, Medicaid, government, and other insurance with private insurance as the base group).

Table 2.4: Estimates from difference-in-differences models: HCC risk scores

	Full sample		Obstetric		Surgical		Medical	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PA × performance period 1	0.002	-0.004	-0.001	0.015***	-0.001	-0.027***	0.001	-0.007
	(0.006)	(0.006)	(0.002)	(0.004)	(0.008)	(0.008)	(0.008)	(0.008)
PA × performance period 2	-0.003	-0.011	-0.003	0.020***	-0.002	-0.038***	-0.008	-0.019
	(0.007)	(0.009)	(0.002)	(0.007)	(0.009)	(0.011)	(0.010)	(0.012)
PA × post payment	-0.008	-0.018	-0.121***	-0.090***	0.128***	0.080***	-0.009	-0.025
	(0.008)	(0.012)	(0.030)	(0.023)	(0.022)	(0.020)	(0.012)	(0.017)
Observations	10,079,975	10,079,975	1,008,487	1,008,487	2,451,087	2,451,087	5,445,624	5,445,624
Number of hospitals	184	184	150	150	184	184	184	184

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors are clustered at hospital level. Each column represents a model. Other control variables include 5-year age categories, gender, race and ethnicity (black, Hispanic, Asian, and other race with white as the base group), type of admission (urgent, trauma, elective, and other with emergency as the base group), insurance type (self-pay, Medicare, Medicaid, government, and other insurance with private insurance as the base group).

Table 2.6: Estimates from difference-in-differences models: process of care measures

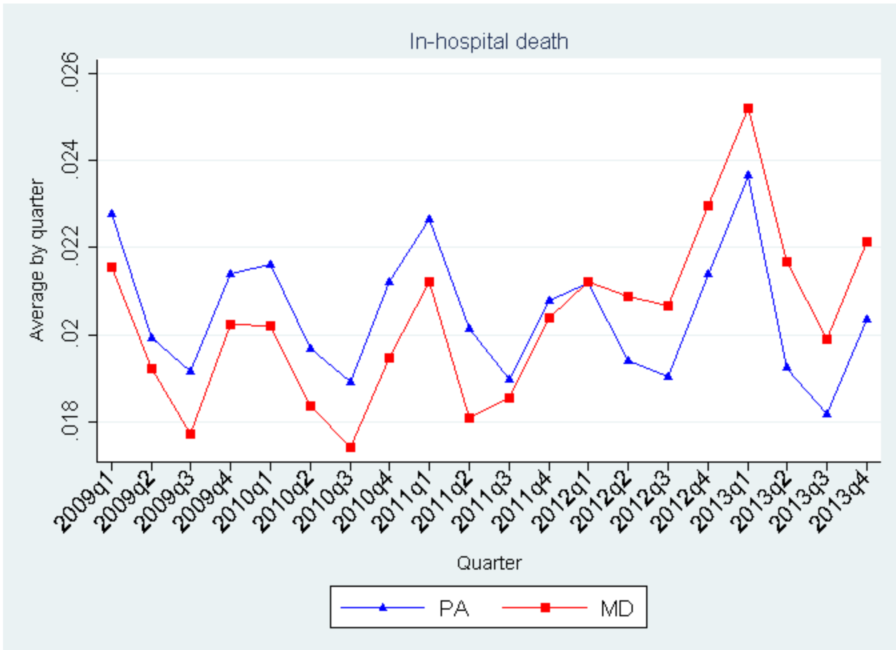
Measures	(1)	(2)	(3)	(4)
HF-1	0.051	0.051	-0.106	-0.045
	(0.209)	(0.209)	(0.238)	(0.296)
Pn-3b	-0.110	-0.110	0.147	0.547
	(0.181)	(0.181)	(0.245)	(0.333)
Pn-6	-0.189	-0.189	-0.001	0.229
	(0.127)	(0.127)	(0.169)	(0.313)
SCIP-Card-2	-0.007	-0.007	-0.021	-0.007
	(0.200)	(0.200)	(0.204)	(0.255)
SCIP-VTE-2	-0.215	-0.215	-0.280	-0.257
	(0.190)	(0.190)	(0.219)	(0.274)
SCIP-Inf-1	-0.162	-0.162	-0.319	-0.381
	(0.203)	(0.204)	(0.204)	(0.257)
SCIP-Inf-2	0.155	0.155	-0.194	-0.379
	(0.189)	(0.189)	(0.245)	(0.317)
SCIP-Inf-3	-0.221	-0.221	-0.213	-0.203
	(0.166)	(0.166)	(0.187)	(0.240)
Linear trend	**X**	**X**	**X**	**X**
Linear trend sq		**X**	**X**	**X**
State-specific linear trend			**X**	**X**
State-specific linear trend sq				**X**
Hospital fixed-effects	**X**	**X**	**X**	**X**

Notes:***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors are clustered at hospital level. Each cell represents a model.

Figures

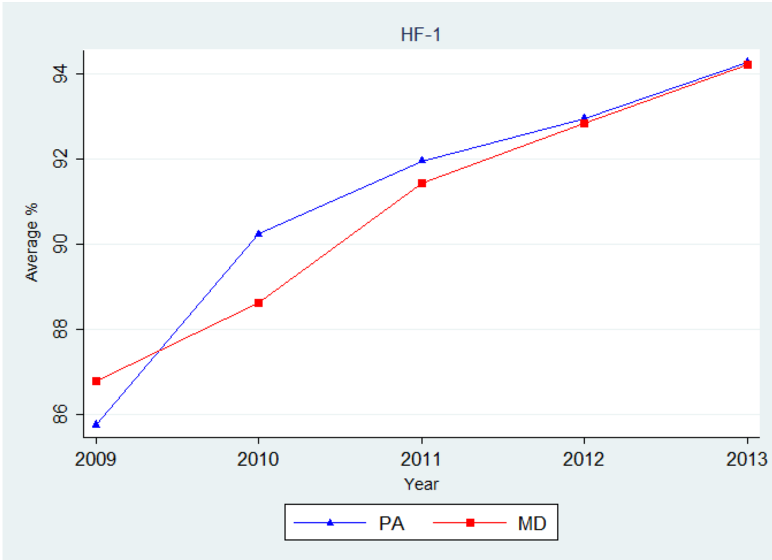
Figure 2.1: Trend in health outcomes

(a) In-hospital mortality

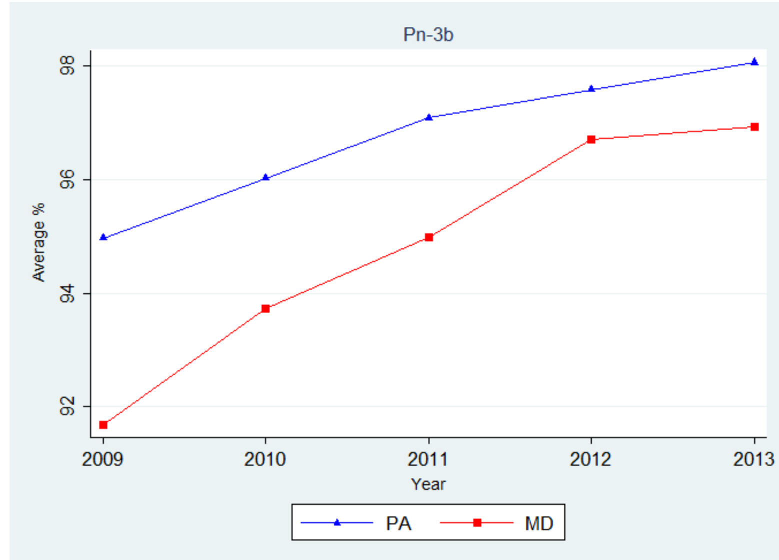


(b) HCC risk scores

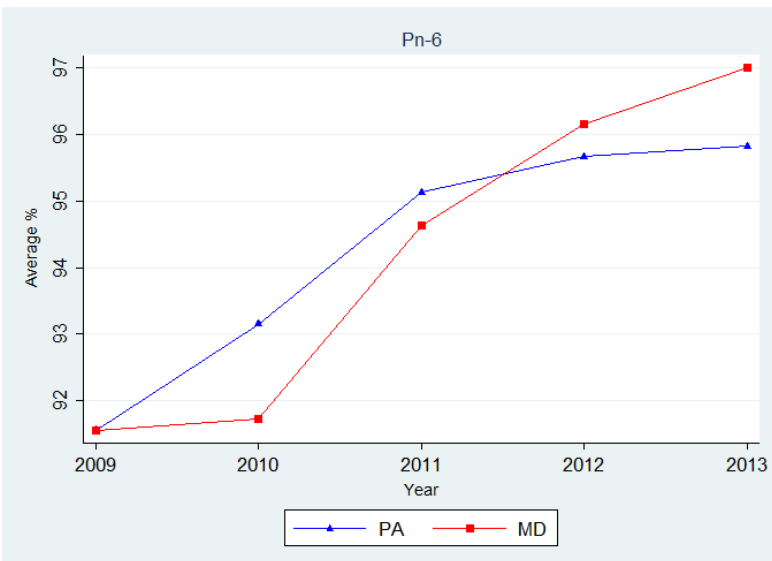
Figure 2.2: Hospital process of care measures



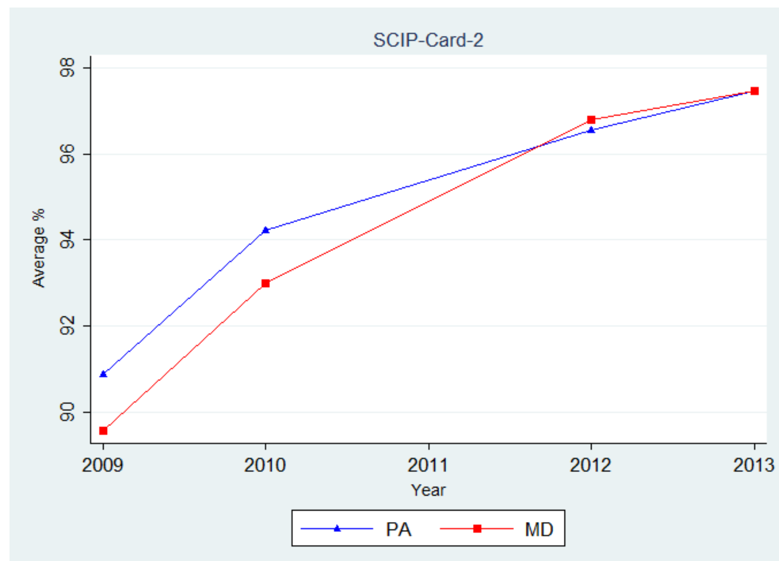
(a) HF-1



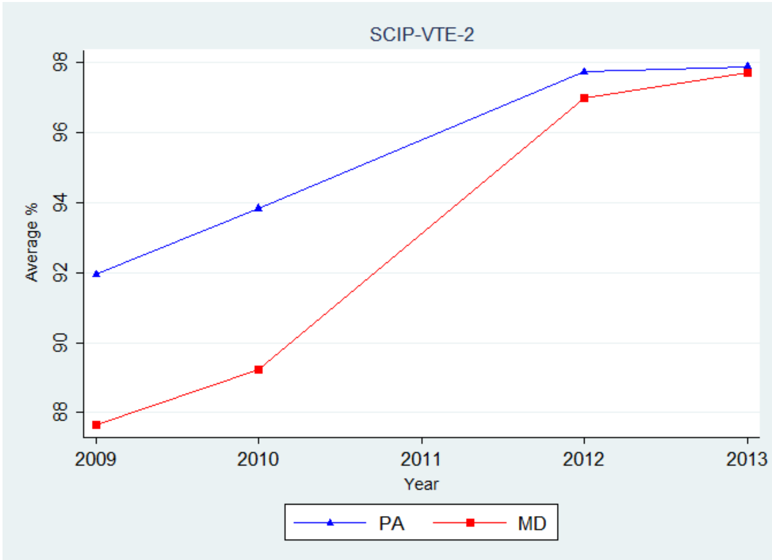
(b) Pn-3b



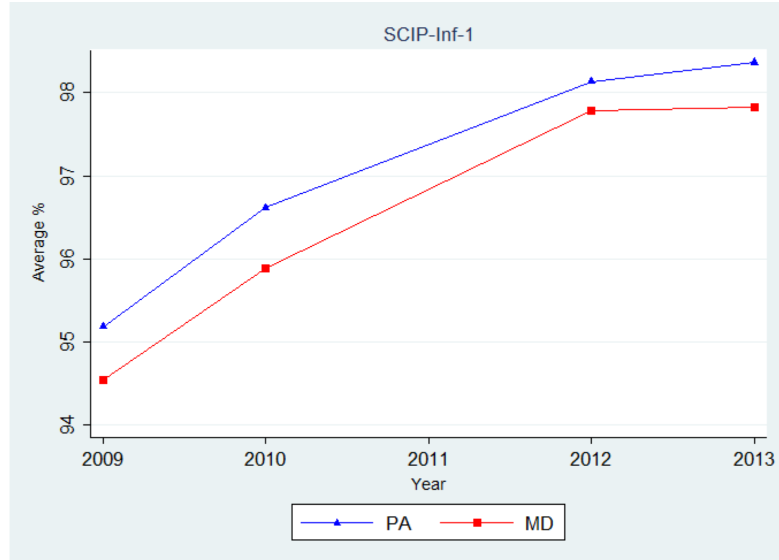
(c) Pn-6



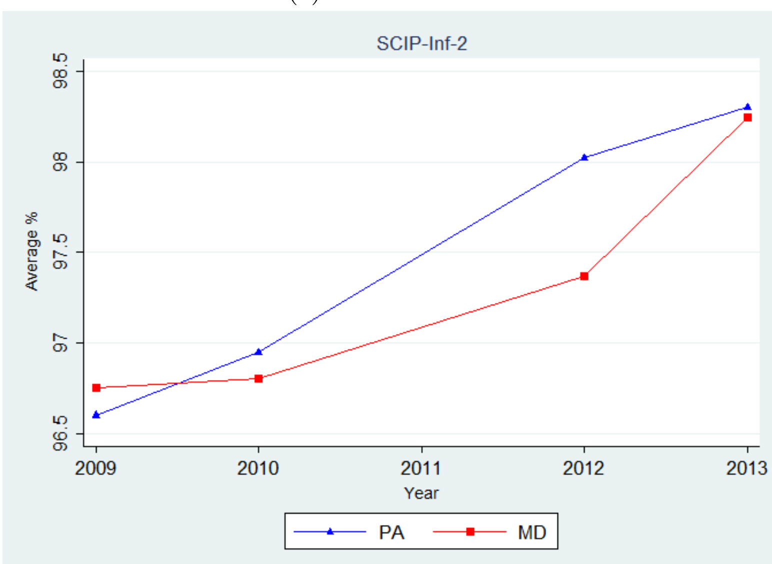
(d) SCIP-Card-2



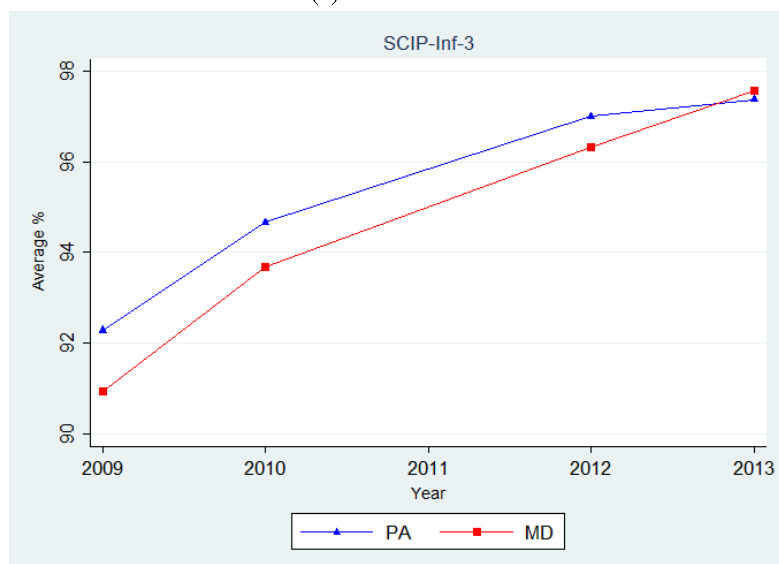
(e) SCIP-VTE-2



(f) SCIP-Inf-1

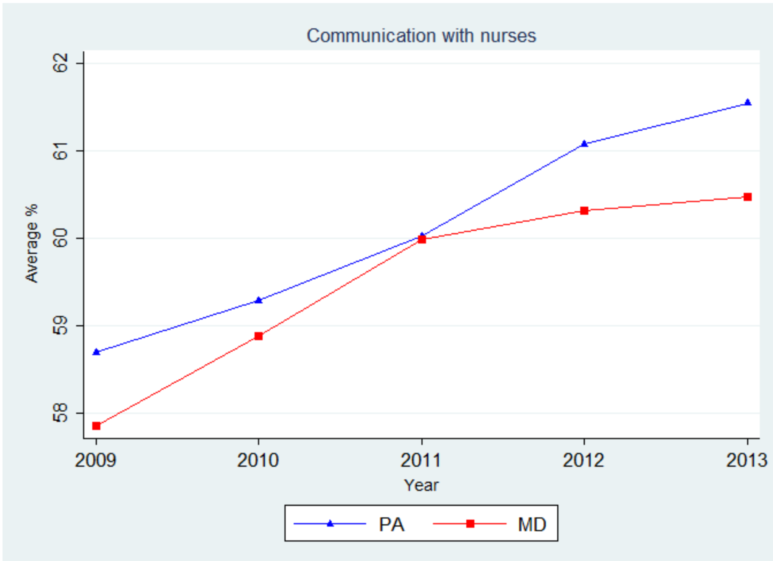


(g) SCIP-Inf-2

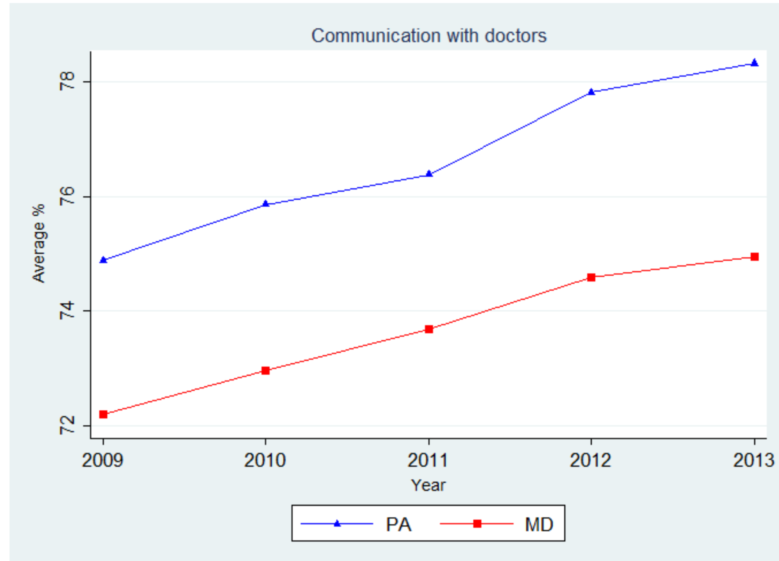


(h) SCIP-Inf-3

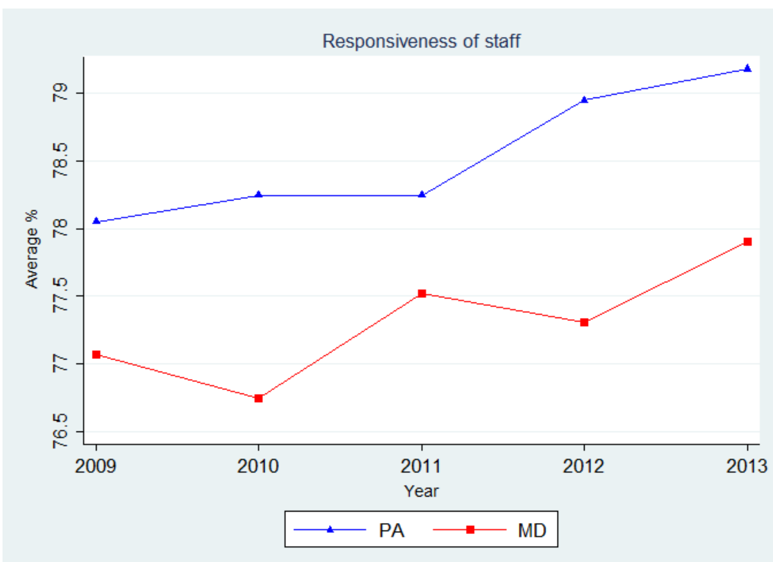
Figure 2.3: HCAHPS patient experience of care measures



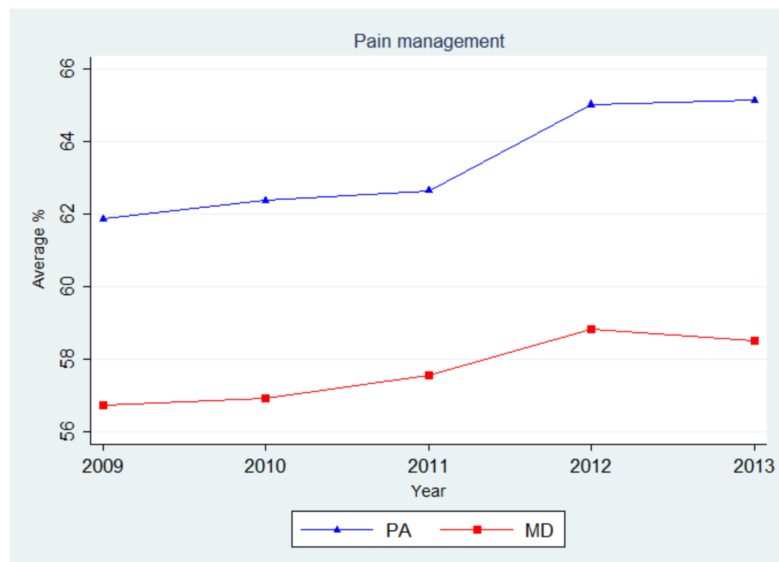
(a) Communication with nurses



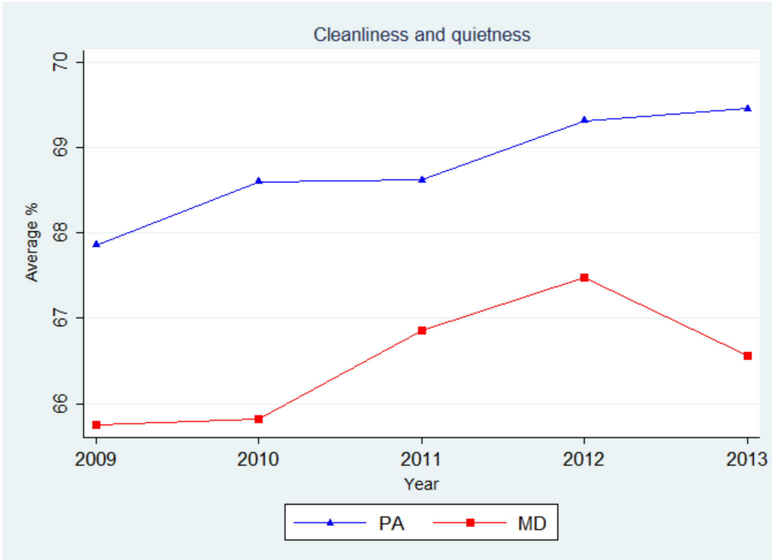
(b) Communication with doctors



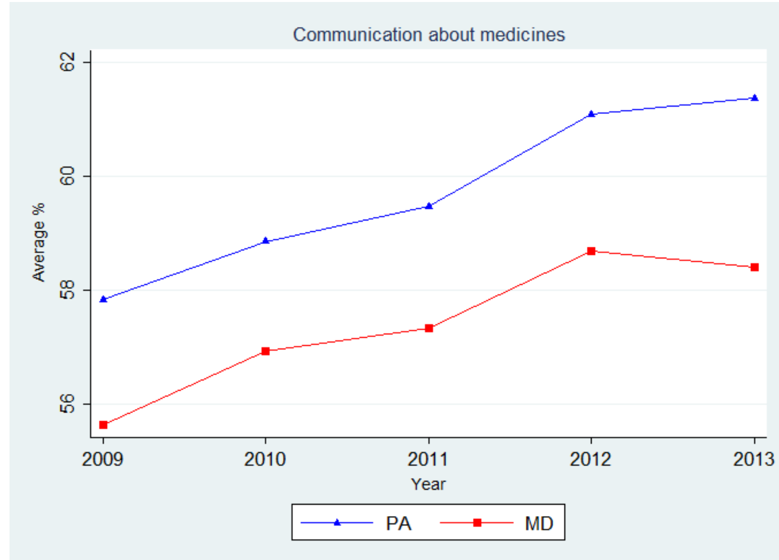
(c) Responsiveness of staff



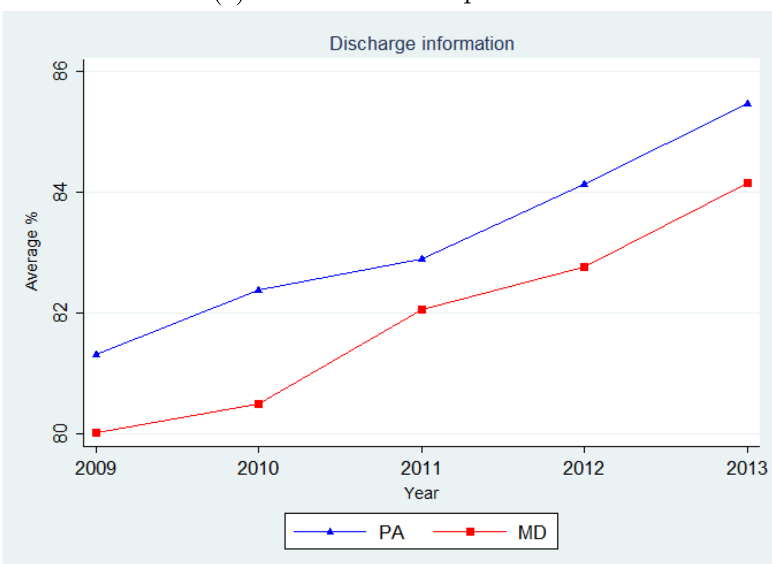
(d) Pain management



(e) cleanliness and quietness



(f) Communication about medicines



(g) Discharge information

(h) Overall rating

2.7 Appendix

2.7.1 Constructions of HCC risk scores

I utilize the CMS HCC Version 12 Model to compute the HCC risk scores. The algorithm I use is based on the SAS programs CMS provides on its website.[41] In addition to the demographic (mainly age and gender) and diagnostic information, the HCC model also utilizes information on patients' Medicaid eligibility and disability status. Since the inpatient discharge data I use lack these variables, the risk scores used in this paper do not account for these factors.

In constructing the risk scores, I first assign each patient to 70 HCCs based on the principle and up to 7 secondary ICD-9 codes using a HCC-ICD-9 crosswalk. Note that a patient can be categorized into more than one HCCs if he has multiple comorbidities. For example, a patient can both have congestive heart failure (HCC 80) and pneumonia (HCC 112). Therefore, the risk score for this patient reflects both conditions. Hierarchies are then applied to the condition categories to make sure that only the most severe form of a condition is coded and a type of co-morbidity is not counted more than once for each patient. For example, if a patient is initially coded to have AMI (HCC 81), then he will be coded as not having "unstable angina and other acute ischemic heart disease" (HCC 82) and "angina pectoris/old myocardial infarction" (HCC 83) because HCC 81 represents the most severe form of the heart condition among the three. In addition to diagnoses, patients are categorized into a number of age-gender cells, each of which is associated with a coefficient that represents patient risk in that cell. As a last step, a patient's risk score is computed as

$$\text{Risk score} = \sum_i HCC_i \times \alpha_i + \lambda,$$

where HCC_i is a binary indicator that equals 1 if the patient has a condition in HCC_i, with

[41] These SAS programs are available at `http://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors.html`.

α_i being the coefficient (weight) for that condition category. The constant term λ is the risk-adjustment factor for the particular age-gender cell the patient belongs to. In computing the risk scores, I use the coefficients from the *CMS HCC-Community Model*, which are based on expenditures of Medicare enrollees residing in communities.[42]

2.7.2 CMS patient case mix adjustment for HCAHPS

I briefly describe the method CMS uses for patient case mix adjustment in this section. For a detailed description of case mix adjustment methodology for HCAHPS, please refer to O'Malley et al. (2005). The purpose of case-mix adjustment is to counter the tendency of patients to respond more positively and negatively (for example, a patient in fair health is less likely to answer the top-box responses than a patient in excellent health). It is performed quarterly after data cleaning and before mode adjustment. Currently, CMS uses 15 case-mix adjustment factors, including self-assessed health status, level of education, age, service line, patient response percentile, interactions between age and service line, and survey language. Hospitals' HCAHPS scores are adjusted to quarterly national means of these adjustment factors. For a hospital, the patient case-mix adjusted score on a given measure is computed using the following formula:

$$\tilde{y_j} = y_j + \sum_i \alpha_{ij}(m_i - \delta_i),$$

where y_j is the unadjusted raw score on measure j, α_{ij} is the coefficient of adjustment factor i along measure j, and δ_i is the national mean of adjustment factor i. The formula shows clearly that the direction of adjustment for a particular measure is determined by both coefficients of adjustment factors and hospitals' patient case mix relative to the national average.[43]

[42] Currently, there are 4 different versions of the CMS HCC risk-adjustment models, each being applicable for different populations. Other versions include *Institutional Model*, *New Enrollee Model*, and *Special Need Model*.

[43] For an example of case mix adjustment coefficients, see `http://www.hcahpsonline.org/modeadjustment.aspx`.

Table 2.7: Domain weights for HVBP: FY 2013 - FY 2015

Domain	FY 2013 Weight	FY 2014 Weight	FY 2015 Weight
Clinical process of care	70%	45%	20%
Patient experience of care	30%	30%	30%
Outcome of care	N/A	25%	30%
Efficiency of care	N/A	N/A	20%

Table 2.8: Baseline and performance periods for HVBP: FY 2013 - FY 2015

	FY 2013		FY 2014		FY 2015	
	Baseline	Performance	Baseline	Performance	Baseline	Performance
Clinical process of care	7/1/09-3/31/10	7/1/11-3/31/12	4/1/10-12/31/10	4/1/12-12/31/12	1/1/11-12/31/11	1/1/13-12/31/13
Patient experience of care	7/1/09-3/31/10	7/1/11-3/31/12	4/1/10-12/31/10	4/1/12-12/31/12	1/1/11-12/31/11	1/1/13-12/31/13
Outcome of care	N/A	N/A	7/1/09-6/30/10	7/1/11-6/30/12	10/1/10-6/30/11	10/1/12-6/30/13
Efficiency of care	N/A	N/A	N/A	N/A	5/1/11-12/31/11	5/1/13-12/31/13

Table 2.9: Clinical process of care domain

Measure ID	Description
Heart conditions	
AMI-7a	Heart attack patients given Fibrinolytic medication with 30 min of arrival
AMI-8a	Heart attack patients given PCI within 90 minutes of arrival
HF-1	Heart failure patients given discharge instructions
Pneumonia	
PN-3b	Pneumonia patients whose initial emergency room blood culture was performed prior to the administration of the first hospital dose of antibiotics
PN-6	Pneumonia patients given the most appropriate appropriate initial antibiotic(s)
Surgical care	
SCIP-Card-2	Surgery patients who were taking heart drugs called Beta blockers before coming to the hospital, who were kept on the Beta blockers just before and after their surgery
SCIP-VTE-1*	Surgery patients whose doctors ordered treatments to prevent blood clots after certain types of surgeries
SCIP-VTE-2	Patients who got treatment at the right time (with 24 hours) before or after their surgery to help prevent blood clots after certain types of surgeries
Healthcare associated infections	
SCIP-Inf-1	Surgery patients who are given an antibiotic at the right time (within 1 hour) before surgery to help prevent infection
SCIP-Inf-2	Surgery patients given the right kind of antibiotic to help prevent infection
SCIP-Inf-3	Surgery patients whose preventive antibiotics are stopped at the right time (within 24 hours after surgery)
SCIP-Inf-4	Heart surgery patients whose blood sugar is kept under good control in the days right after surgery
SCIP-Inf-9**	Surgery patients whose urinary catheters were removed on the first or second day after surgery

Notes: *SCIP-VTE-1 is excluded for FY 2015; **SCIP-Inf-9 is added for FY 2014. All process measures are in proportions.

Table 2.10: Experience of care domain

Measure	Description
Communication with nurses	Percentage of patients who reported that their nurses “Always” communicated well
Communication with doctors	Percentage of patients who reported that their doctors “Always” communicated well
Responsiveness of hospital staff	Percentage of patients who reported that hospital staff were “Always” responsive to their needs
Pain management	Percentage of patients who reported that their pain was “Always” well controlled
Cleanliness and quietness of hospital environment	Percentage of patients who reported that the hospital environment was “Always” clean and quiet
Communication about medicines	Percentage of patients who reported that staff “Always” explained about medicines
Discharge information	Percentage of patients who reported they were given information about what to do during their recovery at home
Overall rating	Percentage of patients whose overall rating of the hospital was ’9’ or ’10’ on a scale from 0 (low) to 10 (high)

Notes: All experience measures are in proportions.

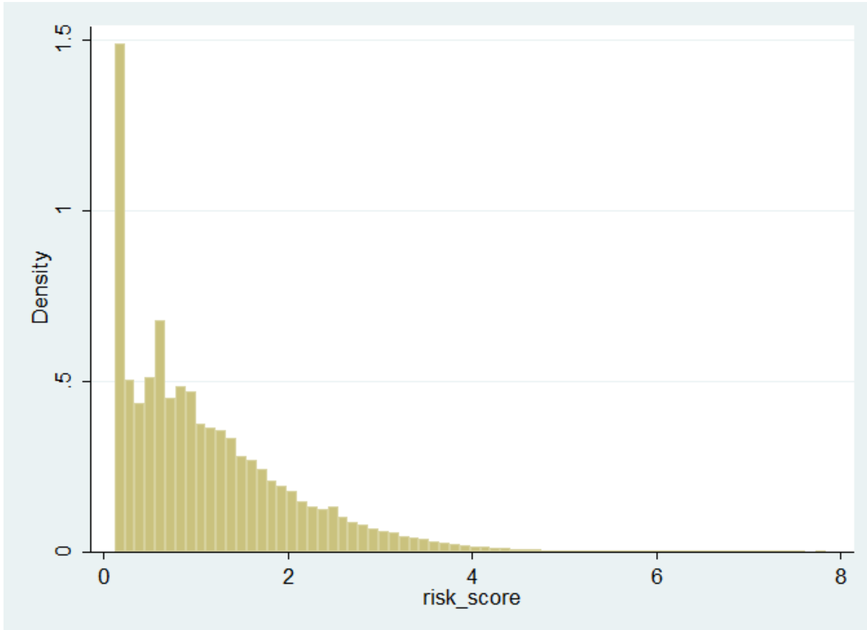
Table 2.11: Estimates from HCAHPS regressions: self-assessed health status

	Communication with nurses		Communication with doctor		Staff responsiveness		Pain management	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Self-assessed health	-0.034***	-0.038***	-0.044***	-0.049***	-0.047***	-0.053***	-0.062***	-0.072***
	(0.006)	(0.007)	(0.006)	(0.008)	(0.009)	(0.011)	(0.009)	(0.012)
Self-assessed health × obstetric		-0.051		0.021		-0.050		0.012
		(0.032)		(0.029)		(0.041)		(0.048)
Self-assessed health × surgical		0.017		0.015		0.025		0.021
		(0.012)		(0.012)		(0.018)		(0.019)
Observations	3,363	3,363	3,354	3,354	3,042	3,042	2,155	2,155

	Communication about medicine		Cleanliness and quietness		Discharge information		Overall rating	
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Self-assessed health	-0.035***	-0.031***	-0.037***	-0.024**	0.003	0.006	-0.069***	-0.074***
	(0.008)	(0.009)	(0.009)	(0.011)	(0.004)	(0.005)	(0.008)	(0.010)
Self-assessed health × obstetric		-0.018		-0.080*		-0.020		-0.108**
		(0.039)		(0.044)		(0.018)		(0.048)
Self-assessed health × surgical		-0.007		-0.027		-0.006		0.024
		(0.017)		(0.019)		(0.008)		(0.017)
Observations	3,359	3,359	2,363	2,363	3,252	3,252	3,334	3,334

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Robust standard errors are reported in parenthesis. Each column represents a model. Other control variables include 10-year age categories, survey language (English), service line (maternal, surgical, and medical is the reference group), level of education, interactions between age category and service line.

Figure 2.4: Distribution of HCC risk scores



2.7.3 HCAHPS survey instrument

Your care from nurses

1. During this hospital stay, how often did nurses treat you with courtesy and respect?

    1. Never
    2. Sometimes
    3. Usually
    4. Always

2. During this hospital stay, how often did nurses listen carefully to you?

    1. Never
    2. Sometimes
    3. Usually
    4. Always

3. During this hospital stay, how often did nurses explain things in a way you could understand?

    1. Never
    2. Sometimes
    3. Usually
    4. Always

4. During this hospital stay, after you pressed the call button, how often did you get help as soon as you wanted it?

    1. Never
    2. Sometimes
    3. Usually
    4. Always

Your care from doctors

5. During this hospital stay, how often did doctors treat you with courtesy and respect?

    1. Never
    2. Sometimes
    3. Usually
    4. Always

6. During this hospital stay, how often did doctors listen carefully to you?

1. Never
2. Sometimes
3. Usually
4. Always

7. During this hospital stay, how often did doctors explain things in a way you could understand?

    1. Never
    2. Sometimes
    3. Usually
    4. Always

The hospital environment

8. During this hospital stay, how often were your room and bathroom kept clean?

    1. Never
    2. Sometimes
    3. Usually
    4. Always

9. During this hospital stay, how often was the area around your room quiet at night?

    1. Never
    2. Sometimes
    3. Usually
    4. Always

Your experiences in the hospital

10. During this hospital stay, did you need help from nurses or other hospital staff in getting to the bathroom or in using a bedpan?

    1. Yes
    2. No

11. How often did you get help in getting to the bathroom or in using a bedpan as soon as you wanted.

    1. Never
    2. Sometimes
    3. Usually
    4. Always

12. During this hospital stay, did you need medicine for pain?
    1. Yes
    2. No
13. During this hospital stay, how often was your pain well controlled?
    1. Never
    2. Sometimes
    3. Usually
    4. Always
14. During this hospital stay, how often did hospital staff do everything they could to help you with your pain?
    1. Never
    2. Sometimes
    3. Usually
    4. Always
15. During this hospital stay, were you given any medicine that you had not taken before?
    1. Yes
    2. No
16. Before giving you any new medicine, how often did hospital staff tell you what the medicine was for?
    1. Never
    2. Sometimes
    3. Usually
    4. Always
17. Before giving you any new medicine, how often did hospital staff describe possible side effects in a way you could understand?
    1. Never
    2. Sometimes
    3. Usually
    4. Always

When you left the hospital

18. After you left the hospital, did you go directly to your own home, to someone else's home, or to another health facility?

    1. Own home
    2. Someone else's home
    3. Another health facility

19. During this hospital stay, did doctors, nurses, or other hospital staff talk with you about whether you would have the help you needed when you left the hospital?

20. Before giving you any new medicine, how often did hospital staff describe possible side effects in a way you could understand?

    1. Yes
    2. No

21. During this hospital stay, did you get information in writing about what symptoms or health problems to look out for after you left the hospital?

    1. Yes
    2. No

Overall rating of hospital

22. Using any number from 0 to 10, when 0 is the worst hospital possible and 10 is the best hospital possible, what number would you use to rate this hospital during your stay?

23. Would you recommend this hospital to your friends and family?

    1. Definitely no
    2. Probably no
    3. Probably yes
    4. Definitely yes

Understanding you care when you left the hospital

24. During this hospital stay, staff took my preferences and those of my family or caregiver into account in deciding what my health care need would be when I left.

    1. Strongly disagree
    2. Disagree
    3. Agree
    4. Strongly agree

25. When I left the hospital, I had a good understanding of the things I was responsible for in managing my health.
    1. Strongly disagree
    2. Disagree
    3. Agree
    4. Strongly agree
26. When I left the hospital, I clearly understood the purpose for taking each of my medications.
    1. Strongly disagree
    2. Disagree
    3. Agree
    4. Strongly agree
    5. I was not given any medication when I left the hospital

About you

27. During this hospital stay, were you admitted to this hospital through the Emergency Room?
    1. Yes
    2. No
28. In general, how would you rate your overall health?
    1. Excellent
    2. Very good
    3. Good
    4. Fair
    5. Poor
29. In general, how would you rate your overall mental or emotional health?
    1. Excellent
    2. Very good
    3. Good
    4. Fair
    5. Poor
30. What is the highest grade of school that you have completed?
    1. 8th grade or less

2. Some high school, but did not graduate
3. High school graduate or GED
4. Some college or 2-year degree
5. 4-year college graduate
6. More than 4-year college degree

31. Are you of Spanish, Hispanic or Latino origin or descent?

    1. No, not Spanish/Hispanic/Latino
    2. Yes, Puerto Rican
    3. Yes, Mexican, Mexican American, Chicano
    4. Yes, Cuban
    5. Yes, other Spanish/Hispanic/Latino

32. What is your race? Please choose one or more.

    1. White
    2. Black or African American
    3. Asian
    4. Native Hawaiian or other Pacific Islander
    5. American Indian or Alaska Native

33. What language do you mainly speak at home?

    1. English
    2. Spanish
    3. Chinese
    4. Russian
    5. Vietnamese
    6. Portuguese
    7. Some other language (please print): ____________

Chapter 3

The Short-Term Effect of Depression on Labor Market Outcomes

3.1 Introduction

Mental health problems are prevalent among U.S. adults. It is estimated that about 26 percent of adults have some type of mental disorder broadly defined, and 6 percent suffer from severe mental illness (Kessler et al, 2005). The costs associated with mental illness include direct medical costs for mental health treatment and indirect costs in the form of time away from work (absenteeism) and reduced productivity while at work (presenteeism). The National Institute of Mental Health (NIMH) estimated that the total annual cost for people with severely debilitating mental disorders exceeds $300 billion each year, a large fraction of which falls on payers of medical services and employers (NIMH, 2002).

Among various mental health conditions, depression is of interest for two reasons. First, it is one of the most common mental health problems among adults. In 2011, 6.6 percent of adults aged 18 or older in the U.S. (15.2 million people) experienced at least one major depressive episode during the past year (National Survey on Drug Use and Health, 2011). Second, the social cost of depression is surprisingly high: in high-income countries, unipolar depression (also known as major depressive disorder or major depression) is the leading cause of burden of disease and accounts for 8.2 percent of total disability-adjusted life years (DALYs). In contrast, heart diseases only account for 6.3 percent of total DALYs

(WHO, 2008).

A correct understanding of the economic costs associated with mental health problems has important policy implications. Historically, mental health conditions are covered less generously than physical illness by private insurance, which has contributed to under-treatment of mental health problems.[1] Legislators at both the federal and state level responded by enacting mental health parity laws that aim to expand coverage for mental health conditions. However, it is still unclear whether these parity laws influence health outcomes (GAO, 2011), and if mental health conditions do not significantly affect individuals labor market opportunities, the benefits of expanding mental health coverage may not justify the costs.

The main difficulty in attributing poor labor market outcomes to the presence of mental health problems in a cross section of individuals is endogeneity. As pointed out by Chatterji et al (2011), mental health status is endogenous in both a structural and statistical sense: reverse causality poses a threat to identification if mental health and labor market outcomes are determined simultaneously. For example, job loss could trigger an episode of depression for an individual with a predisposition. In addition, unobserved physical and mental health status as well as unobserved productivity may lead to omitted variable bias. In most cases, this type of endogeneity will lead to an over-estimation of the negative impact of poor mental health on labor market outcomes, such as employment, wages, work hours and work loss days, in non-causal analyses.[2] To address this endogeneity problem, researchers have primarily relied on instrumental variables or exclusion restrictions in multiple equation models (Ettner et al, 1997; Chatterji et al, 2011; Alexandre and French, 2001). Nonetheless, it is difficult to find valid instruments that affect mental health but have no impact on labor market outcomes. The instruments that have been used in the literature fall into two broad categories: personal characteristics and social support.[3] However, the exogeneity of

[1] According to the National Survey on Drug Use and Health in 2007, among the 24.3 million adults with serious psychological distress, only 44.6 percent used mental health services in the past year.

[2] It is possible that certain types omitted variables could result in the under-estimation of the impact of poor mental health for some outcomes. For example, unobserved generosity in mental health coverage could bias the effect of poor mental health towards zero in a wage equation.

[3] See Chatterji et al (2007) for a comprehensive review of the instruments that have been used in the

these instruments is difficult to validate. For example, the same personal characteristics that affect mental health may be correlated with productivity and labor market outcomes, and greater support from families or communities may also be related to better labor market opportunities.

We exploited the longitudinal dimension of the Medical Expenditure Panel Survey (MEPS) to estimate the causal relationship between depression and labor market outcomes. The MEPS is well suited to this analysis because it contains a clinically validated measure of depression that is collected contemporaneously with several labor market outcomes. We estimated fixed effects and correlated random effects (CRE) models (Chamberlain, 1980), which require no exclusion restrictions for identification and allow unobserved heterogeneity to be correlated with the regressors. This is an attractive alternative method of measuring causal effects when the primary threats to causal inference are unobservable dimensions of productivity and health status.

3.2 Theoretical framework and empirical methods

3.2.1 A simple theoretical model of labor supply

We follow the standard labor supply model with health capital (Currie and Madrian, 1999). The inter-temporal utility function for individual i is

$$U_i = \sum_{t=1}^{T} \left(\frac{1}{1+\delta} \right)^t U_{it}, \tag{3.1}$$

where δ is a constant discount rate. We further assumed that the utility in period t takes the form:

$$U_{it} = U_1(C_{it}) + U_2(H_{it}, L_{it}), \tag{3.2}$$

where C_{it} is a composite commodity with price normalized to one, H_{it} is the current health capital, and L_{it} is the leisure consumed by the individual in that period. The individual

literature.

maximizes his overall utility subject to the following constraints:

$$
\begin{aligned}
&H_{it} = H(H_{i,t-1}, V_{it}, \mu_i), \\
&Q_{it}\omega_{it} + A_{it} = C_{it}, \\
&L_{it} + V_{it} + Q_{it} + S_{it} = \Omega, \\
&\omega_{it} = \omega(H_{it}, X_{it}, M_{it}, m_i), \\
&S_{it} = S(H_{it}, Z_{it}, \theta_i).
\end{aligned}
\tag{3.3}
$$

The first constraint is the health production function where current health capital, H_{it}, depends on past health capital as well as time spent in health production, V_{it}. μ_i is unobserved heterogeneity in health production. The second and third constraints are a budget and time constraint, respectively. The budget constraint limits expenditure on the composite commodity to no more than the sum of unearned income A_{it} and labor income $Q_{it}\omega_{it}$, while the time constraint ensures that the sum of sick time, S_{it} , hours worked, Q_{it} , time devoted to health production, V_{it}, and leisure, L_{it}, equal total time available (Ω). The fourth constraint specifies the wage is a function of health capital, a vector of individual characteristics, X_{it}, observed job and employer characteristics, M_{it}, and unobserved productivity, m_i. In the last constraint, sick time S_{it} is a function of current health capital, a vector of exogenous determinants, Z_{it}, individual-specific propensity to get sick, θ_i. The individual chooses the optimal consumption path, C_{it}, hours worked, $Q_{it} \geq 0$, time inputs into health production, $V_{it} \geq 0$, to maximize his lifetime utility subject to (3.3). The first-order conditions with respect to Q_{it} are given by

$$
\frac{\partial U_2}{\partial L_{it}} \geq (1+\delta)^t \omega_{it}\lambda_{it}, \quad t = 1, \ldots, T, \tag{3.4}
$$

where λ_{it} is the marginal utility of wealth in period t. It follows that the conditional labor supply function in period t is

$$Q_{it} = Q\left(\lambda_{it}, (1+\delta)^t, H_{it}, \omega\left(H_{it}, X_{it}, M_{it}, m_i\right)\right). \tag{3.5}$$

One of the distinct features of this model is that the health stock is an endogenous choice. Individuals can affect the health stock by altering time devoted to health production. From the quasi-reduced form equation of labor supply as a function of health, it is clear that there are several determinants of labor supply that are unobserved to the econometrician. Some of these, such as unobserved determinants of sickness and health production, and unmeasured productivity have a large genetic component, while unobserved preferences may be established prior to adulthood. Under these assumptions, one can further decompose all the unobserved factors into time-invariant and time-varying components. Denote α_i and ε_{it} as all the time-invariant and time-varying determinants of labor supply, respectively, where $\alpha_i = \mu_i + m_i$. Equation (3.5) then becomes:

$$Q_{it} = Q(X_{it}, H_{it}, M_{it}, \alpha_i, \varepsilon_{it}) \tag{3.6}$$

which serves as the basis of our empirical analysis.

3.2.2 Empirical models

One basic estimating equation for (3.6) is:

$$Y_{it} = f(X_{it}'\beta + D_{it}\gamma + \alpha_i + \varepsilon_{it}), \tag{3.7}$$

where Y_{it} is the labor market outcome of interest for individual i in time period t, X_{it} is a vector of covariates including demographics, human capital, and observed physical health status. We use the binary indicator D_{it} to denote whether the individual is depressed in time period t. The time-invariant parameter α_i captures unobserved components of physical and mental health status and unmeasured productivity. The error term ε_{it} is assumed to be uncorrelated with all regressors from all time periods (strong exogeneity). The choice of $f(\cdot)$ is dependent on the outcome Y_{it}. For employment status, we specify (3.7) as

$$\Pr(Y_{it} = 1) = \Phi(X_{it}'\beta + D_{it}\gamma + \alpha_i), \tag{3.8}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. In the case of continuous outcomes such as hourly wage and weekly hours worked, (3.7) is the identity function. Because hourly wage is highly skewed to the right, we used the log transformation of the hourly wage as the dependent variable. In addition, we estimated a set of ordered probit models using the conditional sample to investigate whether depression affects full-time versus part-time work status. We created three categories of weekly hours worked: less than 30 hours a week, between 30 and 40 hours a week, and at least 40 hours a week (full-time). The reason why we made a distinction between those who work part-time for 30 or more hours per week is that some employers offer fringe benefits (such as health insurance) to those who work at least 30 hours a week.[4]

We are also interested in absenteeism, which is captured in the MEPS using a count variable indicating the number work loss days during the reference period. We modeled work loss days using the zero-inflated ordered probit (ZIOP) specification developed by Harris and Zhao (2007). This model is more appropriate than a standard count data model for two reasons. First, the distribution of work loss days is highly skewed, with over 70 percent of employed individuals reporting zero work loss days and only around 10 percent reporting more than 7 annual work loss days. The ZIOP specification allowed us to account for this right skewness by combining work loss days at the higher end of its distribution into discrete categories (Meyerhoefer and Zuvekas, 2010). Second, the ZIOP model facilitates

[4]According to the data from National Compensation Survey, 24 percent of the part-time workers in private industry have access to medical care benefits (Employee Benefits in the United States- March 2012). URL: `http://www.bls.gov/news.release/archives/ebs2_07272010.pdf`

the investigation of the effect of depression at both the extensive and intensive margins by separately modeling two different latent processes that generate the zero counts of work loss days. In particular, zero counts were reported because individuals remained healthy during the sample period, or because they did get sick, but remained at work while they were ill (i.e. presenteeism).

Let $\tilde{r}$ be a latent variable underlying the first data generating process determining sickness:

$$\tilde{r}_{it} = X'_{1it}\beta + D_{it}\gamma + \alpha_i + \eta_{it}. \tag{3.9}$$

The individual is healthy if $\tilde{r} \leq 0$, but when the individual becomes sick $\tilde{r} > 0$. If the error term η_{it} follows a standard normal distribution, then the probability of sickness is

$$\Pr(r_{it} = 1|X_{1it}, D_{it}) = \Pr(\tilde{r}_{it} > 0|X_{1it}, D_{it}) = \Phi(X'_{1it}\beta + D_{it}\gamma + \alpha_i). \tag{3.10}$$

Conditional on illness, the actual work loss days of each individual is a discrete variable ($\tilde{y} = 0, 1, \ldots, J$), which is assumed to be determined by the latent process:

$$\tilde{y}^* = X'_{2it}\pi + D_{it}\varphi + \alpha_i + \nu_{it}, \tag{3.11}$$

where $\tilde{y} = j$ when $c_j < \tilde{y}^*_{it} < c_{j+1}$ ($j = 0, 1, \ldots, J$), which are cutoffs estimated jointly with $(\beta', \gamma, \pi', \varphi)'$, and ν_{it} is standard normally distributed. Therefore, presenteeism occurs when $\tilde{y} = j = 0$. For notational simplicity, we suppress individual and time subscripts as well as unobserved heterogeneity, and denote $\tilde{X}_1 = (X'_1, D)'$ and $\tilde{X}_2 = (X'_2, D)'$. Then the two latent processes map to the observed data as $y = r\tilde{y}$, such that the density of y is given by

$$\Pr(y) = \begin{cases} \Pr(y = 0|\tilde{X}_1, \tilde{X}_2) = \left[1 - \Phi(\tilde{X}'_1\tilde{\beta})\right] + \Phi(\tilde{X}'_1\tilde{\beta})\Phi(-\tilde{X}'_2\tilde{\pi}), \\ \Pr(y = j|\tilde{X}_1, \tilde{X}_2) = \Phi(\tilde{X}'_1\tilde{\beta}) \left[\Phi(c_{j+1} - \tilde{X}'_2\tilde{\pi}) - \Phi(c_j - \tilde{X}'_2\tilde{\pi})\right], \\ \Pr(y = J|\tilde{X}_1, \tilde{X}_2) = \Phi(\tilde{X}'_1\tilde{\beta}) \left[1 - \Phi(c_J - \tilde{X}'_2\tilde{\pi})\right], \end{cases} \tag{3.12}$$

where $j = 1, \cdots, J-1$. We estimated this model using the conditional sample of workers and computed an overall marginal effect of depression on work loss days as

$$ME = \sum_i ME_i \times L_i, \tag{3.13}$$

where L is the weighted average of work loss days in each category, and ME_i is the marginal effect of depression on the probability of being in category i.

Although one could in principle model α_i using fixed effects, this estimation strategy does not yield consistent parameter estimates of most nonlinear specifications when the time series dimension of the panel is fixed (Cameron and Trivedi, 2005). As an alternative Chamberlain (1980) proposed a random effects estimator that is consistent when T is fixed and allows α_i to be correlated with all regressors in all time periods. Again, denote $\tilde{X} = (X', D)'$ and $\tilde{\beta} = (\beta', \gamma)'$, and note that α_i can be modeled using a linear projection containing the regressors from all time periods:

$$\alpha_i = \tilde{X}'_{i1}\lambda_1 + \tilde{X}'_{i2}\lambda_2 + \cdots + \tilde{X}'_{iT}\lambda_T + u_i, \tag{3.14}$$

where u_i is orthogonal to $\tilde{X}$ by construction, and is assumed to be distributed as $N(0, \sigma_u^2)$. Substituting (3.14) into (3.7), we have

$$Y_{it} = f(\tilde{X}'_{i1}\lambda_1 + \cdots + \tilde{X}'_{it}(\tilde{\beta} + \lambda_t) + \cdots + \tilde{X}'_{iT}\lambda_T + u_i + \varepsilon_{it}). \tag{3.15}$$

Jakubson (1988) described a two-step procedure for consistent estimation of $\tilde{\beta}$ in labor supply models. First, consistent estimates of the reduced-form parameters are obtained from equation-by-equation estimation of (3.15), followed by identification of the structural parameters $\tilde{\beta}$, through minimum distance estimation:

$$\min D(\psi) = (\hat{\pi} - H\psi)'\hat{\Omega}^{-1}(\hat{\pi} - H\psi), \tag{3.16}$$

where ψ denotes the vector of structural parameters, $\hat{\pi}$is the vector of reduced-form parameters obtained from the first step, $\hat{\Omega}$ is the variance-covariance matrix of $\hat{\pi}$, and H is a design matrix that maps the structural parameters to the reduced-form parameters. Where possible we confirmed the estimates of our nonlinear CRE models with linear models containing individual fixed effects.

3.3 Data

To estimate our empirical models we used the 2004-2009 Medical Expenditure Panel Survey (MEPS), subset to individuals aged 18-64 who were not full-time students. The MEPS is a nationally representative overlapping panel survey designed to provide estimates of health care use, expenditures, and health insurance coverage for the U.S. civilian non-institutionalized population. The MEPS contains detailed information on respondents health status, demographic and socio-economic characteristics, and employment information. Each panel of respondents was interviewed in 5 rounds covering 2 calendar years.

Our indicator variable for major depression was calculated using the Patient Health Questionnaire (PHQ-2) index, collected during MEPS interview rounds 2 and 4. The PHQ-2 is a validated screening tool for depressive disorders (Kroenke et al, 2003) that is derived from two questions that ask whether the respondent was bothered by ”having little interest or pleasure in doing things” and whether he was ”feeling down, depressed and hopeless” during the past two weeks. For each question, respondents rate themselves on scale of 0-3 based on the frequency of depressed mood and decreased interest in usual activities. The PHQ-2 index is the sum of scores from the above two items and ranges from 0-6. As suggested by Kroenke et al (2003), we coded individuals as suffering from major depression if the value of the PHQ-2 index was greater or equal to 3.[5] Some respondents had missing values for the

[5]Based on this screening tool, 9.1 percent of the individuals in the 2004-2009 MEPS suffered from major depression. Kessler et al. (2005) report that the 12-month prevalence rate of major depressive disorder was 6.7 percent between 2001-2003 based on the DSM diagnostic criteria in the National Comorbidity Survey Replication. Our prevalence rate is higher because the PHQ-2 also captures short-term depressive symptoms. In particular, Kroenke et al (2003) found that PHQ-2 3 had a sensitivity of 83% and a specificity of 92% for major depression.

PHQ-2 index either because they were ineligible for the Self-Administrated Questionnaire (SAQ) containing the PHQ-2 questions, or they did not respond to the SAQ.[6] We excluded individuals ineligible for the SAQ from our estimation sample, and for those where were eligible but did not respond, we imputed a PHQ-2 index value using a generalized linear model.[7]

Respondents were asked to report on labor market outcomes in all 5 rounds, but since the SAQ is only administrated in rounds 2 and 4 (in the middle of the each year), we used the outcome variables collected during these same rounds. We considered individuals with a current main job on the interview date as employed, but excluded the self-employed from our estimation sample because the MEPS does not contain their wage information.[8] We converted hourly wage data and all monetary measures of productivity costs to constant 2009 dollars using the urban CPI.

The MEPS question designed to measure workplace productivity loss asks respondents "the number of times the person lost half-day or more from work because of illness, injury, mental or emotional problems" during the interview round. We constructed work loss days assuming that respondents lost a full day of work. While we believe this to be the most likely duration of work loss, it means that our estimates represent the upper bound of the effect of depression on absenteeism. However, when we computed aggregate productivity loss estimates from the work loss day models, we did so assuming only half a day was lost to identify the lower bound, and a full day was lost to identify the upper bound estimate.

[6] All adults above 18 as of the interview date were asked to fill out the SAQ, with the exception of those made ineligible because they were deceased, institutionalized, moved out of the U.S, or moved to military facilities.

[7] We used a GLM with a log link to account for the right skewness of the distribution of the PHQ-2 index. In the imputation regressions, we included socio-demographic characteristics, the Physical Component Summary (PCS) scores, and Mental Component Summary (MCS) scores. Both PCS and MCS scores were calculated based on the questions in SF-12 Health Survey. The correlation between MCS score and the PHQ-2 index is quite high (the correlation coefficient is -0.72, with the negative sign the result of opposite scaling on the MCS).

[8] Employed individuals who had missing hours or wage were dropped from the conditional sample. Those who reported working more than 120 hours per week were also excluded due to concerns over reporting error. When estimating the aggregate cost of worker absenteeism due to depression we imputed the wages of the self-employed and included them in our calculations.

Due to the fact that the length of reference period varies across respondents, we normalized work loss days to a 12-month period.

We controlled for a full set of socio-demographic and employment characteristics in our models. Our main control variables include age and its square, gender, race and ethnicity, marital status, region (Northeast, South, Midwest, West), urban residence, years of education completed before entering the survey, union status, employer size (less than 25 employees, between 25-99 employees, between 100-500 employees, more than 500 employees), benefits provided by the employer (sick pay, retirement plan, and paid vacations), occupation and industry indicators,[9] number of children under 5 or 18 in the household, log of income earned by other family members (normalized by household size), and year dummies. In order to control for physical health status, we included the Physical Component Summary (PCS) score, which is derived from the questions in SF-12 Health Survey contained in the SAQ. Higher values of the PCS score indicate better physical health status. In the regressions where the dependent variable is weekly hours worked, we also included the log of hourly wages as a control to ensure our specification is consistent with the standard labor supply model. Table 3.1 contains descriptive statistics of all the variables we used in the analysis. Note that all means were weighted using the MEPS SAQ weight which adjusts for SAQ non-response. In our estimation sample, 75 percent of all the individuals aged 18-64 were employed and 9.1 percent suffered from depression based on the PHQ-2 index.

3.4 Results

3.4.1 Employment

We report complete estimation results for the CRE probit models of employment in columns (6) and (8), and the cross-sectional probit models for comparison purposes in columns (2)

[9] The industry indicators include: 1. natural resources/mining/construction/manufacturing; 2. wholesale and retail trade/transportation and utilities; 3. professional and business services/education, health, and social services; 4. other services/public administration/military/unclassifiable industry. We also included an indicator for whether the reported occupation required professional training.

and (4), of Table 3.7 in Appendix 3.7. In all of our models, the random effect capturing unobserved heterogeneity was specified to be correlated with marital status, log of family income (excluding the individual's own income), the PCS score, and the depression measure (index or indicator). Across all specifications, we found a negative and statistically significant association between depression and employment. However, the magnitude of the CRE estimates is substantially smaller than the cross-sectional estimates for both the continuous PHQ-2 depression index and the dichotomous indicator for index values ≥ 3.

The average marginal effects derived from these estimates are reported in Table 3.2. The cross-sectional estimates imply that depression is associated with a 17.6 percentage point reduction in the probability of employment, which is similar in magnitude to the effect found in previous studies (Ettner et al, 1997; Alexandre and French, 2001; Chatterji et al, 2011). In contrast, the CRE estimates imply depression lowers the probability of employment by only 2.6 percentage points, or 3.3 percent. This is similar to the marginal effect we obtain from a linear probability model containing individual fixed effects.

The sizable and significant correlation parameters in Table 3.7 indicate that depression, along with marital status, physical health status, and income earned by other family members, are endogenous in the reduced form estimating equation. This endogeneity is presumably caused by the failure to fully measure worker productivity and health status, which are captured by the random effect in the CRE model. The downward bias on the effect of depression on employment in the cross sectional model is consistent with a positive correlation between unobserved productivity (or good health) and employment, and a negative correlation between productivity and depression.

3.4.2 Hourly wage and weekly hours worked

Panel A of Table 3.3 contains marginal effects of depression on weekly work hours and the hourly wage rate estimated using the conditional sample of workers. The cross-sectional estimates imply that depression is associated with an 8.3 percent reduction in the hourly wage, but this effect is small and imprecisely estimated using both CRE and individual fixed effects to control for unmeasured productivity. We failed to find a relationship between depression and weekly hours worked using either the cross-sectional or panel data specifications. The correlation parameters from the CRE models reported in appendix Table 3.8 indicate that the random effect is not correlated with the depression measure in the work hours model, but that it is significantly correlated with depression in the wage equation.[10] The latter is consistent with the downward bias of depression on wages due to unmeasured worker productivity in the cross sectional model. The marginal effects from both cross-sectional and CRE ordered probit models for part-time versus full-time employment are reported in panel B of Table 3.3. In the cross section, depression reduces the probability of working full time by 1.5 percentage points while the CRE estimates again suggest that depression has no effect on full-time work status.

3.4.3 Absenteeism and aggregate productivity costs of depression

Unlike wages or hours, which are relatively rigid due to labor contracts, work loss days are potentially more responsive to the presence or onset of contemporaneous depression. We report the coefficient estimates of the ZIOP model in Table 3.10. Note that the inflation part of the model captures the effect of depression on the propensity to be sick enough to take days off, while the ordered probit part of the model captures the effect of depression on the length of sick leave. The coefficient estimates from the inflation equation indicate that the depressed are more likely to take days off from work in general than are the non-depressed, although this effect is not precisely estimated. For both the continuous and dichotomous measure, the coefficient estimates of the PHQ-2 indexes are positive and statistically significant in the ordered probit equation, indicating that conditional on taking days off from work, depression increases the length of sick leave.

Marginal effects constructed from these coefficients are reported in Table 3.4. The

[10]Note that the wage rate in the CRE work hours models is assumed to be endogenous, and as a result, is specified to be correlated with the random effect.

treatment effect specification suggests that depression increases work loss days by 1.4 days per year. This represents a one-third increase in work loss days relative to the mean in the conditional sample of workers. Likewise, the marginal effect for the continuous PHQ-2 index indicates that a one unit change in the index value increases work loss days by 0.6 days per year. Although these are relatively large effects, the estimate from the cross-sectional treatment effects model that fails to account for productivity differences across workers is two and half times larger, while the estimate based on the continuous index is nearly twice as large.

In addition, we examined whether depression has a differential impact on absenteeism across worker groups. We stratified the sample by whether the individual had a retirement plan as a proxy for salaried versus hourly paid jobs, and found that depression is associated with a 43.3 percent increase in work loss days by workers without retirement plans (hourly paid), and a 29.3 percent increase among workers with retirement plans (salaried). Given that we controlled for whether the employer provided sick leave or paid time off to visit medical care professionals, these estimates suggest that the cost of depression is greater for lower paid workers.

Using our CRE estimates of the impact of depression on work loss days and information on the weekly work hours and wage rates, we were able to calculate annual estimates of the aggregate cost of depression due to workplace absenteeism for the U.S. working population.[11] In Table 3.5 we report the aggregate cost from 2005-2008 using our CRE estimates. Since the MEPS asks the respondents to report whether they missed work for a half day or more due to sickness or injury, we calculated two different estimates: Our upper bound estimate assumes individuals always missed a full day of work, while our lower bound estimate

[11]For individuals who worked less than 70 hours per week, we assumed that they worked 5 days a week; for those who worked more than 70 hours per week, we assumed that they worked 6 days a week. We then calculated the daily wage by (weekly hours worked*hourly wage/number of days worked per week). The total annual cost of depression is therefore the weighted sum of (daily wage*overall marginal effect) among the depressed individuals. We imputed wages for the self-employed and included them in the aggregate cost estimates. Since the 2003-2004 and 2009-2010 cohorts are excluded from the estimation sample, we did not calculate the aggregate costs for 2004 and 2009.

assumes they always missed a half day. We found that the annual total cost of workplace absenteeism due to depression is relatively constant over time, and ranges from 0.7 to 1.4 billion in 2009 dollars.

3.5 Sensitivity Analysis

We performed several sensitivity checks of our main results. First, we re-estimated all of our models after excluding control variables for union status, industry, occupation, employer size, and employer benefits (and wages). These variables could be endogenous if those with depression seek employment only at firms with certain characteristics. While the cross-sectional estimates of the impact of depression are sensitive to the exclusion of these variables, the CRE estimates are not. Second, we tested whether the differences between the cross-sectional and CRE estimates could be due to functional form misspecification rather than unobservables. To address this, we re-estimated cross-sectional models after including quadratic and cubic terms of all continuous variables, a full set of pairwise interaction terms, and lags of all the variables that were specified to be correlated with unobserved heterogeneity in the CRE model.[12] In the models of employment and log hourly wage, the estimated effects of depression decreased after we added the interaction and lag terms (see Table 3.6). However, the cross-sectional marginal effects are still sizeable and precisely estimated even using this highly flexible specification.

Third, we used a broader measure of mental illness, the Kessler-6 (K-6) index, to investigate whether our results are sensitive to the scale we use to measure depression. The K-6 index is a standardized and validated measure of non-specific psychological distress, and reflects a diagnosis of anxiety disorder, or mood disorder, or non-affective psychosis (Kessler et al, 2002). Individuals who scored greater or equal to 13 (out of 24) are considered to have serious mental illness.[13] We re-estimated all models using the K-6 index and report

[12]Since for our panel, we only used the observations from the second year to include lags of time-varying controls. As a result, point estimates from baseline models differed from the full sample estimates.

[13]In the MEPS, the K6 index was calculated based on questions that asked whether the respondent felt "nervous", "hopeless", "fidget or restless", "worthless", "so sad that nothing could cheer the person up",

all the results in Appendix 3.7. These estimates are consistent with those from our primary specifications using the PHQ-2, and maintain the pattern of CRE estimates of depression that are smaller than those from the cross-sectional models.

Unfortunately, there is no direct way of assessing the robustness of panel data techniques designed to account for unobserved heterogeneity without a confirmatory randomized experiment or valid instrumental variables. However, one can assess how sensitive the cross-sectional estimates are to selection on unobservables. If the cross-sectional estimates are very sensitive to unobservables, this provides strong motivation for the CRE approach and suggests the estimates are likely to be more reliable. Using the framework proposed by Altonji et al (2005) we determined whether a modest degree of negative selection into depression on unobservables could account for the whole effect of depression in the cross-sectional models. Other studies that have applied the same framework include Altonji et al. (2008), Millimet et al. (2010), and Chatterji et al. (2011). For example, Chatterji et al. (2011) demonstrated the utility of this method in measuring associations between recent psychiatric disorders and the probability of employment; a case where identifying instruments are not generally available. We describe the details of this method in the appendix.

The results from this exercise suggest that the estimates from cross-sectional models are very sensitive to selection on unobservables, and that even a modest level of negative selection into depression can account for the entire estimated effects of depression on employment and hourly wage. Although this is not a direct check of our CRE estimates, it is consistent with the significant differences we observe between the cross-sectional and CRE models when worker productivity and health status are not fully observable. If unobserved low productivity is positively correlated with depression and negatively correlated with employment and wages, we should find a smaller negative impact of depression on employment and wages in the CRE models than the naive cross-sectional models.

---

and"everything is an effort" during the past 30 days.

3.6 Discussion and Conclusions

It is difficult to empirically evaluate the disabling effects of mental health conditions like depression. The main challenge in estimation is the endogeneity of mental health measures in the structural model. In this paper, we achieved identification by accounting for time-invariant unobserved differences in productivity and health that are potentially correlated with both labor market outcomes and depression. We found that depression has a modest negative effect on employment and a larger effect on absenteeism. However, we did not find a statistically significant relationship between depression and wages or hours worked conditional on employment.

Because we used a two period panel and measures of depression that were collected contemporaneously with labor market outcomes, the effect of depression is identified off of individuals who had a change in mental health status over approximately one calendar year. As a result, our models capture the short-term effect of depression on labor market outcomes. In contrast, studies that rely on exclusion restrictions for identification estimate the long-term effect, based on cross-person variation in depression status. Such estimates reflect the impact of suffering from depression for the average duration among individuals in the estimation sample, which may well exceed one year.[14]

In the short run, the effect of depression on employment, wages, and hours worked could be small due to stickiness in the labor market (i.e. fixed labor contracts). Moreover, employers may not be able to immediately observe depression-induced changes in a workers productivity, and may adjust their wages or hours slowly over time, or may delay making a change to the workers employment status. This could explain why we found a smaller negative effect of depression on employment than previous studies that address endogeneity using instrumental variables (For example, Alexander and French (2001) reported a 19 per-

[14]Based on a sample of adults in the Netherlands, Spijker et al (2002) reported that in the general population, the mean duration of major depressive episodes was 3 months; 76 percent of the depressed individuals were able to recover within 12 months; and about 20 percent didn't recover within 24 months.

centage point reduction in the probability of employment). It may also be why we failed to find a relationship between depression and both wages and work hours.

Another plausible explanation for the smaller effect we found is the time period of our analysis. During 2004-2009 the use of mental health care increased substantially compared to the 1990s when the NCS was administered. In particular, Soni (2012) reported that between 1999 and 2009, the number of adults treated for depression increased by 74 percent, and the total health care expenditures for depression increased from 18 to 22.8 billion in 2009 dollars. Given the increasing effectiveness of depression treatment, greater use of mental health services might have dampened the disabling effects of depression in more recent years (Simon, 2002; Gibbons et al, 2012).

There are several limitations of our method. Although our identification strategy is not dependent on exclusion restrictions, it hinges on the assumption that the way we model unobserved heterogeneity fully accounts for the unobserved confounders of depression. In all of our models, we specified the random effect in the structural model to be correlated with marital status, income earned by other family members, physical health status, and depression. This is reasonable if the main confounding factors are time-invariant unobserved productivity and unobserved dimensions of physical and mental health status. We feel this is plausible in the case of health status since variation in the PHQ-2 and PCS scores will capture time-varying dimensions of mental and physical health, while age (conditional on educational attainment) should capture changes in labor market experience, and as a result, time-varying productivity. Nonetheless, if the error term has a time-varying component that is also correlated with depression, our estimates will be biased.

In addition, we note that the CRE specification does not address the endogeneity of depression due to reverse causality. This could be of a particular concern when modeling the impact of depression on employment, since involuntary job loss and subsequent unemployment can be very stressful, and has the potential to trigger an episode of depression. Previous research provides support for negative impact of job loss and unemployment on mental health status (Tefft, 2011; Marcus, 2013). The failure to account for reverse causality would lead us to over-estimate the impact of depression on employment, which suggests the short run effect of depression may be smaller than is indicated by our estimates. In contrast, we expect reverse causality to have a much smaller, if any, impact on our estimates of depression on work loss days, conditional on employment and our controls for physical health.

Another limitation is that individuals with the most severe forms of mental illnesses residing in institutions are not in the scope of the MEPS. One would expect the effect of depression on employment to be larger for the full population. Finally, our structural model captures the contemporaneous rather than the life-cycle effect of depression. While a comprehensive assessment of the impact of depression on labor market outcomes would consider both effects, the MEPS lacks information on the onset or duration of depression necessary to estimate the latter.

In the aggregate, we found that the productivity loss in the U.S. economy due to depression-induced absenteeism was between 700 million and 1.4 billion 2009 USD annually, which is less than estimates from most previous studies. For example, Greenberg et al. (2003) reported that the cost of depression from workplace absenteeism totaled 36.2 billion in 2000 (current dollars). Birnbaum et al. (2010) estimated that monthly depression-related productivity loss from both absenteeism and presenteeism was 2 billion, also in 2000 dollars. Due to methodological differences, caution is needed when comparing our estimates of aggregate costs with those from these studies. First, our aggregate cost estimates are based on the short-term effect of depression on work loss days. Second, we did not attempt to quantify the impact of depression on presenteeism, which cannot be calculated using the MEPS without making an untestable assumption about how depression lowers productivity at work. Third, previous studies may also overestimate the workplace productivity loss due to absenteeism as a result of failure to address the endogeneity of depression.

Our findings have implications, in particular, for the substantial expansions of mental health coverage under the Affordable Care Act (ACA). Garfield et al (2011) estimate that 2.0 million previously uninsured Americans with severe mental disorders would gain coverage under the individual and small group exchange provisions of the ACA.[15] Mental health and substance use services are considered essential benefits and must be covered at parity with other health services in the exchanges. In defining the services as part of the essential benefit package, the government extends mental health and substance abuse coverage well beyond the newly insured. The U.S. Department of Health and Human Services estimates that 3.9 million Americans currently covered in the individual market and 1.2 million currently in small group plans would gain mental health coverage under the essential benefit requirement (Beronio et al, 2013). A further 7.1 million currently insured Americans in the individual market and 23.3 million in the small group market with some form of mental health coverage would now be guaranteed parity coverage (Beronio et al, 2013). Our estimates suggest that some of these coverage expansions may be offset through less productivity loss.

[15] Garfield et al (2011) estimate that another 1.7 million Americans with severe mental disorders would gain coverage under the Medicaid expansions if implemented in every state.

Tables

Table 3.1: Descriptive statistics (N= 76,132; T = 2)

Variable	Mean	Standard deviation	Min	Max
Age	41.878	12.181	18	64
Age squared/100	19.021	10.259	3.24	40.96
Female	0.527	0.499	0	1
Married	0.583	0.493	0	1
White	0.672	0.469	0	1
Hispanic	0.148	0.355	0	1
Black	0.125	0.331	0	1
Other race	0.055	0.227	0	1
Urban	0.837	0.369	0	1
West	0.231	0.421	0	1
Midwest	0.224	0.417	0	1
South	0.367	0.482	0	1
Northeast	0.178	0.383	0	1
Education(years)	13.154	2.952	0	17
Number of children under 5	0.273	0.608	0	5
Number of children between 6 and17	0.573	0.941	0	9
Log family income (2009 USD)	6.953	4.448	0	12.965
Union	0.106	0.307	0	1
Employer size $<$ 25	0.226	0.418	0	1
Employer size 25-99	0.178	0.383	0	1
Employer size 100-500	0.162	0.368	0	1
Employer size $>$ 500	0.138	0.345	0	1
Sick pay	0.508	0.5	0	1
Retirement plan	0.444	0.497	0	1
Paid vacation	0.575	0.494	0	1
Industry construction & manufacturing	0.158	0.365	0	1
Industry professional & education	0.381	0.486	0	1
Industry transportation & utility	0.137	0.344	0	1
White collar occupation	0.564	0.496	0	1
PCS score	50.746	9.754	4.56	73.09
PHQ2	0.748	1.322	0	6
PHQ2 $\geq$ 3	0.091	0.287	0	1
Employed	0.75	0.433	0	1
Hourly wage (2009 USD)	15.463	14.909	0	84.632
Weekly hours worked	30.192	19.666	0	112
Annual work loss days	3.548	19.708	0	365

Notes: Means are weighted to be nationally representative.

Table 3.2: Marginal effects of depression on employment

	Cross-section	Fixed-effects	CRE
PHQ-2 index	-0.042***	-0.012***	-0.011***
	(0.001)	(0.002)	(0.002)
Depression (PHQ-2 ≥ 3)	-0.176***	-0.033***	-0.026***
	(0.008)	(0.007)	(0.007)

Notes: Significance level: ***p<0.01, **p<0.05, *p<0.1. Standard errors in parentheses are adjusted for the complex survey design of the MEPS using balanced repeated replication. Models include cross-sectional and CRE probit regressions and linear fixed-effects regressions. Control variables include age and its square, gender, race and ethnicity, marital status, years of education completed before entering the survey, region, urban residence, number of children under 5 or 18 in the household, log of income earned by other family members (normalized by household size), PCS score, and year dummies.

Table 3.3: Marginal effects of depression on work hours and wages

Panel A	Weekly work hours			Log wage		
	Cross-section	Fixed-effects	CRE	Cross-section	Fixed-effects	CRE
PHQ-2 index	-0.031	0.007	0.006	-0.021***	-0.002	-0.004
	(0.055)	(0.034)	(0.033)	(0.003)	(0.002)	(0.003)
Depression (PHQ-2 $\geq$ 3)	0.053	-0.023	-0.053	-0.083***	-0.006	-0.004
	(0.259)	(0.155)	(0.163)	(0.012)	(0.008)	(0.012)

Panel B: Weekly work hours	Cross-section			CRE		
	Less than 30 hours	Between 30 and 40 hours	More than 40 hours	Less than 30 hours	Between 30 and 40 hours	More than 40 hours
Depression (PHQ-2 $\geq$ 3)	0.010***	0.005***	-0.015***	0.004	0.003	-0.006
	(0.003)	(0.001)	(0.005)	(0.002)	(0.002)	(0.004)

Notes: Significance level: ***p<0.01, **p<0.05, *p<0.1. Standard errors in parentheses are adjusted for the complex survey design of the MEPS using balanced repeated replication. Models include cross-sectional and CRE OLS and ordered probit regressions as well as linear fixed-effects regressions. Control variables include age and its square, gender, race and ethnicity, marital status, years of education completed before entering the survey, region, urban residence, union status, employer size (less than 25 employees, between 25-99 employees, between 100-500 employees, more than 500 employees), benefits provided by the employer (sick pay, retirement plan, and paid vacations), occupation and industry indicators, number of children under 5 or 18 in the household, log of income earned by other family members (normalized by household size), PCS score, and year dummies.

Table 3.4: Marginal effect of depression on work loss days

	Cross-section	CRE
PHQ-2 index	0.926***	0.567***
	(0.081)	(0.083)
Depression (PHQ-2 $\geq$ 3)	3.359***	1.382***
	(0.498)	(0.433)

Notes: Significance level: ***p<0.01, **p<0.05, *p<0.1. Standard errors in parentheses are adjusted for the complex survey design of the MEPS using balanced repeated replication. Models include cross-sectional and CRE zero-inflated ordered probit regressions. Control variables include age and its square, gender, race and ethnicity, marital status, years of education completed before entering the survey, region, urban residence, union status, employer size (less than 25 employees, between 25-99 employees, between 100-500 employees, more than 500 employees), benefits provided by the employer (sick pay, retirement plan, and paid vacations), occupation and industry indicators, number of children under 5 or 18 in the household, log of income earned by other family members (normalized by household size), and year dummies.

Table 3.5: Annual cost of depression-induced absenteeism for employed adults aged 18-64 (Billions of 2009 USD; 90% C.I. in brackets)

Year	Total employed population aged 18-64 (millions)	Total Population (full day estimate)	Total population (half day estimate)
2005	135.3	1.41 [0.66, 2.15]	0.7 [0.33, 1.07]
2006	136.5	1.35 [0.63, 2.06]	0.67 [0.32, 1.03]
2007	139.5	1.38 [0.63, 2.12]	0.69 [0.32, 1.06]
2008	138.7	1.43 [0.67, 2.19]	0.71 [0.33, 1.09]
2005-2008 average	137.6	1.39 [0.67, 2.11]	0.69 [0.33, 1.06]

Notes: All confidence intervals are adjusted for the complex survey design of the MEPS. The population estimates are obtained using the MEPS sampling weights. In order to obtain national representative estimates, we adjusted the sampling weights to account for sample attrition of the panel.

Table 3.6: Marginal effects from cross-sectional models using more flexible specifications

	(1)	(2)	(3)
Employment	-0.183***	-0.155***	-0.107***
	(0.010)	(0.010)	(0.010)
Hour	-0.189	-0.241	-0.397
	(0.355)	(0.362)	(0.368)
Log wage	-0.075***	-0.062***	-0.041**
	(0.015)	(0.014)	(0.015)
Higher order and interactions terms	No	Yes	Yes
Lags of correlated variables	No	No	Yes

Notes: Significance level: ***p<0.01, **p<0.05, *p<0.1; Standard errors in parentheses are adjusted for the complex survey design of the MEPS using balanced repeated replication. Models include cross-sectional probit and OLS regressions. Control variables include age and its square, gender, race and ethnicity, marital status, years of education completed before entering the survey, region, urban residence, union status, employer size (less than 25 employees, between 25-99 employees, between 100-500 employees, more than 500 employees), benefits provided by the employer (sick pay, retirement plan, and paid vacations), occupation and industry indicators, number of children under 5 or 18 in the household, log of income earned by other family members (normalized by household size), and year dummies.

3.7 Appendix

We describe in details how we implemented the method outlined in Altonji et al (2005). Intuitively, the degree of selection on observables can be used as a guide to the degree of selection on unobservables. In the case of binary outcome variable, consider the following bivariate probit model:

$$Y = 1(X'\gamma + \alpha D + \varepsilon > 0), \tag{3.17}$$

$$D = 1(X'\beta + u > 0), \tag{3.18}$$

where $1(\cdot)$ is an indicator function, and it is assumed that u and ε follow a standard bivariate normal distribution with correlation ρ. In the absence of valid exclusion restrictions, identification of the bivariate probit model is achieved through the bivariate normality assumption. Let Y^* be a latent variable underlying outcome Y such that

$$Y^* = \alpha D + W'\Gamma, \tag{3.19}$$

where W is the full set of variables that determine Y^* and D is a binary indicator for depression. In practice, we only observe a subset X of W, so we rewrite (3.19) as

$$Y^* = \alpha D + X'\gamma + \varepsilon. \tag{3.20}$$

Note that in (3.20) γ and ε are defined so that $\mathrm{Cov}(X, \varepsilon) = 0$. Similarly, let D^* be the latent variable underlying D and the linear projection of D^* onto $X'\gamma$ and ε is

$$\mathrm{E}(D^*|X'\gamma, \varepsilon) = \varphi_0 + \varphi_1 X'\gamma + \varphi_2 \varepsilon. \tag{3.21}$$

Then the assumption that selection on unobservables is the same as selection on observables can be re-expressed as $\varphi_1 = \varphi_2$. In the bivariate probit framework, this condition can be replaced by

$$\rho^* = \frac{\mathrm{Cov}(X'\beta, X'\gamma)}{\mathrm{Var}(X'\gamma)} \tag{3.22}$$

Setting $\rho = 0$ generates the cross-sectional estimate of α, and setting $\rho = \rho^*$ generates the lower bound of α (Altonji et al, 2005). We followed a two-step procedure to approximate the value of that satisfies equation (3.22). First, β and γ are estimated from an unconstrained bivariate probit model, and then α is estimated from a constrained bivariate probit model imposing the estimated value of ρ^* from the first step.

We set ρ from 0 to -0.4, ρ^* (-0.52 for our application) and $\rho^*/2$ (-0.26), and estimated the constrained bivariate probit models on employment. The coefficient estimates and marginal effects of the dichotomous PHQ-2 index are reported in Table 3.11. The negative effect of depression on employment decreases dramatically as negative selection on the unobservables gets stronger. When ρ is set to -0.3, the coefficient estimate becomes positive and the negative effect of depression goes away. This provides evidence that the cross-sectional estimate is very sensitive to selection on unobservables, and that even a modest level of negative selection on unobservables can reverse the direction of the estimated effect. Under the assumption that selection along unmeasured factors is equal to selection along measured factors ($\rho = \rho^*$), the coefficient estimate is positive and statistically significant. Under this condition the elements of X are just a random subset of W, and are no more useful than the unobserved elements of W in terms of reducing the bias in $\hat{\alpha}$ (Altonji et al, 2005). However, this assumption is likely to be too strong given that our models contain a large number of socio-demographic controls, some of which (such as education and age) are known to be strong determinants of employment.

Given the significant predictive power of the observables on employment, it is reasonable to assume in our application that selection on unobservables is weaker than selection on observables. The last column of Table 3.11 contains the marginal effect of depression on employment under the constraint: $\rho = 0.5\rho^*$. In this case the marginal effect is slightly smaller than that from our CRE model. Although it is impossible to know the true value of ρ, this suggests that the CRE specification generates estimates that are consistent with a moderate degree of selection on unobservables.

In the case of continuous outcomes such as wages and hours worked, Altonji et al (2005) provide a different method for assessing the role of unobservables. Here, the relevant question is: Relative to the degree of selection on observables, how large should the selection on unobservables be, to explain away the whole effect of depression? We write the linear projection of D onto X as

$$D = X'\beta + \tilde{D}. \tag{3.23}$$

By substituting (3.23) into (3.20), we have

$$Y^* = \alpha\tilde{D}\tilde{X}'(\alpha\beta + \gamma) + \varepsilon. \tag{3.24}$$

Since $\tilde{D}$ is orthogonal to X by construction, the probability limit of the OLS estimator of α can be written as

$$\begin{aligned}\text{plim}\hat{\alpha} &= \alpha + \frac{\text{Cov}(\tilde{D}, \varepsilon)}{\text{Var}(\tilde{D})} \\ &= \alpha + \frac{\text{Var}(D)}{\text{Var}(\tilde{D})}\left[E(\varepsilon|D = 1) - E(\varepsilon|D = 0)\right].\end{aligned} \tag{3.25}$$

The bias term in (3.25) is estimated under the assumption that the normalized degree of selection on observables is equal to the normalized degree of selection on unobservables. More formally, this is equivalent to

$$\frac{E(\varepsilon|D = 1) - E(\varepsilon|D = 0)}{\text{Var}(\varepsilon)} = \frac{E(X'\gamma|D = 1) - E(X'\gamma|D = 0)}{\text{Var}(X'\gamma)} \tag{3.26}$$

Under the null hypothesis that depression has no effect ($\alpha = 0$), it is possible to obtain a consistent estimate of γ in (3.24). In this case, the term $E(\varepsilon|D = 1) - E(\varepsilon|D = 0)$ can be estimated using the variance of residual $\hat{\varepsilon}$ under (3.26). With the sample analog of $\text{Var}(D)$

and $\mathrm{Var}(\tilde{D})$ we can consistently estimate the bias term in (3.25). The ratio between the unconstrained estimate of $\hat{\alpha}$ and the bias term

$$\frac{\hat{\alpha}}{[\mathrm{Var}(D)/\mathrm{Var}(\tilde{D})]E(\varepsilon|D=1)-E(\varepsilon|D=0)} \tag{3.27}$$

is used to gauge how large selection on unobservables must be relative to selection on observables to fully account for the effect of depression. If the ratio is less than 1, then the effect of depression in the OLS model is likely to be biased because it suggests the observed effect of depression could result from a relatively small degree of selection along unobservable dimensions. We conducted this sensitivity analysis only for the hourly wage outcome because depression is not associated with hours worked in either the cross-sectional or panel data model. Our estimates indicate that the normalized shift in the distribution of unobservables only needs to be around 30 percent of the normalized shift in the observables to explain away the entire effect of depression.

Table 3.7: Correlated random effect probit models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	K-6 index	PHQ-2 index	K-6 ≥13	PHQ-2≥3	K-6 index	PHQ-2 index	K-6≥13	PHQ-2≥3
Constant	-3.111***	-3.242***	-3.448***	-3.396***	-3.296***	-3.420***	-3.674***	-3.588***
	(0.127)	(0.128)	(0.126)	(0.126)	(0.131)	(0.129)	(0.129)	(0.127)
Age	0.131***	0.131***	0.128***	0.127***	0.132***	0.133***	0.130***	0.129***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Age squared/100	-0.167***	-0.167***	-0.163***	-0.162***	-0.168***	-0.168***	-0.164***	-0.163***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Female	-0.418***	-0.422***	-0.431***	-0.430***	-0.409***	-0.413***	-0.423***	-0.422***
	(0.018)	(0.018)	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)
Hispanic	-0.047*	-0.043*	-0.038	-0.04	-0.057**	-0.052**	-0.046*	-0.049*
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Black	-0.107***	-0.088***	-0.088***	-0.077***	-0.106***	-0.084***	-0.085***	-0.072***
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.026)	(0.025)	(0.026)
Other race	-0.100**	-0.089**	-0.092**	-0.081**	-0.101***	-0.089**	-0.093**	-0.078**
	(0.040)	(0.039)	(0.039)	(0.039)	(0.039)	(0.038)	(0.038)	(0.038)
Urban	0.050*	0.049*	0.049*	0.049*	0.043*	0.041*	0.041*	0.041*
	(0.026)	(0.025)	(0.025)	(0.025)	(0.024)	(0.024)	(0.024)	(0.024)
West	-0.033	-0.037	-0.04	-0.038	-0.031	-0.035	-0.038	-0.036
	(0.030)	(0.030)	(0.030)	(0.030)	(0.035)	(0.035)	(0.035)	(0.035)
Midwest	0.069**	0.063**	0.064**	0.064**	0.070**	0.062*	0.064*	0.065*
	(0.032)	(0.031)	(0.031)	(0.031)	(0.035)	(0.035)	(0.034)	(0.035)
South	0.005	0.005	0.002	0.002	0.011	0.012	0.009	0.009
	(0.027)	(0.027)	(0.026)	(0.026)	(0.032)	(0.032)	(0.033)	(0.032)
Education(years)	0.051***	0.051***	0.053***	0.052***	0.047***	0.047***	0.049***	0.048***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Number of children under 5	-0.232***	-0.228***	-0.228***	-0.226***	-0.240***	-0.236***	-0.237***	-0.234***
	(0.014)	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)
Number of children between 6-17	-0.080***	-0.079***	-0.078***	-0.078***	-0.083***	-0.082***	-0.081***	-0.080***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Married	0.152***	0.151***	0.167***	0.166***	0.026	0.025	0.027	0.026
	(0.022)	(0.022)	(0.022)	(0.022)	(0.049)	(0.049)	(0.050)	(0.050)

Log family income	-0.028***	-0.027***	-0.026***	-0.026***	-0.006*	-0.006*	-0.006*	-0.006*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
PCS score	0.029***	0.030***	0.032***	0.032***	0.002*	0.002*	0.002**	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Depression measure	-0.050***	-0.160***	-0.720***	-0.574***	-0.012***	-0.037***	-0.119***	-0.091***
	(0.002)	(0.006)	(0.029)	(0.024)	(0.002)	(0.005)	(0.029)	(0.022)
Correlation Parameters								
Married period 1					-0.055	-0.051	-0.048	-0.046
					(0.048)	(0.048)	(0.050)	(0.048)
Married period 2					0.191***	0.187***	0.198***	0.196***
					(0.051)	(0.050)	(0.051)	(0.050)
Log family income period 1					-0.008**	-0.008**	-0.008**	-0.008**
					(0.003)	(0.003)	(0.003)	(0.003)
Log family income period 2					-0.018***	-0.018***	-0.017***	-0.017***
					(0.004)	(0.004)	(0.004)	(0.004)
PCS score period 1					0.018***	0.019***	0.020***	0.020***
					(0.001)	(0.001)	(0.001)	(0.001)
PCS score period 2					0.014***	0.014***	0.016***	0.015***
					(0.001)	(0.001)	(0.001)	(0.001)
Depression measure period 1					-0.024***	-0.080***	-0.448***	-0.346***
					(0.003)	(0.008)	(0.037)	(0.032)
Depression measure period 2					-0.022***	-0.072***	-0.353***	-0.320***
					(0.003)	(0.008)	(0.044)	(0.033)

Notes: The coefficients of the year dummies are omitted. Significance level: ***p<0.01, **p<0.05, *p<0.1. Standard errors in parentheses are adjusted for the complex survey design of the MEPS using balanced repeated replication. Base categories are white and northeast.

Table 3.8: CRE linear regression models of weekly work hours and hourly wages

	Weekly hours worked				Log hourly wage			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	K-6 index	PHQ-2 index	K-6≥13	PHQ-2≥3	K-6 index	PHQ-2 index	K-6≥13	PHQ-2≥3
Constant	25.345***	25.553***	25.486***	25.459***	0.470***	0.439***	0.401***	0.422***
	(1.298)	(1.312)	(1.304)	(1.308)	(0.071)	(0.069)	(0.068)	(0.068)
Age	0.238***	0.239***	0.238***	0.238***	0.040***	0.040***	0.040***	0.040***
	(0.050)	(0.051)	(0.051)	(0.050)	(0.003)	(0.003)	(0.003)	(0.003)
Age squared/100	-0.320***	-0.321***	-0.321***	-0.320***	-0.042***	-0.041***	-0.041***	-0.041***
	(0.059)	(0.059)	(0.059)	(0.059)	(0.003)	(0.003)	(0.003)	(0.003)
Female	-3.402***	-3.397***	-3.399***	-3.400***	-0.198***	-0.200***	-0.202***	-0.201***
	(0.160)	(0.161)	(0.160)	(0.161)	(0.009)	(0.009)	(0.009)	(0.009)
Hispanic	1.084***	1.072***	1.074***	1.073***	-0.062***	-0.061***	-0.059***	-0.060***
	(0.194)	(0.195)	(0.195)	(0.196)	(0.015)	(0.015)	(0.015)	(0.015)
Black	-0.208	-0.225	-0.222	-0.227	-0.120***	-0.115***	-0.114***	-0.114***
	(0.193)	(0.194)	(0.195)	(0.195)	(0.010)	(0.010)	(0.010)	(0.010)
Other race	0.151	0.148	0.148	0.141	-0.024	-0.022	-0.022	-0.021
	(0.341)	(0.313)	(0.312)	(0.313)	(0.018)	(0.018)	(0.018)	(0.018)
Urban	-0.694***	-0.691**	-0.693***	-0.690**	0.157***	0.156***	0.156***	0.156***
	(0.244)	(0.245)	(0.244)	(0.246)	(0.015)	(0.015)	(0.015)	(0.015)
Union	-0.762***	-0.764***	-0.763***	-0.759***	0.087***	0.087***	0.087***	0.087***
	(0.255)	(0.255)	(0.255)	(0.256)	(0.014)	(0.014)	(0.014)	(0.014)
West	0.530**	0.534**	0.534**	0.533**	0.022	0.022	0.022	0.022
	(0.246)	(0.247)	(0.247)	(0.247)	(0.016)	(0.016)	(0.016)	(0.016)
Midwest	0.705***	0.708***	0.706***	0.709***	-0.057***	-0.059***	-0.058***	-0.058***
	(0.226)	(0.226)	(0.226)	(0.266)	(0.014)	(0.014)	(0.014)	(0.014)
South	1.675***	1.677***	1.676***	1.674***	-0.063***	-0.063***	-0.063***	-0.063***
	(0.220)	(0.220)	(0.221)	(0.220)	(0.014)	(0.014)	(0.014)	(0.014)
Education(years)	-0.068**	-0.069**	-0.068**	-0.069**	0.071***	0.071***	0.071***	0.071***
	(0.032)	(0.032)	(0.032)	(0.032)	(0.002)	(0.002)	(0.002)	(0.002)
Number of children under 5	-0.044	-0.045	-0.046	-0.046	0.020**	0.020**	0.021**	0.021**
	(0.121)	(0.121)	(0.121)	(0.121)	(0.009)	(0.009)	(0.009)	(0.009)
Number of children between 6-17	-0.261***	-0.262***	-0.262***	-0.262***	-0.005	-0.005	-0.005	-0.005

	(0.084)	(0.084)	(0.084)	(0.084)	(0.004)	(0.004)	(0.004)	(0.004)
Employer size < 25	0.188	0.191	0.189	0.191	-0.120***	-0.121***	-0.119***	-0.121***
	(0.231)	(0.232)	(0.231)	(0.231)	(0.011)	(0.011)	(0.011)	(0.011)
Employer size 25-99	0.921***	0.920***	0.922***	0.919***	-0.118***	-0.119***	-0.118***	-0.119***
	(0.210)	(0.210)	(0.210)	(0.209)	(0.012)	(0.012)	(0.012)	(0.012)
Employer size 100-500	0.293	0.296	0.298	0.294	-0.071***	-0.072***	-0.072***	-0.072***
	(0.220)	(0.200)	(0.200)	(0.201)	(0.011)	(0.011)	(0.011)	(0.011)
Sick pay	2.113***	2.116***	2.107***	2.114***	0.142***	0.143***	0.144***	0.144***
	(0.203)	(0.202)	(0.203)	(0.203)	(0.012)	(0.012)	(0.012)	(0.012)
Retirement plan	2.011***	2.012***	2.010***	2.021***	0.214***	0.214***	0.215***	0.214***
	(0.188)	(0.188)	(0.188)	(0.188)	(0.010)	(0.010)	(0.010)	(0.010)
Paid vacation	3.909***	3.898***	3.909***	3.895***	0.064***	0.065***	0.065***	0.065***
	(0.248)	(0.248)	(0.248)	(0.249)	(0.013)	(0.013)	(0.013)	(0.013)
Industry construction & manufacturing	1.974***	1.975***	1.978***	1.978***	0.121***	0.121***	0.121***	0.122***
	(0.260)	(0.260)	(0.259)	(0.262)	(0.015)	(0.015)	(0.015)	(0.015)
Industry professional & education	-0.103	-0.098	-0.101	-0.097	0.003	0.002	0.001	0.002
	(0.222)	(0.221)	(0.222)	(0.222)	(0.014)	(0.014)	(0.014)	(0.014)
Industry transportation & utility	0.995***	0.997***	0.997***	0.999***	-0.033**	-0.033**	-0.034**	-0.033**
	(0.268)	(0.267)	(0.267)	(0.267)	(0.014)	(0.014)	(0.014)	(0.013)
White collar occupations	-0.942***	-0.938***	-0.937***	-0.934***	0.067***	0.066***	0.066***	0.066***
	(0.222)	(0.221)	(0.220)	(0.222)	(0.012)	(0.012)	(0.012)	(0.012)
Log wage	0.761*	0.763*	0.762*	0.725*				
	(0.427)	(0.434)	(0.429)	(0.432)				
Married	0.164	0.156	0.164	0.13	-0.009	-0.008	-0.008	-0.009
	(0.223)	(0.228)	(0.223)	(0.238)	(0.014)	(0.014)	(0.014)	(0.014)
Log family income	-0.013	-0.013	-0.013	-0.012	0.000	0.000	-0.001	0.000
	(0.014)	(0.014)	(0.014)	(0.014)	(0.002)	(0.002)	(0.002)	(0.002)
PCS score	-0.01	-0.01	-0.010*	-0.01	0.000	0.000	0.000	0.000
	(0.006)	(0.006)	(0.006)	(0.007)	(0.000)	(0.000)	(0.000)	(0.000)
Depression index	-0.009	0.006	-0.148	-0.053	0.000	-0.004	-0.003	-0.004
	(0.012)	(0.033)	(0.185)	(0.163)	(0.001)	(0.003)	(0.014)	(0.012)
Correlation parameters								
Log wage period 1	-0.021	-0.027	-0.016	-0.001				

	(0.398)	(0.402)	(0.398)	(0.400)				
Log wage period 2	1.802***	1.799***	1.793***	1.813***				
	(0.522)	(0.522)	(0.521)	(0.525)				
Married period 1	-0.746**	-0.744**	-0.745**	-0.739**	0.024	0.024	0.023	0.024
	(0.338)	(0.338)	(0.337)	(0.338)	(0.025)	(0.025)	(0.025)	(0.025)
Married period 2	0.800**	0.796**	0.794**	0.823**	0.079***	0.078***	0.081***	0.080***
	(0.364)	(0.361)	(0.364)	(0.365)	(0.024)	(0.025)	(0.025)	(0.025)
Log family income period 1	-0.104***	-0.103***	-0.103***	-0.104***	-0.002	-0.002	-0.002	-0.002
	(0.029)	(0.029)	(0.029)	(0.029)	(0.002)	(0.002)	(0.002)	(0.002)
Log family income period 2	-0.004	-0.005	-0.005	-0.005	0.000	0.000	0.000	0.000
	(0.027)	(0.026)	(0.027)	(0.027)	(0.002)	(0.002)	(0.002)	(0.002)
PCS score period 1	0.005	0.005	0.005	0.005	0.003***	0.003***	0.003***	0.003***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.001)	(0.001)	(0.001)	(0.001)
PCS score period 2	0.027**	0.025**	0.025**	0.026**	0.002**	0.002**	0.002***	0.002***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.001)	(0.001)	(0.001)	(0.001)
Depression index period 1	-0.003	0.038	0.221	0.46	-0.006***	-0.014***	-0.068***	-0.069***
	(0.024)	(0.074)	(0.495)	(0.343)	(0.001)	(0.004)	(0.024)	(0.019)
Depression index period 2	0.037	-0.034	0.392	-0.006	-0.004***	-0.011**	-0.070***	-0.055***
	(0.031)	(0.099)	(0.643)	(0.413)	(0.001)	(0.004)	(0.025)	(0.017)

Notes: The coefficients of the year dummies are omitted. Significance level: ***p<0.01, **p<0.05, *p<0.1. Standard errors in parentheses are adjusted for the complex survey design of the MEPS using balanced repeated replication. Base categories are white, northeast, employer size > 500, and other services/public administration/military/unclassifiable industry.

Table 3.9: CRE ordered probit model models of work hour categories

	(1)	(2)	(3)	(4)
	K-6 index	PHQ-2 index	K-6≥13	PHQ-2≥3
Constant	-0.248	-0.244	-0.258	-0.237
	(0.208)	(0.205)	(0.206)	(0.203)
Age	0.043***	0.043***	0.042***	0.043***
	(0.007)	(0.008)	(0.007)	(0.007)
Age squared/100	-0.054***	-0.054***	-0.054***	-0.055***
	(0.009)	(0.009)	(0.009)	(0.009)
Female	-0.576***	-0.577***	-0.577***	-0.576***
	(0.026)	(0.027)	(0.026)	(0.026)
Hispanic	0.264***	0.264***	0.263***	0.264***
	(0.039)	(0.039)	(0.039)	(0.039)
Black	0.017	0.016	0.017	0.016
	(0.035)	(0.035)	(0.035)	(0.035)
Other race	0.239***	0.239***	0.239***	0.239***
	(0.058)	(0.058)	(0.058)	(0.058)
Urban	-0.03	-0.03	-0.031	-0.03
	(0.038)	(0.037)	(0.037)	(0.037)
Union	-0.140***	-0.140***	-0.140***	-0.140***
	(0.040)	(0.040)	(0.040)	(0.040)
West	0.150***	0.150***	0.151***	0.150***
	(0.039)	(0.039)	(0.039)	(0.039)
Midwest	0.129***	0.129***	0.129***	0.129***
	(0.036)	(0.036)	(0.036)	(0.036)
South	0.336***	0.336***	0.336***	0.336***
	(0.036)	(0.036)	(0.036)	(0.036)
Education(years)	-0.022***	-0.022***	-0.022***	-0.022***
	(0.005)	(0.005)	(0.005)	(0.005)
Number of children under 5	0.004	0.004	0.004	0.004
	(0.022)	(0.022)	(0.022)	(0.022)
Number of children between 6-17	-0.028**	-0.028**	-0.028**	-0.028**
	(0.012)	(0.012)	(0.012)	(0.012)
Employer size < 25	-0.067**	-0.067**	-0.067**	-0.067**
	(0.031)	(0.031)	(0.031)	(0.031)
Employer size 25-99	0.074**	0.073**	0.074**	0.073**
	(0.032)	(0.032)	(0.032)	(0.032)
Employer size 100-500	0.045	0.045	0.045	0.046
	(0.031)	(0.031)	(0.031)	(0.032)
Sick pay	0.396***	0.395***	0.396***	0.396***
	(0.036)	(0.036)	(0.036)	(0.036)
Retirement plan	0.318***	0.318***	0.318***	0.318***
	(0.029)	(0.029)	(0.029)	(0.029)
Paid vacation	0.626***	0.625***	0.625***	0.625***

	(0.032)	(0.032)	(0.032)	(0.031)
Industry construction & manufacturing	0.519***	0.518***	0.520***	0.518***
	(0.054)	(0.055)	(0.054)	(0.055)
Industry professional & education	-0.183***	-0.183***	-0.182***	-0.183***
	(0.040)	(0.040)	(0.040)	(0.040)
Industry transportation & utility	-0.072	-0.072	-0.072	-0.072
	(0.045)	(0.045)	(0.045)	(0.045)
White collar occupation	-0.163***	-0.163***	-0.162***	-0.163***
	(0.039)	(0.039)	(0.039)	(0.039)
Log wage	0.161**	0.161**	0.161**	0.162**
	(0.066)	(0.067)	(0.067)	(0.067)
Married	0.065*	0.065*	0.065*	0.064*
	(0.036)	(0.036)	(0.036)	(0.036)
Log family income	-0.004	-0.004	-0.004	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)
PCS score	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Depression measure	-0.002	-0.001	-0.032	-0.027
	(0.002)	(0.005)	(0.032)	(0.022)
Cutoffs				
μ_1	0.723***	0.723***	0.723***	0.723***
	(0.015)	(0.015)	(0.015)	(0.015)
Correlation parameters				
Log wage period 1	-0.044	-0.044	-0.044	-0.044
	(0.050)	(0.050)	(0.050)	(0.050)
Log wage period 2	0.159**	0.159**	0.161**	0.158**
	(0.068)	(0.069)	(0.068)	(0.068)
Married period 1	-0.140**	-0.141**	-0.142**	-0.141**
	(0.071)	(0.071)	(0.071)	(0.071)
Married period 2	0.056	0.057	0.058	0.057
	(0.072)	(0.072)	(0.072)	(0.072)
Log family income period 1	-0.010**	-0.010**	-0.010**	-0.010**
	(0.005)	(0.005)	(0.005)	(0.005)
Log family income period 2	0.008*	0.008*	0.008*	0.008*
	(0.004)	(0.004)	(0.004)	(0.004)
PCS score period 1	-0.001	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
PCS score period 2	0.003	0.003*	0.003*	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Depression measure period 1	-0.002	-0.007	0.083	-0.03
	(0.003)	(0.010)	(0.066)	(0.041)
Depression measure period 2	0.004	0.007	0.005	0.031

	(0.004)	(0.014)	(0.087)	(0.059)

Notes: The coefficients of the year dummies are omitted. Significance level: ***p<0.01, **p<0.05, *p<0.1; Standard errors in parentheses are adjusted for the complex survey design of the MEPS using balanced repeated replication. Base categories are white, northeast, employer size > 500, and other services/public administration/military/unclassifiable industry.

Table 3.10: CRE zero-inflated ordered probit models of work loss days

	K-6 index		PHQ-2 index		K-6≥13		PHQ-2≥3	
	Inflation	Ordered probit	Inflation	Ordered probit	Inflation	Ordered probit	Inflation	Ordered probit
Constant	0.006	3.230***	0.112	3.335***	0.440	3.562***	0.207	3.497***
	(0.500)	(0.599)	(0.465)	(0.523)	(0.535)	(0.531)	(0.464)	(0.507)
Age	0.010	-0.022	0.010	-0.024	0.011	-0.022	0.013	-0.022
	(0.012)	(0.016)	(0.012)	(0.016)	(0.011)	(0.016)	(0.012)	(0.016)
Age squared/100	-0.025	0.029*	-0.026*	0.031*	-0.025*	0.028	-0.029**	0.029
	(0.015)	(0.018)	(0.014)	(0.018)	(0.014)	(0.017)	(0.014)	(0.018)
Female	0.182***	0.175***	0.191***	0.168***	0.209***	0.168**	0.197***	0.175***
	(0.059)	(0.067)	(0.053)	(0.060)	(0.052)	(0.066)	(0.048)	(0.056)
Hispanic	-0.247***	-0.006	-0.248***	-0.021	-0.250***	-0.025	-0.254***	-0.018
	(0.053)	(0.059)	(0.055)	(0.062)	(0.046)	(0.059)	(0.051)	(0.063)
Black	-0.224***	0.108	-0.239***	0.090	-0.223***	0.086	-0.238***	0.070
	(0.073)	(0.078)	(0.064)	(0.085)	(0.064)	(0.079)	(0.061)	(0.085)
Other race	-0.216*	-0.088	-0.217*	-0.102	-0.223	-0.080	-0.213*	-0.097
	(0.111)	(0.152)	(0.121)	(0.147)	(0.111)	(0.151)	(0.113)	(0.141)
Union	-0.044	0.160	-0.043	0.161	-0.028	0.163**	-0.040	0.158
	(0.074)	(0.102)	(0.073)	(0.107)	(0.071)	(0.080)	(0.067)	(0.098)
Urban	0.083	0.007	0.082	0.013	0.074	0.015	0.080	0.021
	(0.061)	(0.064)	(0.059)	(0.061)	(0.054)	(0.060)	(0.060)	(0.060)
Education(years)	0.035**	-0.036***	0.033**	-0.036***	0.027	-0.036***	0.033*	-0.039***
	(0.015)	(0.010)	(0.017)	(0.010)	(0.018)	(0.011)	(0.017)	(0.010)
Number of children under 5	-0.086	0.212***	-0.092	0.210***	-0.059	0.188***	-0.088	0.207***
	(0.101)	(0.071)	(0.096)	(0.060)	(0.093)	(0.069)	(0.093)	(0.059)
Number of children between 6-17	-0.027	-0.016	-0.026	-0.020	-0.026	-0.016	-0.025	-0.021
	(0.022)	(0.028)	(0.025)	(0.027)	(0.020)	(0.026)	(0.024)	(0.028)
West	0.116**	-0.028	0.115**	-0.026	0.103*	-0.025	0.123**	-0.033
	(0.058)	(0.074)	(0.056)	(0.069)	(0.056)	(0.065)	(0.056)	(0.068)
Midwest	0.089	-0.038	0.098	-0.038	0.087	-0.038	0.102	-0.045
	(0.069)	(0.066)	(0.066)	(0.062)	(0.065)	(0.063)	(0.068)	(0.064)

South	0.113	-0.066	0.116*	-0.071	0.097	-0.062	0.119*	-0.070
	(0.070)	(0.056)	(0.065)	(0.055)	(0.072)	(0.056)	(0.067)	(0.057)
Employer size < 25	-0.024	-0.126**	-0.029	-0.115*	-0.043	-0.113**	-0.032	-0.110*
	(0.059)	(0.063)	(0.055)	(0.065)	(0.052)	(0.056)	(0.053)	(0.060)
Employer size 25-99	-0.022	-0.071	-0.027	-0.055	-0.026	-0.070	-0.032	-0.050
	(0.047)	(0.056)	(0.051)	(0.059)	(0.044)	(0.056)	(0.050)	(0.061)
Employer size 100-500	0.080	-0.003	0.074	0.013	0.074	0.006	0.075	0.014
	(0.056)	(0.057)	(0.060)	(0.060)	(0.053)	(0.058)	(0.057)	(0.058)
Sick pay	0.142*	0.025	0.128*	0.037	0.109	0.027	0.120*	0.030
	(0.075)	(0.109)	(0.068)	(0.110)	(0.067)	(0.093)	(0.063)	(0.100)
Retirement plan	0.057	-0.004	0.066	-0.010	0.051	-0.013	0.061	-0.014
	(0.049)	(0.072)	(0.054)	(0.073)	(0.045)	(0.067)	(0.054)	(0.072)
Paid vacation	0.146***	0.040	0.139**	0.038	0.141***	0.030	0.143**	0.031
	(0.055)	(0.093)	(0.059)	(0.092)	(0.049)	(0.083)	(0.058)	(0.086)
Industry construction & manufacturing	-0.073	-0.193	-0.066	-0.188	-0.096	-0.168	-0.057	-0.210*
	(0.153)	(0.158)	(0.147)	(0.153)	(0.153)	(0.155)	(0.148)	(0.143)
Industry professional & education	-0.077	-0.108	-0.070	-0.105	-0.087	-0.092	-0.058	-0.123
	(0.098)	(0.117)	(0.102)	(0.117)	(0.103)	(0.126)	(0.103)	(0.110)
Industry transportation & utility	-0.184***	-0.024	-0.179**	-0.025	-0.166***	-0.029	-0.164**	-0.041
	(0.069)	(0.095)	(0.070)	(0.100)	(0.060)	(0.093)	(0.065)	(0.093)
White collar occupation	0.172	-0.204***	0.175*	-0.196***	0.147	-0.190***	0.177*	-0.197***
	(0.121)	(0.067)	(0.104)	(0.061)	(0.117)	(0.063)	(0.099)	(0.059)
Married	0.094	0.119	0.099	0.113	0.078	0.114	0.108	0.103
	(0.121)	(0.082)	(0.124)	(0.078)	(0.115)	(0.074)	(0.116)	(0.079)
Log family income	-0.007	0.004	-0.007	0.004	-0.005	0.003	-0.007	0.004
	(0.008)	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)
PCS score	-0.018**	-0.035***	-0.018**	-0.035***	-0.021**	-0.033***	-0.019**	-0.035***
	(0.006)	(0.005)	(0.008)	(0.005)	(0.008)	(0.006)	(0.007)	(0.005)
Depression measure	0.018	0.028***	0.040	0.090***	0.109	0.325***	0.043	0.222**
	(0.013)	(0.009)	(0.036)	(0.030)	(0.116)	(0.093)	(0.120)	(0.092)
Cutoffs								
μ_1	0.609**		0.600**		0.674*		0.603***	

	(0.298)	(0.240)	(0.345)	(0.226)
μ_2	0.926**	0.914***	1.008**	0.917***
	(0.369)	(0.302)	(0.421)	(0.284)
μ_3	1.142***	1.128***	1.231***	1.131***
	(0.403)	(0.333)	(0.457)	(0.314)
μ_4	1.309***	1.294***	1.403***	1.298***
	(0.424)	(0.352)	(0.478)	(0.332)
μ_5	1.443***	1.427***	1.539***	1.430***
	(0.437)	(0.364)	(0.491)	(0.343)
Correlation parameters				
Married period 1	-0.083	-0.082	-0.071	-0.083
	(0.076)	(0.075)	(0.065)	(0.065)
Married period 2	-0.055	-0.054	-0.065	-0.063
	(0.060)	(0.059)	(0.059)	(0.060)
Log family income period 1	-0.003	-0.004	-0.004	-0.003
	(0.005)	(0.005)	(0.004)	(0.005)
Log family income period 2	0.000	-0.001	-0.001	-0.001
	(0.004)	(0.004)	(0.004)	(0.004)
PCS score period 1	0.001	0.000	-0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
PCS score period 2	-0.005***	-0.005***	-0.006***	-0.006***
	(0.002)	(0.002)	(0.002)	(0.002)
Depression measure period 1	0.010***	0.032***	0.144***	0.146***
	(0.003)	(0.009)	(0.049)	(0.046)
Depression measure period 2	0.009	0.025**	0.146**	0.128***
	(0.005)	(0.010)	(0.057)	(0.040)

Notes: The coefficients of the year dummies are omitted. Significance level: ***p<0.01, **p<0.05, *p<0.1; Standard errors are adjusted for the complex survey design of the MEPS using balanced repeated replication. Base categories are white, northeast, employer size > 500, and other services/public administration/military/unclassifiable industry.

Table 3.11: Sensitivity analysis of coefficient on depression in the bivariate probit model of employment

	$\rho =$ -0.4	$\rho =$ -0.3	$\rho =$ -0.2	$\rho =$ -0.1	$\rho = 0$	$\rho = \rho^* =$ -0.052	$\rho = \rho^*/2$=-0.026	CRE
Depression (PHQ-2≥3)	0.205***	0.009	-0.187***	-0.381***	-0.574***	0.432***	-0.074***	-0.091***
	(0.022)	(0.023)	(0.023)	(0.023)	(0.023)	(0.021)	(0.023)	(0.022)
	[0.053***]	[0.002]	[-0.053***]	[-0.113***]	[-0.176***]	[0.104***]	[-0.020***]	[-0.026***]

Notes: Significance level: ***p<0.01, **p<0.05, *p<0.1. Standard errors in parentheses; Marginal effects in brackets; All standard errors are adjusted for the complex survey design of the MEPS. Models include cross-sectional bivariate probit regressions. Control variables include age and its square, gender, race and ethnicity, marital status, years of education completed before entering the survey, region, urban residence, number of children under 5 or 18 in the household, log of income earned by other family members (normalized by household size), and year dummies.

Chapter 4

The Heath Implications of Unconventional Natural Gas Development in Pennsylvania

4.1 Introduction

Natural gas has become a key source of energy in the United States. Over the past decades, technological advancements in horizontal drilling and hydraulic fracturing (often referred to as fracking) have made natural gas trapped beneath various shale formations more economically accessible. The contribution of shale gas to total U.S. natural gas production increased drastically from less than 2 percent in 2000 to over 20 percent in 2010; it is also projected that 46 percent of natural gas supply will come from shale gas by 2035 (EIA, 2013). With this rapid expansion of shale gas development, the potential health risks have drawn attention from the public and regulators at various levels.

Typically, the process to develop shale gas wells involves well pad preparation and construction, drilling and well construction, hydraulic fracturing, flaring of excess natural gas, and gas extraction and compression. Air pollution can occur during each stage of the process, while water contamination mostly occurs during wellbore drilling and hydraulic fracturing (see Appendix 4.6.1 for a more detailed description of the stages of shale gas development). Numerous studies have documented that emissions of greenhouse gases (predominantly water vapor, carbon dioxide, methane, and ozone), volatile organic compounds

(VOCs), other air pollutants, and hazardous chemicals increase as a result of unconventional natural gas development (GAO, 2012). Based on data from a natural gas emissions inventory recently created by the Pennsylvania Department of Environmental Protection (PADEP), levels of air pollutant emissions (CO, NOX, PM2.5, PM10, etc.) attributable to unconventional natural gas drilling mostly increased between 2011 and 2012 as the number of gas wells in the state rose by nearly 30% (see Table 4.1).[1]

Despite the fact that air pollution has clear adverse health effects, there is little scientific research on the impact of shale gas development on human health (see Appendix 4.6.2 for a detailed discussion of the link between air pollution and human health). We know of only one study that investigates the direct impact of shale gas development on health. In particular, Hill (2012) found a higher incidence of low birth weight among babies born to mothers living in the vicinity of shale gas wells in Pennsylvania. We examine the impact of unconventional natural gas drilling in Pennsylvania on several air-pollution-sensitive medical conditions. The state of Pennsylvania is rich in Marcellus shale reserves[2] and has witnessed a significant expansion of unconventional natural gas development in the past decade, making it a good location to study the effects of drilling. We analyze detailed information on natural gas drilling and production activities and all inpatient hospital admissions in the state from 2001-2013, which encompasses the recent expansion in unconventional natural gas extraction. We estimate a set of fixed effects regressions models to control for unobserved time-invariant county-level attributes that might otherwise confound our analyses, and identify the effect of Marcellus well development on the county-level hospitalization rates for acute myocardial infarction (AMI), chronic obstructive pulmonary disease (COPD), asthma, pneumonia, and upper respiratory infections (URI). We find that unconventional natural gas development was associated with a significant increase in rates of URI among adults of all ages, and high

[1] These data are released by Pennsylvania Department of Environmental Protection (PADEP). URL: http://www.portal.state.pa.us/portal/server.pt/community/emission_inventory/21810/marcellus_inventory/1829

[2] Marcellus shale (also known as the Marcellus formation) is a geological formation found in eastern North America, spanning 6 states in the northeast U.S.. It is a unit of marine sedimentary rock that contains largely untapped natural gas reserves.

hospitalization rates for AMI and asthma among young adults. We also find that shale gas development increased the hospitalization rate for pneumonia among both young adults and the elderly. These findings are consistent with higher levels of air pollution.

4.2 Data

We obtained data on natural gas wells from the PADEP Oil and Gas Reports. The Spud Data Report contains information on the drilling commencement date (i.e. spud date), location, operator, and configuration of all conventional and unconventional natural gas wells drilled in Pennsylvania between 2001 and 2013. Unfortunately, these data do not include the well completion date, which indicates when the well is ready to produce natural gas.[3] We also obtained data on total annual gas production from the PADEP statewide well production database, which contains this information for all active natural gas wells. We linked spud dates to the gas production data using a unique well permit number, allowing us to determine the annual gas production for each active well after its spud date. Table 4.2 contains a list of all Pennsylvania counties that had unconventional natural gas wells drilled during the timeframe of this study. Our health outcome measures are derived from the Pennsylvania Health Care Cost Containment Councils (PHC4) compilation of all inpatient hospital admission records in the state during 2001-2013. We use ICD-9-CM codes to identify the main diagnosis for each inpatient admission and then group related diagnoses into clinically meaningful categories using the Clinical Classification Software (CCS) developed by the Agency for Healthcare Research and Quality (AHRQ) (see Appendix Table 4.4 for how each condition is defined).[4] We focus our analysis on the following five health conditions that are sensitive to air pollution: acute myocardial infarction (AMI), chronic obstructive

[3]For wells that begin producing immediately after drilling, the spud date and completion date are the same.

[4]The Clinical Classifications Software (CCS) for ICD-9-CM is developed as part of the Healthcare Cost and Utilization Project (HCUP); a federal-state-industry partnership sponsored by the Agency for Healthcare Research and Quality. It is a diagnosis and procedure categorization scheme that can be employed to collapse ICD-9-CM codes into a smaller number of clinically meaningful categories. The CCS can be accessed at http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp.

pulmonary disease (COPD), asthma, pneumonia, and upper respiratory infections (URI). Although many known or possible carcinogenic chemicals are used during the well development process (such as the BTEX compound), we do not consider cancer in this analysis due to the short time frame of our data.[5] Persistent correlations between the above five conditions and air pollution have been established in the epidemiological literature (for example, see Lee, Kim and Lee (2014); Garshick (2014); Neupane et al. (2010); Schwartz et al. (1993); Dockery and Pope (1994); Dominici et al. (2006); Brunekreef and Holgate (2002); Eder, Ege and von Mutius (2006)).

4.3 Empirical methods

We stratify the sample into four age groups: 5-19, 20-44, 45-64, and above 65 because the pollution caused by shale gas development might have a more pronounced effect on one age group than others. For example, elderly individuals with a pre-existing respiratory illness are more sensitive to air pollution, and are more likely than younger individuals to contract pneumonia (Ryan et al. (2013)). Because hospitalization rates for AMI and COPD are either very low or zero among children ages 5-19, we only estimate models for asthma, pneumonia, and URI for this age group. In each year, we construct county-level measures of the five health conditions by first aggregating the individual-level PHC4 data into county-year cells, and then normalizing the total number of inpatient admissions for each condition by the population of that county in each age group,[6] so that the measures reflect hospitalizations per one thousand people.

The average hospitalization rates for AMI, COPD, asthma, pneumonia, and URI are higher in counties containing unconventional natural gas wells than those without wells (Appendix Figures 4.2-4.6 contain trends in hospitalization rates over time in well and non-

[5]For example, a recent study measuring the impact of diesel exhaust exposure on the incidence of lung cancer among miners used a 15-year lag between the time of exposure and diagnosis of the illness (see Attfield et al. (2012)).

[6]These data are obtained from the Census Bureau Population Estimates Program, which can be accessed at `http://www.census.gov/popest/`

well counties for each condition and age group). This is likely because poverty rates are 28% higher in well counties, on average, and the fact that these counties have much higher levels of extraction of other natural resources, such as coal (see Table 4.3). Furthermore, the hospitalization rates for certain conditions and age groups exhibit strong secular trends over time in both well and non-well counties. In order to address both of these issues, we estimate the impact of unconventional natural gas development on county-level hospitalization rates using linear regressions containing fixed effects for county and year. This is sometimes referred to as a difference-in-difference specification because the estimates represent changes in hospitalization rates over time in well counties relative to changes in hospitalization rates over time in non-well counties. To assess the robustness of our results, we additionally estimate models that contain: 1) a linear time trend; or 2) linear, quadratic and cubic time trends instead of year fixed effects (see Appendix 4.6.3 for a more detailed discussion of the econometric model).

Another important consideration is that air pollution may come from gas well development activities (site construction, drilling, and initial hydraulic fracturing) as well as ongoing extraction activities (gas production, compression, and fuel transportation).[7] Developed activities occur prior to and in the same year as the well spud date, while extraction activities occur after the spud date. Therefore, we include in our regressions a set of binary indicator variables representing whether there is at least one active unconventional well in the county in the preceding (lag), current, or following (lead) year. Since we only know the spud date for each well, the 1-year lead indicator captures the adverse health effects of well site preparation and construction. The current-year indicator captures both the contemporaneous effect of drilling and hydraulic fracturing and the lagged effect of site preparation and construction; and the 1-year lag indicator captures the lagged response to all development activities. In order to measure the impact of ongoing extraction activities, we aggregate

[7]It may be necessary to re-stimulate (also known as re-fracturing or well workovers) the wells throughout the production period by repeating the hydraulic fracturing process, the frequency of which depends on the characteristics of geologic formation and production phase of a particular well.

natural gas production from unconventional wells to the county level, and include in our models the log of county-level total output[8] in preceding and current year.

Our models also contain control variables for county-level demographic composition, economic conditions and activities, as well as measures of patient characteristics at each countys hospitals (see Table 4.3). Specifically, we include the county-level unemployment rate, poverty rate, the log of county population density, quartiles of county median household income,[9] and the percentage of the county population in each five-year age category, from 0-4 to 85 and above.[10] These measures are derived from Small Area Income and Poverty Estimates (SAIPE) and the Population Estimates Program (PEP) data made available by the U.S. Census Bureau.

Greater availability and lower prices of natural gas may reduce the demand for other fuels that generate pollution during their extraction and use (EPA, 1999; NRC, 2010). As unconventional natural gas extraction in Pennsylvania has increased, the extraction of coal for electricity generation has decreased. In order to capture this substitution of fuels we include in our models the log of annual county-level production of coal from surface and underground mining reported to the U.S. Energy Information Administration (EIA).

Finally, we construct county-level measures of patient characteristics from the PHC4 inpatient admission records and include these in our models. These are the county-level proportion of female patients, the proportion of patients of different racial and ethnic categories (white, black, Asian, Hispanic, and other races), the proportion of different types of admissions (elective, urgent, emergency, and other types), and the proportion of patients in mutually exclusive insurance categories (private insurance, Medicaid, Medicare, other gov-

[8] We add 1 to the level of output in counties without wells.

[9] We use the annual quartiles of household income due the methodological changes in SAIPE. For the series of SAIPE state and county estimates, notable differences include the break between 2004 and 2005 due to the switch from Current Population Survey Annual Social and Economic Supplement (CPS ASEC) to American Community Survey (ACS) data in SAIPE modeling. For that reason, estimates for these particular years are directly comparable.

[10] The Intercensal Population Estimates released by the U.S. Census Bureau provide estimates of county-level population by five-year age groups. We include percentages of population in each age group for each county-year in all of our models to account for the demographic changes in the counties.

ernment insurance, self-pay, and all other payers), and the county average Charlson index (Charlson et al. (1987)).[11] In Table 4.3 we report descriptive statistics for all variables used in this analysis by age group. As expected, the county-level hospitalization rates for pollution-sensitive conditions, and the Charlson index increase with age.

4.4 Results

Overall, the estimates from our empirical models suggest that unconventional natural gas development is positively and consistently associated with significantly higher rates of hospitalization for all but one of the five pollution-sensitive conditions. There are, however, differences across the conditions in which age groups are most affected and at what point in time gas development translates into higher hospitalization rates. We summarize the results of our each model below (see Figure 4.1), focusing on our preferred specification with county and year fixed effects, and report our full set of estimates, including models with a linear time trend, and a linear, quadratic, and cubic time trend in Appendix Table 4.5:

AMI: Unconventional well development in the previous year is associated with an increase of 0.11 hospital admissions per one thousand people aged 20-44, which is a 24 percent increase relative to the state average hospitalization rate for AMI for this age group. We do not find any statistically significant effects of variables indicating wells in the previous, current, or following year for the other age groups. We do find that higher levels of gas output in the previous year are associated with higher rates of AMI hospitalizations among those aged 45-64 in the models that contain time trend controls, but this result is not statistically significant in our preferred specification with year fixed effect.

COPD: Among individuals aged 20-44, we find that well development measured in the current year is associated with an increase of 0.07 admissions for COPD per one thousand people, while well development in the previous year is associated with a decrease of 0.08 admissions per one thousand people. Because these two effects offset each other, this

[11] To avoid endogeneity, we only use contemporaneous secondary diagnoses when computing the Charlson index.

result may reflect a temporary shift in hospitalizations one year earlier (i.e. well development accelerates COPD hospitalizations, but does not cause a higher overall hospitalization rate). For individuals aged 65 and above, we find that well development in the following year is associated with an increase of approximately 0.6 admissions per one thousand people, which is a 5 percent increase relative to the state average. Although this estimate is not statistically significant in our preferred specification, the magnitude of the estimate increases slightly and becomes significant at the 5% level in the two specifications containing time trend variables instead of year fixed effects.

Asthma: Although we find a statistically significant association between previous year well development and higher hospitalizations for individuals aged 20-44, the sensitivity analyses we present below indicate that this result could be spurious. We do not find any consistent associations among other age groups.

Pneumonia: We find that unconventional well development in the following year is associated with an increase of 0.1 admissions per one thousand people, or 10 percent, among individuals aged 20-44. In contrast, it is well development in the previous year that is associated with an increase of 1.6 admissions per one thousand people among the elderly. This is a 9 percent increase relative to the state average.

URI: We find a positive and consistent correlation between unconventional gas well development and URI hospitalizations across all age groups, except for children aged 5-19. Well development in the following year is associated with an increase of 0.05 and 0.04 admissions per one thousand people for individuals aged 20-44 and 45-64, respectively. For the elderly, well development in the previous year is associated with an increase of 0.2 admissions per one thousand people. When compared to the state average of hospitalization rates for URI in each age group, these effects represent increases of 38 percent, 14 percent, and 17 percent, respectively.

In order test to robustness of our results, we conduct a falsification test and simulation designed to measure the likelihood of spurious correlation between the well development indicators and the outcome variables. Importantly, our estimates are reliable only if model either captures or removes county-level factors that are correlated with both unconventional well development and hospitalization rates. Our falsification test considers of hospitalizations for trauma-related disorders[12] as an alternative outcome. These disorders should not be related to pollution, so finding a statistically significant impact of well development on trauma hospitalizations would indicate the presence of uncontrolled unobservable factors that are confounding the relationship between well development and the other health outcomes. In addition, hospitalizations for trauma are very common, so this falsification test has good statistical power. In support of our empirical approach, we do not find any statistically significant correlations between the well development or gas output variables and the hospitalization rate for trauma-related disorders. We also estimated several alternative specifications, including one set of model that excludes county characteristics and another set of models containing county-specific-specific log-linear time trends. Our results are generally not sensitive to either of these alternative specifications.[13]

Next, we conduct a Monte Carlo simulation designed to detect spurious correlations between the well development indicators and outcome variables, which is described in Appendix section 4.6.4 and Table 4.6. For all of our statistically significant estimates except those of the impact of well development in the previous year on asthma among individuals age 20-44, the simulations indicate the probability of a Type I error (incorrectly rejecting the null hypothesis of no effect) is no higher than the standard levels of statistical significance we report to indicate the strength of the associations (1%, 5%, or 10%). However, in the case of asthma, the simulation suggests that the probability of a Type I error is not 10%, but closer to 18%, which is higher than conventional standards.

[12]These conditions include joint dislocations, fractures, intracranial injury, crushing injury or internal injury, open wounds, burns, and other conditions due to external cause.

[13]These results are available upon request.

4.5 Discussion and conclusions

We examine the impact of unconventional natural gas development, which is known to cause air pollution, on human health. Our results show that activities surrounding horizontal drilling and hydraulic fracturing into Marcellus shale in the state of Pennsylvania over the past decade are associated with significant increases in hospitalization rates for AMI (aged 20-44), COPD (aged 20-44 and 65 and above), pneumonia (aged 20-44 and 65 and above), and URI (aged 20-44, 45-64, and 65 and above). Notably, we do not find any impact of gas well development on asthma, pneumonia, or URI among children aged 5-19. Because children spend more time outdoors, breath more rapidly than adults, and breath through their mouths rather than filtering air through their nose, their exposure to air pollution is typically assumed to be higher than adults (OEHHA and ALA, 2003). It is possible that the impact of air pollution from well development has a longer term impact on children through the development of respiratory and other illnesses that we are unable to detect during the limited timeframe of our analysis. In contrast, the effects we find among adults may reflect the acute aggravation of pre-existing conditions. Nonetheless, differences in exposure within the group of adults could explain why, for example, unconventional gas well development has a statistically impact on AMI hospitalizations among young adults aged 20-44, but not older adults who spend relatively less time outdoors.

Another noteworthy finding is that nearly all of the adverse effects we identify are due to well development and not natural gas extraction and compression. The only exception is a significant association between lagged gas output and AMI hospitalizations among 45-64 year olds, which is not precisely estimated in our preferred model. A likely source of air pollution during well development are the diesel engines that power heavy equipment used to build roads, clear well sites, construct wells, drill, and inject fracking fluid into the wells. Horizontal drilling followed by hydraulic fracturing is more energy intensive than traditional vertical drilling, and the diesel engines used to pump fracking fluid commonly exceed 2000 bhp (Treida, 2010).

Our results indicate that the pollution caused by well development takes a longer time to impact the health of some age groups than others. In particular, gas development has an immediate effect on pneumonia and URI hospitalization rates among younger adults captured by our one year lead variable, but a lagged effect on the elderly. This could be due to differences in the exposure or susceptibility to air pollution across these different age groups. Identifying the precise mechanisms through which air pollution from unconventional well development impacts the health of different age groups is beyond the scope of this analysis, but certainly worthy of further investigation.

Our analysis does have some limitations that should be kept in mind when interpreting our results. The large-scale development of Marcellus shale has inevitably caused migration into affected communities by gas workers and other individuals seeking jobs created by greater economic activity associated with growth in the gas industry. Likewise, shale gas development has resulted in out-migration by individuals that have sold their land to gas companies, or have been displaced by rising cost of housing in well counties. We have included variables in our empirical models that capture changes in the county-level age distribution over time. Nonetheless, if the net impact of migration was to increase (decrease) the number of individuals with pollution-sensitive diseases in well counties, or decrease (increase) the number of individuals with these conditions in non-well counties, our estimates will be upwardly (downwardly) biased. Despite these limitations, our study is the first to establish a consistent link between unconventional natural gas extraction and higher rates of disease. Our results have important implications for public policy because they provide evidence of an adverse impact of shale gas development on health which is currently of concern to policy makers. For example, in April 2012, the U.S. Environmental Protection Agency (EPA), Department of Energy, and Department of Interior agreed to collaborate on research in order to improve the scientific understanding of hydraulic fracturing. The EPA also launched a national study (still ongoing) to investigate the potential impact of hydraulic fracturing on drinking water resources. In 2010 and 2011, the Pennsylvania DEP conducted three short-term studies to determine whether shale gas development affects air quality in the southwestern, northeastern, and northcentral regions of the state. In all three studies, natural gas constituents and associated compounds were detected in the air near Marcellus shale drilling operations, but the DEP concluded that none of the compounds reached a level of concentration that could cause air-related health issues (PADEP, 2011a; 2011b; 2011c). However, a recent study conducted by the Southwest Pennsylvania Environmental Health Project (SWPA-EHP), a non-profit environmental health organization, found short-term high values of particulate matter in the air and concluded that current methods of collecting and analyzing air pollutants emission data are not sufficiently accurate for evaluating the health risks of unconventional natural gas development (Brown et al., 2014*a*).

In the absence of strong scientific evidence on the relationship between shale gas development and health, states in the mid-Atlantic region have demonstrated conflicting regulatory objectives. In February 2012 the Pennsylvania General Assembly passed Act 13, which was a major overhaul of the states oil and gas law. According to the new law, municipal governments are not allowed to impose stricter regulations on drilling activities than other industries and must allow oil and gas operations in all zoning districts (The General Assembly of Pennsylvania, 2011). This portion of the law resulted in disputes between local communities and state government and was subsequently ruled unconstitutional by the Pennsylvania Supreme Court (Cusick, 2013). In contrast, the state of New York banned natural gas extraction activities that involve hydraulic fracturing in 2008 through a de facto moratorium on shale development. The ban was renewed in June 2014 for the second time (first time in 2013), and prohibits hydraulic fracturing in the state until 2017 (New York State Assembly, 2013).

We seek to inform future regulatory policy on unconventional natural gas development by providing evidence on the link between Marcellus shale gas development and health. Because we find that unconventional natural gas well development has a stronger link to poor health than post-development gas production, there is more limited justification for natural gas extraction taxes based on pollution-related externalities than per-well fees (for example, under Pennsylvania's Act 13 of 2012, operator have to pay an annual "impact fee" on every Marcellus well they drill). However, our results demonstrate a clear need for additional studies to confirm the precise causal pathways between unconventional gas well development, elevated levels of air pollution, and adverse health effects among different age groups.

Tables

Table 4.1: Statewide air emissions from unconventional natural gas development in Pennsylvania, 2011-2012

Emissions (tons)	2011	2012	change (%)
Air pollutants			
CO	6852	7350	7.26%
NOx	16542	16361	-1.09%
PM-10	577	600	4.05%
PM-2.5	505	548	8.56%
SOx	122	101	-17.32%
VOC	2820	4024	42.69%
Benzene	20	25	26.97%
Ethyl Benzene	5	6	11.25%
Formal dehyde	251	374	48.93%
n-Hexane	51	98	92.96%
Toluene	34	33	-2.64%
Xylenes	26	34	32.30%
2,2,4-Trimethyl pentane	4	19	439.11%
Greenhouse gases			
CO2	N/A	4291316	N/A
Methane	N/A	123884	N/A
Nitrous Oxide	N/A	209.3	N/A
Number of active unconventional wells	4052	5253	29.64%

Notes: These emission data are only for drilling and production phases, and do not include the emissions from the well pad construction phase. Emissions for CO2, methane, and nitrous oxide are not available for 2011.

Table 4.2: List of counties with unconventional natural gas wells

County name	Drilling Year	County name	Drilling Year
Allegheny	2008	Susquehanna	2006
Armstrong	2009	Tioga	2006
Beaver	2009	Venango	2011
Bedford	2010	Warren	2007
Blair	2010	Washington	2002
Bradford	2005	Wayne	2008
Butler	2006	Westmoreland	2003
Cambria	2009	Wyoming	2009
Cameron	2008	**Non-well counties**	
Centre	2007	Adams	Philadelphia
Clarion	2007	Berks	Pike
Clearfield	2007	Bucks	Schuylkill
Clinton	2008	Carbon	Snyder
Columbia	2010	Chester	Union
Crawford	2012	Cumberland	York
Elk	2005	Dauphin	
Fayette	2006	Delaware	
Forest	2009	Erie	
Greene	2006	Franklin	
Huntingdon	2010	Fulton	
Indiana	2003	Juniata	
Jefferson	2008	Lancaster	
Lackawanna	2009	Lebanon	
Lawrence	2011	Lehigh	
Luzerne	2010	Mifflin	
Lycoming	2007	Monroe	
Mckean	2006	Montgomery	
Mercer	2012	Montour	
Potter	2007	Northampton	
Somerset	2004	Northumberland	
Sullivian	2010	Perry	

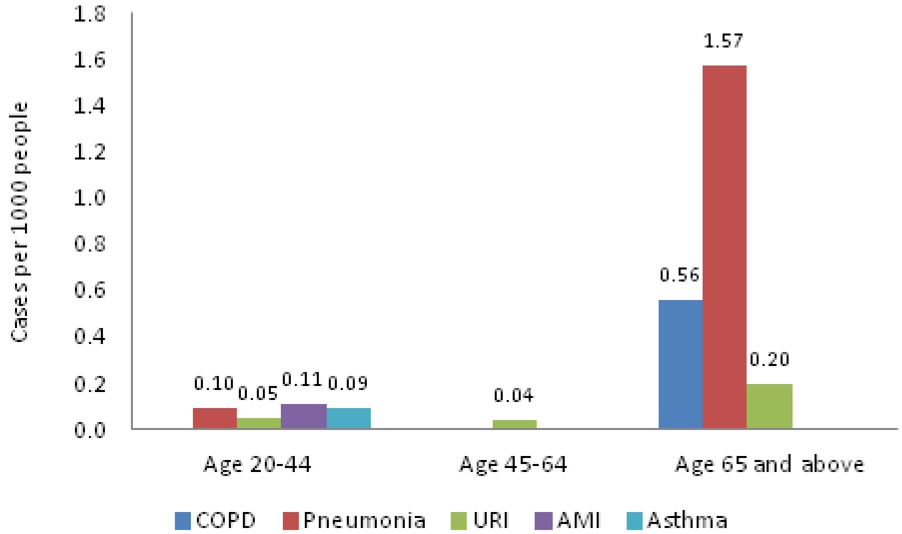
Table 4.3: Summary statistics, county-level, 2001-2013

	Age 5-19		Age 20-44		Age 45-64		Age 65 and above	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Patient characteristics								
Age	14.424	0.430	32.011	0.616	55.179	0.402	77.982	0.705
Female	0.570	0.046	0.701	0.033	0.495	0.024	0.562	0.026
White	0.847	0.142	0.881	0.121	0.909	0.102	0.947	0.079
Black	0.071	0.098	0.053	0.081	0.045	0.080	0.022	0.053
Asian	0.004	0.005	0.005	0.007	0.002	0.003	0.001	0.002
Other race	0.078	0.076	0.061	0.063	0.043	0.050	0.030	0.053
Hispanic	0.038	0.059	0.027	0.049	0.015	0.040	0.009	0.040
Private	0.589	0.088	0.545	0.079	0.574	0.082	0.060	0.028
Medicare	0.004	0.008	0.083	0.023	0.227	0.047	0.924	0.034
Medicaid	0.364	0.090	0.308	0.065	0.153	0.046	0.004	0.005
Government insurance	0.014	0.011	0.013	0.011	0.015	0.012	0.005	0.006
Self-pay	0.021	0.016	0.045	0.020	0.025	0.014	0.002	0.010
Other insurance	0.007	0.011	0.006	0.009	0.005	0.008	0.005	0.012
Admission: emergency	0.415	0.152	0.322	0.137	0.427	0.168	0.479	0.205
Admission: urgent	0.286	0.140	0.241	0.133	0.261	0.165	0.283	0.205
Admission: elective	0.293	0.111	0.431	0.098	0.308	0.058	0.235	0.058
Admission: other type	0.000	0.002	0.000	0.001	0.000	0.001	0.000	0.001
Charlson index	0.157	0.064	0.274	0.060	1.065	0.156	1.484	0.263
County-level hospitalization rate (cases per 1000 people)								
Acute myocardial infarction (AMI)	0.002	0.011	0.447	0.243	3.577	1.090	12.080	3.720
Chronic obstructive pulmonary disease (COPD)	0.018	0.044	0.223	0.201	3.001	1.390	12.450	4.474
Asthma	0.719	0.749	0.780	0.449	1.296	0.791	2.075	1.226
Pneumonia	0.737	0.435	1.008	0.466	3.237	1.269	17.686	5.484
Upper respiratory infection (URI)	0.058	0.104	0.129	0.140	0.276	0.229	1.164	0.765
County Characteristics (across all age groups)								

	Well counties		Non-well counties	
	Mean	S.D.	Mean	S.D.
Poverty rate* (percent)	13.351	2.719	10.401	3.759
Household median income* (thousand, in 2013 dollars)	44.515	4.452	56.191	11.573
Unemployment rate* (percent)	6.832	1.775	6.133	1.936
Population density* (Number of people per square mile)	4.520	1.019	5.744	1.209
Log of coal production (underground, short tons)**	1.199	5.115	0.007	0.030
Log of coal production (surface, short tons)**	0.276	0.593	0.034	0.173
Has a well in following year (0/1)***	0.568	0.496		
Has a well in current year (0/1)***	0.491	0.500		
Has a well in previous year (0/1)***	0.414	0.493		
Log of output in current year*** (million cubic feet)	5.144	6.971		
Log of output in previous year*** (million cubic feet)	4.087	6.410		

Notes: *U.S. Census Bureau, Small Area Income and Poverty Estimates (SAIPE); **U.S Energy Information Administration; ***Pennsylvania Department of Environmental Protection; Natural gas output is measured in thousand cubic feet (MCF); Household median income is unadjusted due to the methodological change of SAIPE after 2005. County level incidences are calculated by first aggregating the individual-level PHC4 data to county-year cells for each age group, and then normalizing by the population in that age group. All patient characteristics are county-level means from the PHC4 data.

Figure 4.1: Additional hospitalizations per 1,000 people due to unconventional natural gas well development



4.6 Appendix

4.6.1 Shale gas development process

Unlike conventional natural gas development, extracting natural gas from unconventional formations relies heavily on horizontal drilling and hydraulic fracturing. Typically, operators first construct a well pad at the location suitable for drilling, build infrastructure, and transport equipment to the drilling site. In the next stage, a hole (wellbore) is drilled into the earth through a combination of vertical and horizontal drilling. Casing[14] and cement are inserted into the wellbore in order to isolate it from the surrounding formation. Finally, hydraulic fracturing is used to stimulate the shale formation. This involves the injection of highly pressurized fracturing fluid through the holes created by a perforating tool inserted in the casing and cement. As fracturing fluid is forced into the surrounding formation, fractures or cracks are created or expanded in the target formation. The underlying gas is then released and collected. It is worth noting that throughout the production period it may be necessary to re-stimulate (also known as re-fracturing or well workovers) the wells by repeating the hydraulic fracturing process, the frequency of which depends on the characteristics of geologic formation and production phase of a particular well (GAO, 2012; DOE, 2009). When estimating annual green house gas emission from natural gas production, the Environmental Protection Agency use the assumption that 10 percent of unconventional wells need re-stimulation every year (EPA, 2012).

4.6.2 Potential public health risks

Shale gas development and production may pose a threat to public health through air pollution (NRDC, 2014). First, the construction of infrastructure at the drilling site requires massive transportation of water, sand, chemicals, and heavy machinery. Air pollutants such

[14] Casing is a metal pipe that is inserted inside the wellbore to prevent high pressure fluids outside the formation from entering the well and to prevent drilling mud inside the well from fracturing fragile sections of the wellbore.

as nitrogen oxides (NOX) and particulate matters (PM) contained in the engine exhaust brought about by increased traffic are released into the atmosphere. In addition, the development and production process requires substantial amount of power, which is often supplied by diesel engines. The burning of diesel fuel also generates exhaust. Second, for operational reasons, flaring (burning) or venting (direct release into the atmosphere) of natural gas during the development and production process is sometimes necessary, which leads to emissions of carbon dioxide and the release of methane and volatile organic compounds (VOCs). Third, evaporation of fracturing fluid and produced water may also emit hazardous chemicals into the atmosphere. 4, 5 Some of the air pollutants and chemicals from the drilling and gas production activities may be harmful to human health and even carcinogenic. NOX can form small particles through reactions with ammonia, moisture, and other compounds. These particles penetrate deeply into the sensitive part of lungs and cause or worsen respiratory diseases (EPA, 1998). In addition, when reacting with VOCs in the presence of heat and sunlight, NOX can form ground-level Ozone (smog), which irritates the respiratory system, reduces lung function, aggravates chronic conditions such as asthma and chronic bronchitis, and potentially results in permanent lung damage (EPA, 2009).

Particulate matter (PM) is also harmful. Short-term exposure to fine particles can cause asthma attacks and acute bronchitis, and increases the risk of heart attacks and arrhythmias among people with heart disease (EPA, 2003). There are a multitude of studies that attempt to uncover the link between air pollution and adverse health outcomes. In general, researchers have found consistent evidence that air pollution is associated with respiratory problems. For example, Ko et al. (2007) find that levels of major air pollutants (NO2, O3, PM10, and PM2.5) in Hong Kong were associated with increased hospital admissions, with O3 being the most important contributor. Likewise, Zanobetti and Schwartz (2006) find that air pollution in the greater Boston area was associated with a higher risk of hospitalization for pneumonia among individuals aged 65 and older .

Colborn et al. (2011) compile a list of 632 chemicals used during the fracturing and drilling stages of natural gas development and report that many of them could have a negative impact on human health. In particular, more than 75 percent of the chemicals could affect the respiratory system; about half could affect the immune and cardiovascular systems; and 25 percent could cause cancer. A similar analysis was conducted in a congressional report by the Committee on Energy and Commerce of the U.S. House of Representatives. The report reviews the type and volume of hydraulic fracturing products used by 14 leading oil and gas companies between 2005 and 2009 and finds that the most widely used chemical during that period was methanol, which is a hazardous air pollutant, and that more than 650 hydraulic fracturing products contain 29 chemicals that are known or possible human carcinogens (Committee on Energy and Commerce, U.S. House of Representatives, 2011). These chemicals are either regulated under the Safe Drinking Water Act for their risks to human health, and/or listed as hazardous air pollutants under the Clean Air Act. For instance, the BTEX compounds (benzene, toluene, xylene, and ethylbenzene) were found in many of the hydraulic products. Each BTEX compound is a regulated contaminant under the Safe Drinking Water Act and a hazardous air pollutant under the Clean Air Act. Benzene alone is also known to be carcinogenic.

4.6.3 Empirical methods

The baseline econometric model for our empirical analysis is

$$y_{ct} = X'_{ct}\beta + \alpha_1 W_{c,t-1} + \alpha_2 W_{ct} + \alpha_3 W_{c,t+1} + \gamma_1 L_{ct} + \gamma_2 L_{c,t-1} + \psi_c + \zeta_t + u_{ct}, \tag{4.1}$$

where y_{ct} is the health outcome of interest in county c and year t, X_{ct} is a vector of county characteristics, $W_{c,t-1}$ (W_{ct} or $W_{c,t+1}$) is is an indicator variable that equals to one if there are active unconventional wells in the county in the previous (current or following) year,[15] $L_{c,t-1}$ (or $L_{c,t+1}$) is the log of natural gas output[16] from all active unconventional wells in

[15]Once an unconventional well is drilled, a county remains "treated" for the rest of the sample period.
[16]Natural gas output is measured in thousand cubic feet (MCF).

the county in the previous (current) year, and u_{it} is a random error term assumed to be uncorrelated with X_{ct}, $W_{c,t-1}$, W_{ct}, $W_{c,t+1}$, $L_{c,t-1}$, and L_{ct} (we add 1 to the level of output in counties without wells because the output is equal to 0 for them). In order to control for time invariant unobserved county-level heterogeneity and overall secular trends in the outcomes, we include a set of county fixed effects, ψ_c, and time effects, ζ_t. For each outcome, we estimate three models with different time effects: (1) year fixed effects; (2) a linear time trend; and (3) linear, quadratic and cubic time trends.

The parameters of interest, α_1, α_2, α_3, γ_1, and γ_2 capture the reduced-form effects of Marcellus shale gas development and production on the county-level hospitalization rate for each medical condition. Since α_1, α_2, and α_3 are coefficients on variables that indicate the presence of unconventional wells (i.e. treatment effects), they measure the change in hospitalization rates over time in counties with unconventional wells relative to the change in hospitalization rates over time in non-well counties when the specification includes county and year fixed effects. When the specification includes continuous time trend variables, these coefficients measure the within-county change in hospitalizations relative to the overall trend in hospitalization rates across all counties.

4.6.4 Monte Carlo simulation

Using a Monte Carlo simulation, we investigate whether our findings could result from spurious correlation between drilling activity and county-level disease trends. The simulation results indicate that the potential for unobservable county-level attributes to confound our findings in this manner is small.

In order to conduct the simulation we first subset the sample to 2000 - 2005, the period before unconventional drilling in Marcellus shale began, and then randomly assign the treatment of unconventional wells to counties that contained wells in the post-2005 period. These wells are assigned by drawing the year of drilling in the pre-well period from a uniform distribution. We then estimate equation (1) for each condition with the lead, current, and lagged placebo treatment indicators and year fixed effects, but without the output variables.

Provided there are no unobserved determinants of hospitalization rates in well counties, the coefficients on the placebo treatment variables should be statistically significant under a t-test at the same rate at the α level of the test (i.e. the Type 1 error rate). If, however, we find a spurious correlation between the placebo treatment and hospitalization rates at a higher rate than the α level of the test, it indicates that there are unobservable differences in factors determining hospitalization rates that are not accounted for by our models, which may lead us to incorrectly conclude that there is a statistically significant relationship between shale gas development and health. We report the 1%, 5%, and 10% rejection rates for the placebo treatments for the models corresponding to our health outcomes in Table 4.6.

In general, our simulation results indicate that we are no more likely to commit a type 1 error than under the classical linear regression assumptions. For example, we find an effect of the 1-year lag indicator for well development on pneumonia among individuals aged 65 and above that is statistically at 1 percent level, and in our simulation the rejection rate at the same level of significance for this variable is 0.88 percent. However, there is one case where the rejection rate from our simulation is noticeably different from the probability of a type I error; this is for the lagged effect of well development on asthma among individuals aged 20-44. Although our estimated suggest the coefficient on this variable is statistically significant at the 10 percent level, our simulations show that the probability of finding a "false positive" for asthma in this case is in fact 18 percent. As a result, we believe this particular estimation result is unreliable. But overall, the simulations lend credibility to our main findings.

Table 4.4: ICD-9-CM diagnosis codes for all five conditions

AMI
4100 41000 41001 41002 4101 41010 41011 41012 4102 41020 41021 41022 4103 41030 41031 41032 4104 41040 41041 41042 4105 41050 41051 41052 4106 41060 41061 41062 4107 41070 41071 41072 4108 41080 41081 41082 4109 41090 41091 41092

COPD
490 4910 4911 4912 49120 49121 49122 4918 4919 4920 4928 494 4940 4941 496

Asthma
49300 49301 49302 49310 49311 49312 49320 49321 49322 49381 49382 49390 49391 49392

Pneumonia
00322 0203 0204 0205 0212 0221 0310 0391 0521 0551 0730 0830 1124 1140 1144 1145 11505 11515 11595 1304 1363 4800 4801 4802 4803 4808 4809 481 4820 4821 4822 4823 48230 48231 48232 48239 4824 48240 48241 48242 48249 4828 48281 48282 48283 48284 48289 4829 483 4830 4831 4838 4841 4843 4845 4846 4847 4848 485 486 5130 5171

URI
4660 4661 46611 46619 0320 0321 0322 0323 0340 460 4610 4611 4612 4613 4618 4619 462 4640 46400 46401 46410 46411 46420 46421 46430 46431 4644 46450 46451 4650 4658 4659 4730 4731 4732 4733 4738 4739 78491

Table 4.5: Impact of shale gas development on county-level hospitalization rates, 2001-2013

	Age 5-19			Age 20-44			Age 45-64			Age 65 and above		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Acute myocardial infarction (AMI)												
Well lead				0.000	-0.004	-0.004	-0.076	0.016	-0.042	-0.054	-0.074	-0.173
				(0.036)	(0.033)	(0.035)	(0.116)	(0.119)	(0.119)	(0.332)	(0.310)	(0.345)
Well current				-0.044	-0.044	-0.044	-0.048	-0.104	-0.056	0.051	-0.068	0.114
				(0.047)	(0.045)	(0.049)	(0.127)	(0.134)	(0.132)	(0.339)	(0.344)	(0.349)
Well lag				0.110**	0.114**	0.114**	0.007	0.017	0.003	0.350	0.392	0.386
				(0.049)	(0.051)	(0.051)	(0.133)	(0.131)	(0.134)	(0.352)	(0.337)	(0.344)
Log output current				-0.000	-0.001	-0.001	-0.003	-0.005	-0.004	-0.002	0.006	0.005
				(0.005)	(0.005)	(0.005)	(0.014)	(0.014)	(0.014)	(0.046)	(0.046)	(0.046)
Log output lag				-0.001	-0.000	-0.000	0.016	0.022**	0.020*	0.066	0.056	0.061
				(0.003)	(0.003)	(0.003)	(0.010)	(0.010)	(0.010)	(0.040)	(0.040)	(0.040)
Chronic obstructive pulmonary (COPD)												
Well lead				-0.039	-0.027	-0.027	0.102	0.137	0.161	0.557	0.725**	0.850**
				(0.025)	(0.022)	(0.023)	(0.133)	(0.119)	(0.120)	(0.348)	(0.336)	(0.370)
Well current				0.077**	0.071**	0.069**	-0.058	-0.077	-0.113	0.086	-0.135	-0.213
				(0.029)	(0.028)	(0.028)	(0.154)	(0.155)	(0.154)	(0.376)	(0.406)	(0.428)
Well lag				-0.080*	-0.081*	-0.081*	-0.250	-0.224	-0.218	0.729	0.826	0.841*
				(0.042)	(0.042)	(0.042)	(0.170)	(0.182)	(0.180)	(0.444)	(0.508)	(0.503)
Log output current				0.001	0.001	0.001	0.021	0.022	0.021	-0.046	-0.046	-0.049
				(0.003)	(0.003)	(0.003)	(0.015)	(0.016)	(0.016)	(0.048)	(0.054)	(0.053)
Log output lag				-0.002	-0.001	-0.001	-0.006	-0.004	-0.004	0.005	0.019	0.024
				(0.002)	(0.003)	(0.003)	(0.016)	(0.015)	(0.016)	(0.037)	(0.039)	(0.038)
Asthma												
Well lead	-0.057	-0.045	-0.040	0.023	0.025	0.032	0.089	0.104	0.135*	-0.004	-0.001	0.020
	(0.052)	(0.049)	(0.048)	(0.046)	(0.042)	(0.041)	(0.074)	(0.076)	(0.075)	(0.112)	(0.109)	(0.111)
Well current	0.028	0.040	0.027	-0.028	-0.019	-0.025	-0.035	-0.050	-0.069	0.035	-0.028	0.001
	(0.072)	(0.073)	(0.072)	(0.046)	(0.048)	(0.047)	(0.083)	(0.078)	(0.080)	(0.168)	(0.180)	(0.175)
Well lag	-0.040	-0.048	-0.041	0.093*	0.088*	0.089*	0.077	0.080	0.087	0.113	0.128	0.132

	(0.060)	(0.059)	(0.060)	(0.048)	(0.049)	(0.048)	(0.076)	(0.079)	(0.081)	(0.127)	(0.132)	(0.131)
Log output current	-0.002	-0.002	-0.002	-0.008	-0.009	-0.009	0.005	0.004	0.004	-0.012	-0.013	-0.015
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.016)	(0.016)	(0.016)
Log output lag	-0.006	-0.004	-0.005	0.006	0.008	0.008	-0.005	-0.004	-0.003	-0.001	0.001	0.005
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.014)	(0.014)	(0.014)
Pneumonia												
Well lead	0.020	0.002	0.025	0.095**	0.103**	0.112**	0.056	0.107	0.090	0.453	0.698	0.571
	(0.050)	(0.055)	(0.054)	(0.047)	(0.046)	(0.044)	(0.111)	(0.107)	(0.106)	(0.464)	(0.460)	(0.443)
Well current	0.023	0.044	0.017	-0.008	-0.023	-0.024	0.075	0.027	0.060	0.374	-0.079	0.179
	(0.075)	(0.075)	(0.075)	(0.052)	(0.058)	(0.056)	(0.165)	(0.163)	(0.166)	(0.510)	(0.531)	(0.507)
Well lag	0.073	0.072	0.081	0.002	0.013	0.012	0.185	0.190	0.185	1.573***	1.609***	1.603***
	(0.077)	(0.078)	(0.078)	(0.073)	(0.076)	(0.078)	(0.197)	(0.196)	(0.193)	(0.410)	(0.430)	(0.418)
Log output current	-0.010	-0.008	-0.009	0.001	0.003	0.002	-0.013	-0.014	-0.014	0.005	0.002	0.001
	(0.008)	(0.008)	(0.009)	(0.007)	(0.007)	(0.008)	(0.016)	(0.015)	(0.015)	(0.066)	(0.067)	(0.068)
Log output lag	-0.001	-0.001	-0.000	0.011*	0.010	0.011*	0.009	0.009	0.009	-0.036	-0.029	-0.020
	(0.007)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)	(0.016)	(0.015)	(0.015)	(0.058)	(0.060)	(0.058)
Upper respiratory infection (URI)												
Well lead	0.002	-0.000	0.001	0.051*	0.052*	0.049*	0.040*	0.045**	0.041*	-0.107	-0.091	-0.100
	(0.013)	(0.013)	(0.012)	(0.028)	(0.026)	(0.027)	(0.023)	(0.021)	(0.022)	(0.082)	(0.076)	(0.076)
Well current	-0.016	-0.011	-0.011	0.012	0.008	0.011	0.002	-0.011	-0.001	0.055	0.035	0.053
	(0.019)	(0.016)	(0.017)	(0.027)	(0.028)	(0.027)	(0.040)	(0.040)	(0.042)	(0.096)	(0.098)	(0.104)
Well lag	0.021	0.019	0.019	-0.051	-0.048	-0.049	-0.005	0.004	0.003	0.195**	0.219***	0.219***
	(0.018)	(0.018)	(0.018)	(0.044)	(0.044)	(0.044)	(0.035)	(0.034)	(0.034)	(0.081)	(0.076)	(0.076)
Log output current	-0.000	-0.000	-0.000	0.003	0.002	0.003	0.000	0.000	0.000	-0.020*	-0.019*	-0.019*
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.010)	(0.011)	(0.011)
Log output lag	-0.000	0.000	0.000	-0.001	-0.001	-0.001	0.001	0.001	0.001	-0.002	-0.004	-0.003
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.009)	(0.010)	(0.009)
Trauma-related disorders												
Well lead	0.045	0.059	0.075	-0.016	-0.023	-0.030	-0.050	-0.067	-0.037	-0.327	-0.089	-0.258
	(0.125)	(0.123)	(0.128)	(0.113)	(0.107)	(0.116)	(0.079)	(0.083)	(0.085)	(0.342)	(0.324)	(0.328)
Well current	0.164	0.128	0.123	-0.251	-0.274	-0.259	-0.038	-0.037	-0.059	0.250	0.010	0.226

	(0.145)	(0.149)	(0.151)	(0.176)	(0.184)	(0.189)	(0.158)	(0.154)	(0.161)	(0.485)	(0.468)	(0.481)
Well lag	0.071	0.088	0.086	0.090	0.099	0.095	0.060	0.062	0.069	-0.097	-0.089	-0.104
	(0.130)	(0.124)	(0.127)	(0.151)	(0.154)	(0.155)	(0.142)	(0.138)	(0.138)	(0.390)	(0.398)	(0.388)
Log output current	-0.018	-0.017	-0.018	-0.007	-0.007	-0.007	-0.003	-0.001	-0.001	0.002	-0.000	0.001
	(0.016)	(0.016)	(0.016)	(0.019)	(0.019)	(0.018)	(0.013)	(0.013)	(0.013)	(0.038)	(0.040)	(0.040)
Log output lag	0.020*	0.016	0.017	0.018	0.013	0.014	-0.000	-0.001	-0.000	0.042	0.049	0.051
	(0.011)	(0.012)	(0.012)	(0.016)	(0.017)	(0.017)	(0.015)	(0.015)	(0.015)	(0.047)	(0.048)	(0.048)

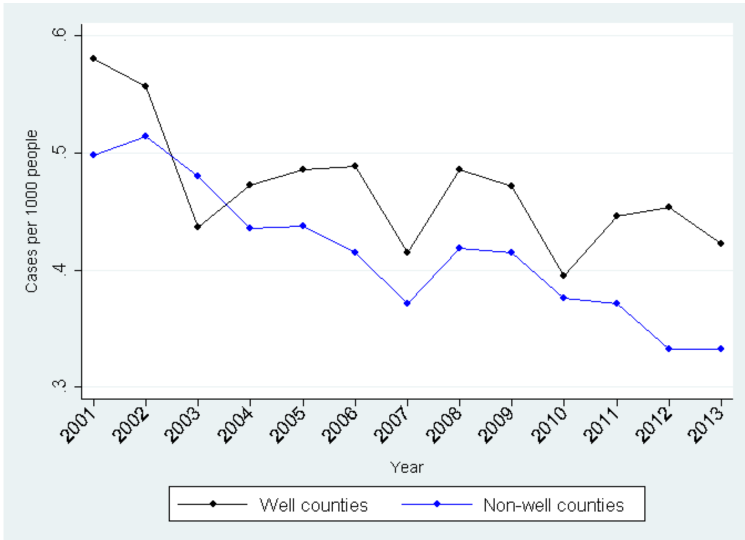
Notes: *p<0.1, **p<0.05, ***p<0.01. Standard errors are clustered at county level. Total of number of observations is 871. All models include county fixed effects. Each column represents a separate model. For each outcome in each age group, the first model controls for a set of year dummies, the second model controls for an overall linear time trend, and the third model controls for more flexible linear time trends (squared and cubic). Other control variables include average age, the share of different types of insurance (Medicare, Medicaid, private, self-pay, government, and other insurance), the share of female patients, the share of different race and ethnicity groups (white, black, Asian, Hispanic, and other race), the share of different types of admission (emergency, urgent, elective, and other types), average Charlson index, county-level unemployment rate, poverty rate, annual quartiles of median household income, log of population density, log of annual coal production (both surface and underground) and entire county-level age distribution.

Table 4.6: Rejection rates for H_0 from Monte Carlo simulations of placebo treatments, 2000-2005

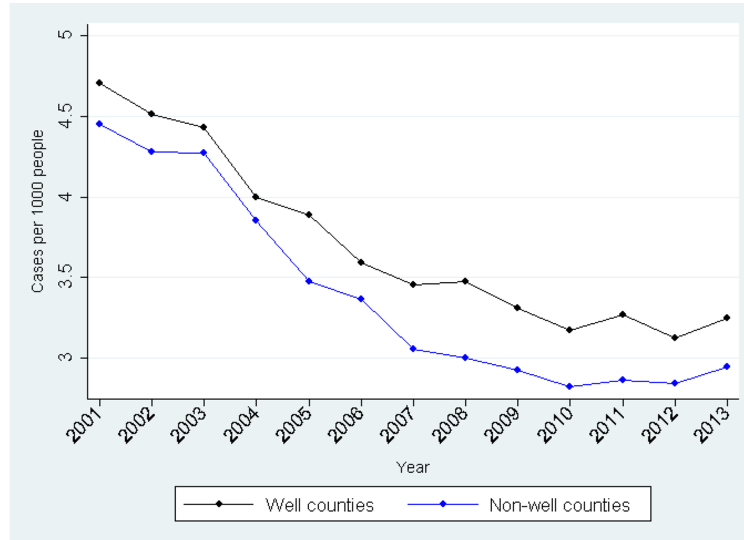
	Age 5-19			Age 20-44			Age 45-64			Age 65 and above		
	< 0.01	< 0.05	< 0.1	< 0.01	< 0.05	< 0.1	< 0.01	< 0.05	< 0.1	< 0.01	< 0.05	< 0.1
AMI												
Well lead				1.35%	4.50%	10.10%	1.48%	7.04%	12.88%	1.44%	6.920%	13.29%
Well current				1.26%	5.83%	11.90%	1.01%	5.03%	10.29%	1.00%	5.990%	11.52%
Well lag				1.19%	4.63%	9.86%	1.13%	5.24%	11.18%	1.42%	7.020%	12.99%
COPD												
Well lead				0.53%	4.26%	9.56%	1.56%	6.44%	12.14%	0.48%	3.430%	8.13%
Well current				1.20%	5.09%	10.75%	1.67%	6.58%	11.80%	0.84%	4.840%	10.14%
Well lag				0.76%	4.22%	9.90%	1.89%	7.54%	13.43%	0.54%	3.280%	7.63%
Asthma												
Well lead	3.18%	10.82%	18.75%	2.07%	8.48%	15.31%	2.62%	9.60%	16.74%	1.27%	6.350%	12.24%
Well current	1.27%	5.94%	11.26%	1.47%	6.36%	12.00%	0.94%	5.19%	10.69%	0.92%	5.510%	11.35%
Well lag	2.63%	9.75%	16.79%	3.22%	10.70%	17.99%	1.63%	7.23%	13.88%	1.45%	7.060%	13.13%
Pneumonia												
Well lead	1.06%	5.20%	10.45%	0.79%	4.67%	9.92%	0.41%	3.57%	9.74%	1.10%	5.380%	10.65%
Well current	1.11%	5.61%	10.90%	0.79%	5.40%	11.30%	0.50%	4.08%	9.77%	0.89%	5.310%	11.06%
Well lag	0.80%	4.75%	9.71%	0.81%	5.14%	10.81%	0.57%	4.68%	10.77%	0.88%	4.280%	9.40%
URI												
Well lead	0.92%	5.28%	10.88%	1.00%	4.71%	9.74%	0.57%	3.96%	8.68%	0.73%	4.870%	10.46%
Well current	1.17%	5.84%	11.50%	1.01%	5.06%	10.02%	0.64%	4.11%	9.25%	0.69%	5.100%	11.02%
Well lag	0.75%	4.66%	9.61%	0.82%	4.14%	9.11%	0.37%	3.15%	7.63%	0.78%	4.790%	9.96%

Notes: Results are based on 10,000 replications. All models include county and year fixed effects. Control variables include average age, the share of different types of insurance (Medicare, Medicaid, private, self-pay, government, and other insurance), the share of female patients, the share of different race and ethnicity groups (white, black, Asian, Hispanic, and other race), the share of different types of admission (emergency, urgent, elective, and other types), average Charlson index, county-level unemployment rate, poverty rate, annual quartiles of median household income, log of population density, log of annual coal production (both surface and underground), and the entire county-level age distribution.

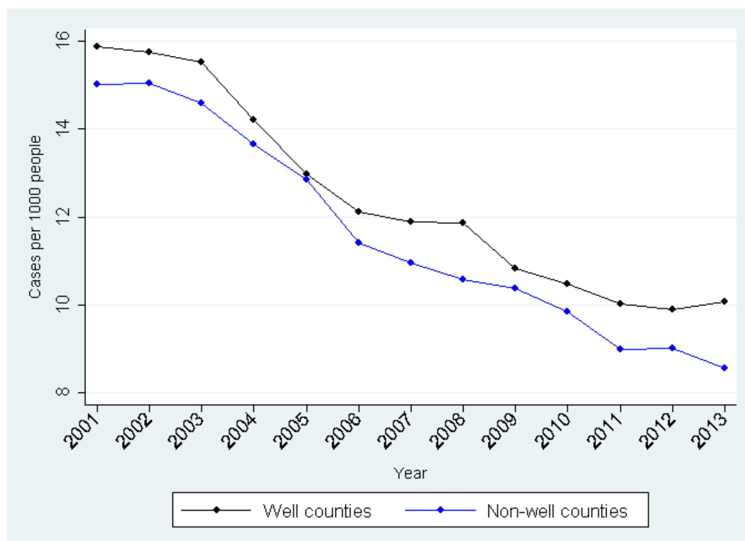
Figure 4.2: Hospitalization rates for AMI, 2001-2013



(a) Age 20-44

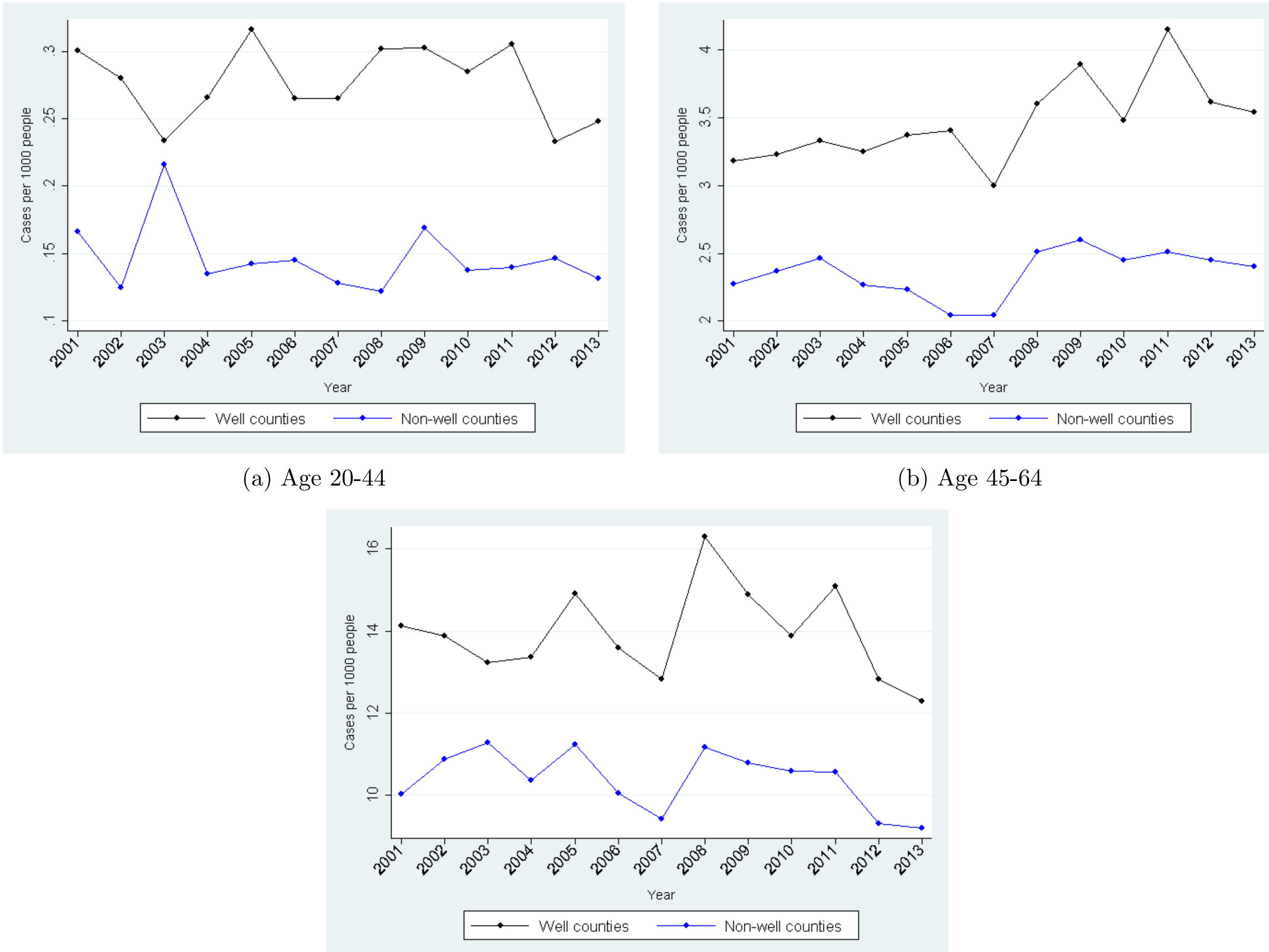


(b) Age 45-64



(c) Age 65 and above

Figure 4.3: Hospitalization rates for COPD, 2001-2013

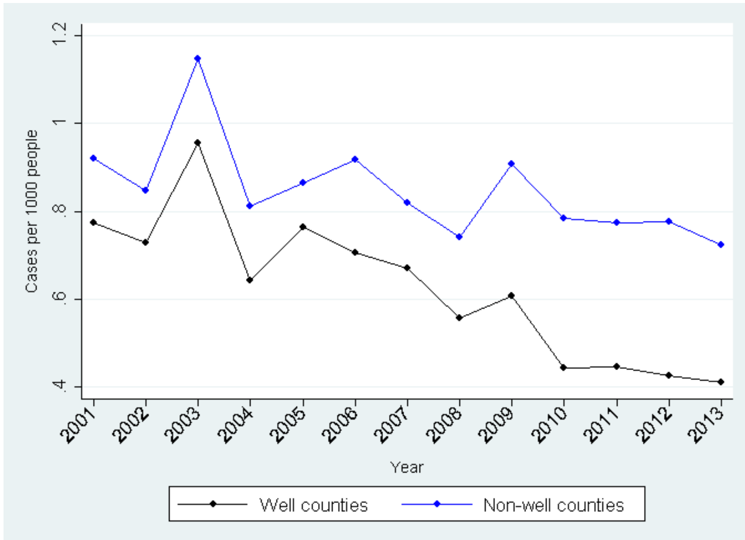


(a) Age 20-44

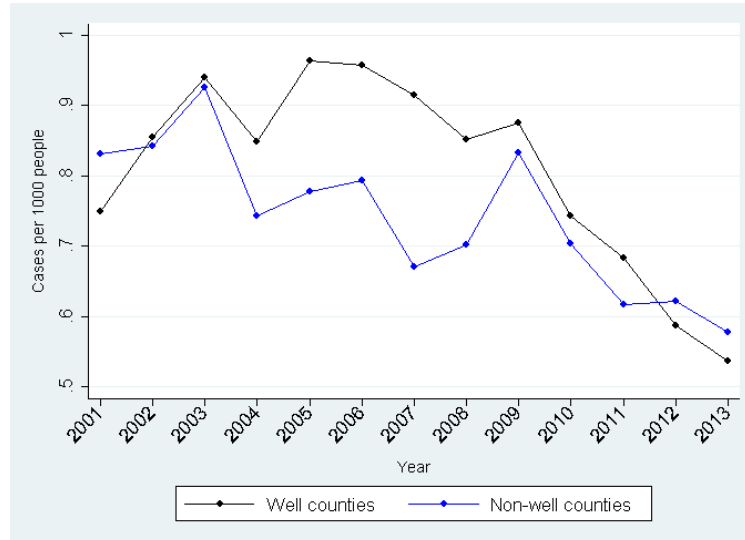
(b) Age 45-64

(c) Age 65 and above

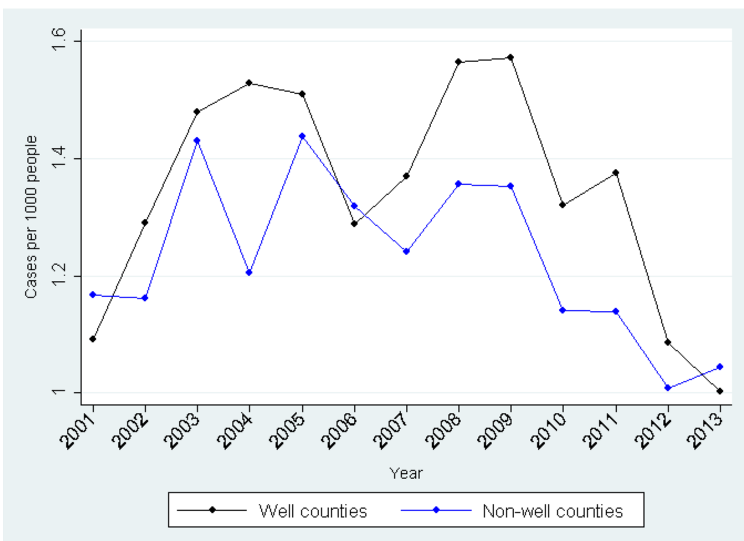
Figure 4.4: Hospitalization rates for asthma, 2001-2013



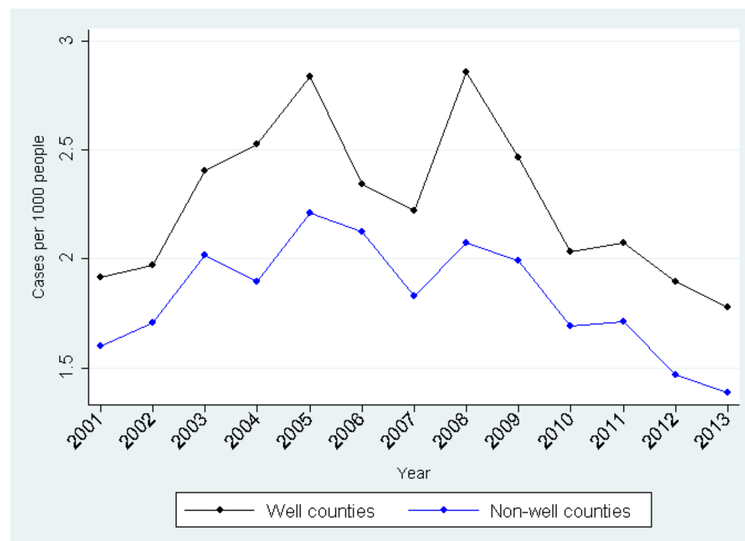
(a) Age 5-19



(b) Age 20-44

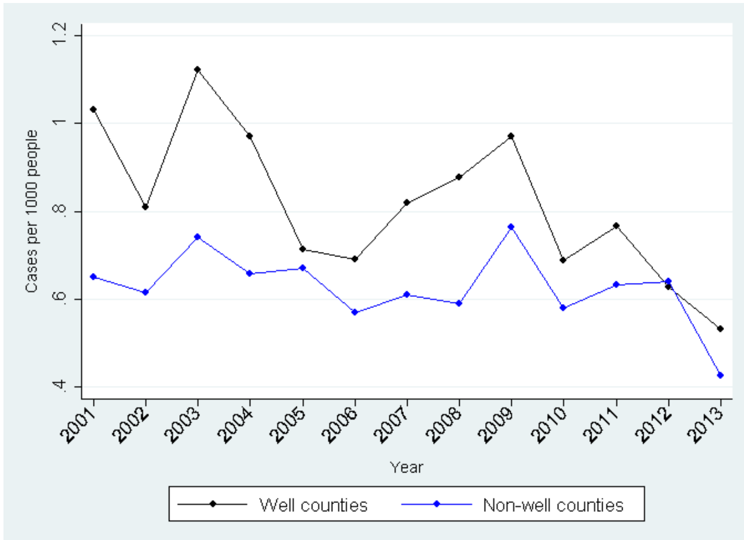


(c) Age 45-64

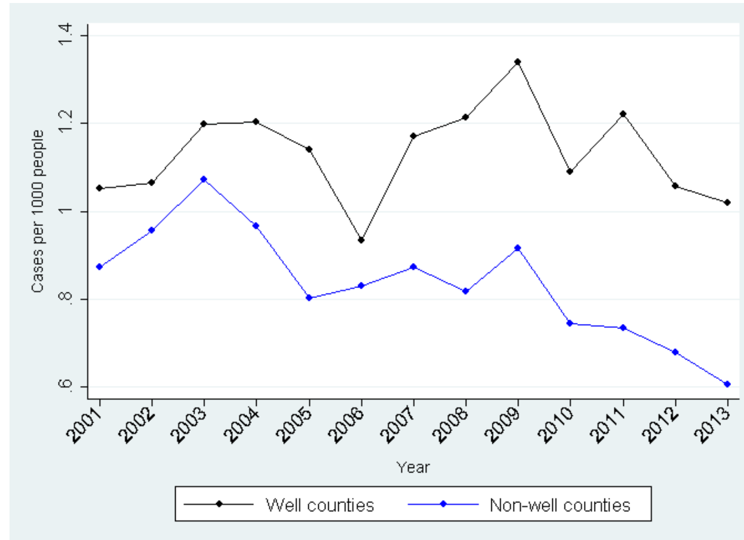


(d) Age 65 and above

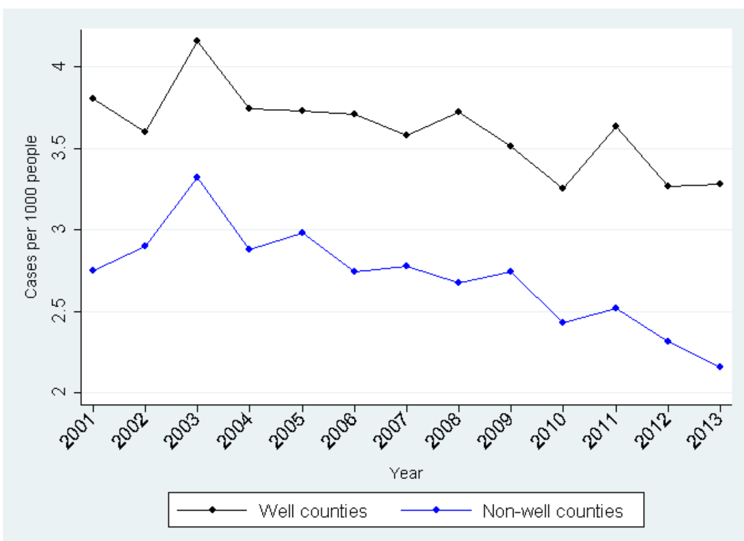
Figure 4.5: Hospitalization rates for pneumonia, 2001-2013



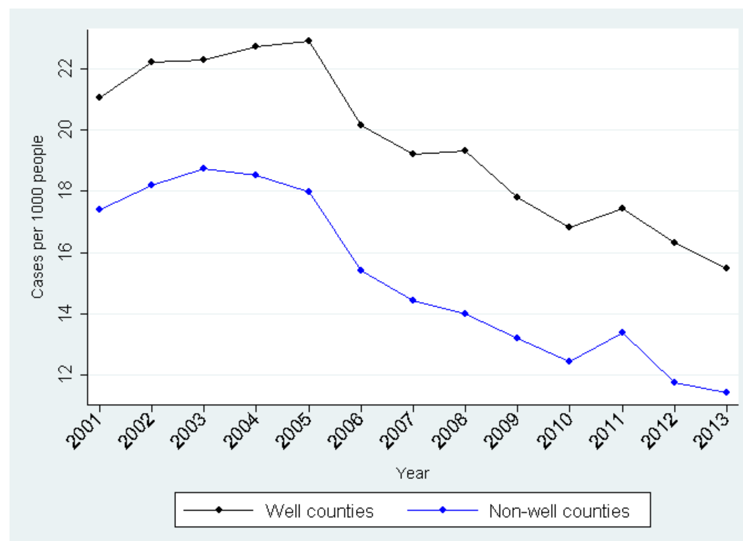
(a) Age 5-19



(b) Age 20-44

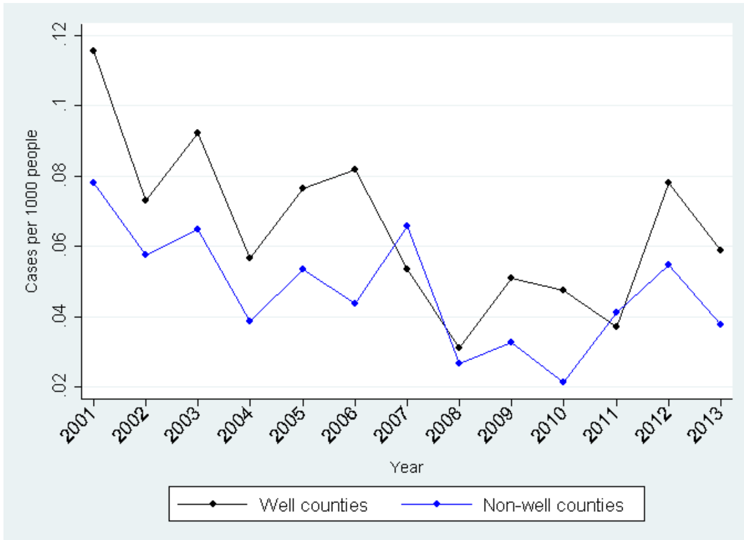


(c) Age 45-64

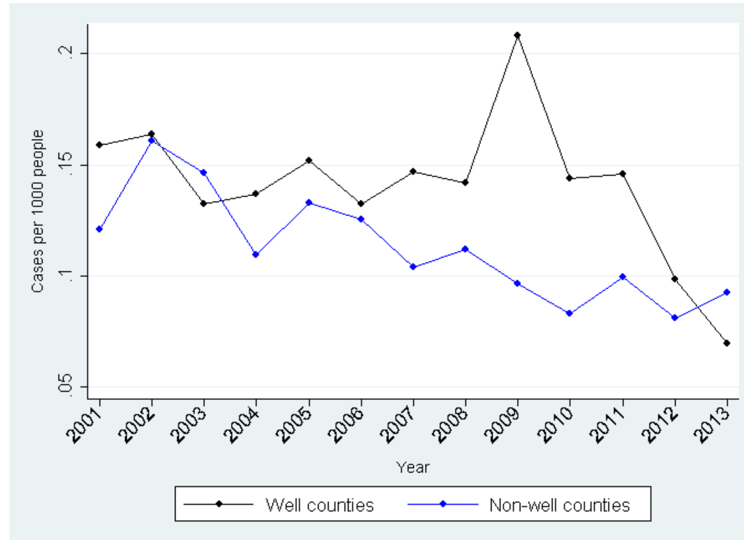


(d) Age 65 and above

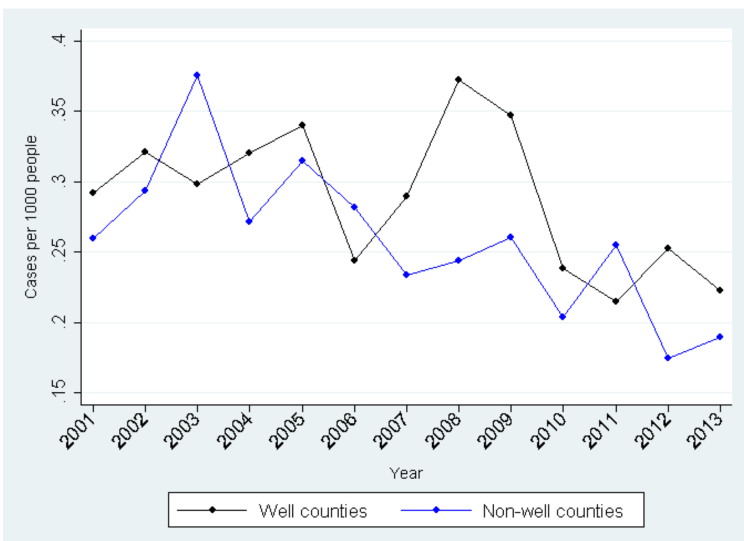
Figure 4.6: Hospitalization rates for URI, 2001-2013



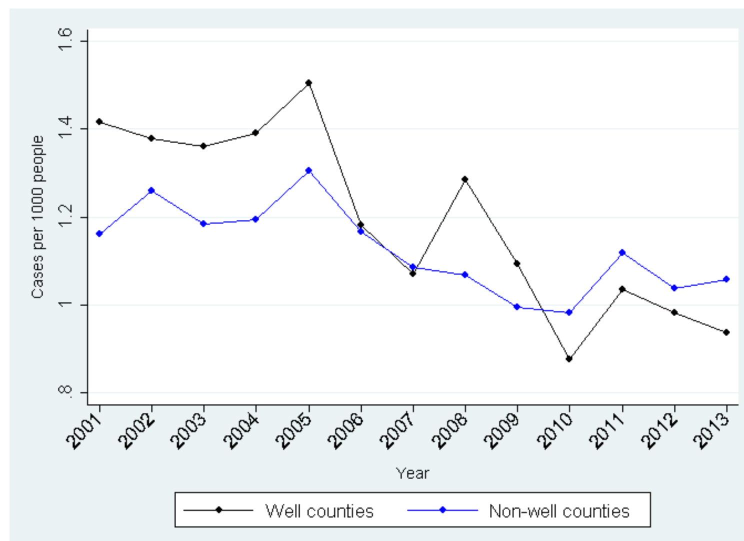
(a) Age 5-19



(b) Age 20-44



(c) Age 45-64



(d) Age 65 and above

Bibliography

Alexandre, P.K., and M.T. French. 2001. "Labor supply of poor residents in metropolitan Miami, Florida: The role of depression and the co-morbid effects of substance use." *Journal of Mental Health Policy and Economics*, 4: 161–173.

Altonji, J.G., T.E. Elder, and C.R. Taber. 2005. "Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools." 113: 151–184.

Altonji, J.G., T.E. Elder, and C.R. Taber. 2008. "Using selection on observed variables to assess bias from unobservables when evaluating Swan-Ganz catheterization." *The American Economic Review: Papers and Proceedings*, 98: 345–350.

Attfield, M.D., P.L. Schleiff, J.H. Lubin, A. Blair, P.A. Stewart, R. Vermeulen, J.B. Coble, and D.T. Silverman. 2012. "The Diesel Exhaust in Miners study: a cohort mortality study with emphasis on lung cancer." *Journal of the National Cancer Institute*, 104(11): 869–883.

Beronio, K., R. Po, L. Skopec, and S. Glied. 2013. "Affordable Care Act Will Expand Mental Health and Substance Use Disorder Benefits and Parity Protections for 62 Million Americans." *ASPE Research Brief.* February 2013.

Birnbaum, H.G., R.C. Kessler, D. Kelley, R. Ben-Hamadi, V.N. Joish, and P.E. Greenberg. 2010. "Employer burden of mild, moderate, and severe major depressive disorder: Mental health services utilization and costs, and work performance." *Depression and Anxiety*, 27: 78–89.

Brown, D., B. Weinberger, C. Lewis, and H. Bonaparte. 2014*a*. "Understanding exposure from natural gas drilling puts current air standards to the test." March 29.

Brown, J., M. Duggan, I. Kuziemko, and W. Woolston. 2014*b*. "How does risk selection respond to risk adjustment." *American Economic Review*, 104(10): 3335–3364.

Brunekreef, B., and S. T. Holgate. 2002. "Air pollution and health." *The Lancet*, 360(9341): 1233–1242.

California Office of Environmental Health Hazard Assessment (OEHHA) and the American Lung Association (ALA). 2003. "Air Pollution and Children's Health." URL: `http://oehha.ca.gov/public_info/facts/airkids.html`, accessed on March 19, 2015.

Cameron, A.C., and P.K. Trivedi. 2005. Cambridge and New York:Cambridge University Press.

Centers for Medicare & Medicaid Services (CMS). 2012. "Frequently asked questions: Hospital Value-Based Purchasing Program." URL: `http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/FY-2013-Program-Frequently-Asked-Questions-about-Hospital-VBP-3-9-12.pdf`, accessed on May 5, 2014.

Chamberlain, G. 1980. "Analysis of covariance with qualitative data." *Review of Economic Studies*, 47: 225–238.

Charlson, M.E., P. Pompei, K.L. Ales, and C.R. Mackenzie. 1987. "A new method of classifying prognostic comorbidity in longitudinal studies: development and validation." *Journal of Chronic Diseases*, 40(5): 373–383.

Chatterji, P., M. Alegria, and D. Takeuchi. 2011. "Psychiatric disorders and labor market outcomes: Evidence from the national comorbidity survey-replication." *Journal of Health Economics*, 30: 858–868.

Chatterji, P., M. Alegria, M. Lu, and D. Takeuchi. 2007. "Psychiatric disorders and labor market outcomes: Evidence from the national Latino and Asian American study." *Health Economics*, 16: 1069–1090.

Chien, Alyna T.and K. Wroblewski, C. Damberg, T.R. Williams, D. Yanagihara, Y. Yakunina, and L.P. Casalino. 2012. "Do physician organizations located in lower socioeconomic status areas score lower on pay-for-performance measures?" *Journal of General Internal Medicine*, 27(5): 548–554.

Colborn, Theo, Carol Kwiatkowski, Kim Schultz, and Mary Bachran. 2011. "Natural gas operations from a public health perspective." *Human and Ecological Risk Assessment: An International Journal*, 17(5): 1039–1056.

Committee on Energy and Commerce, U.S. House of Representatives. 2011. "Chemicals Used in Hydraulic Fracturing." URL: `http://democrats.energycommerce.house.gov/sites/default/files/documents/Hydraulic-Fracturing-Chemicals-2011-4-18.pdf`, accessed on May 5, 2014.

Currie, J., and B.C. Madrian. 1999. "Health, health insurance, and the labor market." In *Handbook of Labor Economics*. Vol. 3, , ed. O. Ashenfelter and D. Card, 3309–3415.

Cusick, M. 2013. "Pennsylvania Supreme Court strikes down controversial portions of Act 13. Stateimpact Pennsylvania." URL: `http://stateimpact.npr.org/pennsylvania/2013/12/19/state-supreme-court-strikes-down-act-13-local-zoning-restrictions/`, accessed on August 6, 2014.

Dockery, D. W., and C. A. Pope. 1994. "Acute Respiratory Effects of Particulate Air Pollution." *Annual Review of Public Health*, 15(1): 107–132.

Dominici, F., R.D. Peng, Bell M.L., and et al. 2006. "FIne particulate air pollution and hospital admission for cardiovascular and respiratory diseases." *JAMA*, 295(10): 1127–1134.

Dranove, D., D. Kessler, M. McClellan, and M. Satterthwaite. 2003. "Is more information better? The effects of "report cards" on health care providers." *Journal of Political Economy*, 111(3): 555–588.

Eder, W., M. J. Ege, and E. von Mutius. 2006. "The asthma epidemic." *New England Journal of Medicine*, 355(21): 2226–2235.

Elixhauser, A, C Steiner, D.R. Harris, and R.M. Coffey. 1998. "Comorbidity measures for use with administrative data." *Medical Care*, 36(1): 8–27.

Elliott, M.N., A.M. Zaslavsky, E. Goldstein, W. Lehrman, K. Hambarsoomians, M.K. Beckett, and L. Giordano. 2009. "Effects of survey mode, patient-mix, and nonresponse on CAHPS hospital survey scores." *Health Services Research*, 44(2): 501–518.

Ellis, R.P. 1998. "Creaming, skimping, and dumping: provider competition on the intensive and extensive margins." *Journal of Health Economics*, 17: 537–555.

Environmental Protection Agency (EPA). 1998. "NOX: How Nitrogen Oxides affect the Way We Live and Breathe." URL: `http://www.nchh.org/Portals/0/Contents/EPANitrogen_Oxides.pdf`, accessed on May 5, 2014.

Environmental Protection Agency (EPA). 1999. "The Benefits and Costs of the Clean Air Act 1990-2010." URL: `http://www.epa.gov/oar/sect812/1990-2010/chap1130.pdf`, accessed on March 10, 2015.

Environmental Protection Agency (EPA). 2003. "Particle Pollution and Your Health." URL: `http://www.epa.gov/airnow/particle/pm-color.pdf`, accessed on May 5, 2014.

Environmental Protection Agency (EPA). 2009. "Ozone and Your Health." URL: `http://www.epa.gov/airnow/ozone-c.pdf`, accessed on May 5, 2014.

Environmental Protection Agency (EPA). 2012. "Well Completions and Workovers with Hydraulic Fracturing." URL: `http://www.epa.gov/climatechange/Downloads/ghgemissions/2012Workshop/EPA-Well_Completions_and_Workovers_with_Hydraulic_Fracturing.pdf`, accessed on March 10, 2015.

Ettner, S.L., R.G. Frank, and R.C. Kessler. 1997. "The impact of psychiatric disorders on labor market outcomes." *Industrial and Labor Relations Review*, 51: 64–81.

Garfield, R.L., S.H. Zuvekas, J.R. Lave, and J.M. Donohue. 2011. "The impact of national health care reform on adults with severe mental disorders." *The American Journal of Psychiatry*, 168: 486–494.

Garshick, E. 2014. "Effects of short- and long-term exposures to ambient air pollution on COPD." *European Respiratory Journal*, 44(3): 558–561.

Gibbons, R.D., K. Hur, C.H. Brown, J.M. Davis, and J.J. Mann. 2012. “Benefits from antidepressants: Synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine.” *Archives of General Psychiatry*, 69: 572–579.

Government Accountability Office (GAO). 2011. “Mental Health and Substance Use: Employers Insurance Coverage Maintained or Enhanced Since Parity Act, but Effect of Coverage on Enrollees Varied.” URL: `http://www.gao.gov/new.items/d1263.pdf`, accessed on May 1, 2013.

Government Accountability Office (GAO). 2012. “Oil and Gas: Information on Shale Resources, Development, and Environmental and Public Health Risks.” URL: `http://www.gao.gov/products/GAO-12-732`, accessed on July 6, 2014.

Greenberg, P.E., R.C. Kessler, H.G. Birnbaum, S.A. Leong, S.W. Lowe, P.A. Berglund, and et al. 2003. “The economic burden of depression in the United States: how did it change between 1990 and 2000.” *Journal of Clinical Psychiatry*, 64(12): 1465–1475.

Harris, M.N., and X. Zhao. 2007. “A zero-inflated ordered probit model, with an application to modelling tobacco consumption.” *Journal of Econometrics*, 141: 1073–1099.

HCUP State Inpatient Databases (SID). 2014. “Healthcare Cost and Utilization Project (HCUP), 2009-2012.” Agency for Healthcare Research and Quality, Rockville, MD.

Hill, E. 2012. “Shale gas development and infant health: evidence from Pennsylvania.” Cornell Dyson School working paper.

Issac, T., A.M. Zalslavsky, P.D. Cleary, and B.E. Landon. 2010. “The relationship between patients’ perception of care and measures of hospital quality and safety.” *Health Services Research*, 45(4): 1024–1040.

Jakubson, G. 1988. “The Sensitivity of Labor Supply Parameter Estimates to Unobserved Individual Effects: Fixed and Random Effects Estimates in a Non-Linear Model Using Panel Data.” *Journal of Labor Economics*, 6(3): 302–329.

James, J. 2012. “pay-for-performance.” *Health Affairs*. Health policy brief, October 11, 2012.

Jha, A.K., E.J. Orav, and A.M. Epstein. 2011. “Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares Of elderly black, Hispanic, and Medicaid patients.” *Health Affairs*, 30(10): 1904–1911.

Jha, A.K., E.J. Orav, J. Zheng, and A.M. Epstein. 2008. “Patients’ perception of hospital care in the United States.” *The New England Journal of Medicine*, 359: 1921–1931.

Kessler, R.C., G. Andrews, L.J. Colpe, E. Hiripi, D.K. Mroczek, S.T. Normand, and et al. 2002. "Short screening scales to monitor population prevalences and trends in non-specific psychological distress." *Psychological Medicine*, 32: 959–976.

Kessler, R.C., W.T. Chiu, O. Demler, and E.E. Walters. 2005. "Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication." *Archives of General Psychiatry*, 62: 617–627.

Ko, F. W. S., W. Tam, T. W. Wong, C. K. W. Lai, G. W. K. Wong, T.-F. Leung, S. S. S. Ng, and D. S. C. Hui. 2007. "Effects of air pollution on asthma hospitalization rates in different age groups in Hong Kong." *Clinical and Experimental Allergy*, 37(9): 1312–1319.

Kroenke, K., R.L. Spitzer, and J.B.W. Williams. 2003. "The patient health questionnaire-2: Validity of a two-item depression screener." *Medical Care*, 41: 1284–1292.

Lee, B.J., B. Kim, and K. Lee. 2014. "Air pollution exposure and cardiovascular disease." *Toxicological Research*, 30(2): 71–75.

Li, J., J. Hurley, P. Decicca, and G. Buckley. 2014. "Physician response to pay-for-performance: evidence from a natural experiment." *Health Economics*, 23: 962–978.

Li, P., M.M. Kim, and J.A. Doshi. 2010. "Comparison of the performance of the CMS Hierarchical Condition Category (CMS-HCC) risk adjuster with the charlson and elixhauser comorbidity measures in predicting mortality." *BMC Health Services Research*, 10: 245–245.

Maddala, G.S. 1983. *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge:Cambridge University Press.

Marcus, J. 2013. "The effect of unemployment on the mental health of spouses-evidence from plant closures in Germany." *Journal of Health Economics*, 32: 546–558.

Maryland Hospital Services Cost Review Commission (Maryland HSCRC). 2011. "Maryland Report on Hospital Payments linked with Peformance Initiatives and Hopsital Value Based Purchasing Program Exemption Request Pursuant to Section 1886(o)(1)(C)(iv) of Social Secutiy Act." URL: `http://www.hscrc.state.md.us/documents/HSCRC_Initiatives/QualityImprovement/11-11/MarylandHSCRC_CostOutcomesRptVBP_ExemptionRequest2011-09-30.pdf`, accessed on September 15, 2014.

Meyerhoefer, C.D., and S.H. Zuvekas. 2010. "New estimates of the demand for physical and mental health treatment." *Health Economics*, 19: 297–315.

Miilunpalo, Seppo, Ilkka Vuori, Pekka Oja, Matti Pasanen, and Helka Urponen. 1997. "Self-rated health status as a health measure: the predictive value of self-reported health status on the use of physician services and on mortality in the working-age population." *Journal of clinical epidemiology*, 50(5): 517–528.

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

Mossey, Jana M, and Evelyn Shapiro. 1982. "Self-rated health: a predictor of mortality among the elderly." *American journal of public health*, 72(8): 800–808.

Mullen, K.J., R.G. Frank, and R.G. Rosenthal. 2010. "Can you get what you pay for? Pay-for-performance and the quality of healthcare providers." *RAND Journal of Economics*, 41: 64–91.

National Institute of Mental Health (NIMH). 2002. "Annual total direct and indirect costs of serious mental illness." URL: `http://www.nimh.nih.gov/statistics/4COST_TOTAN.shtml`, accessed on April 1, 2013.

National Research Council (NRC). 2010. "Hidden Cost of Energy: Unpriced Consequences of Energy Production and Use." URL: `http://www.nap.edu/catalog/12794/hidden-costs-of-energy-unpriced-consequences-of-energy-production-and`, accessed on March 10, 2015.

National Resources Defense Council (NRDC). 2014. "Fracking Fumes: Air pollution from Hydraulic Fracturing Threatens Public Health and Communities." URL: `http://www.nrdc.org/health/files/fracking-air-pollution-IB.pdf`, accessed on March 10, 2015.

National Survey on Drug Use and Health. 2011. "Results from the 2011 National Survey on Drug Use and Health: Mental Health Findings." URL: `http://www.samhsa.gov/data/NSDUH/2k11MH_FindingsandDetTables/2K11MHFR/NSDUHmhfr2011.pdf`, accessed on March 3, 2013.

Neupane, B., M. Jerrett, R.T. Burnett, T. Marrie, A. Arain, and M. Loeb. 2010. "Long-term exposure to ambient air pollution and risk of hospitalization with community-acquired pneumonia in older adults." *American Journal of Respiratory and Critical Care Medicine*, 181: 47–53.

New York State Assembly. 2013. "Bill No.A05424." URL: `http://assembly.state.ny.us/leg/?default_fld=&bn=A05424&term=2013&Summary=Y&Text=Y`, accessed on August 6, 2014.

O'Malley, A.J., A.M. Zaslavsky, M.N. Elliott, L. Zaborski, and P.D. Cleary. 2005. "Case-mix adjustment of the CAHPS hospital survey." *Health Services Research*, 40(6, part 2): 2162–2181.

Papke, L.E., and J.M. Wooldridge. 1996. "Econometric methods for fractional response variables with an application to 401(K) plan participation rates." *Journal of Applied Econometrics*, 11: 619–632.

Pennsylvania Department of Environmental Protection (PA DEP). 2011a. "North-central Pennsylvania Marcellus shale short-term ambient air sampling report." URL: `http://www.dep.state.pa.us/dep/deputate/airwaste/aq/aqm/docs/Marcellus_NC_05-06-11.pdf`, accessed on June 6, 2014.

Pennsylvania Department of Environmental Protection (PA DEP). 2011b. "Northeastern Pennsylvania Marcellus shale short-term ambient air sampling report." URL: `http://www.dep.state.pa.us/dep/deputate/airwaste/aq/aqm/docs/Marcellus_NE_01-12-11.pdf`, accessed on June 6, 2014.

Pennsylvania Department of Environmental Protection (PA DEP). 2011c. "Southwestern Pennsylvania Marcellus shale short-term ambient air sampling report." URL: `http://www.dep.state.pa.us/dep/deputate/airwaste/aq/aqm/docs/Marcellus_SW_11-01-10.pdf`, accessed on June 6, 2014.

Pope, G.C., J. Kautter, M.J. Ingber, S. Freeman, R. Sekar, and C. Newhart. 2011. "Evaluation of the CMS-HCC risk adjustment model." URL: `http://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/downloads/evaluation_risk_adj_model_2011.pdf`, accessed on June 20, 2014.

Pope, G.C., John Kautter, R.P. Ellis, A.S. Ash, J.Z. Ayanian, L.I. Iezzoni, M.J. Ingber, J.M. Levy, and J. Robst. 2004. "Risk adjustment of Medicare capitation payments using the CMS-HCC model."

Rosenthal, M.B., R.G. Frank, H. Li, and A.M. Epstein. 2005. "Early experience with pay-for-performance: from concept to practice." *Journal of the American Medical Association*, 294(14): 1788–1793.

Ryan, A. M., J. Blustein, and L.P. Casalino. 2012. "Medicares flagship test of pay-for-performance did not spur more rapid quality improvement among low-performing hospitals." *Health Affairs*, 31(4): 1932–1940.

Ryan, M., J.A. Suaya, J.D. Chapman, W.B. Stason, D.S. Shepard, and C.P. Thomas. 2013. "Incidence and cost of pneumonia in older adults with COPD in the United States." *PLoS One*, 8(10): e75887.

Schone, E., and R. Brown. 2013. "Risk adjustment: what is the current state-of-the-art, and how can it be improved." URL: `http://www.rwjf.org/content/dam/farm/reports/reports/2013/rwjf407046/subassets/rwjf407046_1`, accessed on August 10, 2014.

Schwartz, J., D. Slater, T.V. Larson, W.E. Pierson, and J.Q. Koenig. 1993. "Particulate air pollution and hospital emergency room visits for asthma in Seattle." *American Review of Respiratory Disease*, , (4): 826–831.

Shen, Y. 2003. "Selection incentives in a performance-based contracting system." *Health Services Research*, 38(2): 535–552.

Simon, G.E. 2002. "Evidence review: efficacy and effectiveness of antidepressant treatment in primary care." *General Hospital Psychiatry*, 24: 213–224.

Soni, Anita. 2012. "Statistical Brief# 377." Agency for Healthcare Research and Quality, Rockville, MD.

Spijker, J., R. de Graaf, R.V. Bijl, A.T. Beekman, J. Ormel, and W.A. Nolen. 2002. "Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS)." *The British Journal of Psychiatry*, 181: 208–213.

Tefft, N. 2011. "Insights on unemployment, unemployment insurance, and mental health." *Journal of Health Economics*, 30: 258–264.

The General Assembly of Pennsylvania. 2011. "House bill No.1950, session of 2011." URL: `http://www.legis.state.pa.us/WU01/LI/LI/US/HTM/2012/0/0013..HTM`, accessed on August 6, 2014.

Treida, M. 2010. "The evolution of hydraulic fracturing and its effect on frac pump technology." URL: `http://www.pumpsandsystems.com/topics/upstream-pumping/evolution-hydraulic-fracturing-and-its-effect-frac-pump-technology?page=2`, accessed on March 19, 2015.

U.S. Department of Energy (DOE). 2009. "Modern Shale Gas Development in the United States: a Primer." URL: `http://www.netl.doe.gov/technologies/oil-gas/publications/EPreports/Shale_Gas_Primer_2009.pdf`, accessed on July 6, 2014.

U.S. Energy Information Administration. 2013. "Annual Energy Outlook 2013 with Projections to 2040." URL: `http://www.eia.gov/forecasts/aeo/pdf/0383(2013).pdf`, accessed on July 6, 2014.

Werner, R., J.T. Kolstad, E.A. Stuart, and D. Polsky. 2011. "The effect of pay-for-performance in hospitals: lessons for quality Improvement." *Health Affairs*, 30(4): 690–698.

Werner, R.M., and D.A. Asch. 2005. "The unintended consequences of publicly reporting quality information." *The Journal of the Amercian Medical Association*, 293(10): 1239–1244.

World Health Organization (WHO). 2008. "The global burden of disease: 2004 update." URL: `http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf`, accessed on Januray 28, 2013.

Wu, Shunquan, Rui Wang, Yanfang Zhao, Xiuqiang Ma, Meijing Wu, Xiaoyan Yan, and Jia He. 2013. "The relationship between self-rated health and objective health status: a population-based study." *BMC Public Health*, 13(1): 320.

Zanobetti, A., and J. Schwartz. 2006. "Air pollution and emergency admissions in Boston, MA." *Journal of Epidemiology and Community Health*, 80: 890–895.

Curriculum Vitae

(2015/4)

Lizhong Peng

Rauch Business Center
621 Taylor Street
Bethlehem, PA 18015
lip210@lehigh.edu
(484) 838-9200
`http://www.lehigh.edu/~lip210/`

EDUCATION

Lehigh University — 2010-present
Ph.D. in Economics (expected May 2015)

Shanghai Jiao Tong University — 2006-2010
B.A. in Economics

RESEARCH INTERESTS

Primary Fields: Health Economics, Applied Econometrics
Secondary Fields: Labor Economics, Health Policy and Management

WORKING PAPERS

1. Peng, L., "Patient Selection Under Incomplete Case Mix Adjustment: Evidence from the Hospital Value-based Purchasing Program".

2. Peng, L., C. Meyerhoefer, and S. Zuvekas,"The Short-term Effect of Depression on Labor Market Outcomes", NBER working paper 19451, revised and resubmitted to *Health Economics*.

3. Peng, L., C. Meyerhoefer, and S-Y. Chou "The Health Implications of Shale Gas Development in Pennsylvania", revised and resubmitted to *Health Affairs*.

4. Meyerhoefer, C., M. Deily, S. Sherer, S-Y. Chou, L. Peng, M. Sheinberg, and D. Levick, "The Consequences of Electronic Health Record Adoption for Physician Productivity and Birth Outcomes", revise and resubmit at *Industrial and Labor Relations Review*.

5. Sherer, S., C. Meyerhoefer, and L. Peng, "Applying Institutional Theory to the Adoption of Electronic Health Records", submitted to *Information & Management*.

6. Meyerhoefer, C., S. Sherer, M. Deily, S-Y. Chou, L. Peng, T. Hu, M. Nihen, M. Sheinberg, and D. Levick, "A Mixed Method Study of Information Availability on Pregnancy Outcomes".

7. Peng, L., C. Meyerhoefer, and J. Dearden, "Selection incentives under Pay-for-performance" (in progress).

PROFESSIONAL ACTIVITIES

Conference Presentations

1. Peng, L., C. Meyerhoefer, and S. Zuvekas, "The Impact of Depression on Labor Market Outcomes", Lehigh University Academic Symposium, Bethlehem, PA, April 2013.

2. Peng, L., C. Meyerhoefer, and S. Zuvekas, "The Impact of Depression on Labor Market Outcomes", Eastern Economics Association Annual Conference, New York City, NY, May 2013.

3. Peng, L., C. Meyerhoefer, and S. Zuvekas, "The Short-term Effect of Depression on Labor Market Outcomes", 5th ASHE Biennial Conference, Los Angeles, CA, June 2014.

4. Peng, L., "Patient Selection Under Quality Incentives: Evidence from the Hospital Value-based Purchasing Program", 5th ASHE Biennial Conference, Los Angeles, CA, June 2014.

Other Papers Accepted at Conferences (presenter)

1. Meyerhoefer, C., M. Deily, S. Sherer, S-Y. Chou, L. Peng, M. Sheinberg, and D. Levick, "The Impact of Electronic Health Record Adoption and Integration on Physician Productivity and Health Outcomes", Temple University, Department of Risk, Insurance, and Healthcare Management Seminar, Philadelphia, PA, November 2013.

2. Meyerhoefer, C., M. Deily, S. Sherer, S-Y. Chou, L. Peng, M. Sheinberg, and D. Levick, "The Impact of Electronic Health Record Adoption and Integration on Physician Productivity and Health Outcomes", ILR Review Special Conference on Employment Relations in Healthcare, New Brunswick, NJ, March 2014.

3. Meyerhoefer, C., M. Deily, S. Sherer, S-Y. Chou, L. Peng, M. Sheinberg, and D. Levick, "The Impact of Electronic Health Record Adoption and Integration on Physician Productivity and Health Outcomes", 5th ASHE Biennial Conference, Los Angeles, CA, June 2014.

4. Meyerhoefer, C., S. Sherer, M. Deily, S-Y. Chou, L. Peng, T. Hu, M. Nihen, M. Sheinberg, and D. Levick, "A Mixed Method Study of Information Availability on Pregnancy Outcomes", 5th Annual Workshop on Health IT and Economics (WHITE), Washington, D.C, October 2014.

Other Presentations or Activities

1. Peng, L., C. Meyerhoefer, and S. Zuvekas, "The Effect of Depression on Labor Market Outcomes", CFACT Seminar, Agency of Healthcare Research and Quality, Rockville, MD, November 2013.

2. National Bureau of Economic Research Summer Institute, Attendee (Invited), Cambridge, MA, July 2012.

3. Referee for *Review of Economics of the Household.*

EXPERIENCE

Instructor 2013-present
Department of Economics, Lehigh University

- *Principles of Economics*: Summer 2013, Summer 2014
- *Applied Microeconomic Analysis*: Fall 2014
- *Statistical Methods*: Spring 2015

Research Assistant 2012-2014
Department of Economics, Lehigh University

Teaching Assistant
Department of Economics, Lehigh University
- *Principles of Economics*: Fall 2010
- *Money, Banking, and Financial Markets*: Spring 2011, Fall 2011, Spring 2012

AWARDS Warren York Dissertation Fellowship, Lehigh University, Spring 2013

AFFILIATIONS American Economic Association, American Society of Health Economists

SKILLS *Computer Softwares:* STATA, SAS, LaTeX, Microsoft Excel
Languages: English (fluent) and Mandarin (native)

REFERENCES

Chad Meyerhoefer
Associate Professor of Economics,
Department of Economics
Lehigh University
and NBER
Rauch Business Center,
621 Taylor Street,
Bethlehem, PA 18015
Email: chm308@lehigh.edu
(Tel) 610-758-3445

James Dearden
Professor and Chair of Economics,
Department of Economics
Lehigh University
Rauch Business Center,
621 Taylor Street,
Bethlehem, PA 18015
Email: jad8@lehigh.edu
(Tel) 610-758-5129

Samuel Zuvekas
Economist,
Division of Social and Economic Research
Center for Financing, Access and Cost Trends
Agency for Healthcare Research and Quality
540 Gaither Road,
Rockville, MD 20850
Email: Samuel.Zuvekas@ahrq.hhs.gov
(Tel) 301-427-1673